Article

Study of dynamics of structured knowledge: Qualitative analysis of different mapping approaches

Veslava Osinska
Institute of Information Science and Book Studies
Nicolaus Copernicus University
Torun, Poland

Piotr Bala
Computer Science Modeling Center
Warsaw University, Poland

Abstract
The authors compared three methods of mapping, considering the maps as a visual interface for exploration of scientific articles related to computer science. Data were classified according to the original CCS classification and co-categories were used for similarity metrics calculation. The authors' approach based on MDS was enriched by algorithm mapping to spherical topology. Three other methods were based on VOS, VxOrd and SOM mapping techniques. Visualisation of the classified collection was made for three different decades. Tracking the changes in visualisation patterns, the authors sought which method would reveal the real evolution of the CCS scheme which is still being updated by the editorial board. Comparative analysis is based on qualitative methods. Changes of those properties over two decades were evaluated for the benefit of the authors' method of mapping. The qualitative analysis shows clustering of proper categories and overlapping of other ones in the authors' approach, which corresponds to the current changes of the classification scheme and computer science literature.

Keywords
Information visualisation, knowledge domain mapping, visualisation evaluation, qualitative analysis.

1. Introduction

In the representation of a large collection of data regardless of their origin, specificity and format, practical knowledge about modern methods of information visualisation (Infovis) is indispensable. A variety of mapping and clustering techniques provide 2D or 3D graphical representations which can be used for the study of science or particular science domain(s) – knowledge domain visualisation [1]. Such research is oriented towards seeking the main research fields within a domain, analysing the essential structural changes over the time as well as collaborative network within and between academic societie.

Using a series of chronologically sequential maps, it is possible to study how knowledge advances and how knowledge organization change. E. Garfield first introduced the term longitudinal mapping to describe this method of domain analysis [3]. He emphasized that longitudinal maps become forecasting tools because main trends can be detected through observing changes in time. By means of software package HistCite, Garfield analyzed chronological maps of subject (topical) collections – historiographs, resulting from bibliographic database [3]. In order to recognize and track essential changes in scientific landscape, researchers use animation techniques for sequence of static images - snapshots [4] or online fluently animated networks [5].

This paper is a continuation of the study on visualisation and visual interface of classified articles using a new spherical topology. The presented approach is compared with traditional mapping techniques (VxOrd, VOS and SOM) in terms of quality of output visualisation. Authors constructed an interactive interface for exploring scientific articles. Historical changes in computer science literature organization can be analyzed due to longitudinal maps, generated for three time periods. The general objective is to compare different mapping approaches with emphasis on structural and dynamic content. Original classification is permanently updated by editors, what can be the key in visualisation evaluation. Finally, the series of maps which depict the real changes of primary classification scheme indicate the most effective mapping technique in this case.

In a parallel work the authors argue that a sphere surface is the most convenient shape for users' perception and cognition [6]. The authors focused on the way of how different groups of users, including experts can observe and identify the changes and evolution in a particular field of computer science.

2. Mapping challenges

2.1. Analysis

Science mapping is rapidly developed research field which gave a new (perceptual) perspective in the study of science evolution. It has emerged from established disciplines such as: scientometrics, computer science, statistics and network science and incorporates varying techniques and methods from relatively young research areas including: information visualisation and human computer interaction. Such landscapes display analysis units and semantic relations between them by visual and spatial metaphors. The overall objective of these representation is to deliver more user insight into science development and history of human activity. As described Ch. Chen:

> …visualisations enable us to augment our perceptual and cognitive abilities in exploring complex patterns more easily and more vividly than what can be done in traditional science studies [4]

Science maps are symbolic representations of measured objects, which are positioned so that other ones with close or similar features are located in their vicinity, while those object that are dissimilar are positioned at distant locations. These distances at the distribution are main characteristics that affect the user can read and understand complex pattern of data. But subjective reading and processing information can be accompanied by perceptual biases [7]. The possible problem mentioned by H. Small and E. Garfield [8], concerning geographical interpretation of maps. When we project to a plane high dimensional networks of associations (links), weakly related objects (nodes) can be placed close together. This can evoke mistakes in ultimate interpretation. Another difficulty in visual analysis of maps concerns measurement of dynamical characteristics. Lingitudinal mapping method requires collecting bibliographic data at equal time intervals. Scientific landscape is changing due to data gain, but final distribution is not always continuous. Researchers look for critical changes in patterns and expect there scientific paradigms or discovery like events. The main problem is how to measure dynamics by mean of several static representations. Are the spatial differences between maps can be considered as progressive changes of the system under study? Infovis scientists harvest data for many time series (publication years), what allows to use qualitative methods (generally statistical) to evaluate dynamics. L. Leydesdorff formulates this problem as a methodological gap between multi-variate and time-series analysis [9]. He proposed to construct the "bridge" by using entropy statistics and recent developments in multi-dimensional scaling.

2.2. Algorithms

The most popular technique in information visualisation, particularly in scientific domain mapping, is multidimensional scaling (MDS), which is a way to "rearrange" objects in efficient manner, so as to arrive at a configuration that best approximates the observed distances. It actually moves objects around in the space defined by a requested number of dimensions and checks how well the distances between the objects can be reproduced by a new configuration. It is based on minimization algorithm that evaluates different configurations with the goal of maximizing the goodness-of-fit. MDS has been widely applied for constructing knowledge maps of authors, articles, journals, and keywords [1, 10]. Such visualisation maps can be used as information retrieval interface [11].

Network models are commonly used for representing similarities between large-scale data [12, 13]. An intuitive approach to network layout is to model a graph as a physical system. Nodes are considered as charged particles which repel each other. The links between them act as dampened springs (or weighted arcs) that pull related nodes together. This algorithm is called force-directed and it is a fundamental concept for understanding the structure of a general undirected graph. A common feature of the algorithm or its modifications is balancing between attraction and repulsion force.

