Document Visualization Overview and Detail by Selectively Varying Density

Richard H. Fowler
Department of Computer Science, University of Texas – Pan American, Edinburg, TX, USA

Abstract - *The challenges of network data display are among the most longstanding in visualization. Current approaches typically rely on some form of network abstraction to reduce network density, or "clutter", to reveal structural elements salient for the task at hand. Any reduction of complexity must consider the competing goals of providing visual representations that reveal overall, global relations, as well as detailed, local relations. The present work addresses this issue by utilizing the Pathfinder network representation, which allows systematic variation in density, to provide differing, but formally complementary, visual representations. These representations can provide both global context and local detail by selectively increasing structural complexity in a way that overcomes some of the limitations of other techniques. The paper's examples are drawn from the domain of document relations, though the techniques are broadly applicable.*

Keywords: visualization, networks, information visualization, Pathfinder networks, document relations, document maps

1 Introduction

A key element of visualization is the ability to provide users techniques for reducing complexity in data sets in order to reveal underlying structure. For scientific visualization this reduction may be in support of tasks ranging from hypothesis formation to interpretation of statistical results. For information visualization the range is no less broad, and typically includes the use of visual representations in the performance of an externally oriented task.

Visual representations of document sets have received particular attention. This is in large part due to the widespread electronic availability of documents through public Internet search engines, as well as digital libraries. Because documents are textual it is not possible to directly provide a visual display, especially of relations among documents. To provide a visual representation of document relations it is necessary to use documents' textual form to create a quantitative form suitable for visual display and interaction.

Among the most widely studied representations for visualization of document relations are networks [25]. This is due to the power of graph representations and because graphs are generally well understood computationally. Network visualization techniques are among the earliest examined, e.g., SemNet for representing semantic units [13], the Information Visualizer, which introduced cone trees and other seminal visualization techniques [3], and the Hyperbolic Browser, which uses mapping to non-Euclidean space to enhance display effectiveness [28].

The current work examines the use of Pathfinder networks (PfNets) [36] as a mechanism to reduce complexity in visual representation, while preserving salient structural elements. The work extends PfNets' use to provide both global context for document examination, as well as local detail, by selectively increasing structural complexity in areas of interest. The paper's examples are drawn from the domain of document relations, though the techniques are broadly applicable.

2 Document Network Visualization

A fundamental challenge of network data visualization is reducing the "clutter", or link density, of the network in order to provide a visual representation that allows users to extract salient structural elements relevant to the task at hand. Network visualization should supply a view that reveals overall, global structure, as well as representations that allow the user to examine individual elements in detail. Providing a global view suggests a sparse network that is visually presented in a manner that facilitates extraction of global arrangement. Providing mechanisms for inspection of individual elements or neighborhoods in detail suggests that not only should that element be placed globally in the context of the larger network, as is provided by a global view, but that further representations be provided to reveal more detailed relations than provided by a sparse graph. Clearly, the two goals compete, and the usual solution has been to emphasize global relations using mechanisms to create relatively sparse graphs at the expense of visual representations that reveal more detailed local relations among elements, as would be provided by a more dense network.

2.1 Density Reduction

The visual display of dense networks is particularly challenging. Though force directed and spring embedded graph drawing methods [12, 27] work reasonably well in providing aesthetically pleasing and comprehensible layouts for sparse graphs, they are much less effective for more dense graphs. Thus, link reduction becomes a primary goal, and a number of techniques exist for accomplishing this.

One link reduction technique creates a link weight threshold whereby only links with weights above the threshold are included [41]. The threshold technique has the advantage that it is straightforward to implement and efficient. However, the link inclusion threshold, which is typically based on a single value for all nodes, does not take into account any metric of global network structure, and so the derived network may not retain essential structural properties of the original. Other link reduction techniques that have been used include minimum spanning trees and Pathfinder networks.

Minimum spanning trees (MSTs) have the advantage for link reduction that they contain the fewest number of links possible (n-1) to form a connected graph. Also, in contrast to threshold graphs MSTs consider global structure, as reflected in graphs' connectivity, in the reduction of links, rather than just single link values, as does the threshold technique. This element of structure preservation is important in providing a representation likely to convey salient information after transformation to a visual representation.

