

# Pathfinder Network Scaling

## Description

Pathfinder Network Scaling is a structural modeling technique originally developed for the analysis of proximity data in psychology by Schvaneveldt, Durso & Dearholt (1998). Pathfinder algorithms take estimates of the proximity's between pairs of items as input and define a network representation of the items that preserves only the most important links. The resulting Pathfinder network (PFNET) consists of the items as nodes and a set of links (which may be either directed or undirected for symmetrical or non symmetrical proximity estimates) connecting pairs of the nodes.

Graphical representations of Pathfinder networks are generated using force directed graph drawing algorithms.

## Pros & Cons

The value of Pathfinder Network Scaling for IV lies in its ability to reduce the number of links in meaningful ways, often resulting in a concise representation of clarified proximity patterns. According to Chen, 1999, p. 408, Pathfinder Network Scaling provides "a fuller representation of the salient semantic structures than minimal spanning trees, but also a more accurate representation of local structures than multidimensional scaling techniques."

## Applications

Pathfinder Network Scaling is commonly used to visualize complex (semantic) structures. For example, research by Chen (1999) aims to support people explore and navigate intuitively in a semantically organized information space. To generate the figure below, he selected 169 long papers from the ACM SIGCHI Conference Proceedings (1995-1997), and all papers from the ACM Hypertext Conference Proceedings (1987-1998).

pathf-chen.jpg A document-document similarity matrix was generated via Latent Semantic Analysis. Each document was indexed by its title, authors' affiliations, abstract, and list of keywords. Pathfinder Network Scaling was used to extract underlying patterns in the similarity matrix and to present them spatially in a class of 'pathfinder networks' (Chen & Carr, 1999). The approach was implemented in the StarWalker a system that displays citation networks as a set union of all possible minimum spanning trees. Visually, important papers (according to a citation threshold) are represented by spheres. Links between two spheres indicate that those two authors have been cited together to an extent determined by the author citation analysis. Each sphere is linked to the original full text version in ACM's electronic proceedings on the WWW and can be downloaded and displayed for detailed study.

## Details

Pathfinder network scaling relies on the so-called triangle inequality to eliminate redundant or counter-intuitive links. Given two links or paths in a network that connect two nodes, the link/path is preserved that has a greater weight defined via the Minkowski metric. It is assumed that the link/path with the greater weight better captures the interrelationship between the two nodes and that the alternative link/path with less weight is redundant or even counter-intuitive and should be pruned from the network.

Two parameters  $r$  and  $q$  influence the topology of a pathfinder network. The  $r$ -parameter influences the weight of a path based on the Minkowski metric. The  $q$ -parameter defines the number of links in alternative paths (=length of a path) up to which the triangle inequality must be maintained. A network of  $N$  nodes can have a maximum path length of  $q=N-1$ . With  $q=N-1$  the triangle inequality is maintained throughout the entire network. For details on the method and its applications see (Schvaneveldt, 1990).

## Usage Hints

The Pathfinder Network Scaling code is part of the IVC Software Framework. It accepts network data in dense matrix format, e.g., a similarity matrix generated using LSA. Networks generated via one of the many network modeling algorithms need to be converted into matrix format before they can be analyzed via PFNET. Simply use 'Converters > Graph <--> Matrix'.

For example, you can read in .... and then use 'Analysis > PFNET' to prune your network. Hover with your mouse over the question marks to receive guidance on what  $r$  and  $q$  parameters are valid. Examine your network before and after pruning to see the effect of PFNET and experiment with different parameter settings.

Pathfinder Network Scaling can also be conducted using the commercial Knowledge Network Organizing Tool PCKNOT software package. See also Technical Information on KNOT and Pathfinder. It was originally designed to run under DOS and it is still running under DOS.

## References

- Schvaneveldt, R. (Ed.) (1990). Pathfinder Associative Networks: Studies in Knowledge Organization. Norwood, NJ: Ablex.
- Chen, C. (2004) Searching for intellectual turning points: Progressive knowledge domain visualization. Proceedings of the National Academy of Sciences of the United States of America (PNAS), 101 (Suppl. 1), 5303-5310.
- Chen C. (1999) Visualizing semantic spaces and author co-citation networks in digital libraries. Information Processing and Management 35(3), 401-420.
- Chen, C., & Carr, L. (1999) Trailblazing the literature of hypertext: Author co-citation analysis (1989-1998). In Proceedings of the 10th ACM Conference on Hypertext, pp. 51-60.
- KNOT tools for Pathfinder Network Analysis <http://www.geocities.com/interlinkinc/home.html>
- Many more Pathfinder references.
- KU-Mapper by Roy Clariana for collecting PFnet proximity data in 3 different ways.