

# COMPARATIVE VISUAL ANALYSIS OF “CARTOGRAPHY” USING NETWORK-BASED AND CONTENT-BASED APPROACHES

**André Skupin**

Department of Geography  
San Diego State University  
San Diego, CA 92182-4493  
USA

skupin@mail.sdsu.edu

**Wyson J. Pang**

Department of Geography  
San Diego State University  
San Diego, CA 92182-4493  
USA

wpang@rohan.sdsu.edu

## ABSTRACT

Visual analytics relies on a combination of multiple computational and human interface approaches, which frequently evolved in separate fields and in different application contexts. Intelligence analysis of documents may combine text processing methods developed in information science with geographically informed clustering methods and present findings using interfaces rigorously tested in terms of cognitive performance. However, there still exist a number of distinct methods which have not been observed in a comparative setting. That is for example the case in the visual analysis of knowledge domains, which is increasingly relying on various geographic metaphors in trying to make sense of the volume of intellectual output generated across the sciences. It aims at giving users visual access to the content and structures of the documents and communities through which different sciences communicate their intellectual progress. This approach has been identified as useful to a wide array of applications, from science policy making to drug discovery, where it has already saved pharmaceutical companies millions of dollars in research expenditures.

In this quest, the structure and development of knowledge domains, such as biology or anthropology, are typically analyzed based on network-type information extracted from documents and respective metadata. For example, citation networks can be derived based on papers citing other papers. Such networks are thought to reflect the transfer and transformation of knowledge. Similar processes are captured by authorship networks, which are based on overlapping co-authorship of documents. Quantitative analysis of citation networks is the basis of numerical measures of the impact of different documents and authors and several examples of visual display of these networks likewise exist. Such analysis is almost always based on large bibliometric databases, especially those produced by the *Institute for Scientific Information* (ISI), such as the database underlying the *Web of Science*. On the other hand, content-based approaches have been proposed that visually depict structures and relationships among text documents based their content, for example using vector-space models and such dimensionality reduction techniques as self-organizing maps. However, these have traditionally not been applied to identical source data and a more objective judgment of their relative value within a visual analytics framework is thus difficult.

In this paper it is demonstrated that bibliometric data could be visualized using both *network structures* and *text content*, and with geographic technologies playing an integrative role. Specifically, the experiment described here involves first performing a keyword search for “cartography” within a database containing several million journal articles. Matching records are extracted and two visualizations are derived. First, the network of citations is visualized to derive major intellectual structures and the dominant authors and publications. Second, the abstracts of all retrieved documents are transformed into a spatialization reflecting semantic structures among the documents. Finally, the two visualizations are juxtaposed in a quest to identify relative strengths and weaknesses, and future research directions, including the integration of spatial and non-spatial semantic components, are discussed.