2.2 Pathfinder Networks

The Pathfinder algorithm was developed to model semantic memory in humans and to provide a paradigm for scaling psychological similarity data [35]. A number of psychological and design studies have compared PfNets with other scaling techniques and found that they provide a useful tool for revealing conceptual structure [29].

PfNets can be conceptualized as path length limited minimum cost networks. Algorithms to derive minimum cost spanning trees (MCSTs) have only the constraints that the network is connected and cost, as measured by the sum of link weights, is a minimum. For PfNets, an additional constraint is added: not only must the graph be connected and minimum cost, but also the longest path length to connect node pairs, as measured by number of links, is less than some criterion. To derive a PfNet direct distances between each pair of nodes are compared with indirect distances, and a direct link between two nodes is included in the PfNet unless the data contain a shorter path satisfying the constraint of maximum path length. Since its initial development [36] there have been continued efforts in algorithm development, e.g., parallelizing implementation [23], and recent work has reduced time complexity for some cases to cubic [22, 32, 33].

In constructing a PfNet two parameters are incorporated: r determines path weight according to the Minkowski r-metric and q specifies the maximum number of edges considered in finding a minimum cost path between nodes. The weight of a path with k links is determined by weights w_1, w_2, ..., w_k of each individual link:

$$W(P) = \left[\sum_{i=1}^{k} w_i^r\right]^{\frac{1}{r}} \quad (1)$$

The q-parameter specifies that triangle inequalities must be satisfied for paths with $k \leq q$ links:

$$w_{n_i n_k} = \left[\sum_{i=1}^{k} w_{n_i n_k}^r\right]^{\frac{1}{r}} \forall\, k \leq q \quad (2)$$

As either r or q is manipulated, edges in a less complex network form a subset of the edges in a more complex network. Thus, the algorithm generates two families of networks, controlled by r and q. The least complex network is obtained with $r = \infty$ and $q = n$-1, where n is the total number of nodes in the network. In this work the containment property has provided a particularly useful technique for systematically varying network density to provide both relatively sparse networks (the PfNet derived with $r = \infty$ and $q = n\text{-}1$ is the union of MCSTs) with for global navigation, as well as more dense networks for local inspection.

2.3 Document Network Visualizations using Pathfinder Networks

PfNets have been used in a variety of document visualization systems, primarily as a technique to create sparse network representations effective in providing global views of document network structure. Previous uses of PfNets in document relation visualization have focused on the use of the most sparse network PfNet ($r = \infty$, n-1). AuthorLink and ConceptLink are interactive author co-citation analysis maps based on PfNets and self organizing maps [39]. The Pathtrieve [14], Information Navigator [15], and Document Explorer [16] systems provide PfNet based visual representations of document networks for information retrieval. Other examples of PfNet use include StarWalker for social navigation [9], author co-citation analysis [7] and visualizations of knowledge domains [8].

Generalized Similarity Analysis [5, 6] is a framework for structuring and visualizing distributed information resources in hypertext systems. It is based on PfNets to organize its structural modeling and visualization facilities, including representations of inter document relations, inter author similarity, as well as interrelationships between hypertext documents based on a combination of hypertext connectivity, word occurrence based similarity, and traversal based similarity.

3 Varying Display Density and Structural Detail using Pathfinder Networks

Exploration and utilization of large visual information spaces requires both global context and local detail. Many techniques are available for providing this context plus focus, including fisheye views [17, 19] and distortion techniques generally [10]. Such techniques are widely used in information visualization and typically employ a spatial viewing transformation in which elements close to the focus are magnified and those farther away are made to appear relatively smaller. Yet, other approaches to providing detail in context manipulate not only spatial size as a function of distance, but also which elements are to be displayed. For example, in Furnas' original account of fisheye views [17] both an element's distance from the view focus and the element's "a priori importance" are used to calculate a degree of interest (doi) value for each element. The doi is calculated as a function of distance and a priori importance, and only if the value exceeds some threshold is the element displayed. Hence, important items, as indicated by value of a priori importance, are visible at longer distances from the focus than are items of lower a priori importance. Furnas details the value of providing more than just spatial distortion in overview plus context viewing in his later account of information display [18].

