Supporting Methodology Transfer in Visualization Research with Literature-Based Discovery and Visual Text Analytics

Doctoral Dissertation to obtain the

Degree of Doctor of Philosophy in Computer Engineering

by Alejandro Benito-Santos

supervised by **Roberto Therón Sánchez, PhD**

27th November 2020

Visual Analytics and Visualization Information Group

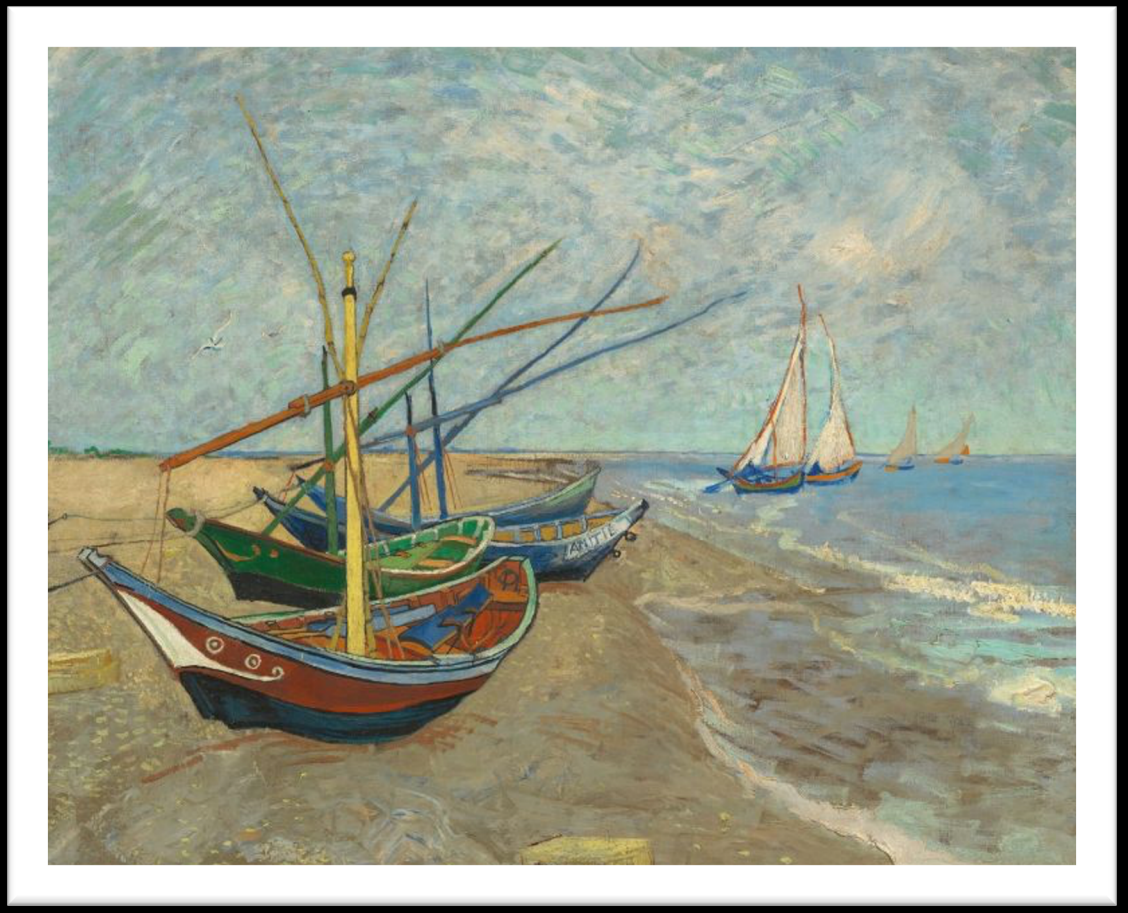
Universidad de Salamanca

Research Group Interaction and e-Learning

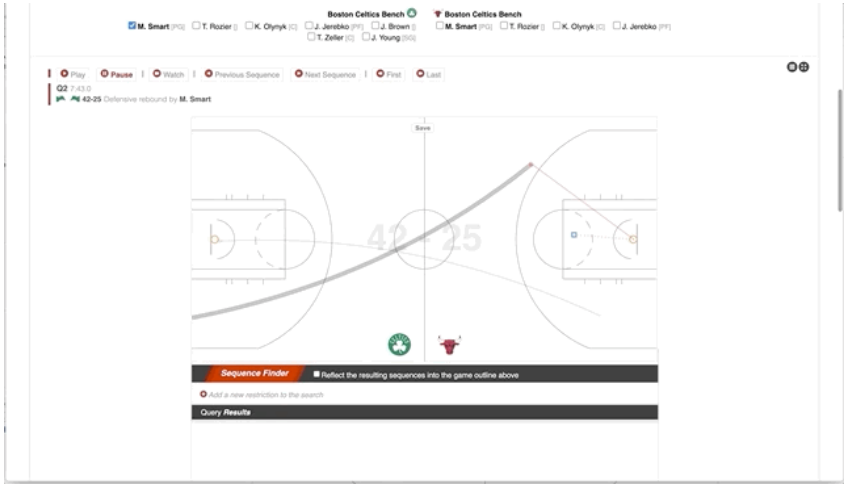
Outline

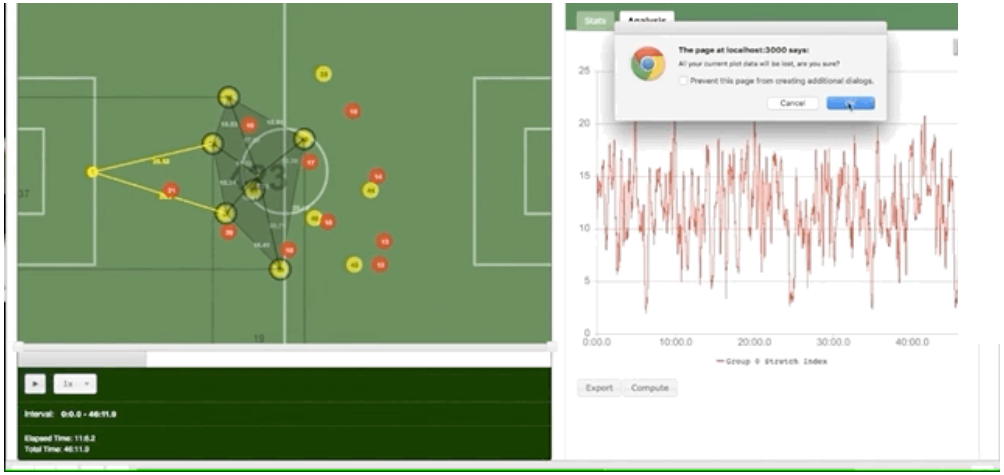
1. Introduction
2. Background
3. Methodology
4. Objectives and Main Hypotheses
5. Data collection
6. Visualization and visual analysis techniques
7. Applicability of the findings to knowledge defragmentation
8. Conclusions
    1. Main contributions
    2. Future work

