3

Multivariate Social Network Visual Analytics

Chris Muelder, Liang Gou, Kwan-Liu Ma, and Michelle X. Zhou

One of the key research topics in Social Science and Sociology is to understand and analyze various social networks. Like any other types of networks, a social network consists of a set of nodes and links. Here, each node often represents a social entity, such as an individual or a group, and each link represents a particular relationship between two social entities. In a *multivariate* social network, each node/link can be associated with a set of properties, or there can even be multiple sets of heterogenous nodes or edges.

3.1 Data Characteristics

In addition to understanding the behavior of individual social entities, Sociology is also concerned with the behavior of *groups*, in particular, how these groups interact with each other [49]. In this context, a multivariate social network is composed of the entities of groups. The connections between them depend on the task being pursued and the information that is available, but are generally a set (or multiple sets) of relationships between the entities. These relationships can be directed or undirected, weighted or unweighted. Additionally, the nodes can carry any additional properties.

In particular, social networks are imbued with a number of properties. The size and complexity of the topology itself can be overwhelming for many traditional approaches. Additionally, both the nodes and links can carry any number of properties, including nominal, ordinal, and continuous measures: for instance, nodes can often be broken down into classes both ordered (age, grade, etc.) and unordered (gender, race, etc.), or there can be multiple classes of edges on the same set of nodes (e.g., both friendship and aggression ties between the same group of actors), or the nodes or edges could contain weights, or even multiple weight metrics, and the edges could be directional if the network is not symmetric. Also, many social networks evolve over time, so while static analysis can reveal some insights, in many cases the evolution of dynamic social networks could be of importance.

A. Kerren et al. (Eds.): Multivariate Network Visualization, LNCS 8380, pp. 37–59, 2014.

To make things more complicated, a multivariate social network may not be homogeneous as it may contain multiple types of nodes, which represent both individuals and groups. For example, an enterprise social network may include nodes representing individual employees as well as those representing companies or organizations that are customers, suppliers, or partners of the enterprise.

All of the mentioned properties are often compounded by a difficulty in acquiring the network data. Next, we briefly describe different approaches to acquire social network data.

3.1.1 Traditional Data Collection

Traditionally, social network data are acquired by polling small groups of people, where people report on their social ties via questionnaires. This introduces numerous points for the introduction of uncertainty. Not only do the subjects' responses depend on the questions asked, but even on how they are worded. The accuracy of the data also relies on the honesty of the subjects. In addition, temporal resolution of such networks is extremely low, since it is very difficult to get a large number of subjects to willingly and dutifully fill out one questionnaire, let alone repeated (e.g., weekly, monthly, or even annual) questionnaires. To address these challenges, alternative data acquisition approaches have been considered, particularly in social network analysis of non-human actors such as herds of animals. In these cases, the actors are tagged and tracked, but such tracking obfuscates most details of the network, as the animals are unable to communicate the specifics of their relationships, e.g., a proximity test could determine two animals met, but not inherently determine if the meeting was amicable or hostile. In either case, most traditional social network data collection methods result in data sets that are small, incomplete, noisy, vague, or even unreliable.

3.1.2 Data Collection from Online Social Media Networks

Conversely, the advent of online social media, such as Facebook and Twitter, has created the ultimate data source for social network analysis, as an incredible number of users are readily willing to divulge both explicit and implicit social connectivity information in order to benefit from the service that social media provides. This has resulted in an explosion of social network data in recent years. The result of this is that sociologists now often have to deal with massive data challenges, such as handling extremely large networks or real-time trend detection and analysis. But the emergence of social media has also introduced privacy issues that often limit the third party access to these networks and also limit what can be done with them.

In parallel to collecting social network data from public social media sites, an alternative approach is to build social networks from people's communication data, including emails, online chats, phone calls, and meeting invites, especially within the context of enterprise [59]. Not only can such data connect one social entity to another, but they can also be used to characterize the relationships between any two connected entities, including their tie strength, topic of interest, and style or type of communications. Furthermore, such information can the be used to better understand the characteristics of an individual's as well as an organization's social network [59].

3.2 Task Characteristics

Given a multivariate social network, the typical tasks of understanding such a network are to analyze its different social entities, and the properties of the entities or the network as a whole.

3.2.1 Understanding Social Network Nodes

As mentioned in the previous section, a node of a multivariate social network represents an individual or a group, which is often associated with a set of traits describing the individual or group. With the emergence of social media and advances in data analytics, much information can be inferred from one's social media footprints to describe various traits of the individual or group. In particular, there is much research on understanding various traits of an individual, from demographics to political orientation to personality traits [21, 39, 47]. Similarly, there is also much work on extracting the properties of a group, including aggregated properties of a group such as the level of expertise [46] or the discovery of latent communities/groups along with their properties [48, 58].

3.2.2 Understanding Social Network Links

Since a link represents the relationship between two social entities, understanding a link is often to characterize such a relationship (e.g., type and strength) and predict its properties (e.g., likelihood to last). The relationship between two social entities can be characterized in many different ways. For example, between two individuals, such a relationship can be used to describe what, how, and when such a relationship is established [59]. Besides understanding the characteristics of a relationship, there is also research on predicting the properties in particular the existence of a particular relationship between two entites [26].

3.2.3 Understanding Social Networks

Understanding a network as a whole is a complex task as it depends on the purposes of the analysis as well as the analytic technologies. Because of the challenges, visualization is often developed to accompany the analytics technologies to help users better understand various properties of a network.

Graph-Based Analysis

Many sociological research works draw on graph analytic algorithms and statistics. These analyses can range in scale from looking at small scale patterns such as dyads (pairs of entities) [49] or triads (groups of 3) [17], to centrality metrics that measure nodes' importance to the network as a whole [18], and up to large scale analysis of the high-level relationships that find and compare large groups of entities and how they interact.

Sociograms Analysis

One key element of social network research has been visualization of node-link diagrams, which sociologists often refer to as "sociograms" [19]. While statistical metrics can be quite succinct, it can be difficult to know a priori what metric will produce the right result, and it can be difficult to directly verify that the results are correct. Pictorial representations of social networks can help to both directly communicate the content of the network such as structural patterns, as well as to guide and confirm the choices of statistical metrics. Nevertheless, traditional visual diagrams of social networks often suffer from a range of problems, the most common of which being the high density of edges and complex structures in large networks, yielding sociograms that often appear as indecipherable clouds of nodes and edges.