The VxOrd algorithm improves a traditional force-directed approach by employing barrier jumping to avoid trapping of clusters in local minima and a density grid to model repulsive forces [14]. Another advantage of this ordination algorithm is that it determines the number and size of clusters automatically based on the data. These properties allow VxOrd to be used on large graphs consisting of even millions of nodes [15]. Data described by many properties are multivariate and need to be displayed on a two-dimensional surface of either paper or computer screen. The basic problem of visualisation – dimension reduction into a 2D or 3D space – is solved by ordination techniques, the aim of which is to place similar objects in n-dimensions so that they are close to each other and dissimilar objects are far apart.

To reveal the underlying social or structural organization of the bibliographic data, other graph based techniques are used. One of them - pathfinder networks (PFNETs) – is based on graph scaling; it calculates the proximities for pairs of objects and allows to get clusters [1, 16, 17]. F. Moya-Anegon with team successfully grouped ISI categories in major thematic areas by use PFNET and can show the most 'salient' structure of selected domain [18]. Given graphs are drawn in an aesthetically pleasing way and such benefits as symmetry, non-overlapping, and minimized edge crossings are exposed.

Recently, a new popular mapping technique VOS (visualisation of similarities) has been introduced in a series of works by N.J. Van Eck and L. Waltman [19,20] and made freely available in the form of interactive application VOSviewer (www.vosviewer.com). It is intended as an alternative to MDS. VOS and MDS have a close mathematical relationship. The idea of VOS is to minimize a weighted sum of the squared distances between all pairs of items. Those experiments show that maps constructed using MDS approaches may suffer from certain artefacts. Maps constructed using the VOS approach do not have this problem [21].

Among graphs algorithms, the artificial neural networks are widely used in science visualisation. Self organized maps (or Kohonen network) aims to produce a two dimensional representation of the input data using a neighbouring function. During training process similar input vectors (objects) stimulate adjacent neurons; thus output SOM map shows semantic relationships between data, where similar items are mapped close together. Euclidean space between input data is represented by U-matrix [22]. SOM is often considered as alternative to MDS – there are a lot of LIS and Infovis fields papers described this issue. For example, HD. White found that these maps are "in effect the same map, differing mainly in matters of nuance" [23]. Moya-Anegon et al. in LIS domain analysis concluded that MDS is distance preserving while the SOM is topology-preserving technique [24].

2.3. Output space dimension

3D visualisation is a popular trend in graphic design, simulation and modelling. One can find a lot of arguments supporting the systems with dominance of spatial visualisation, i.e. 3D graphs. It is natural to say that we live in a three-dimensional world and our brain is designed to process information in three dimensions. An original image mapped to the spherical retina also shows the characteristics of a spherical 3D structure [25]. It seems reasonable to use spherical data visualisation methods adapted to the natural abilities of the human vision [6].

There is a lot of open source software for graph and network analyses including wide interactive possibilities (Walrus, Gephi, Pajek etc.). A lot of visualisation layouts are limited to regular shapes like circle or sphere. For example, Science–Related Wikipedian Activity map [15] shows the structure and dynamics of the English Wikipedia based on 659,388 articles and their editing activity. The similarity of each article-article pair was calculated as the number of shared links to other articles. The final graph layout generated by the VxOrd routine is circular. The most frequently cited map of global science is Forecasting Large Trends in Science within 3D space [26, 27]. The authors visualized 7.2 million papers and over 16,000 separate journals, proceedings, and series from a five-year period, 2001-2005. The metric was based on a combination of the bibliographic coupling of references and keyword vectors. Using a three-dimensional graph they achieved spatial visualisation of disciplines on a sphere and created a two-dimensional version of the map - the Mercator projection.

But today two-dimensional visualisation is still the easiest way of presenting and exploring results, particularly those having hierarchical structure. In the present work a sphere surface has been selected as a target mapping space. A sphere surface has no edges and therefore it is possible to represent not only local similarities but also large-scale ones regarding the whole space. The benefit of a curved surface in comparison to a plane one is a more capacious exploration space.

3. Scope of Work

Presented mapping method is based on multidimensional scaling with three dimensional topology. Authors innovation to that technique relies on use Morse potential in order to obtain final distribution on a sphere surface. Obtained this way visualisation will be compared with those, generated by 2D mapping algorithms:

(1) VxOrd – as former, traditional one, related to force-directed network
(2) VOS – as a very popular today, perhaps due to on-line application VOSviewer[1]

(3) SOM – as a very distinct origin (neural networks) representation.

Dataset of documents is limited to three time periods. This causes the difficulties in measuring dynamical characteristics of current domain - computer science. The authors intuitively selected qualitative analysis to find the sequence of maps preserving continuity in global arrangement of units. Ultimate visualisations must repeat the real changes of original classification scheme, to be introduced and controlled by humans (editors and authors). This is the second criterion for selection the proper mapping method.

Infovis researchers and scientometricians have put a lot of effort into science domain visualisation. It is generally science, medicine and social sciences are drawn by use ISI or Scopus database. Computer science and/or engineering are visualized very rarely. Perhaps, the reason is global indexes cover the smaller part of CS articles, which are significantly diffused on conference proceedings. Computer and Internet history is a separate CS research field with extensive bibliographical information [28, 29]. In CS history study there is no place for the thematic analysis. Epistemological problem about how the main categories develop, which naturally reflect CS growth and progress in ICT, should be within the scope of interest of many specialists. Of course this knowledge can be useful for librarians and classifiers but first of all for computer scientists and everyone who follows the rapid development of computer technologies.

Authors constructed interactive visualisation interface, accessible online[2]. Application can be used for analyzing the evolution of CS main topics during two last decades. Scholars can also evaluate the emergence of scientific paradigm such as cloud computing [30].

Figure 1 shows schematically the all tasks to be performed.