Introduction



Van Gogh, V. (1888), Fishing Boats on the Beach at
Les Saintes-Maries-de-la-Mer [oil on canvas]





BKViz: A Basketball Visual Analysis Tool

Antonio G. Losada, Roberto Therón and Alejandro Benito

Abstract. The amount of data available nowadays in the sports field is hard to comprehend using classic analytic methods. This calls for the development of systems such as the prototype discussed here, which makes it possible to manipulate chunks of data to then portray them in visual ways, easing their understanding. Based on basketball, this tool helps users in reaching conclusions regarding performances during individual matches. This enables them to
the events occurring at differe
employ based on player chem
integrated system, data can be
and change their analytic pow
ple comprehension based on
the challenge was to provide a
actions in an innovative and

Keywords: Basketball, Data
alytics

1 Introduction

As sports have evolved over the ye
the analysis of players' performanc
ment of statistical metrics. In baske
often presented as text in tables,
ways. The visual analysis of these
attain an important standing alon

Most basketball analysis metho
metrics that depict trends and ave
keep developing new statistical me
performance. The problem with th
to understand, something visual a

Statistical data, normally store
fills up to 26 table rows (maximum
around 35 columns, one per variab
cognitive load. If the analysis also
effort needed escapes human capa

There is a growing need for ne
information and to get to know it
able to reach conclusions regarding
games, and behaviors that show

rs
nology

TECHNOLOGY REPORT
published: 05 December 2018
doi: 10.3389/fpsyg.2018.02416

Check for updates

Data-Driven Visual Performance Analysis in Soccer: An Exploratory Prototype

Alejandro Benito Santos[1], *Roberto Theron*[1], *Antonio Losada*[1], *Jaime E. Sampaio*[2] *and Carlos Lago-Peñas*[3]*

[1] Departamento de Informática y Automática, University of Salamanca, Salamanca, Spain, [2] Department of Sports Sciences and Health, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal, [3] Department of Sports Sciences, University of Vigo, Pontevedra, Spain

In soccer, understanding of collective tactical behavior has become an integral part in sports analysis at elite levels. Evolution of technology allows collection of increasingly larger and more specific data sets related to sport activities in cost-effective and accessible manner. All this information is minutely scrutinized by thousands of analysts around the globe in search of answers that can in the long-term help increase the performance of individuals or teams in their respective competitions. As the volume of data increases in size, so does the complexity of the problem and the need for suitable tools that leverage the cognitive load involved in the investigation. It is proven that visualization and computer-vision techniques, correctly applied to the context of a problem, help data analysts focus on the relevant information at each stage of the process, and generally lead to a better understanding of the facts that lie behind the data. In the current study, we presented a software prototype capable of assisting researchers and performance analysts in their duty of studying group collective behavior in soccer games and trainings. We used geospatial data acquired from a professional match to demonstrate its capabilities in two different case studies. Furthermore, we successfully proved the efficiency of the different visualization techniques implemented in the prototype and demonstrated how visual analysis can effectively improve some of the basic tasks employed by sports experts on their daily work, complementing more traditional approaches.

Keywords: football association, visual analytics, performance analysis, information visualization, collective tactical behavior

