# PURSUING THE WORK OF JACQUES BERTIN

Nathalie Henry Riche, Microsoft Research, WA 98052 Redmond, U.S.A.
E-mail: nath@microsoft.com

Submitted: <leave for Editor to date>

**Abstract**

Representing networks is a major interest of researchers in information visualization. In this article, I present novel visualization techniques based on Jacques Bertin's *reorderable matrices*.

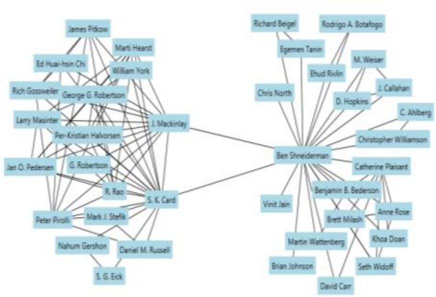
As the human brain is particularly effective at processing visual information, researchers in computer science developed a number of visual exploration systems to analyze graphs and networks. In the last five years, massive online social networks became available for scientists to analyze, pushing researchers to innovate in the field of graph visualization. In this article, I present recent techniques designed to visualize large and dense social networks based on Jacques Bertin's *reorderable matrices*.

## Node-Link Diagrams

Node-link diagrams are the most commonly used representation of graphs and networks. It is well illustrated by Freeman in his survey and history of social network visualization [1]. In this article, Freeman presents a wide variety of social networks and demonstrates that visual representations are a powerful tool to illustrate social network analysis concepts such as central actors or communities. Figure 1 demonstrates how this visual representation can depict two communities as dense groups of nodes and links; and place the actors bridging them in the center of the diagram.

Node-link representations are widely used and familiar to a very large audience, making them a powerful communication tool. However, their readability and the message they convey greatly depends on the positions of their nodes.

**Fig. 1. Node-Link diagram of a social network. Collaboration between researchers.**




| | Beef | Horse | Goat | Porc | Veal |
|---|---|---|---|---|---|
| Belgium | 6 | 6 | 2 | 5 | 3 |
| Germany | 32 | 6 | 5 | 45 | 15 |
| France | 37 | 61 | 69 | 29 | 59 |
| Italy | 16 | 23 | 21 | 9 | 14 |
| Nederland | 9 | 4 | 3 | 12 | 9 |

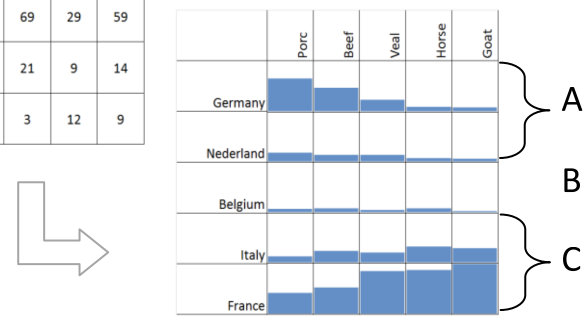



**Fig. 2. Numerical table of the production of 5 types of meat by 5 countries (left) and visual matrix with rows and columns reordered (right).**

Since the 90s, an entire field of research is devoted to the problem of graph drawing, i.e. generating algorithms to place nodes in the space [2]. Scaling to large networks with several thousand or even millions of nodes remains a challenge. Node-link diagrams with more than a few hundred nodes often become an undistinguishable pile of nodes and links, difficult to transform either automatically or manually into a readable representation. A common solution consists in filtering data. However, when removing nodes and links, the topological structure of the network may be damaged and properties lost. For these reasons, researchers such as Jacques Bertin investigated an alternative representation: the matrix.

## Matrix Representations

In the Semiology of graphics [3], Jaques Bertin shows that replacing numerical values by visual indicators and reordering rows and columns dramatically improve the readability of tables and matrices. Figure 2 shows an example of the *reorderable matrix*, representing the production of 5 types of meat in 5 countries. While the numerical table makes it possible to read any value, it remains difficult to grasp higher-level insights about this data. However, once values are transformed into graphical indicators and rows and columns reordered, one can discover more insights about the data. For example, an analyst can quickly spot three different profiles of production (marked A, B and C in the Figure) and identify that France has the highest production overall.