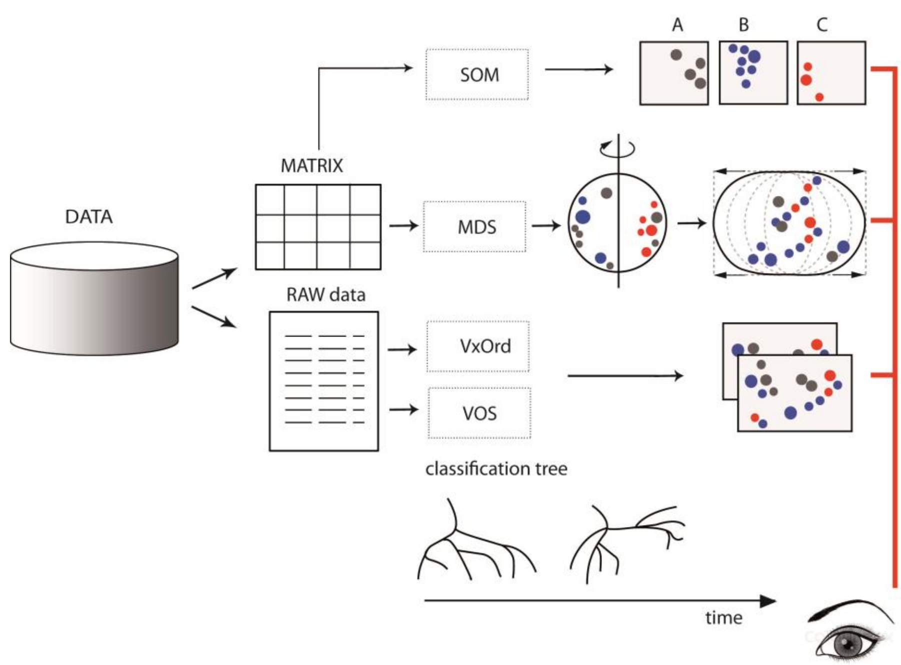

Figure 1. The scheme of tasks.

4. Methodology

4.1. Material and structure

Datasets constituted documents derived from the Association for Computing Machinery (ACM) Digital Library (www.acm.org/dl). The ACM digital library stores an enormous collection of articles and conference proceedings which are organized according to the ACM Computing Classification System (CCS). The CCS as the approved subject classification was initially based on a traditional taxonomy for computer science and engineering. It consisted of 11 main classes indicated by letters: A. General Literature; B. Hardware; C. Computer Systems Organization; D. Software; E. Data; F. Theory of Computation; G. Mathematics of Computing; H. Information Systems; I. Computing Methodologies; J. Computer Applications; K. Computing Milieux. Since 2012 ACM has developed a new version of classification with a poly-hierarchical ontology in response to current requirements of semantic computing. This system "relies on a semantic vocabulary as the single source of categories and concepts that reflect the state of the art of the computing discipline and is receptive to structural change as it evolves in the future"[3].The upper level has been extended to 13 main categories instead of 11 and the alphabetic coding has been removed. Table 1 shows the new and renewed categories as bold.

Table 1. Main classes of CCS today and their basic changes relative to previous classification version.

Current name	Class in previous version
General and reference	
Hardware	
Computer systems organization	
Networks	C
Software **and its engineering**	
Theory of computation	
Mathematics of computing	
Information systems	
Security and privacy	C, K
Human-centered computing	H, I
Computing methodologies	
Applied computing	J
Social and professional topics	K
Proper nouns: People, technologies and companies	

.
The category of **Networks** has been separated from Computer Systems Organization and new important categories have been created. In the age of an increased role of information users, the item **Human-centered computing** concerns the interface design. The knowledge about **Security and privacy** of computing became very important in Web 2.0 epoch. Social sciences have begun to intensively exploit artificial intelligence, complex networks and other advanced CS methods; thus, the proper category (**Social and professional topics**) gathers papers about management of information systems, ICT education and computing profession issues.

Computer applications has been renamed to a more precise **Applied computing**. The classification tree can still be expanded to the lower third level. For the better subject description the last category of Proper nouns has been added to the upper level.

The authors have analysed the classification collection and its dynamics on the basis of the traditional CCS taxonomy as well as its changes since 2012. They harvested data from the digital library in the publishing period from 1968 (when the classification started) to 2009, with a ten-year step. For the analysis purposes the visualisations with critical changes i.e. of 1988, 1998 and 2009 were chosen. There was a technical problem downloading the crucial traits of documents, such as (sub)classes, since the new release of CCS had emerged. Although ACM - the first scientific and educational computing society in the world, declares a free availability to "full CCS classification tree", their interests clash with an open science idea. In that case the authors constrained the current research to the former data but at the same time they developed a mapping methodology and inserted a comparison element. The results of the presented classification visualisation do not contradict the current CCS scheme and can show the validity of the current approach.

4.2. Experiment

The documents of the digital library were attributed to more than one category (class). Every article was ascribed to one main thematic category and some additional ones. Overlapping classes and subclasses therefore appeared simultaneously throughout the documents collection. A co-occurrence matrix was not symmetrical and weights between the main and additional classifications had to be introduced. The final number of all possible classes and subclasses in the collection varied from 270 (1988) to 353 (2009). Those numbers defined the dimension of similarity matrixes of co-classes. For symmetrisation and normalization the IC-cosine measure was used [10]:

$$cos_{i,j} = cos_{j,i} = \frac{(RAW_{i,j})}{\sqrt{\sum_{k=1}^{n} c_{i,k} \sum_{k=1}^{n} c_{j,k}}} \quad (1)$$

where:

$$RAW_{i,j} = WC_i + (1 - W)C_j \quad (2)$$

$C_{i,j}$ - matrix elements, W – appointed weight.

To decrease data dimension, the MDS-based scatter plot selecting a sphere as output space was used. The reason was that the nodes were considered as single particles under the Morse potential, where the energy minimum corresponded to the radius of the sphere [29]. Alternatively, co-occurrence data among classes and subclasses were implemented in the VxOrd and VOS mapping. Finally, the output space (sphere or plane) was also occupied by documents nodes using classic geometrical rules and an appropriate variant of weights: **0.6:0.4** (used for plots presented in this paper), 0.7:0.3 and 0.5:0.5. All documents nodes were marked by their main class colour, thus the final visualisation palette consisted of 11 colours. For a convenient image of the analysis, cartographic projections of MDS-sphere visualisation layouts were used.