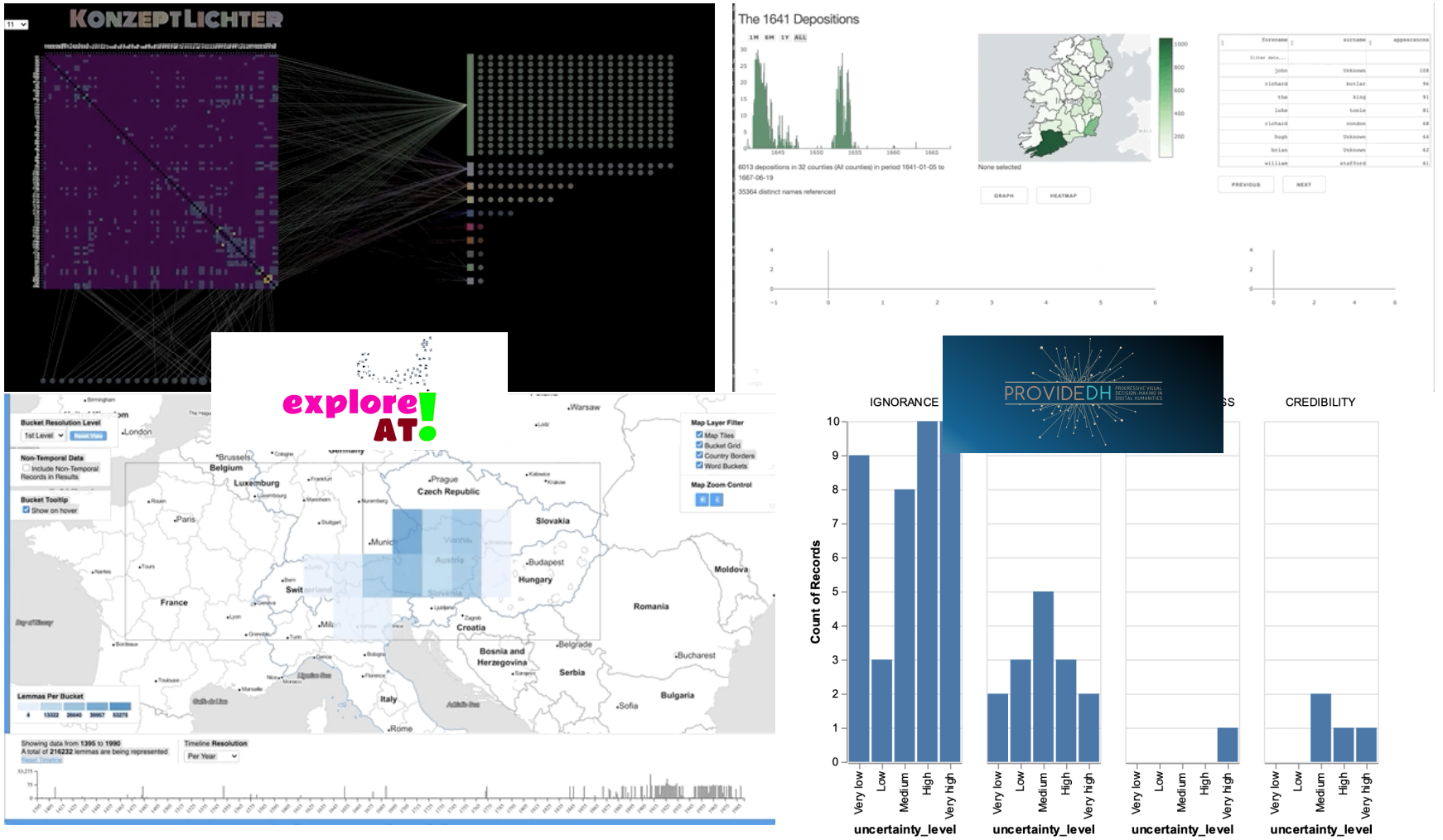
INTRODUCTION

Soccer performance is a multifactorial process requiring high-level interaction analysis within physiological, technical, and tactical performances. At the elite level, technical staffs need to capture, process, and analyze great amounts of data, in order to measure performance in their respective teams and opponents, as well as assess potential prospects. Recently, the time pressure of this process and the constant increase in the amount of available data has demanded for a major emphasis in the visualization methods. In fact, technical staffs are nowadays expected to capture, process, analyze, and visualize data to provide fast assimilated information for coaching purposes. Current technology allows capturing data from players' positions, either in competition or training scenarios, with very acceptable degrees of accuracy. These technological advances can use radio

N ACCESS
Edited by:
Reviewed by:
Correspondence:
ecialty section:
Citation:

[1] A. G. Losada, R. Therón, and A. Benito, 'BKViz: A Basketball Visual Analysis Tool', IEEE Computer Graphics and Applications, vol. 36, no. 6, pp. 58–68, 2016.

[2] A. Benito-Santos, R. Theron, A. Losada, J. E. Sampaio, and C. Lago-Peñas, 'Data-Driven Visual Performance Analysis in Soccer: An Exploratory Prototype', Front. Psychol., vol. 9, 2018, doi: 10.3389/fpsyg.2018.02416



[3] R. Therón Sánchez, A. Benito-Santos, R. S. Santamaría Vicente, and A. Losada Gómez, 'Towards an Uncertainty-Aware Visualization in the Digital Humanities', Informatics, vol. 6, no. 3, p. 31, Sep. 2019, doi: 10.3390/informatics6030031.

[4] A. Dorn et al., 'A project review under the focus of complexities on the example of ExploreAT!', presented at the Digital Humanities 2019, Utrecht, 2019.

[5] A. Benito, A. Dorn, R. Therón, E. Wandl-Vogt, and A. Losada, 'Shedding Light on Indigenous Knowledge Concepts and World Perception through Visual Analysis', in Digital Humanities 2018 Book of Abstracts, Mexico City, Mexico, 2018, pp. 537–538.

[6] A. Benito, R. Therón, A. Losada, E. Wandl-Vogt, and A. Dorn, 'Exploring Lemma Interconnections in Historical Dictionaries', presented at the 2nd Workshop on Visualization for the Digital Humanities (VIS4DH), 2017.

[7] A. Benito Santos and R. Therón Sánchez, 'Visualización de Datos en Humanidades Digitales', Tesis de Master, Universidad de Salamanca, España, 2016.

Interdisciplinarity and language

- As a result of the **increasing specialization** in the sciences, many researchers have turned their attention to other disciplines, producing novel inter-disciplinary collaborations.

- Within these collaborations, the **use of language** and the **acquisition of communication skills** has been identified as key in the success of these research endeavors [8].

[8] Bracken, L.J., Oughton, E.A.: `What do you mean?' The importance of language in developing interdisciplinary research. Transactions of the Institute of British Geographers 31(3), 371{382 (Jul 2006). https://doi.org/10.1111/j.1475-5661.2006.00218.x

Document search

- One or more **keywords** are matched against **documents metadata**.
- Current interfaces limit the **browsing experience** to receiving a series of **keyhole views** of a corpus.
- **Wrongly assume** a **research focus** can always be expressed with a **combination of keywords.**
- Cold-start problem.

Excess of information

- Scientists devote a substantial amount of their work to the querying and browsing of large online collections of research papers.
- The number of available scientific publications doubles every 9 years [9].