3.3 Examples of Technologies

3.3.1 Clustering

Another way to simplify large, complex networks is to cluster tightly connected groups of nodes together and consider the resulting abstracted super-network. Many current clustering algorithms are based on the modularity metric, such as the Louvain clustering method [3] or the "Fast Community" clustering algorithm of Clauset, Newman, and Moore [8]. These clustering algorithms have been shown to be effective on real-world networks, as the modularity metric is demonstrably comparable to force directed energy functions [45]. Modularity is a metric that evaluates a specific proposed clustering of a network by measuring the density of cluster interiors and the sparsity of inter-cluster connections. Specifically, given a network with a proposed clustering, the modularity Q is defined as:

$$Q = \frac{1}{2|E|} \sum_{i,j} \left[A_{i,j} - \frac{k_i k_j}{2|E|} \right] \delta_{i,j} \tag{3.1}$$

where $|E|$ is the number of edges in the network, k_i, k_j are the degrees of nodes i and j, $A_{i,j}$ is 1 if there is an edge between nodes i and j and 0 otherwise, and $\delta_{i,j}$ is 1 if nodes i and j are in the same cluster and 0 otherwise. Recent efforts have also been shown to make such approaches produce more balanced hierarchies [31] or to parallelize the clustering calculation [36].

3.3.2 Network Centralities

Centrality metrics are commonly applied to the analysis of social networks, such as Eigenvector [6, 35], Markov [57], Betweenness [18, 38], and Closeness [32, 44] centrality. Each of these measure vertices' overall importance with respect to the whole network. Rather than basing the importance of a node solely on how many connections it has, eigenvector centrality also takes into account the weights of connections to other nodes; a single connection to a highly important node can carry more weight than many connections to nodes of low importance. Eigenvector centrality sensitivity extends this notion to derive the importance of nodes relative to each other.

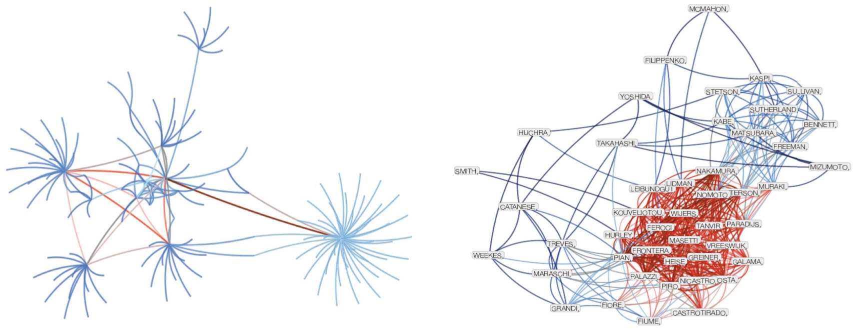
3.3.3 Centrality Derivatives

While centrality gives one value per node, centrality sensitivity analysis measures a vertex's importance to the structure of the network relative to other vertices in the graph [9]. These metrics are essentially derivatives of centrality, and as such can be calculated similarly for any type of centrality. To calculate a reference node's sensitivity to a target node, the reference node's initial centrality is calculated, each edge of the target node is removed one at a time, and the centrality of the reference node is recalculated after each removal. The negative changes in centrality of the reference node give a measure of how important the target node is to the reference node—in other words, how sensitive the reference node's centrality is to the target node. For instance, if removing a target node's edges results in large decreases in the reference node's centrality, then the reference node is said to be highly sensitive to the target node—that is, the target node has high importance relative to the reference node. This can be summarized in the following equation [9]:

$$\frac{\partial x}{\partial t_i} = -Q^+ \frac{\partial Q}{\partial t_i} x \tag{3.2}$$

where x is the centrality, t_i is the degree of vertex i, Q is the subtraction of the identity matrix from the adjacency matrix of the network ($Q = A - I$) (A is the adjacency matrix, and I is the identity matrix), and Q^+ is the pseudoinverse of Q.

One application of these sensitivities is to evaluate the roles of edges in the graph. If two nodes impact each other negatively, then they have a competitive relationship, whereas other nodes have mutually beneficial relationships. This can be shown as simply as using color, as in Fig. 3.1. Alternately, since every node has a centrality derivative with respect to every other node, centrality sensitivity can be thought of as a complete, weighted network. From



(a) Freindster network (b) Coauthorship

Fig. 3.1. Centrality sensitivity analysis can indicate how collaborative (blue) or competitive (red) relationships are along edges in the network

this network, it is possible to derive a skeleton network based on edge existence, high centrality derivatives, and overall connectivity (e.g., using a spanning tree). This skeleton network can then be thresholded to be as sparse or as dense as needed, and can be used for a wide variety of purposes, such as simplifying/clarifying layouts (as in Fig. 3.2), visualizing only the most important connections, or finding important relationships between nodes with no direct connections.

3.3.4 Traditional Network Layouts

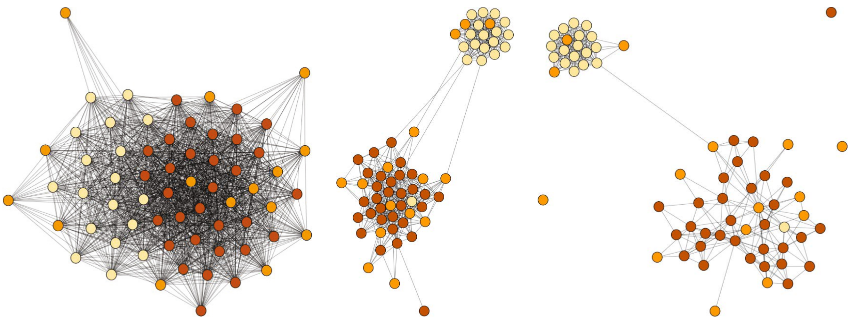
One key task in creating visual images of networks is to determine the appropriate geometrical layout of the nodes and edges. There are several well-defined criteria for assessing the accuracy and validity of a particular graph layout [13]. Some common criteria [2, 4] include, but are not limited to:

1. edges of the same approximate length;
2. vertices distributed over the area;
3. reduction of the number of edge crossings.

Nevertheless, optimization of such criteria can be intractable and often contradictory [4]. For surveys of many modern graph layout algorithms see Battista, Eades, Tamassia, and Tollis [55] or Hachul and Jünger [24].

The most traditional and commonly used layout algorithm for social network analysis are force-directed layouts [33], often referred to as "spring embedders" [15]. In this well-known procedure, nodes in a network graph are positioned iteratively, where the edges connecting them are treated like springs that push and pull on them until the system converges to an equilibrium. However, spring embedder techniques do not always scale nicely to large graphs [5]. Thus, a common problem that faces many existing visualizations of large social networks (most of which use force-directed layouts) is that they often result in a tangled mess of incomprehensible lines; this is often referred to as the "hair-ball" problem (Fig. 3.2(a) shows an example).

Other approaches have been developed with the goal to improve network layout in terms of quality and algorithmic efficiency, especially for large graphs. One such technique [4] is based on a variant of dimension-reduction methods, referred to as multidimensional scaling [10], in which the goal is to minimize stress. In this approach, the purpose of stress minimization is to determine positions for every node such that the Euclidean distances in the n-dimensional space resemble the given distances between the nodes, as determined by graph-theoretic measures, such as the shortest paths (i.e., geodesics). However, such geodesic based layouts tend to fail on networks with small diameter, as is common among social networks.