All three methods are based on top down data processing: primarily we establish classes topology, whereupon visualized the documents. SOM is usually used as complementary visualisation technique because of Cartesian transformations problems [24]. The authors applied SOM visualisation for ultimate collation with maps of "winner" method. Input data vector was designed to represent documents properties and thus down –top mapping approach is presented.

4.3. Interactive sphere

Documents nodes were indicated by the main class colour and thus Figure 2 shows the complex pattern formed by all categories. Users can switch 11 checkboxes and/or year buttons and explore how organization of particular theme change. Complete navigation through computer science literature is allowed by rotation function. Overlapping patterns point the categories mutually integrate and articles at that location are similar semantically. No edge of visual layout does not evoke the reduction dimensions problem in nodes arrangements (see 2.1). Visualisation of classified documents reveals both organization of digital library content and allows to track how it change over time.

5. Visualisation maps – qualitative analysis

For qualitative comparison we selected the areas which play a key role in CS growth: networks, methodologies and, social computing. For simplification relative categories (see Table 1) were divided into two groups. To compare the dynamics of patterns visually, a set of original pictures miniatures was prepared. The human perception system can rapidly estimate essential and subtle differences of images [7]. The resolution and the size of objects play important role when we want to direct our perception not to the structure but to the shape contour. Small sharp pictures placed together, make a perfect composition for catching figures contours and their instant comparison.

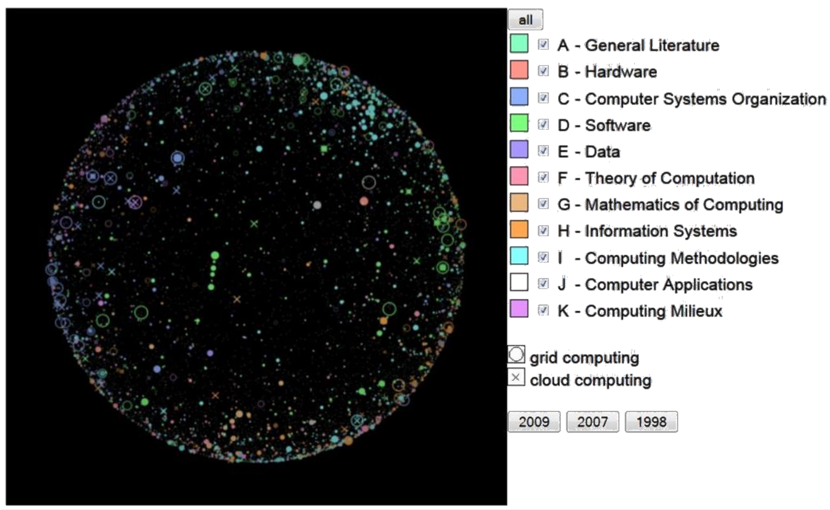

Figure 2. Interface screenshot[3]

5.1. Information systems organization (H), Computing methodologies (I), Computer application (J)

The first group consists of **H**, **I** and **J** classes. *Information systems organization* (H) includes a wide spectrum of information resources management and application. There are articles describing data mining, information retrieval, clustering and Web services issues. Many terms deal with the study of interaction between users and computers: *Human computer interface*. Knowledge management, semantic web and artificial intelligence topics are assigned to *Computing methodologies* (I). The label *Computer applications* of **J** category has become obsolete: what is presented in today's information technology without computer? The more appropriate *Applied computing* is used in the new CCS version. It is related to social and professional topics branch (*Enterprise information systems*).

In the modified classification scheme the three classes have been split into more categories. User study, information interface and HCI issues have been moved to the new *Human-centered computing*. The revised *Computing methodologies* is now focused on artificial intelligence and expert systems. **J** category has been reorganized the most significantly because of the established position of computing in science, engineering, medicine, art and education.

MAIN DIFFERENCES

The first row in Figure 3 demonstrates the results of mapping according to the authors' approach. Each column corresponds to publishing years 1988, 1998 and 2009, so that it is possible to compare changes of every 10 years. The presented distributions are the most even and may form a ergonomic interface for further exploration (indeed, the largest distortion is in the top/ bottom parts of the image, but it is the same for all first row images). Classes **H, I** and **J** reveal a close mutual configuration. In 1988, the distribution of nodes is sparse in comparison with the next decades. Ten and twenty years later the nodes tend to group closer and concentrate around class **J** (Computer applications, grey points). The strong clustering of **J** class articles means that it needs to keep self-separation. It has been preserved in a new CCS. The sequence of mini maps allows to notice that *Information systems* and *Computing methodologies* develop towards applications using modern computational methods. Overlapping between these classes progressively increases with a strong tendency to merge. Thus changes in the distribution correspond to the real ICT dynamics.

VOS maps evolution is presented in the second row of Figure 3. The largest integration between methodologies, information systems and applications is observed two *decades* ago – in 1988. Ten years later *Computer applications* (grey points) becomes invisible. Could they be really useless? Nowadays (2009), Information systems (orange) has a tendency to separate from *Computing methodologies* (cyan) which consists of topics such as: artificial intelligence, knowledge management and representations, ontology, semantic Web, and graphic design.