This work presents a method for providing overview of global network structure together with detail of local relations that overcomes some of the shortcomings of techniques based solely on spatial distortion. This is accomplished *by selectively increasing structural complexity in areas of interest*. The technique exploits the inclusion property of PfNets, whereby in the family of PfNets that can be created for a dataset by varying r or q links in a less dense network are included in the more dense network. For example, the links of PfNet ($r = \infty$, $q = n\text{-}1$) are a subset of the links of PfNet ($r = 2$, $q = n\text{-}1$). As detailed above, previous work has found the most sparse PfNet, PfNet ($r = \infty$, $q = n$-1), effective in providing network display that conveys overall structure using well known display techniques.

The work reported below extends past visualizations using PfNets by creating a hybrid PfNet representation that combines a sparse PfNet with a second, denser PfNet. This creates a representation that is effective in displaying both overall, global relations, as well as detailed, local relations.

3.1 Creating Document Networks

The typical stages in creating a visual representation for a document collection entail, first, the transformation of text documents to a form suitable for quantitative analysis. This quantitative representation is then used to create a representation, e.g., tree, network, suitable for spatial display. The final stage is the mapping of the representation to the display and its visual presentation, e.g., a graph layout algorithm.

For this work the transformation of document text to a quantitative representation uses the Vector Space Model [34]. This technique utilizes a document's words to transform each document to a vector representation that captures the document's content. Comparisons among documents are then employed to provide inter document distances used to derive the PfNets upon which visual representations are based.

3.2 Quantitative Representation of Documents

To use the Vector Space Model a set of indexing terms is defined for a particular document collection. These terms can come from a fixed vocabulary or be derived for individual document sets, as is the case in this work. The text of each document is analyzed, and a vector representing each document is created in which the presence or absence of each indexing terms is noted. In a space of dimensionality equal to the number of index terms, each document can be considered as a single point.

Having created a quantitative representation for individual documents, it is then necessary to consider relations among documents. In this work inter document similarity uses the number of terms common to a document pair as represented by their document vectors, normalized by number of terms in documents. In order to increase the discriminative power of terms, terms are first weighted by the inverse frequence, the $tf \times idf$ model [38]. In equation 3 below, tf is a term's frequency in the document collection, idf is the inverse document frequency, and tf_{ik} is the number of occurrences of term T_k in D_i. N is the number of documents in the collection and n_k represents the number of documents containing term T_k. Again, each document is represented by an array of terms T and each term is associated with a weight determined by the $tf \times idf$ model. The weight of term T_k in document D_i, is:

$$w_{ik} = \frac{tf_{ik} \times log\left(\frac{N}{n_k}\right)}{\sqrt{\sum_{j=1}^{Tn} \left(tf_{if}\right)^2 \times log\left(\frac{N}{n_j}\right)^2}} \qquad (3)$$

Finally, to calculate inter document similarity in the system used as input to the Pathfinder algorithm document similarity is computed based on corresponding document vectors for documents i and j, $D_i = (w_{i1}, w_{i2}, \ldots, w_{iT})$ and $D_j = (w_{j1}, w_{j2}, \ldots, w_{jT})$:

$$sim_{ij} = \sum_{k=1}^{T} w_{ik} \times w_{jk} \qquad (4)$$

Reduction of vector space dimensionality, necessary to create a representation than can be realized visually, is considered next. Among the approaches used for dimensionality reduction in visualization [1] are eigenvalue/eigenvector decomposition [11], factor analysis and principal component analysis [31, 40], multidimensional scaling [4, 37, 26], latent semantic analysis [2, 5], self-organizing maps [29, 30], and various network abstractions, including Pathfinder networks, which are used here.

3.3 Sparse Pathfinder Network Representation

The work examined document sets created using Citeseer [21]. Pathfinder networks of varying densities were created using these datasets with results similar to Chen [9]. For PfNet ($r = \infty$, $q = n\text{-}1$) node-link ratios (n nodes / n links) varied from .92 to .97. The prefuse system [24] is used to display the PfNets.

Figure 1 shows the least dense PfNet ($r = \infty$, $q = n$-1) for the document set described here. The organization shows the characteristic "hubs" and "stars" reported by Chen [9] and others and yields a representation well suited to providing a global, readily comprehended view of the document set. The network has a node–link ratio of .97.