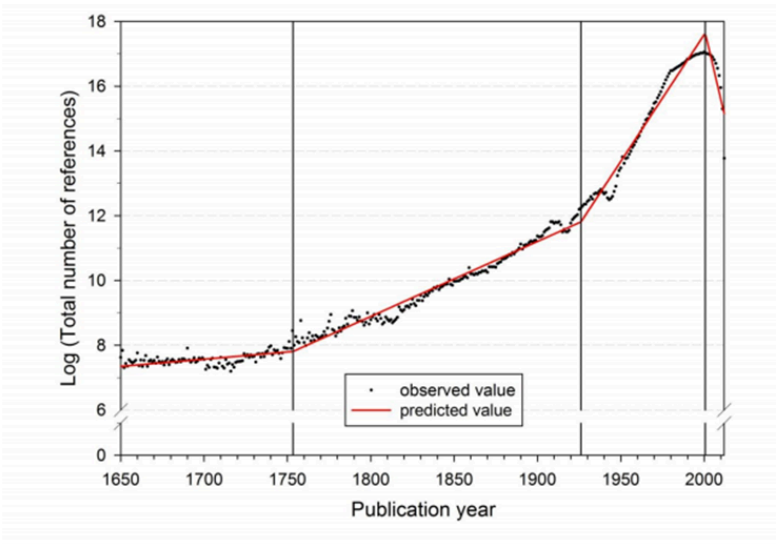
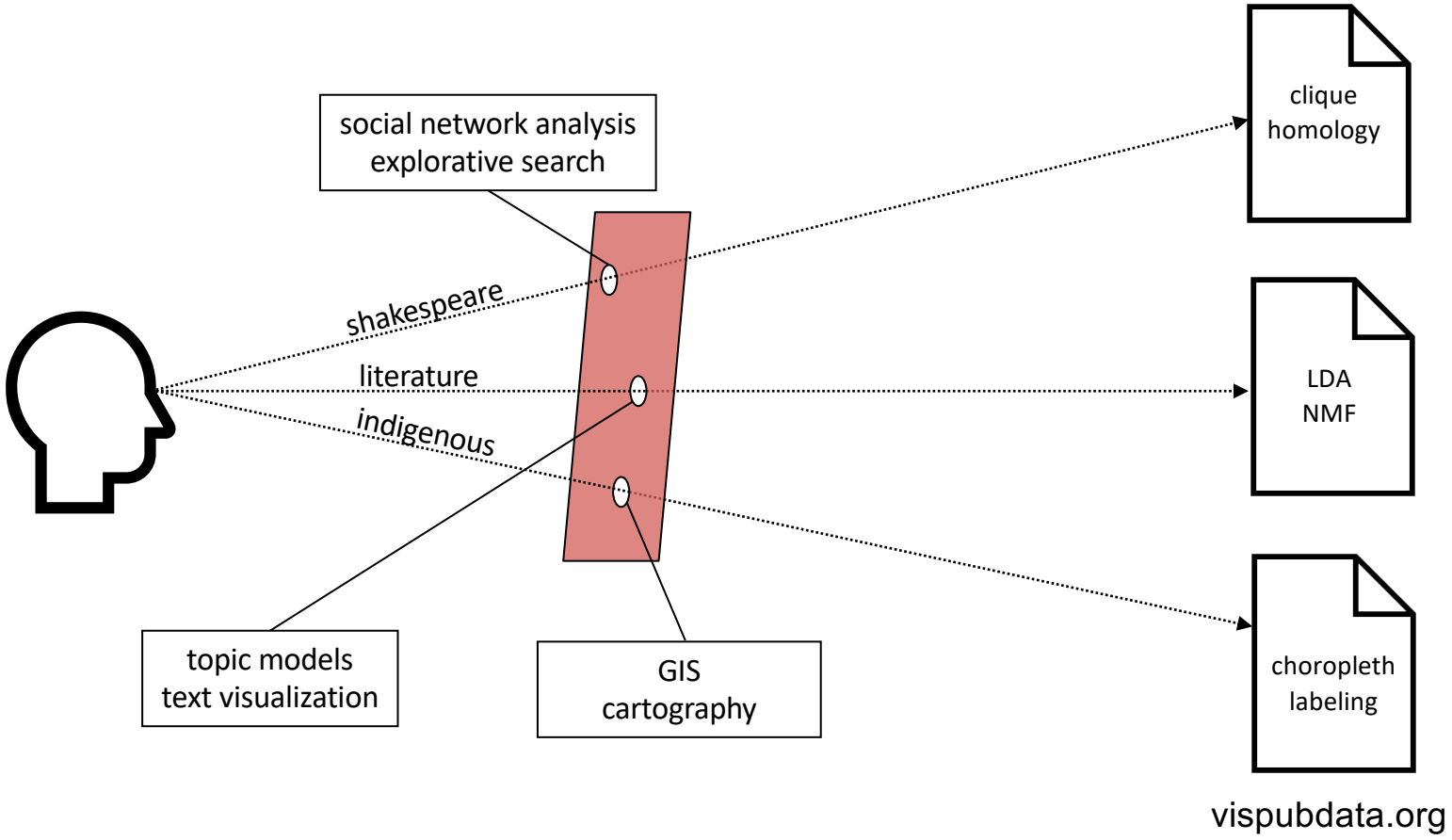


Figure 2. Segmented growth of the annual number of cited references from 1650 to 2012 (citing publications from 1980 to 2012)

[9] 'Global scientific output doubles every nine years : News blog'. http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html (accessed Nov. 25, 2020).

Interdisciplinary documental search



Keywords

- **Descriptive terms**, given as metadata, that authors freely assign to their papers to make them discoverable and distinguishable from others.
  - They make up a tiny, highly-condensed language of science.
  - They encode expert knowledge by the authors.
- They are **concise summarizations** of the concepts described in a paper.
- The **process** by which **humans extract keywords** from academic texts remains mostly **unknown** to this day [10].