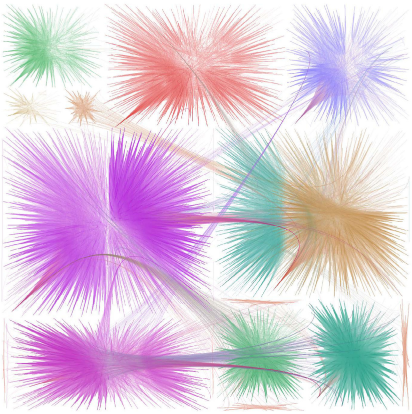
(a) All edges (b) Filtered (c) Further filtered

Fig. 3.2. MIT reality data set. Trying to lay out the whole network can yield an unintelligible hairball (a), but filtering out the less important edges via sensitivity analysis reveals two clusters (b), and further filtering starts to dissolve one of them while the other remains strong (c).

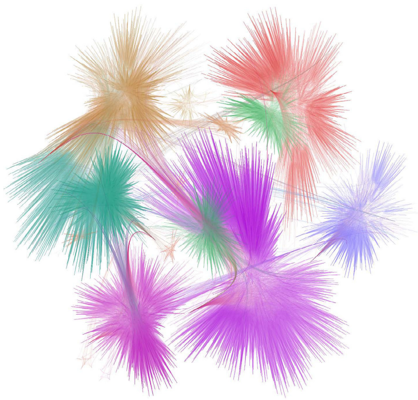
3.3.5 Improved Network Layouts

One method for improving the layout of dense social networks is to trim the network of its less essential connections to reduce it to a core network consisting of just the most important connections. A naïve way to do this would be to simply take a spanning tree of the network itself, but this is not always ideal for preserving the centralities of the nodes, which sociologists are often concerned with. Instead, the edge filtering can be weighted according to the centrality derivatives, so that the edges that are removed are the ones that affect the centralities the least. This produces a core network that preserves as much of the critical structure of the network, which can then be used to create an improved layout of the graph that reveals more detailed structures, as in Figs. 3.2(b) and 3.2(c). Once this reduced network is laid out, the original edges can optionally be reintroduced.

As social networks tend to exhibit strong community structures, cluster-based layouts based on hierarchical structures have proven useful, such as the treemap layout [41] or space-filling curve based layouts [42], as shown in Fig. 3.3. A treemap defines a hierarchical decomposition of screen space, where the whole screen is recursively subdivided according to the tree, i.e., the root of the tree takes up the whole screen, each branch subdivides the screen at each level of the tree, and finally each leaf of the tree is allotted its own region of the screen. When applied to a graph's clustering hierarchy, each node in the graph is a leaf in the hierarchy, and can thus be placed in the corresponding region to define the layout. In the space-filling curve layout, the nodes are ordered in 1 dimension, and then mapped to the screen using a recursively defined fractal curve, such as the well known Hilbert or Gosper curves. Any such clustering-based layout can provide clear boundaries between communities—particularly when combined with edge bundling techniques. And since a clustering is already computed, hierarchical edge bundling is a good fit [28].



(a) Treemap-based layout

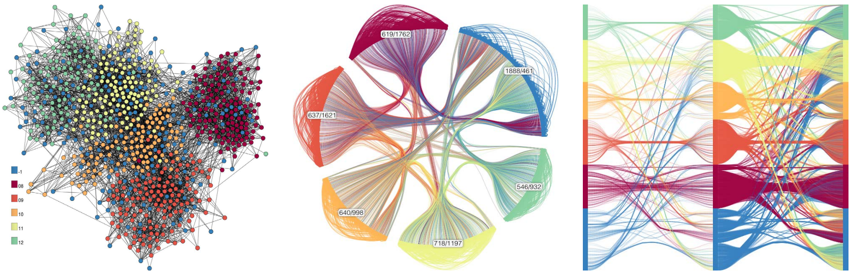


(b) Space filling curve-based layout

Fig. 3.3. Clustering-based graph layouts using trees (a) or space filling curves (b) can be used to show explicit separation between communities

3.3.6 Multivariate Social Networks

Sometimes, merely improving the layout algorithm is insufficient for showing particular aspects of a network. Specifically, social networks can often be divided into groups according to discrete properties besides connectivity, such as gender, race, school grade, or others. However, the density of ties in most traditional node-link diagrams make it difficult to distinguish in inter-group patterns from intra-group patterns, as in Fig. 3.4(a). One approach to address this is a modified radial representation that arranges nodes according to



(a) Traditional layout (b) Radial layout, divided by node category (grade) (c) Parallel layout, with multiple edge sets (aggression vs. friendship)

Fig. 3.4. Some networks can be divided up categorically, with multiple node categories or edge sets. Here we show a social network of students colored by grade. Traditional node-link diagrams (a) can be too cluttered to read, but explicitly dividing nodes by category (b) or showing multiple edge sets in parallel (c) can better show how these categories or edge sets interact. (Images from [12] with permission.)

categorical properties in addition to connectivity, as shown in Fig. 3.4(b). Nodes are placed around a circle, grouped into discrete arcs based on the selected data attribute, and ordered within each group by connectivity with the use of modularity clustering. This new representation also delegates the two kinds of connections to separate regions of space: intra-group edges are displayed outside the circle while inter-group edges are drawn in the middle. The label on each group shows the number of inter-group and intra-group connections, respectively.

In addition to node categories, social networks can also contain more than one kind of edge, defining two or more unique networks on the same set of nodes. In such cases, a layout that is good for one set of edges might not be good for another. Alternately, with one unified layout, sparser networks may get lost inside denser ones. One approach to address this is a representation based on n-partite network layouts, where layers of nodes are laid out parallel to each other, similar to the dynamic graph approach of Burch et al. [7]. This concept has been applied to multiple edge sets on the same set of nodes by replicating the nodes in each layer, and considering each edge set as a bipartite graph from the full set of nodes to a duplicate set of nodes, which creates an n-partite network where n is one more than the number of edge sets. This n-partite network is then laid out in a series of columns by evenly spacing the nodes in each column. Edge directionality is also shown in this representation, since all edges proceed from left to right. While hierarchical layouts such as Sugiyama [53] or Dig-Cola [14] could be used, here each layer of nodes is identical, and thus it is more natural for each column to have the same ordering. Thus, the afore calculated categorical modularity clustering is applied to cluster the nodes, and the resulting clustering is traversed to define a universal ordering. An example of this parallel layout is shown in Fig. 3.4(c).

3.3.7 Dynamic Social Networks

In real world applications, social networks are often intrinsically time-varying: New friendships can be made, or old friendships lost. While the problem of visualizing static networks has been studied quite extensively, work on dynamic network visualization is less mature.