A critical element of the readability of this matrix is its ordering. Reordered matrices provide opportunities to discover more insights about the data than unordered ones. Many reordering algorithms exist ranging from graph linearization algorithms in graph theory to microarray clustering algorithms in bioinformatics. While empirical evidence shows that different reordering algorithms make different patterns emerge in the data, assessing their quality remains an open question.

## Graphs as Matrices

In a matrix representation of a graph, vertices are placed both in rows and columns. If two vertices are connected, the corresponding cell of the table is marked. Since vertices are represented both in rows and columns, two cells correspond to a pair of vertices, making it possible to represent directed edges. This type of matrix is an *adjacency matrix*.

Adjacency matrices are particularly useful for dense or large networks. Its representation does not degrade as drastically in visual clutter (e.g. edge crossing) and remains efficient to display on screen. Indeed, the cells representing the links do not cross by overlapping each other and the time to draw the representation is low since the whole list of actors is placed linearly. However, *matrices use a large amount of space* since they represent all possible pairs of vertices. A user study [4] also showed that *path-related tasks* (e.g. find a shortest path between two vertices) are difficult. My research explored the design of hybrid representations and novel interaction techniques to alleviate these drawbacks [5].

## Beyond Node-Link and Matrix

Far less compact than node-link diagrams, matrices can be made more usable with effective navigation techniques. For example, when exploring large matrices, it is essential to be able to read labels of rows and columns while identifying interesting connection patterns. This is why in Mélange [6], we provide a technique for splitting the view while retaining context. Figure 3 shows an example of the Mélange technique, folding the visualization as one would fold paper. Further inspired by online map applications such as BingMaps, we created a zoomable matrix explorer [7], capable of handling much larger datasets. Navigation through multiple aggregation levels makes the analysis of 6 million Wikipedia pages possible.

Ghoniem et al. [4] compared the advantages and drawbacks of matrices over node-link diagrams for simple readability tasks. In particular, they reported that node-link diagrams performed better for small and sparse networks, while matrices performed better for dense ones. Since matrices do not suffer from edge crossing or node overlap, they are more readable for dense networks. However it has been shown that they poorly support path-following tasks. To alleviate this latter shortcoming, we designed MatLink [8], which augments a standard matrix with links between its labels (Figure 3). These links, appearing on demand, provide a dual encoding of the connections between actors which proved effective in user experiments.

We can conclude that node-link diagrams are more suited for representing sparse networks whereas matrices work best for denser ones. However, it is difficult to decide which representation is best suited for *small-world networks*, since these networks have a globally sparse structure with locally dense communities.

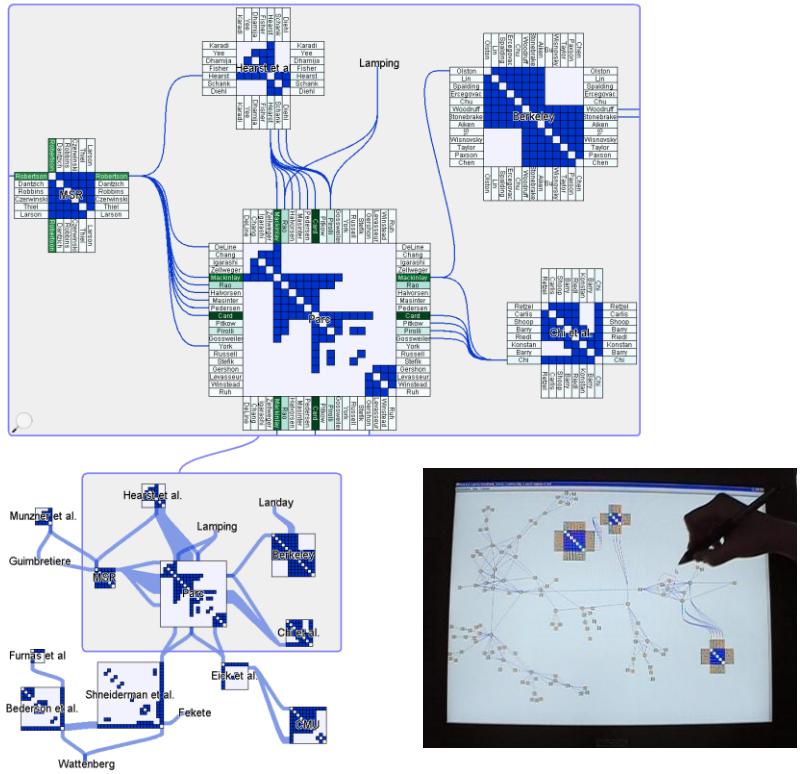


**Fig. 4. NodeTrix, a hybrid representation visualizing the global structure as a node-link diagram and dense areas of the network as matrices.**