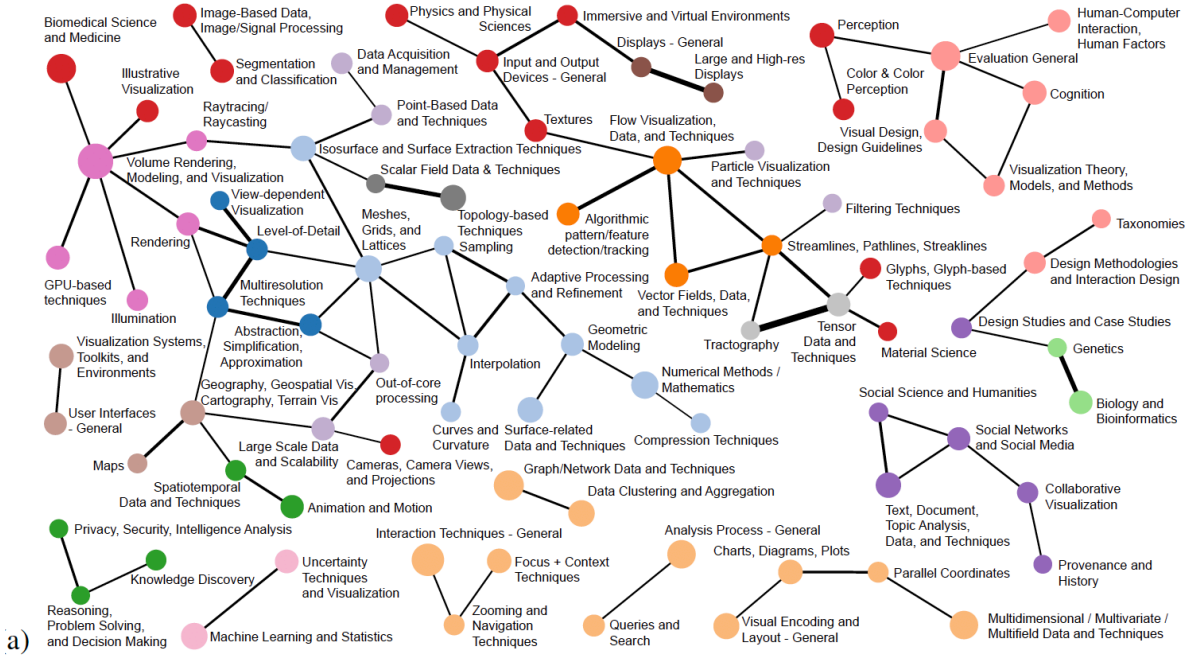
[10] J. Chuang, C. D. Manning, and J. Heer, '"Without the Clutter of Unimportant Words": Descriptive Keyphrases for Text Visualization', ACM Trans. Comput.-Hum. Interact., vol. 19, no. 3, p. 19:1–19:29, Oct. 2012, doi: 10.1145/2362364.2362367.

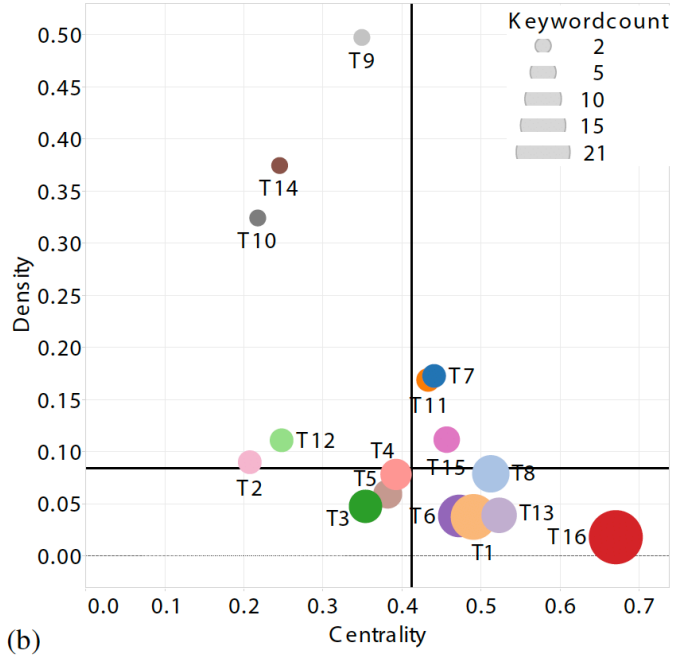
Keywords

- Keywords are often used to build **literature reviews**, **mapping studies**, or **surveys** on a given field of science.
- However, its use in certain situations **has negative implications on document discoverability** [11].
- This calls for novel approaches that address these limitations while exploiting the **beneficial properties** of keywords, understood as **valid, human-produced summarizations** of a piece of text.

[11] K. El-Arini and C. Guestrin, 'Beyond Keyword Search: Discovering Relevant Scientific Literature', in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2011, pp. 439–447, doi: 10.1145/2020408.2020479.

Keywords





- **Keywords have been employed by other visualization researchers in the past**

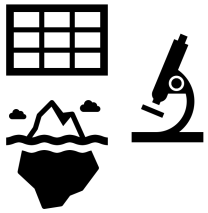
1. To improve communication among different visualization sub-groups, and
2. To facilitate the process of understanding differences and commonalities of the various research sub-fields in visualization.

[12] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller, 'Visualization as Seen through its Research Paper Keywords', *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 771–780, Jan. 2017

Background

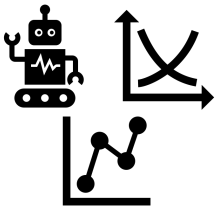
Firle, W. (1900), The Fairy Tale [Oil on canvas]