A common method for visualizing dynamic graphs is to animate the transitions between time steps. This approach yields dynamic visualization with nodes appearing, disappearing and moving to produce a readable layout for each time step. Alternatively, multiple time steps can be statically placed next to each other using "Small Multiples" [56]. This eases the comparison of distant time steps but limits the area devoted for each time step which reduces the legibility of each graph. Archambault et al. [1] have done an empirical study to compare the advantages and drawbacks of these approaches (i.e., "Animation" vs. "Small Multiples"). In either case, when creating a node-link diagram for a dynamic graph, not only does the layout need to consider graph topology, but also the stability between time steps. Hu et al. [30] proposed a method based on a geographical metaphor to visualize a summary of clustered dynamic graphs. An alternate visualization approach for dealing with dynamic large directed graphs is to directly represent time as an axis. In the work of Burch et al. [7], vertices are ordered and positioned on several vertical parallel lines, and directed edges connect these vertices from left to right. Each time-step's graph is thus displayed between two consecutive vertical axes.

Storyline visualizations have become popular in recent years for showing evolution of interactions such as clusterings or networks [54]. Sallaberry et al. [51] use a globally optimized dynamic graph clustering approach to both extend the SFC layout method [42] and create a storyline-like timeline representation of the network. An example of such a timeline is shown in Fig. 3.5.

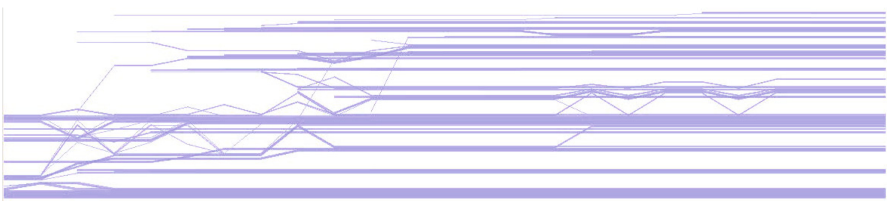


Fig. 3.5. Evolution of a small social network collected off the Rimzu social media site

3.3.8 Egocentric Approaches

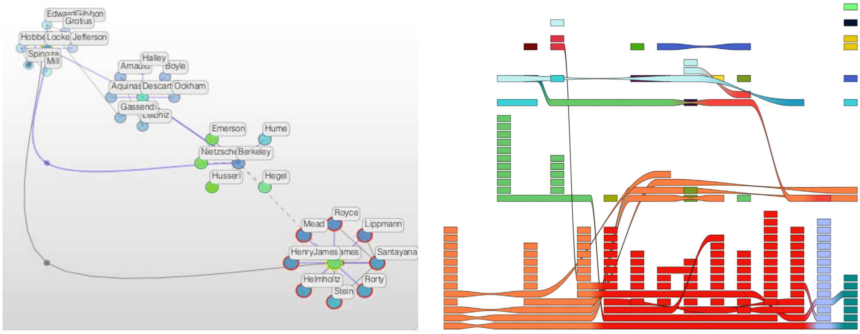
Due to screen and retinal resolution limits, and a psychological limit on attentiveness, there is a finite maximum amount of information that can be conveyed by any one view. Thus, as datasets get even bigger, an overview of the dataset will show proportionally less and less of the underlying data. As a means to address this, researchers have introduced several bottom-up techniques, which bypass or supplement the overview with a detailed view that starts at the lowest level of the data (i.e., a single selected node and its immediate context). Additional relevant nodes and connections are revealed only on demand, based on graph structure or specialized degree-of-interest (DOI) functions.

"Link Sliding" and "Bring & Go" are two such DOI functions for navigating large networks [40]. Heer and Boyd [27] presented a visualization method which only shows a focus node's neighboring nodes up to a certain level. Similarly, Elmqvist and Fekete [16] described a bottom-up system based on hierarchy traversal methods. These methods are useful when the inherent graph structure is more important than other properties for the task at hand. For other applications, where node/edge attributes are the focus of analysis, researchers create specialized DOI functions. Furnas [20] introduced a DOI function to evaluate the importance of a selected node based on distance and a priori interest. Van Ham and Perer [25] extended this function to operate on embedded attributes and graph topology, as well as user-generated search actions. Crnovrsanin et al. [11] combine this concept with an interaction history based importance similar to Amazon's item-to-item collaborative filtering [37]. The result of this is a visual recommendation system that takes into account not only the underlying topology, but also the users' interaction histories. An example of a path in a user's exploration is shown in Fig. 3.6(a).

In dynamic networks, not only will importance depend on the interactive selection of focal points, but also on the temporal history of the network. Muelder et al. [43] have extended the DOI functions for dynamic networks by using computing a DOI that takes into account not only static topology, but also temporal topological history and interaction history. This is then used along with dynamic clustering to create focused, egocentric storylines, as shown in Fig. 3.6(b).

3.3.9 TreeNetViz: Revealing Patterns of Networks with Hierarchical Attributes

This sample technology demonstrates a new visualization technique, TreeNetViz [22], to help users understand a network with hierarchical attribute information. This technology is built up on a TreeNet graph, a type of multivariate network in which node attribute has hierarchical structure. For example, as shown in Fig. 3.7, a subgraph of a scientific co-author network in Fig. 3.7a has node attribute of affiliation, such as country, university and department,



(a) Interactive visual recommendation (b) Storyline recommendation

Fig. 3.6. Using a recommendation system to focus on the neighborhood around a single focal node over time can show dynamic context of changes in that individual's relation to the network. In static networks, this change can be from interactive focal point changes as the user explores (a). In dynamic networks, this change can also be due to the evolution of the network itself over time (b).

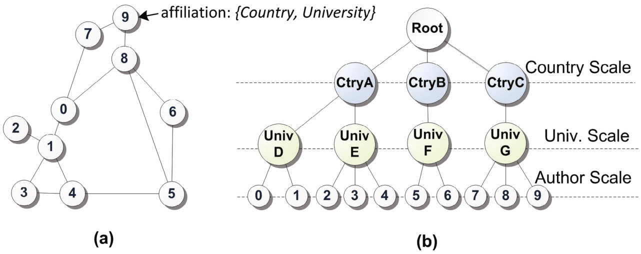


Fig. 3.7. An example of a TreeNet graph. It includes (a) a scientific co-author network, and (b) node affiliation attribute with a hierarchical structure.

and the affiliation attribute has a hierarchical structure shown in Fig. 3.7b. This type of graph, a special type of multivariate network, is called TreeNet graph.

Analysis of this type of network is not a trivial task. It is important to analyze the connectivity, centrality and path patterns at different levels aggregation on the node attribute. For instance, to fully understand the scientific co-author network shown in Fig. 3.7, the collaboration activities can be analyzed through different entities, from individual authors to multiple universities to international collaborations. Th analysis is achieved by aggregating network connections at different levels of node attribute hierarchy. This type of analysis enables us to understand an individual's social activities at different affiliation levels [34].

TreeNetViz Design. TreeNetViz is designed to support various multivariate networks analysis at different levels of node hierarchy for a TreeNet graph.

TreeNetViz uses a Radial, Space-Filling (RSF) [52] technique to show a tree structure of the node attribute in the TreeNet graph (Fig. 3.8a). It then uses a circular layout for an aggregated network and places the aggregated network over the RSF tree (Fig. 3.8b and c). To reduce visual cluttering, it adopts an edge bundling technique based on [29](Fig. 3.8d). It also includes an algorithm to improve circular node placement to reduce the edge crossings with the consideration of various constraints.