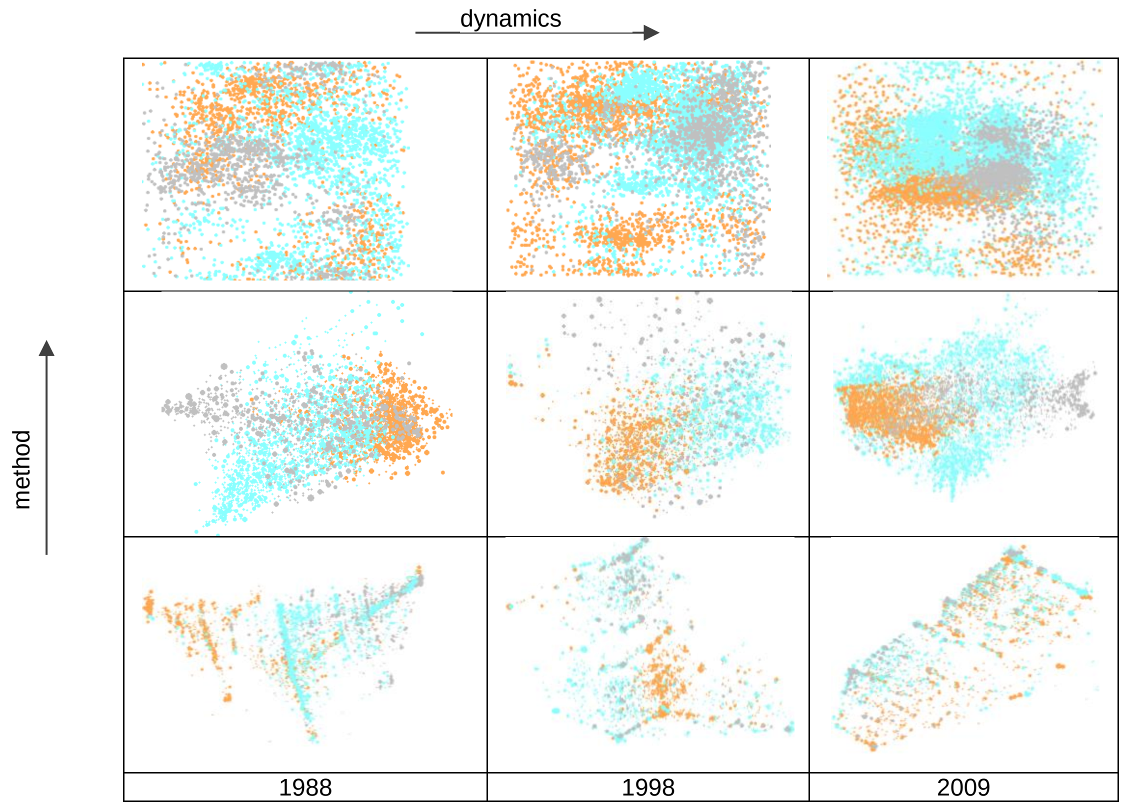

Figure 3. Comparison of H (orange points) I (cyan) J (grey) articles distributions generated by three methods: authors' - 1st, VOS -2nd and VxOrd -3rd row.

This means that data mining, information retrieval, clustering, and Web services do not reflect the cognitive context and are not to be oriented to the human–centred research. Cognitive sciences are not clearly visible in the VOS structure. *Computer applications* nodes are also in isolation from methodologies (right, cyan). Thus, *Computing methodologies* here does not find practical substantiation.

In the case of mapping using the VxOrd procedure (third row, Figure 3), the nodes are distributed along a sharp irregular grid, leaving a lot of white space around. The cluster of H class articles (orange points) is the most distinct in the middle of the image (1998). This could be explained by the fact that information systems developed faster than computer methodologies and formed a separate category. The largest integration between all categories is observed in 2009. However, adequate scattering of nodes follows it. These conclusions are consistent with the history of computer science but the distribution has unacceptable topology for such processes as: browsing, searching and comparison.

5.2. Computer systems organization (C) and Computing Milieux (K)

The new and extremely useful category *Networks* is strongly related to *Computer-communication networks* pulled from class **C** of the previous scheme. However, the renewed *Computer systems organization* more broadly deals with processor architectures, embedded and real-time based systems. The most differentiated category of *Computing Milieux* can be an interesting entity for our analysis. It has been reduced by the subcategory of computers in education which has been moved to *Applied computing* and partly to *Human-centered computing*. Finally, editors grouped collaborative and social computing as well as user characteristics and history of computers within the class *Social and professional topics*.

MAIN DIFFERENCES

The distribution of both **C** and **K** articles shows (first row in Figure 4) networking research and applications development. Their fast growth is observed two decades ago – the nodes are evenly scattered over a fully available space.