Problem-Driven visualization research



- Data
- Driving problems
- Tasks

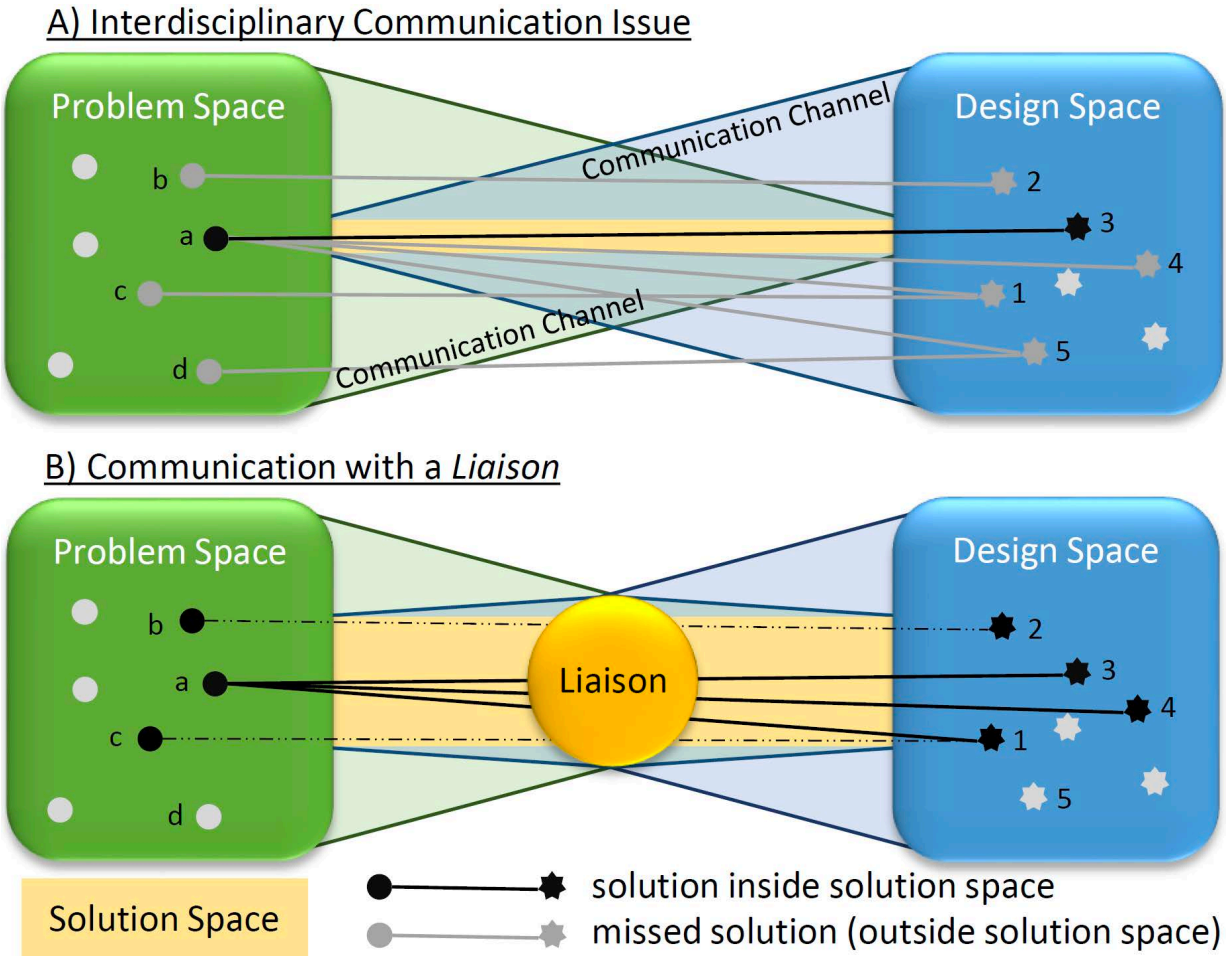
- Algorithms
- Analysis Techniques
- Visualizations



Domain Experts

VIS Experts

Interdisciplinary communication



[13] S. Simon, S. Mittelstädt, D. A. Keim, and M. Sedlmair, 'Bridging the gap of domain and visualization experts with a Liaison', *Eurographics Conference on Visualization (EuroVis 2015, Short Paper)*, Cagliari, Italy, 2015, vol. 2015.

Methodology transfer

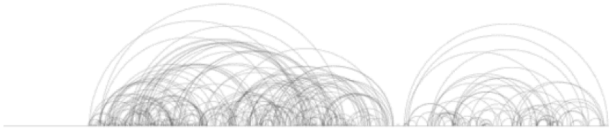


Figure 13. Java class file (bytecode)

Figure 15. DNA sequence

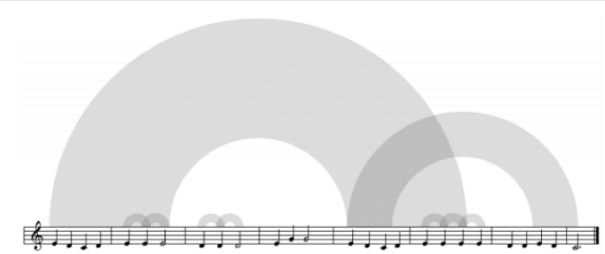
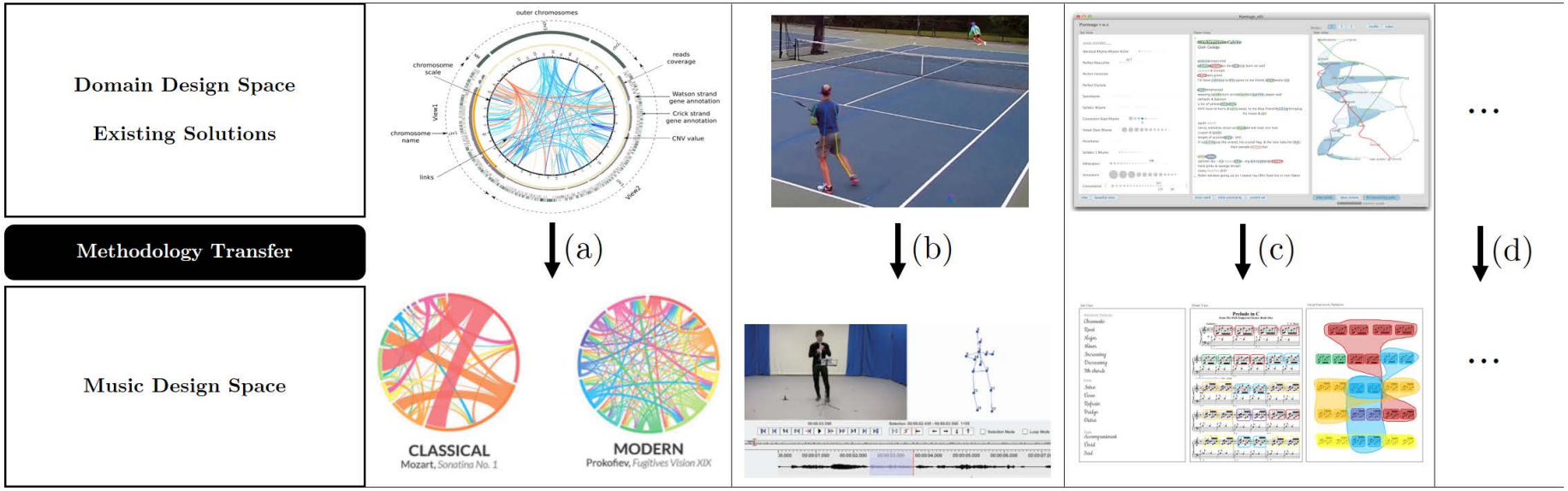


Figure 9. Arc diagram for music

- Many visualizations are ”domain-agnostic” because:
  - they aim to support generic analytical tasks (e.g., comparison or establishing relationships).
  - they efficiently exploit the human visual channel to augment the user’s cognition capabilities.
- For these reasons, other visualizations created to support a certain task in a given domain might be successfully applied to other domains.