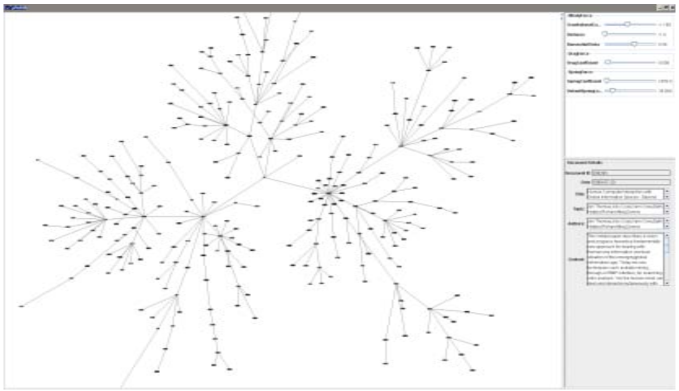

Figure 1. PfNet ($r = \infty$, $q = n$-1). Node-link density is .97. Overall document set relations are readily evident.

3.4 Dense Pathfinder Network Representations

The primary focus was on the structural and visual representations that result from deriving more dense PfNets by varying Pathfinder *r* and *q* metrics, Minkowski metric and path length, respectively. As noted, PfNet ($r = \infty$, $q = n\text{-}1$) is the limiting case with respect to the minimum number of links. It is the union of MCSTs, and for most data sets the number of nodes approaches, *n*-1, the minimum number of links required to connect, or provide a path between, all nodes.

Though the sparse PfNet ($r = \infty$, $q = n\text{-}1$) reveals global structure well, the data from which the network was derived contains much additional information about document relations that can be used to convey more information about local relations than is represented in this sparse network. For example, Figure 1 shows the characteristic hub and spoke structure of document collections, i.e., a set of "central" documents are directly connected to a number of other documents. These are analogous to cluster centroids, and so single link examinations are of utility in looking at similar documents. Though this structure reveals relations of a document to a single topic, or cluster of documents, well, a document also has relations to other topics, as represented by other hub documents, and these relations are not reflected in this sparse structure.

It is a straightforward matter to "enrich" network structure to show a wider range of relations by deriving a denser network in which more links are present. Using the Citeseer data sets, we examined denser networks created with Pathfinder. The limiting case for the maximum number of links is PfNet (r = 1, q = 1) in which the network is the complete graph. The number of links can be systematically varied to the other limiting case of PfNet ($r = \infty$, $q = n\text{-}1$), the union of MCSTs, by varying *r* and *q*.

Unfortunately, the visual representation becomes "cluttered" for even moderately dense networks with node-link ratios of 0.5, twice as many links as nodes. Figure 2 below shows a PfNet ($r < \infty$, $q < n\text{-}1$) for the document data set that has a node-link ratio of .53. The dispay is cluttered, and, except for nodes on the periphery, little information about structure is evident. Certainly, global structure is not portrayed in a way comparable to the sparse network of Figure 1. In isolation, this more dense network has limited utility for visualization.

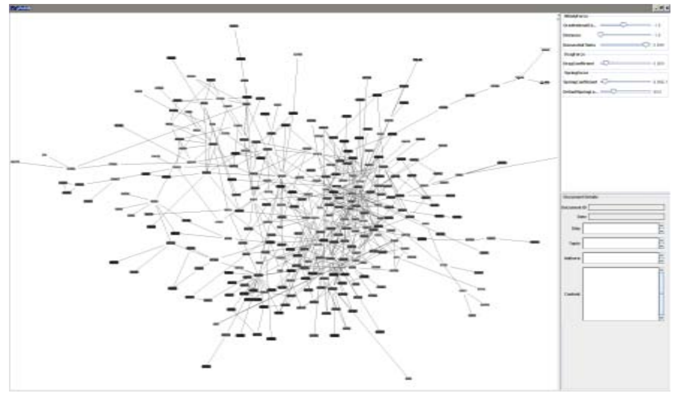

Figure 2. PfNet ($r < \infty$, $q < n\text{-}1$). Node-link density is .53. In isolation there is limited utility for this visualization.

Additionally, even though PfNet ($r = \infty$, $q = n$-1) is included in all PfNets ($r < \infty$, $q < n\text{-}1$), and the same node positioning algorithm used for both graphs, node locations are different when nodes in each PfNet are positioned, as detailed below. Hence, it is not possible to simply remove links from the denser layout to create a structure that effectively reveals global structure, as in Figure 1.