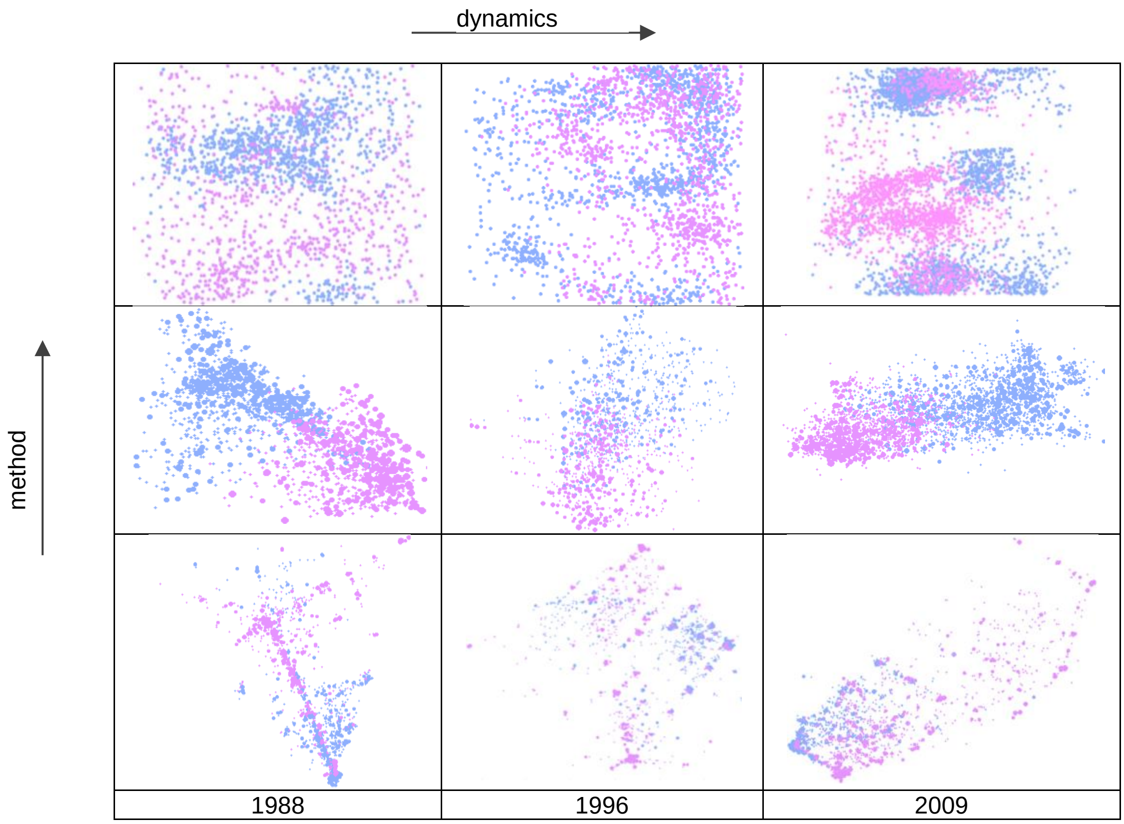

Figure 4. Comparison of CK articles distributions generated by three methods: author's -1st, VOS -2nd and VxOrd -3rd row.

Previously locally overlapping categories *Networks* (blue points) and *Computing Millieux* (magenta colour) start to form separate large groups in 2009, excluding one longitudinal cluster. This one is arranged on a sphere pole and responsible for security networking – a common topic between networks and their social aspects. The well-formed **K** (magenta horseshoe shape) class requires a separate thematic category. CCS editors have accomplished this task and renamed the odd label *Milieux* to *Social and professional topics*.

The second row - VOS in Figure 4 demonstrates recurring changes of **C** and **K** categories. If the middle of the image was erased, it could be assumed that both configurations are almost the same with a different orientation: the first one needs to be rotated at an approximately 110 degree. Backward changes indicate a 20 year-cycle dynamics, which does not correspond to reality. On the contrary, since 2000 a clear social development of networking can be observed.

The geometry of the VxOrd mapping does not represent the ergonomic use of space (third row in Figure 4). It is not convinient for dynamics analysis. In 1988, *Millieux* nodes (magenta points) show a relatively largest disconnection from other categories. In the next decades they lose a topical autonomy and become integrated with *Networks/Computer systems.* **C** class behaviour does not show clustering tendency in each tested period. This demonstrates no need either to form a separate category for *Milieux* computing or to split Computer systems organization. This sequence of maps does not replicate the last classification changes.

6. Interpretation and reference to SOMs

Five original scheme categories: *Computer systems organization*, *Information systems*, *Computing methodologies*, *Computer applications* and *Computing Milieux* have been changed significantly by CCS editors since 2012. Dynamics analysis of **H**, **I**, **J** group shows cohesion in changes of maps made by author's method. From tendency to integrate the classes become to form own clusters close to each other. *Information systems* and *Computing methodologies* drift owards computer applications. Clustering overlapping shows that these limbs of tree need modifications and a new subdivisions. The sequence of **C**, **K** categories depicts the shift (2009) to proliferation clear separation respectively.

Computer systems Organization must be described more detailed (many spots group), while *Computing Millieux* must be reorganized because of large central “horseshoe “ cluster. All Mentioned categories have been modified the most by the ACM.

Critical changes of classification universe could be observed from almost even, relatively sparsely occupied on the map in 1988 to the beginning of categories clustering in 1998. This period provides a close adaptation of the CCS scheme in the ACM Digital Library resources. Using a series of chronologically sequential maps, one can study how knowledge advances and how scientific knowledge organization changes.

Presented in Figure 5, SOM visualisation consists of component planes maps. This way we can visualize relative class distribution of input data. As SOM topology the 15x15 hexagonal grid was selected. Lighter and darker colors represent larger (high data concentration) and smaller neuronal weights (fewer concentration) respectively. SOM maps are in coincidence with the presented approach. In 1988 **H**, **I**, **J** nodes are widely scattered over at the most space (first row, Figure 5). At the end of time range there is visible clustering (particularly *Information Systems* and *Computer Applications* – in the center) and overlapping simultaneously. **C** and **K** reveal noticeable mutual separation in 2009 (second row in Figure 5). The data of both categories are evidently became clustered.