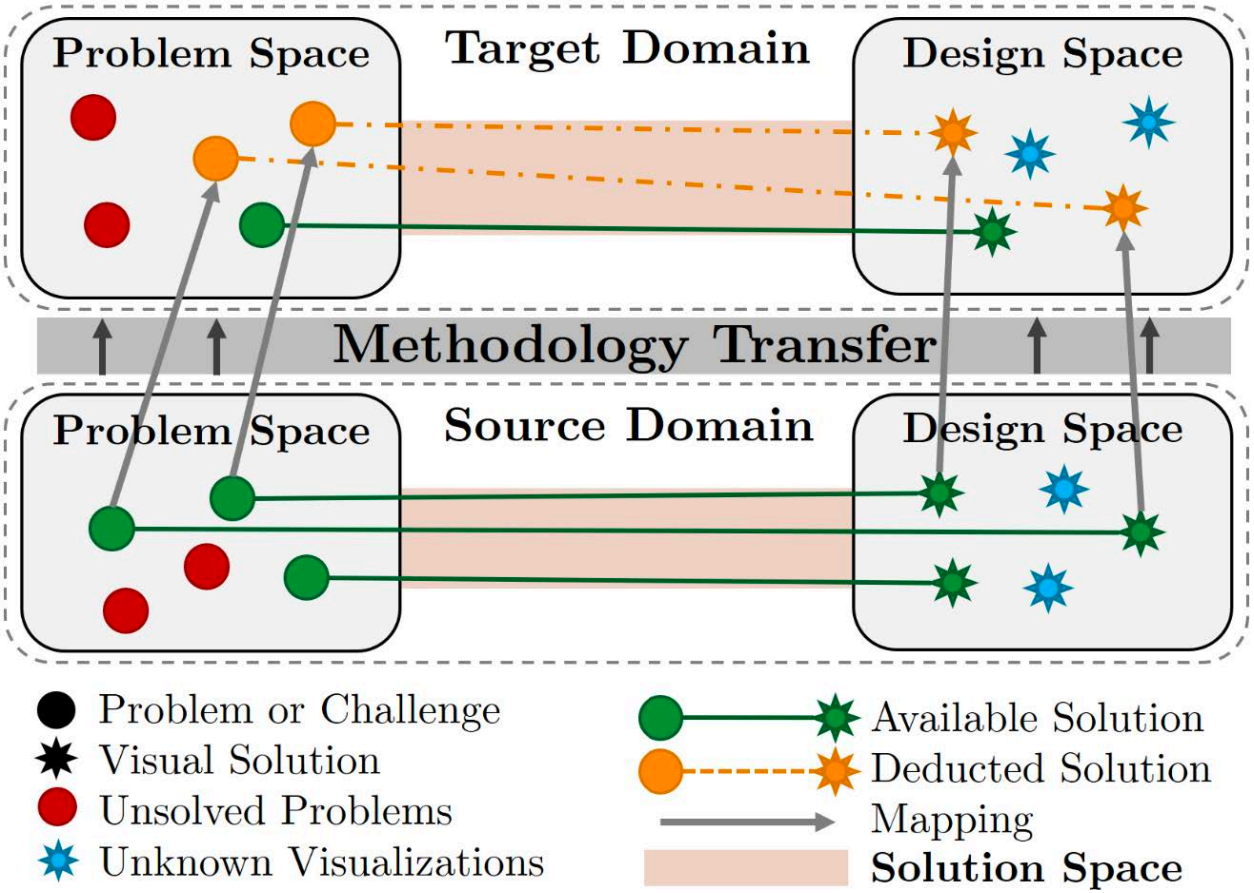
[14] M. Wattenberg, ‘Arc diagrams: Visualizing Structure in Strings’, in IEEE Symposium on Information Visualization, 2002. INFOVIS 2002., Oct. 2002, pp. 110–116, doi: 10.1109/INFVIS.2002.1173155.

Methodology transfer



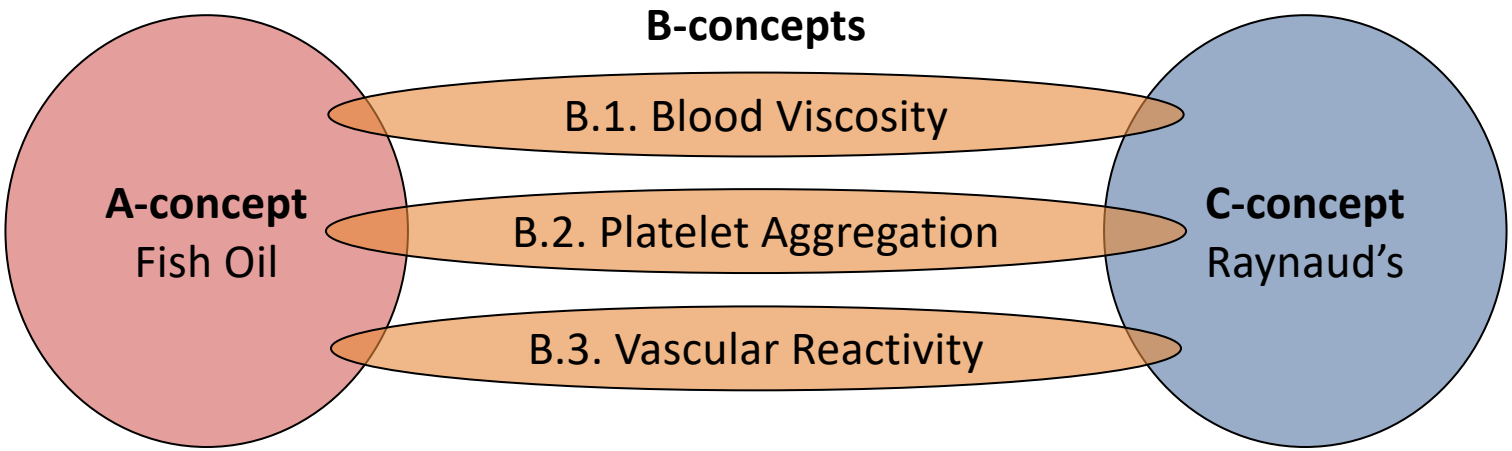
[15] M. Miller, H. Schäfer, M. Kraus, M. Leman, D. A. Keim, and M. El-Assady, 'Framing Visual Musicology through Methodology Transfer', Proceedings of the Workshop on Visualization for the Digital Humanities (VIS4DH) at IEEE VIS 2019, Oct. 2019.