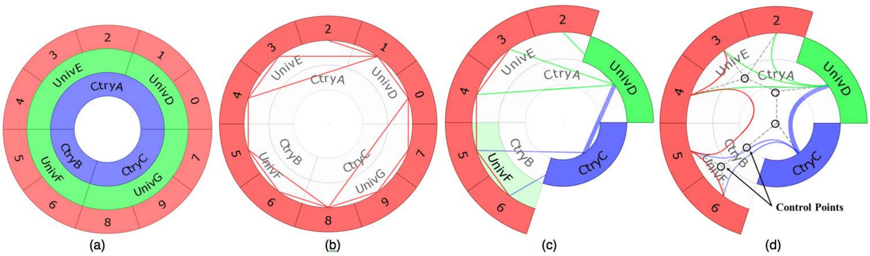


Fig. 3.8. TreeNetViz Visualization Design: (a) a Radial, Space-Filling (RSF) layout of the node attribute structure; (b) the optimized circular layout of the network overlaid on RSF tree; (c) a RSF circular layout of an aggregated network; (d) the view after edge bundling

Treenetviz also includes rich interactions to support network analysis tasks at different levels of aggregation. It enables users to observe network patterns (connectivity, centrality, and reach) among entities of the same type (e.g. the collaboration patterns among all universities or countries in previous example) by controlling the view level. It supports arbitrary aggregation of network by expanding and folding node sector in the visualization. It also enables users find the short paths among nodes of interest in aggregated networks.

An example application of TreeNetViz. A TreeNetViz example is presented to help people understand collaboration patterns among researchers in a co-author network. The collaboration network was extracted from MedLINE research articles published from 2006 to 2010 in the area of diabetes at University of Michigan. The data set includes 614 articles, 847 authors and 2,498 co-author relationships. 10 college-level nodes and 90 department-level nodes are identified.

Fig. 3.9 shows the visualization results of collaboration patterns at three different levels of colleges (Fig. 3.9a), departments (Fig. 3.9b), and individuals (Fig. 3.9c). With this visualization, people can understand network patterns at different scales from the perspectives of the power and status of collaboration resources, and the access control to social groups and individual authors.

TreeNetViz also presents patterns how social actors collaborate with each other from different scales. As shown in Fig. 3.10a, the collaboration patterns of researchers in the “Biochemistry Dept” with other departments in

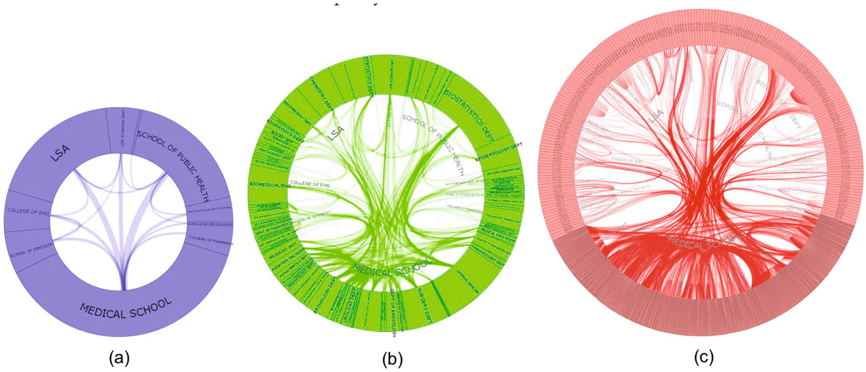


Fig. 3.9. TreeNetViz visualizes collaboration patterns at three different levels: collaborations among colleges(a), departments(b), and individuals(c)

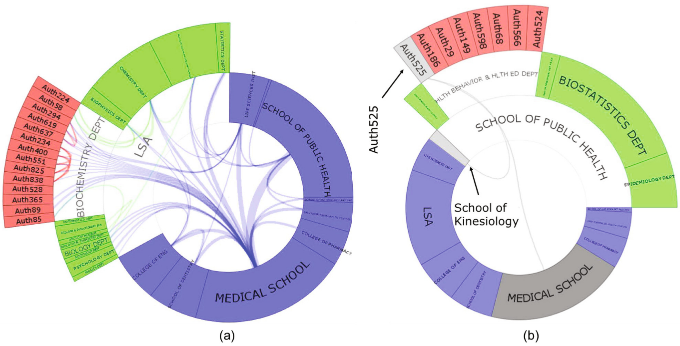


Fig. 3.10. (a) Collaboration patterns across different levels of entities of colleges, departments and individuals; (b) A critical path connecting two colleges by the researcher "Auth525"

"LSA", and other colleges are presented. It is also helpful to identify important people connecting different organizations. Figure 3.10b shows the researcher "Auth525" connects two organizations: "School of Kinesiology" and "Medical School".

3.3.10 SocialNetSense: Making Sense of Multivariate Social Networks

While TreeNetViz is a specific visualization technique to represent and help people to explore a specific type of multivariate network, SocialNetSense [23], on the other hand, is a visual analytics tool to support different analysis tasks on social networks with rich node attributes. SocialNetSense integrates different visualizations of multivariate social networks and supports the analysis process with a sensemaking approach. Here, the social network with rich node attribute information, such as TreeNet graph, is a type of multivariate network of interest.

Sensemaking approach for visual analytics. SocialNetSense adopts a sensemaking approach to support the visual analytics of multivariate social network. Sensemaking is a process to iteratively construct and refine a representation or understanding of data and fit data with the representation to meet the requirements of a task [50]. There are several important sensemaking tasks on social networks with rich node attribute information including understanding network features, the social attribute features and the hybrid features of network and attribute.

Figure 3.11 shows the sensemaking framework for multivariate social network visual analytics. The framework consists of a network exploring loop and a representation building loop. In the network exploring loop, users can explore social attribute features, network features, and hybrid features to collect information based on their tasks and existing knowledge. Various metrics (such as degree, betweenness and closeness), plots (such as degree distribution) and visualization tools are implemented to help users to explore these features. On the other hand, in the representation building loop, users process and comprehend the information collected, build and revise their representation of the data.

The two loops interact with each other with bottom-up and top-down processes. In the top-down process, representations are used to guide users' exploration to look for new evidences. In the bottom-up process, information of interest are collected as evidences to confirm or dis-confirm the representation. As more evidence are collected, representations can be revised and even re-constructed.