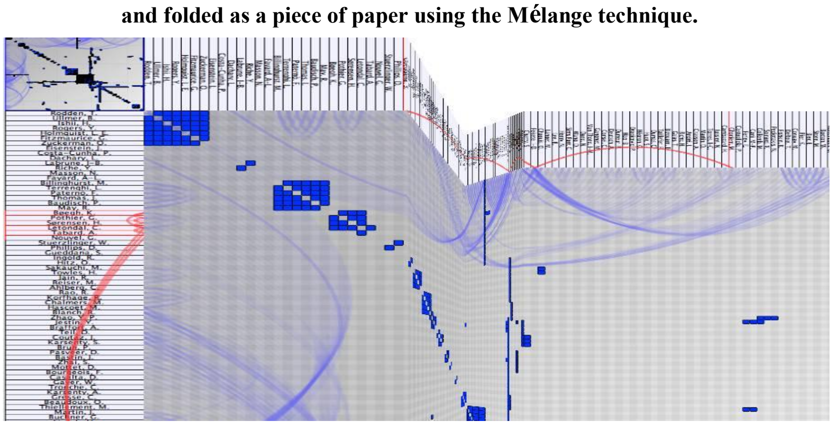
To solve this problem, we invented NodeTrix [9], a hybrid visualization merging node-link diagrams and matrices. NodeTrix represents the global structure of networks using a node-link diagram and allows dense subparts to be represented as matrices. Figure 4 presents an example illustrating how matrices can be created interactively using an interactive pen display. This representation allies best of both worlds and can prove more compact than node-link diagrams for large datasets.

## Conclusion

Following the steps of Jacques Bertin, I believe that interactive matrix-based representations are a promising solution for analyzing large and dense networks. While matrices require a learning phase to identify meaningful patterns, they offer unique opportunities to think differently about the data.

Amongst the many possible directions for future research, creating matrix-based representations that can support multiple types of relations and depict the evolution of networks over time are certainly the most exciting.

**Fig. 3. A matrix representation augmented with links to ease path-following tasks and folded as a piece of paper using the Mélange technique.**



### References and Notes


**1.** L. Freeman, "Visualizing social networks", *Journal of Social Structure* (2000).

**2.** G. Di Battista, P. Eades, R. Tamassia, I. Tollis, *Graph Drawing: Algorithms for the Visualization of Graphs* (Prentice Hall PTR, 1998).

**3.** J. Bertin, *The Semiology of Graphics* (Editions de l'Ecole des Hautes Etudes, 1967).

**4.** M. Ghoniem, J-D Fekete, and P. Castagliola, "On the Readability of Graphs using Node-link and Matrix-based Representations", *Information Visualization Journal* **4,** No 2 (2005)

**5.** N. Henry, *Exploring Social Networks using Matrix-based Representations*, Ph.D. dissertation, Univ. of Paris-Sud, Univ of Sydney (2008)

**6.** N. Elmqvist, N. Henry, Y. Riche and J-D. Fekete, "Melange: Space-folding for Multi-focus Interaction", *Proceedings of SIGCHI'08* (2008)

**7.** N. Elmqvist, T-N, Do, H. Goodell, N. Henry and J-D Fekete, "ZAME: Interactive Large-Scale Graph Visualization", *Proceedings of PacificVis'08* (2008)

**8.** N. Henry and J-D Fekete, "MatLink: Enhanced Matrix Visualziation for Analyzing Social Networks", *Proceedings of INTERACT '07*(2007)

**9.** N. Henry, J-D. Fekete and M. McGuffin, "NodeTrix: Hybrid Representation for Analyzing Social Networks", *IEEE TVCG* **13**, No 6 (2007)