3.5 Hybrid Pathfinder Network Representation for Global Structure and Local Detail

The work's primary focus is the structural and visual representations that result from combining sparse and more dense PfNets by varying Pathfinder *r* and *q* metrics. In particular, the work has sought to create a representation that selectively increases structural complexity in areas of interest, yet retains the ability to convey overall structure.

Clearly, the spatial arrangement of nodes and their edge connectivity in the sparse PfNet ($r = \infty$, $q = n$-1) provides a different and more comprehensible view of the data set than the more dense PfNet ($r < \infty$, $q < n\text{-}1$). Nonetheless, the more detailed view of document relations provided by the denser network can be of use. Figure 3 below is centered on a node in the more dense network that might be of interest to a user wanting to learn more about the relations of a particular document to others in the document collection.

As noted, it is not possible to simply remove nodes from the more dense PfNet to create a visually effective structure such as created with PfNet ($r = \infty$, $q = n$-1). Even though PfNet ($r = \infty$, $q = n$-1) is included in PfNet ($r < \infty$, $q < n$-1) and the same node positioning algorithm used for both graphs, the location of nodes will differ when nodes in each are positioned by a force directed algorithm. The difference in node positioning is due to a force directed algorithm's

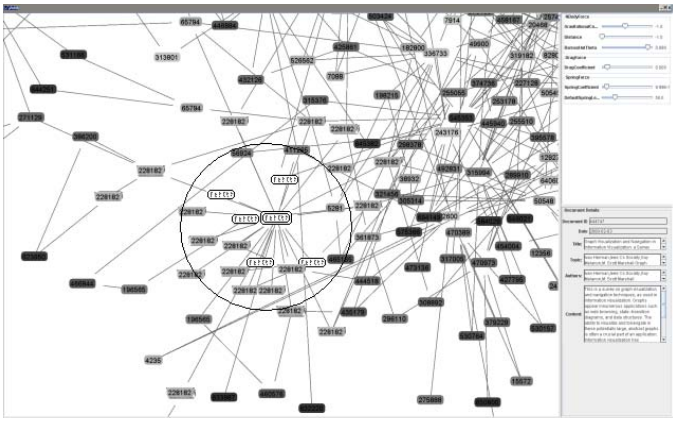

Figure 3. Expanded view of dense PfNet shown in Figure 2. The node of interest and documents directly connected in this PfNet, but not directly connected in PfNet ($r = \infty$, $q = n$-1), are shown as white. They are circled in the figure to assist in distinguishing without color.

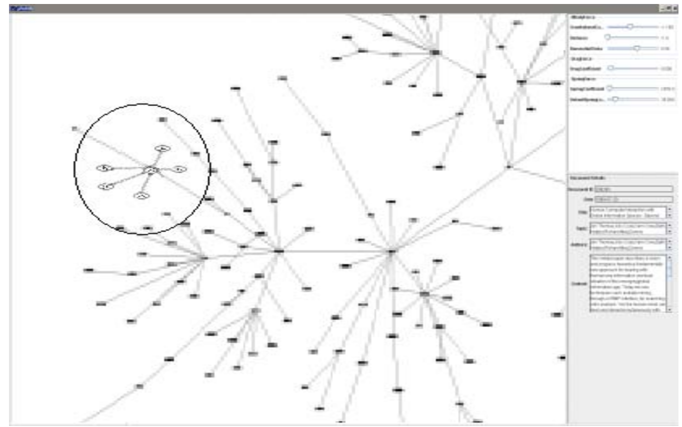

Figure 4. Hybrid network selectively increasing structural complexity in area of interest. A subgraph of the dense PfNet is embedded in PfNet ($r = \infty$, $q = n$-1). In this way local relations from the dense network are placed in the context of the complete sparse network. Only paths of length 1 are shown here, though any length can be used.

consideration of the additional links in the denser network and its derivation of a layout that considers optimality utilizing all links. Because of the consideration of all links in determining node placement, removing links from the denser layout to create a structure that effectively reveals global structure, as in Figure 1, is not effective. The interaction of layout algorithm and structure is among the challenges of network visualization, and no universal solution is available.

The approach this work takes is to create a hybrid representation that exploits the visual efficacy of PfNet ($r = \infty$, $q = n$-1) in revealing overall structure together with the structural property of the Pathfinder algorithm in deriving denser networks that include less dense networks. This inclusion property guarantees that subgraphs of less dense networks can be found in more dense networks. Hence, it is possible within constraints to use nodes of less dense networks to place the links of a subgraph of a more dense in that less dense network.