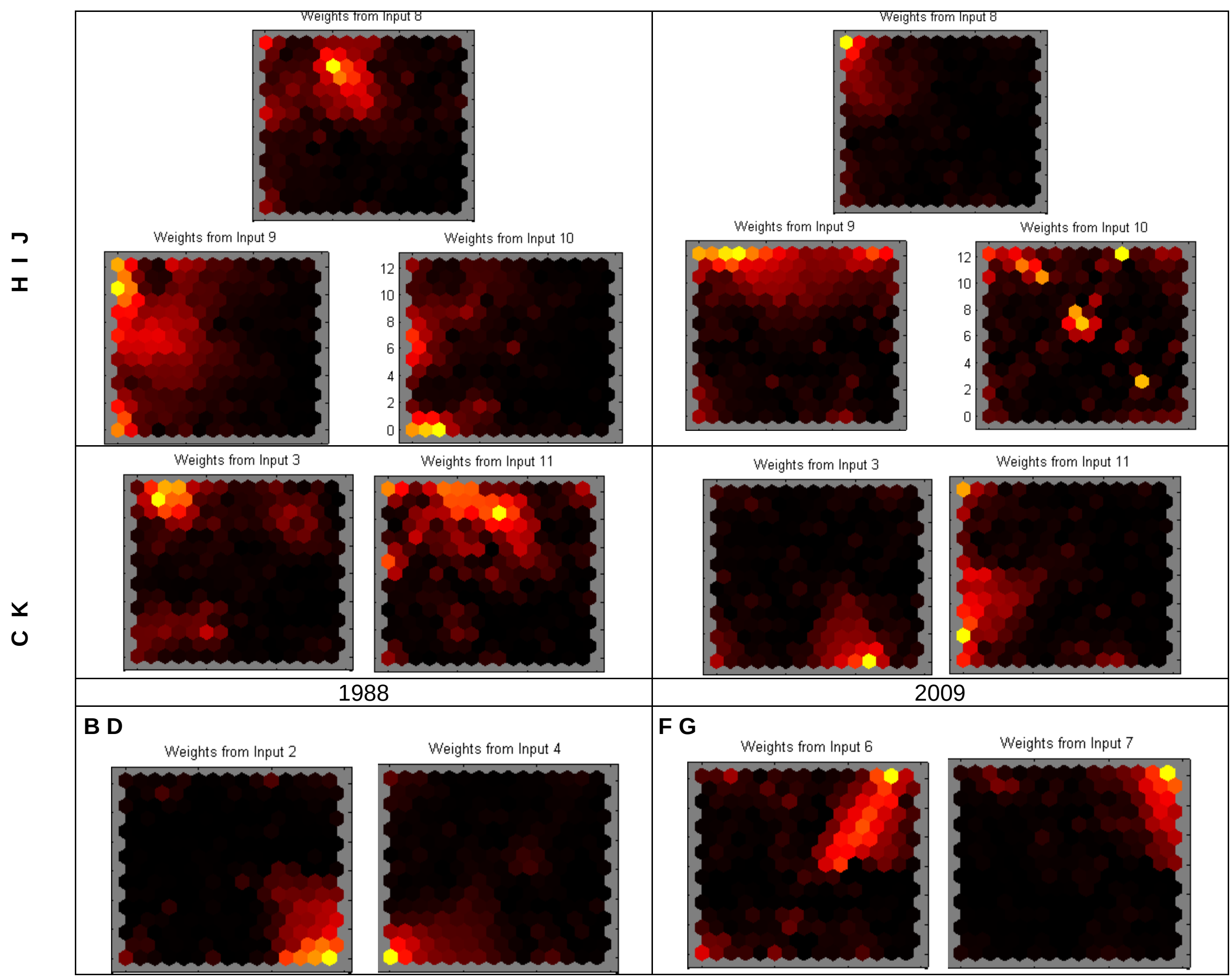

Figure 5. SOM component planes representing groups of classes: H I J (1st row), CK (2nd row) and BD; FG (3rd row)

In this phase we tested the topological properties of categories arrangement instead of dynamical, so SOM maps were generated for two years only. SOM and MDS-sphere (authors) methods show coincidence in distribution of particular category and their mutual correlation. The most typical (presented in 3rd row of Figure 5) is that semantically different hardware (**B**) and software (**D**) locate on the opposite poles on a sphere[3] and are not utterly correlated by SOM representation [31]. However the largest correlation (topological separation) is observed for **F**, **G** classes (the same row). Both refer to mathematical base of computing.

7. Conclusion and future work

In this paper three methods of mapping structured data in respect of possibilities of dynamics analysis have been compared. The subject study was computer science knowledge and its evolution during the last two decades. It has revealed essential changes in computer science literature during the time of the development of the CCS classification.

The critical changes in five main categories of the classification correspond to the newest version of the CCS Scheme, published by the ACM in 2013. Human contribution to modernization is the most certain: editors as a rule have good orientation how to match thematic organisation of scientific articles to computer scientists and engineers needs. Current, scheme of classification was considered as reference system. How the changes have been reflected by every series of map was the main criterion in the selection of the presented mapping approaches.

The findings prove that merely authors' mapping shows continuous changes in classification dynamics which confirm the modifications made by the CCS developers recently. SOM is suitable if we want to conduct not holistic, but piecemeal comparison. The same conclusion was made by Moya-Anegón with co-authors. [24].

Undoubtedly, a sequential series of three maps is not enough to estimate knowledge dynamics. As mentioned in chapter 4.1, truly objective circumstances concerning data gathering were appeared. A more multi-variety dataset to track all changes in fast growing knowledge is needed. Then, longitudinal mapping – the term introduced by E. Garfield – will deliver essential research material. The proposed measures can be considered as promising method in the quantitative evaluation of visualisation of discrete values which still remains a problem today – the era of visual information and media. Authors plan to extend comparative visualisation study by quantitative analysis by use pattern recognition methods. They made preliminary tests based on clusterization degree, fractal dimension and lacunarity and gave original criteria for visualisation evaluation. They guided by several papers oriented on fractal analysis [32, 33].

Looking for an effective interactive visual representation of dataset, an original interface based on spherical topology has been designed. In this aspect the MDS-sphere (authors') method can compete with VOS: the last one demonstrates relatively even distribution and clustering over more layout. Those two techniques were compared by N.V. Van Eck and L. Waltman [34] in visualisation similarities between objects. They found that maps constructed using VOS provide a more satisfactory representation than MDS..

Besides the dynamics analysis, we focused on effective visualisation interface. These two clearly differentiated functions – analysis and interface; their coexistence is the permanent visualisation problem [22]. We strived for reconcile these two aspects through creating visualisation tool that can be used to provide a visual summary of research activity in a computer science to members of expert panels that conduct technology forecasts. Such a tool should help experts to understand changes over time in the technology of interest: identifying specific research areas and showing which areas are growing and which are declining. It is also important for experts to understand the flow of information among research areas: which areas contribute knowledge and which areas are "borrowers" of knowledge.

Notes

1 http://vosviewer.org
2 http://www-users.mat.umk. pl/~garfi/vis2009v3/
3 http://www.acm.org/about/class/2012.

References

[1] Chen Ch. Information Visualisation. Beyond the Horizon. 2nd ed. London: Springer, 2006, pp.143-210.