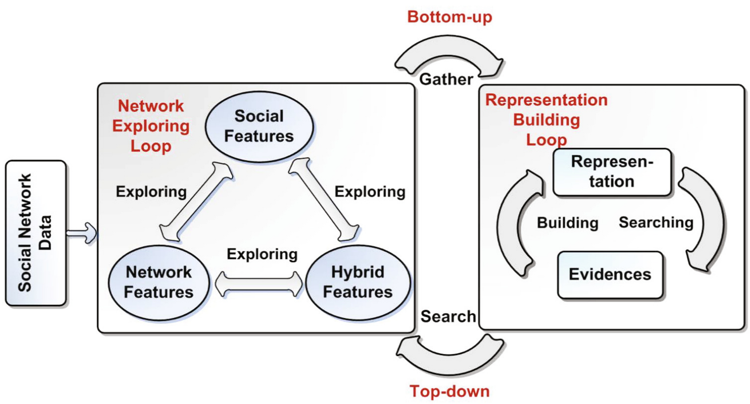


Fig. 3.11. A sensemaking framework for multivariate social network visual analytics in SocialNetSense

SocialNetSense User Interface. Guided by the sensemaking framework, the SocialNetSense user interface includes three main components: a Network Exploring Space (NES), a Representation Building Space (RBS), and a process view.

The interface of NES is shown in Fig. 3.12, including two main panels: a visualization view (Panel 1) showing social networks along with hierarchical social structures, and a control panel (Panel 2) offering a set of analytical tools, such as different aggregation metrics of connectivity and centrality, analytic plots and searching function. View manipulation tools, such as zooming, panning and layouting, are provided in the tool bar above Panel 1. It uses multiple visual representations of networks to facilitate the exploration of social, network and hybrid features. Node-link diagrams (Panel 1 in Fig. 3.12) are used to show social features and network features, and TreeNetViz is used to show hybrid features of aggregated networks over node attributes.

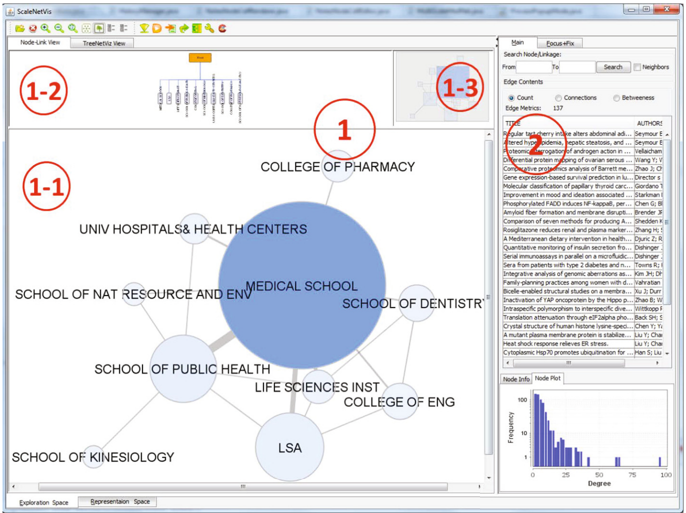


Fig. 3.12. Network Exploring Space (NES) in SocialNetSense: (a) Coordinated node-link views, including a network visualization (Panel 1-1), a tree visualization (Panel 1-2) for social hierarchy, a network overview panel (Panel 1-3), and a control panel (Panel 2) with analytical tools

In RBS, users can organize the evidence collected from the NES to create their representations using editing functions. Figure 3.13 shows the user interface of RBS. The interface consists of a editing space (Panel 1), a process view (Panel 2) and an element list view (Panel 3). The editing space enables users to collect visualization elements from the node-link view and the TreeNetViz view, network metrics of size, centrality, betweenness and closeness, and also plots from the NES. It also provides functions such as grouping/ungrouping, note-taking, and element-linking to build representation.

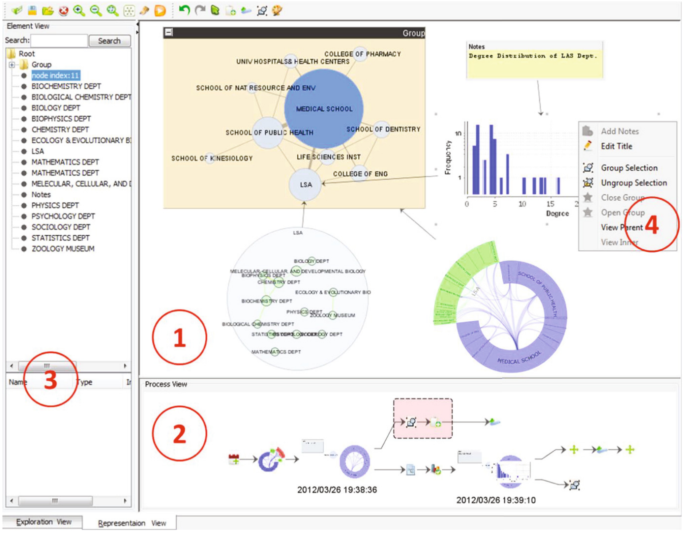


Fig. 3.13. Representation Building Space (RBS): Panel 1 is the main working space; Panel 2 is the process view; Panel 3 lists the elements in the working space; Panel 4 shows tools of representation building, such as add note, group/ungroup elements and link element

Sample analysis with SocialNetSense. The same data set of a diabetes researcher collaboration network is used to demonstrate how SocialNetSense supports the visual analytics of multivariate social networks.

With SocialNetSense, users can build their understanding of the collaboration patterns, such as the power and status of social actors and collaborations, at three different levels (colleges, departments, and individuals). In Fig. 3.14, an example representation is shown for the main network patterns at the level of colleges composed by a user. The strong collaboration among "Medical School", "LSA" (Literature, Science and the Arts) and "Public Health" is captured with detailed notation of network metrics, plots, and co-authored articles. Similarly, it can also help users understand cross-scale patterns such as collaboration among the departments in "LSA" with other colleges.

With SocialNetSense, users can have comprehensive understanding of the analytics process. Figure 3.15 shows how a user makes sense of the network to identify an important actor (Author 525) acting as a "boundary spanner" to connect "Medical School (MS)" and "School of Public Health (SPH)". Compared with the visualization result shown in Fig. 3.10b, SocialNetSense shows the intermediate results and reasoning process.

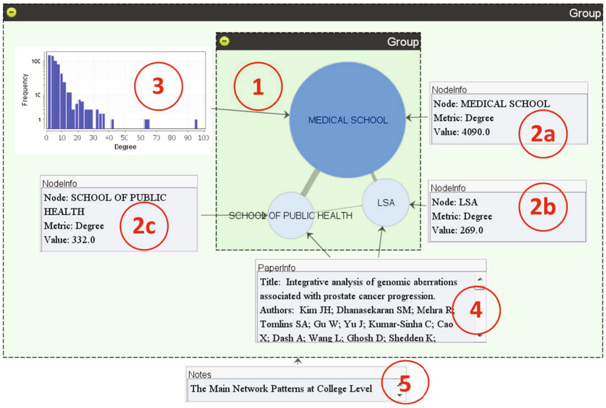


Fig. 3.14. A sample representation of the main network patterns at the level of college