Figure 4 illustrates such a hybrid network constructed for the data set used in the other examples. The subgraph centered on the node of interest of the dense PfNet in Figure 3 has been placed in the sparse PfNet ($r = \infty$, $q = n$-1). Due to the Pathfinder inclusion property, the subgraph can be appropriately positioned and connectivity preserved in creating this hybrid network. This is in contrast to many other density reduction techniques, such as link thresholds, which do not take into account overall structure as does the path based approach of Pathfinder networks, as most clearly seen in the inclusion property.

A further advantage of the hybrid approach of combining networks of differing densities that is not illustrated in the figure is the approach's ability to further increase density beyond a particular starting point. In the example, only nodes that are directly connected, i.e., have a path length of 1, to the node of interest are illustrated. The hybrid technique can also be used to further expand the area of dense connectivity beyond those immediately connected nodes, e.g., to nodes that are connected to the node of interest by path lengths of 2 or 3. This provides a larger area of dense connectivity and so more relations of the node of interest with other documents. This mechanism of increasing path length from a particular node creates a mechanism for measured exploration of the entire data set from the perspective of the node of interest.

4 Conclusions

The current work has utilized Pathfinder networks to create a novel visual representation of document networks that both reveals global structure and provides local detail. As such, it addresses longstanding issues of network display. Selectively increasing structural complexity in areas of interest through hybrid PfNets exploits the Pathfinder algorithm's ability to create networks of varying density and allows the seamless combination of networks of differing density.

5 Acknowledgements

This work was supported by Department of Education grant P116Z0202159 and the University of Texas – Pan American Computing and Information Technology Center. Thanks to Kyle Picou and Yavuz Tor for programming efforts in the use of prefuse.

6 References

[1] Borner, K., Chen, C., & Boyack, K.W. (2003). Visualizing knowledge domains. In Cronin, B. (Ed.), *Annual Review of Information Science and Technology*, 37, 179-255.

[2] Börner, K., Dillon, A., & Dolinsky, M. (2000). *LVis - Digital Library Visualizer.* In *Proceedings of Information Visualisation 2000, Symposium on Digital Libraries*, 77-81.

[3] Card, S. K., Robertson, G. G. & Mackinlay, J. D. (1991). The information visualizer : An information workspace. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CH1 '91)*, 181-188.

[4] Chalmers, M. (1992). *BEAD: Explorations in information visualisation.* In *Proceedings of ACM Special Interest Group in Information Retrieval International Conference (SIGIR '92)*, 330-337.

[5] Chen, C. (1997). Tracking latent domain structures: An integration of pathfinder and latent semantic analysis. *AI & Society*, 11(1-2), 48-62.

[6] Chen, C. (1998). Bridging the gap: The use of pathfinder networks in visual navigation. *Journal of Visual Languages and Computing*, 9(3), 267-286.

[7] Chen, C. (1999). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing and Management*, 35(2), 401-420.

[8] Chen, C. & Paul, R. J. (2001). Visualizing a knowledge domain's intellectual structure. *Computer*, 34(3), 65-71.

[9] Chen, C. & Morris, S. (2003). Visualizing evolving networks: Minimum spanning trees versus pathfinder networks. In *Proceedings of IEEE Symposium of Information Visualization*, 67-74.

[10] Cockburn, A., Karlson, A., & Bederson, B. (2008). A review of overview+detail, zooming, and focus+context. *Computing Surveys*, 41(1), 2:1-2:31.

[11] Davidson, G. S., Hendrickson, B., Johnson, D. K., Meyers, C. E., & Wylie, B. N. (1998). Knowledge mining with VxInsight: Discovery through interaction. *Journal of Intelligent Information Systems*, 11(3), 259-285.

[12] Eades, P. (1984). A heuristic for graph drawing. *Congressus Numerantium*, 42, 149-160.

[13] Fairchild, K., Poltrock, S., & Furnas, G. (1988). SemNet: Three dimensional graphic representations of large knowledge bases. In *Cognitive Science and its Applications for Human-Computer Interaction*, Guidon, R. (Ed.), New York: Lawrence Erlbaum Associates, 201-233.