[2] Garfield E, Sher IH and Torpie RJ. The use of citation data in writing the history of science. In: Report of research for Air Force Office of Scientific Research. Philadelphia: Institute for Scientific Information, 1964.

[3] Garfield E. Historiographic Mapping of Knowledge Domains Literature. Journal of Information Science 2004; 30(2): 119-145.

[4] Chen Ch and Kuljis J. The Rising Landscape: A Visual Exploration of Superstring Revolutions in Physics. Journal of the American Society for Information Science and Technology 2004; 54(5): 435–446.

[5] Leydesdorff L. 'Analysis, Visualisation and Animation of Social Networks using Visone'. Online Referencing, http://www.leydesdorff.net/visone/ (2008, accessed July 2014).

[6] Osinska V, Dreszer-Drogorob J, Osinski G and Gawarkiewicz M. Cognitive Approach in Classification Visualisation. End-Users Study. In: A. Slavic et al (ed.), Proceedings of the Conference on Classification & Visualisation: interfaces to knowledge, Hague, Holland, 23 -25 October 2013 (Würzburg: Ergon Verlag, 2013); pp. 273-283.

[7] Ware C. Information Visualisation: Perception for Design. CA: Morgan Kaufmann, 2004, pp. 11, 188, 227-293.

[8] Small H and Garfield E. The geography of science: disciplinary and national mappings. Journal of Information Science 1986; 11(4): 147-159.

[9] Leydesdorff L. Statistics for the dynamic analysis of scientometric data: the evolution of the sciences in terms of trajectories and regimes. Scientometrics 2013, 96(3): 731-741.

[10] Boyack KW, Klavans R and Börner K. Mapping the backbone of science. Scientometrics 2005; 64(3): 351-374.

[11] Hearst MA. Search User Interfaces. 1st ed. Cambridge University Press, 2009, pp. 234-296

[12] Fortunato S. Community detection in Graphs. Physics Reports 2010; 486: 75-174.

[13] Yang B, Liu Z and Meloche JA. Visualisation of the Chinese academic web based on social network analysis. Journal of Information Science 2010; 36 (2): 131-143.

[14] Boyack KW, Wylie BN and Davidson GS. Domain visualisation using VxInsight for science and technology management. Journal of the American Society for Information Science and Technology 2002;, 53(9): 764-774.

[15] Holloway T, Miran B, and Börner K. Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. Complexity 2007; 12 (3): 30-40.

[16] Kim H. and Lee J.Y. Exploring the emerging intellectual structure of archival studies using text mining: 2001—2004. Journal of Information Science 2008; 34(3): 356-369.

[17] Lee JY, Kim H and Kim PJ. Domain analysis with text mining: Analysis of digital library research trends using profiling methods. Journal of Information Science 2010; 36(2): 144-161.

[18] Vargas-Quesada B, Moya-Anegon F, Chinchilla-Rodriguez Z and Gonzalez-Molina A. Showing the essential science structure of a scientific domain and its evolution. Information visualisation 2010; 9: 288--300.

[19] Van Eck NJ, and Waltman L. VOS: a new method for visualizing similarities between objects. In: HJ Lenz, R Decker (eds) Proceedings of the 30th Annual Conference of the German Classification Society Advances in Data Analysis, 2007, (London: Springer 2007); 299-306.

[20] Van Eck NJ and Waltman L. How to normalize cooccurrence data? An analysis of some well-known similarity measures. Journal of the American Society for Information Science and Technology 2009; 60(8): 1635-1651.

[21] Van Eck NJ, Waltman L, Dekker R and Van den Berg J. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. Journal of the American Society for Information Science and Technology 2010; 61(12): 2405-2416.

[22] Ultsch A. U*-Matrix: A tool to visualize clusters in high dimensional data. In: Technical Report of Department of Computer Science, Marburg: University of Marburg 2003; Nr. 36:1-12.

[23] White HD, Lin X and McCain KW. Two modes of automated domain analysis: Multidimensional scaling v.s Kohonen feature mapping of information science authors. In: M. Haidi, J. Maniez and SA. Polit (eds), Proceedings of the Fifth International ISKO Conference 1998, Würzberg: Ergon Verlag; 57–61.

[24] Moya-Anegón F, Herrero-Solana V and Jiménez-Contreras E. A connectionist and multivariate approach to science maps: the SOM, clustering and MDS applied to library and information science research. Journal of Information Science 2006; 32(1): 63-77.

[25] Read JCA and Cumming BG. Does depth perception require vertical-disparity detectors? Journal of Vision 2006; 6 (12): 1327.

[26] Börner K. Atlas of Science. The MIT Press, 2010, pp. 170-174.

[27] Börner K, Klavans KR, Patek et al. Design and update of a classification system: The UCSD map of science. PLoS ONE 2012, 7.

[28] Ceruzzi PE. A History of Modern Computing (History of Computing). The MIT Press 20013.

[29] Concise Encyclopedia of Computer Science. ED Reilly (ed.). Chichester, UK: Wiley, 2004.

[30] Osinska V, Bala P and Gawarkiewicz M. Information Retrieval across Information Visualisation. In Proceeedings of the Federated Conference on Computer Science and Information (FedCSIS), Wroclaw, September 8-14, 2012 (Wroclaw University of Technology, 2012); 233 – 239.

[31] Osinska V and Bala P. New Methods for Visualisation and Improvement of Classification Schemes – the case of computer science. Knowledge Organization 2010; 37(3).

[32] Smith TG, Lange GD and Marks WB. Fractal Methods and Results in Cellular Morphology, Journal of Neuroscience Methods 1996; 69: 1123-126.

[33] Osinska V. Fractal analysis of Knowledge Organization in Digital Library. In: New Trends in Qualitative and Quantitative Methods in Libraries 2011, pp. 17-23.

[34] Waltman L, Van Eck NJ, and Noyons ECM. A unified approach to mapping and clustering of bibliometric networks. Journal of Informetrics 2010; 4(4): 629-635.