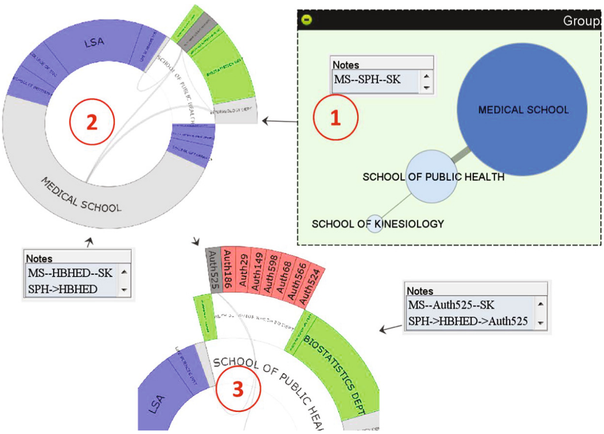


Fig. 3.15. A representation of a boundary spanner connecting "Medical School (MS)" and "School of Public Health (SPH)"

3.3.11 Summary

Many works have proposed methods to accomplish these tasks. Dyadic and triadic analyses often rely on basic statistics, but other metrics require more complex algorithms. There are a number of centrality metrics, each with their own strengths and weaknesses, and a number of additional metrics derived from centrality metrics. Community-scale analyses depend on clusters, so there are also a substantial number of clustering algorithms.

As node-link diagrams are traditional for visually inspecting social networks, node-link diagram layout algorithms are intrinsically applicable. But social networks exhibit certain properties that make many layout algorithms less useful. So, there are additional approaches to improve layout algorithms for social networks, based on relevant statistical analyses such as centralities or clusterings. Some visual approaches even incorporate semantic information, such as node categories [22] or multiple edge sets. There has also been recent work in extending visual analyses to dynamic social networks. And finally, as the size of social networks available to researchers has grown incredibly in recent years, bottom-up visual analytic approaches, such as recommendation and sensemaking-based systems [23], are becoming increasingly popular.

3.4 Challenges and Future Directions

While much work has been done on the visualization and analysis of social networks, many of the key challenges are only getting more important. The size of social network data available has exploded in recent years due to social media, and continues to grow every year. Many of these networks generate complex data in real-time, and real-time analysis offers many unique opportunities and challenges. The kind of information in social networks can also be quite varied, such as social network messages from different devices, different locations, and different social media sites, and may contain various meta-data that could potentially improve analytic results. The validity of the information must also be considered, as most social networks rely on the honesty of the users, and are potentially vulnerable to wildly inaccurate input, missing data, or even spam. And lastly, there is much that can be done to improve the analytic insights to be gained from previous data that has already been collected.

As such, there are numerous opportunities for future work in this area. As there are many social media sites, finding new ways of combining and analyzing networks from various sources would be beneficial to creating a more complete picture of the underlying social trends. Producing useful analytic results while preserving the privacy of the subjects is also important, as many users would be more willing to provide accurate information if they trust the privacy policies. But the resulting networks will still have uncertainty, whether due to the data being sanitized to protect privacy or the users omitting data to protect their own privacy. So incorporating further uncertainty metrics to either measure the validity of the input data or conjecture missing information would aid in improving the analytic results. And lastly, even when a visual analytic process produces an insight, it is important for the analyst to convey the underlying derivation, as tracking and analyzing the provenance of insights is critical for improving the analytic process.

References

1. Archambault, D., Purchase, H.C., Pinaud, B.: Animation, small multiples, and the effect of mental map preservation in dynamic graphs. IEEE Transactions on Visualization and Computer Graphics 17(4), 539–552 (2011)
2. Bertin, J.: Semiology of graphics. University of Wisconsin Press (1983)
3. Blondel, V., Guillaume, J., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008(10), P10008 (2008)
4. Brandes, U., Indlekofer, N., Mader, M.: Visualization methods for longitudinal social networks and stochastic actor-oriented modeling. Social Networks 34(3), 291–308 (2011)
5. Brandes, U., Pich, C.: An experimental study on distance-based graph drawing. In: Tollis, I.G., Patrignani, M. (eds.) GD 2008. LNCS, vol. 5417, pp. 218–229. Springer, Heidelberg (2009)
6. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 30(1-7), 107–117 (1998), `http://linkinghub.elsevier.com/retrieve/pii/S016975529800110X`
7. Burch, M., Vehlow, C., Beck, F., Diehl, S., Weiskopf, D.: Parallel edge splatting for scalable dynamic graph visualization. IEEE Transactions on Visualization and Computer Graphics 17(12), 2344–2353 (2011)
8. Clauset, A., Newman, M.E.J., , Moore, C.: Finding community structure in very large networks. Physical Review E, 1–6 (2004), `http://www.ece.unm.edu/ifis/papers/community-moore.pdf`
9. Correa, C.D., Crnovrsanin, T., Ma, K.L.: Visual reasoning about social networks using centrality sensitivity. IEEE Transactions on Visualization & Computer Graphics 18(1), 106–120 (2012)
10. Cox, T., Cox, M.: Multidimensional Scaling. Chapman & Hall, London (2001)
11. Crnovrsanin, T., Liao, I., Wuy, Y., Ma, K.L.: Visual recommendations for network navigation. In: Proceedings of the 13th Eurographics / IEEE – VGTC conference on Visualization, EuroVis 2011, pp. 1081–1090. Eurographics Association, Aire-la-Ville (2011), `http://dx.doi.org/10.1111/j.1467-8659.2011.01957.x`
12. Crnovrsanin, T., Muelder, C.W., Faris, R., Felmle, D., Ma, K.L.: Visualization of friendship and aggression networks (2012), `http://vidi.cs.ucdavis.edu/projects/AggressionNetworks/`, CNN's AC360 study: Schoolyard bullies not just preying on the weak
13. Demoll, B.S., Mcfarland, D.: The Art and Science of Dynamic Network Visualization. JoSS: Journal of Social Structure 7 (2005), `http://www.cmu.edu/joss/content/articles/volume7/deMollMcFarland/`
14. Dwyer, T., Koren, Y.: Dig-cola: Directed graph layout through constrained energy minimization. In: IEEE Symposium on Information Visualization, pp. 65–72 (2005)
15. Eades, P.: A Heuristic for Graph Drawing. Congressus Numerantium 42, 149–160 (1984)
16. Elmqvist, N., Fekete, J.D.: Hierarchical Aggregation for Information Visualization: Overview, Techniques, and Design Guidelines. IEEE TVCG 16(3), 439–454 (2009)
17. Faust, K.: Triadic configurations in limited choice sociometric networks: Empirical and theoretical results. Social Networks 30, 273–282 (2008)