[14] Fowler, R. H. & Dearholt, D. W. (1990). Information retrieval using pathfinder networks. In *Pathfinder associative networks: Studies in knowledge organization,* Schvaneveldt, R. W. (Ed.), Norwood, NJ: Ablex, 179-196.

[15] Fowler, R. H., Fowler, W. A. L. & Wilson, B. A. (1991). Integrating query, thesaurus, and documents through a common visual representation. In *Proceedings of ACM Special Interest Group in Information Retrieval International Conference (SIGIR '91)*, 142-151.

[16] Fowler, R. H., Fowler, W. A. L., & Williams, J. L. (1997). 3D visualization of WWW semantic content for browsing and query formulation. *Journal of Network and Computer Applications*, 20, 135-149.

[17] Furnas, G. (1986). Generalized fisheye views. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '86)*, 16–23.

[18] Furnas, G. (1997). Effective view navigation. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*, 367–374.

[19] Furnas, G. (2006). A fisheye follow-up: Further reflections on focus+context. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, 999–1008.

[20] G. Salton (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. New York: Addison-Wesley.

[21] Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). CiteSeer: An automatic citation indexing system. *ACM Digital Libraries*, 89-98.

[22] Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M. E., Crisostomo, R. G., & de Moya-Anegon, F. (2006). Binary pathfinder: An improvement to the pathfinder algorithm, *Information Processing and Management*, 42, 1484-1490.

[23] Hauguel, S., Zhai, C. & Han, J. (2009). Parallel pathfinder algorithms for mining structures from graphs. In *Proceeding of IEEE International Conference on Data Mining*, 812-817.

[24] Heer, J., Card, S. K., & Landay, J. A. (2005). prefuse: A toolkit for interactive information visualization. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, 421-430.

[25] Herman, I., Melancon, G., & Marshall, M. S. (2000). Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1), 24-43.

[26] Hetzler, B., Whitney, P., Martucci, L., & Thomas, J. (1998). Multi-faceted insight through interoperable visual information analysis paradigms. In *Proceeding of IEEE Symposium on Information Visualization*, 137-144.

[27] Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7-15.

[28] Lamping, J. & Rao, R. (1996). The hyperbolic browser: A focus plus context technique for visualizing large

hierarchies. *Journal of Visual Languages and Computing*, 7(1), 33-55.

[29] Lin, X. (1997). Map displays for information retrieval. *Journal of the American Society for Information Science,* 48(1), 40-54.

[30] Lin, X., Soergel, D., & Marchionini, G. (1991). A self-organizing semantic map for information retrieval. In *Proceedings of ACM Special Interest Group in Information Retrieval International Conference (SIGIR '91),* 262-269.

[31] McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science,* 41(6), 433-443.

[32] Quirin, A., Cordon, O., Guerrero-Bote, V. P., Vargas-Quesada, B., & Moya-Anegon, F. (2008). A quick MST-based algorithm to obtain pathfinder networks (∞, *n*-1), *Journal of the American Society for Information Science and Technology*, 59(12), 1912-1924.

[33] Quirin, A., Cordon, O., Santamaria, J., Vargas-Quesada, B., & Moya-Anegon, F. (2007). A new variant of the pathfinder algorithm to generate large visual science maps in cubic time. *Information Processing and Management,* 44(4), 1397-1409.

[34] Salton, G., Yang, C., & Wong, A. (1975). A vector space model for automatic indexing. *Communications of the ACM,* 18(11), 613-620.

[35] Schvaneveldt, R. W. (Ed.) (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex.

[36] Schvaneveldt, R. W., Dearholt, D. W., & Durso, F. T. (1989). Network structures in proximity data. *The psychology of learning and motivation: Advances in research and theory*, Bower, G. H. (Ed.), New York: Academic Press, 249–284.

[37] Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science,* 50(9), 799-813.

[38] Sparck-Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation,* 28(1), 11–20.

[39] White, H. D., Buzydlowsky, J., & Xia, L. (2000). Co-cited author maps as interfaces to digital libraries. In *Proceedings of IEEE Information Visualization Conference*, 25-30.

[40] Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V. (1995). Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceeding of IEEE Symposium on Information Visualization*, 51-58.

[41] Zizi, M. & Beaudouin-Lafomn, S. (1994). Accessing hyperdocuments through interactive dynamic maps. In *Proceeding of European Conference on Hypertext (ECHT '94),*126-135.