18. Freeman, L.: Centrality in social networks conceptual clarification. Social Networks 1(3), 215–239 (1979)
19. Freeman, L.C.: The Development of Social Network Analysis: A Study in the Sociology of Science. Booksurge (2004)
20. Furnas, G.W.: Generalized fisheye views. In: Human Factors in Computing Systems CHI, pp. 16–23 (1986)
21. Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting personality from twitter. In: Proc. SocialCom 2011, pp. 149–156 (2011)
22. Gou, L., Zhang, X.: TreeNetViz: revealing patterns of networks over tree structures. IEEE Transactions on Visualization and Computer Graphics 17(12), 2449–2458 (2011)
23. Gou, L., Zhang, X., Luo, A., Anderson, P.: SocialNetSense: supporting sensemaking of social and structural features in networks with interactive visualization. In: 2012 IEEE Conference on Visual Analytics Science and Technology (VAST 2012), pp. 133–142 (2012)
24. Hachul, S., Jünger, M.: An experimental comparison of fast algorithms for drawing general large graphs. In: Healy, P., Nikolov, N.S. (eds.) GD 2005. LNCS, vol. 3843, pp. 235–250. Springer, Heidelberg (2006)
25. van Ham, F., Perer, A.: Search, Show Context, Expand on Demand: Supporting Large Graph Exploration with Degree-of-Interest. IEEE TVCG 15(6), 953–960 (2009)
26. Hasan, M.A., Zaki, M.J.: A survey of link prediction in social networks. In: Aggarwal, C.C. (ed.) Social Network Data Analytics, pp. 243–275. Springer US (2011)
27. Heer, J., Boyd, D.: Vizster: visualizing online social networks. In: IEEE Symposium on Information Visualization, pp. 32–39 (2005)
28. Holten, D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. IEEE Transactions on Visualization and Computer Graphics 12(5), 741–748 (2006)
29. Holten, D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. IEEE Trans. Vis. Comput. Graph. 12(5), 741–748 (2006)
30. Hu, Y., Kobourov, S.G., Veeramoni, S.: Embedding, clustering and coloring for dynamic maps. In: Proceedings of the 5th IEEE Pacific Visualization Symposium, pp. 33–40 (2012)
31. Huang, M.L., Nguyen, Q.V.: A fast algorithm for balanced graph clustering. In: Proceedings of the 2007 IEEE Symposium on Information Visualization (InfoVis), pp. 46–52 (2007)
32. Jacob, R., Koschützki, D., Lehmann, K., Peeters, L., Tenfelde-Podehl, D.: Algorithms for centrality indices. In: Brandes, U., Erlebach, T. (eds.) Network Analysis. LNCS, vol. 3418, pp. 62–82. Springer, Heidelberg (2005)
33. Kamada, T., Kawai, S.: An algorithm for drawing general undirected graphs. Inf. Process. Lett. 31(1), 7–15 (1989)
34. Kilduff, M., Tsai, W.: Social Networks and Organizations. SAGE (September 2003)
35. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. Journal of the ACM 46(5), 604–632 (1999), `http://portal.acm.org/citation.cfm?doid=324133.324140`
36. Langevin, D.G.S., Schretlen, P., Jonker, D., Bozowsky, N., Wright, W.: Louvain clustering for big data graph visual analytics (2013), poster at VIS 2013

37. Linden, G., Smith, B., York, J.: Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing 7, 76–80 (2003)
38. Lister, R.: After the gold rush: toward sustainable scholarship in computing. In: Simon, M., Hamilton (eds.) Tenth Australasian Computing Education Conference (ACE 2008). CRPIT, vol. 78, pp. 3–18. ACS, Wollongong (2008)
39. Mahmud, J., Zhou, M., Megiddo, N., Nichols, J., Drews, C.: Recommending targeted strangers from whom to solicit information on social media. In: Proc. IUI 2013, pp. 37–48 (2013)
40. Moscovich, T., Chevalier, F., Henry, N., Pietriga, E., Fekete, J.-D.: Topology-Aware Navigation in Large Networks. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 2319–2328 (2009), `http://hal.inria.fr/inria-00373679`
41. Muelder, C., Ma, K.L.: A treemap based method for rapid layout of large graphs. In: Proceedings of the IEEE Pacific Visualization Symposium (PacificVis 2008), pp. 231–238 (2008)
42. Muelder, C., Ma, K.L.: Rapid graph layout using space filling curves. IEEE Transactions on Visualization and Computer Graphics 14(6), 1301–1308 (2008)
43. Muelder, C.W., Crnovrsanin, T., Ma, K.L.: Egocentric storylines for visual analysis of large dynamic graphs. In: Proceedings of 1st IEEE Workshop on Big Data Visualization (BigDataVis 2013), pp. 56–62 (October 2013)
44. Newman, M.E.J.: The Structure and Function of Complex Networks. SIAM Review 45(2), 167–256 (2003)
45. Noack, A.: Modularity clustering is force-directed layout. CoRR abs/0807.4052 (2008)
46. Pal, A., Wang, F., Zhou, M., Nichols, J., Smith, B.: Question routing to user communities. In: CIKM 2013 (to appear, 2013)
47. Pennacchiotti, M., Popescu, A.M.: A machine learning approach to twitter user classification. In: ICWSM (2011)
48. Qian, T., Li, Q., Liu, B., Xiong, H., Srivastava, J., Sheu, P.: Topic formation and development: a core-group evolving process. In: WWW 2013, pp. 1–31 (2013)
49. Rivera, M.T., Soderstrom, S.B., Uzzi, B.: Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. Annual Review of Sociology 36, 91–115 (2010)
50. Russell, D.M., Stefik, M.J., Pirolli, P., Card, S.K.: The cost structure of sensemaking. In: Proceedings of the INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems, CHI 1993, pp. 269–276. ACM, New York (1993), `http://doi.acm.org/10.1145/169059.169209`
51. Sallaberry, A., Muelder, C., Ma, K.-L.: Clustering, visualizing, and navigating for large dynamic graphs. In: Didimo, W., Patrignani, M. (eds.) GD 2012. LNCS, vol. 7704, pp. 487–498. Springer, Heidelberg (2013)
52. Stasko, J., Zhang, E.: Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: IEEE Symposium on Information Visualization, InfoVis 2000. pp. 57–65 (2000)
53. Sugiyama, K., Tagawa, S., Toda, M.: Methods for visual understanding of hierarchical systems. IEEE Trans. Systems, Man, and Cybernetics 11, 109–125 (1981)
54. Tanahashi, Y., Ma, K.L.: Design considerations for optimizing storyline visualizations. IEEE TVCG 18(12), 2679–2688 (2012)

55. Tollis, I.G., Di Battista, G., Eades, P., Tamassia, R.: Graph Drawing: Algorithms for the Visualization of Graphs. Prentice Hall (July 1999)
56. Tufte, E.R.: Envisionning Information. Graphics Press (1990)
57. White, S., Smyth, P.: Algorithms for estimating relative importance in networks. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD, pp. 266–275 (2003)
58. Zhao, S., Zhou, M., Zhang, X., Yuan, Q., Zheng, W., Fu, R.: Who is doing what and when: Social map-based recommendation for content-centric social web sites. ACM TIST 3(1), 5–25 (2011)
59. Zhou, M., Zhang, W., Smith, B., Varga, E., Farias, M., Badenes, H.: Finding someone in my social directory whom i do not fully remember or barely know. In: Proc. ACM IUI 2012, pp. 203–206 (2012)