Methodology transfer



[15] M. Miller, H. Schäfer, M. Kraus, M. Leman, D. A. Keim, and M. El-Assady, 'Framing Visual Musicology through Methodology Transfer', Proceedings of the Workshop on Visualization for the Digital Humanities (VIS4DH) at IEEE VIS 2019, Oct. 2019.

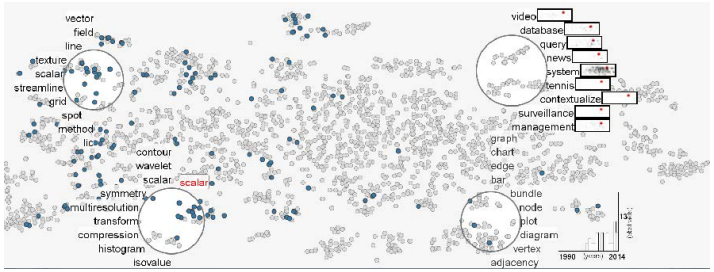
Literature-Based discovery



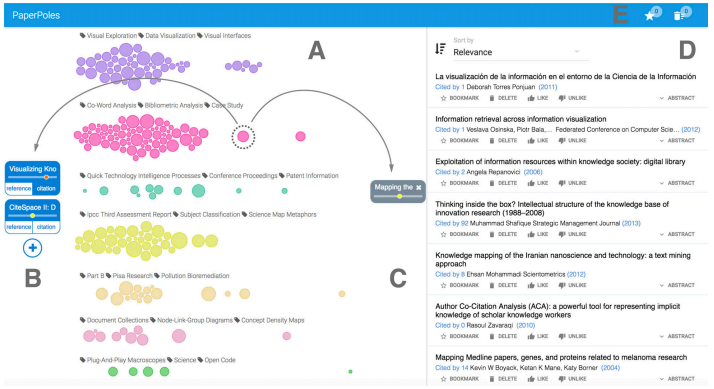
Modern automatic approaches employ similarity scores derived from word embeddings (e.g., word2vec, Glove) to make their findings.

[16] D. R. Swanson, 'Fish Oil, Raynaud's Syndrome, and Undiscovered Public Knowledge', *Perspectives in Biology and Medicine*, vol. 30, no. 1, pp. 7–18, 1986, doi: 10.1353/pbm.1986.0087
[17] M. Thilakaratne, K. Falkner, and T. Atapattu, 'A Systematic Review on Literature-based Discovery', ACM Computing Surveys (CSUR), Dec. 2019
[18] M. Thilakaratne, K. Falkner, and T. Atapattu, 'Automatic Detection of Cross-Disciplinary Knowledge Associations', Jul. 2018, pp. 45–51,

Visual text analytics (VTA)



[19] F. Heimerl, M. John, Qi Han, S. Koch, and T. Ertl. DocuCompass: Effective exploration of document landscapes. In 2016 IEEE Conference on Visual Analytics Science and Technology (VAST), pages 11–20, October 2016.



[20] J. He, Q. Ping, W. Lou, and C. Chen, 'PaperPoles: Facilitating adaptive visual exploration of scientific publications by citation links', Journal of the Association for Information Science and Technology, vol. 70, no. 8, pp. 843–857, 2019, doi: 10.1002/asi.24171.

[21] E. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher, 'Serendip: Topic model-driven visual exploration of text corpora', in 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), Paris, France, Oct. 2014, pp. 173–182, doi: 10.1109/VAST.2014.7042493.

- Novel specialization of a larger research discipline known as visual analytics (VA) that focuses on structured and unstructured textual data.
- Employs NLP, visualization, and text mining techniques to enhance the comprehension of large bodies of text.
- Highly related to the exploration and analysis of scientific corpora (full-texts and metadata).
    - And also to Digital Humanities and cultural heritage!

Open questions in VIS4DH research

(Adapted and extended from [22])

1. How can interactive visualizations support new questions, and new scales of research, in the DH community?
2. How can we encourage DH scholars to seek out visualizations, or collaboration with visualization researchers?
3. How does visualization with a DH focus differ from general research in the visualization community?
4. How can we remove obstacles for humanities scholars wanting to use visual analytics approaches in their research?
5. How can we remove obstacles for visualization practitioners willing to get involved into DH research?

[22] A. J. Bradley et al., 'Visualization and the Digital Humanities: Moving Toward Stronger Collaborations', IEEE Computer Graphics and Applications, vol. 38, no. 6, pp. 26–38, Nov. 2018

Methodology

Bosch, H. (ca. 1494) The Extraction of the Stone of Madness [Oil on board]

Methodology

1. **Observation**: Development of user needs and other insight through the deep study of a novel area of PDVR (visualization for the digital humanities).
2. **Hypothesis formulation**: Design of novel linguistic, text mining, and visual methods that satisfy the user needs gathered in 1.
3. **Observation gathering**: Evaluation of the methods developed in 2 with data originating in the VIS4DH domain.
4. **Contrasting the hypothesis**: Measure the reproducibility of the methods and observations with data from other PDVR domains.
5. **Hypothesis proof or refusal**: Acceptance, rejection, or modification of the developed techniques. Previous steps shall be repeated if necessary.
6. **Scientific thesis**: Synthesis of the results to allow the appropriate communication and reproducibility of the findings obtained in previous steps.

Objectives and Main Hypotheses

Goya, F. (1806) The Death of Julius Caesar [Oil on canvas]

Objectives

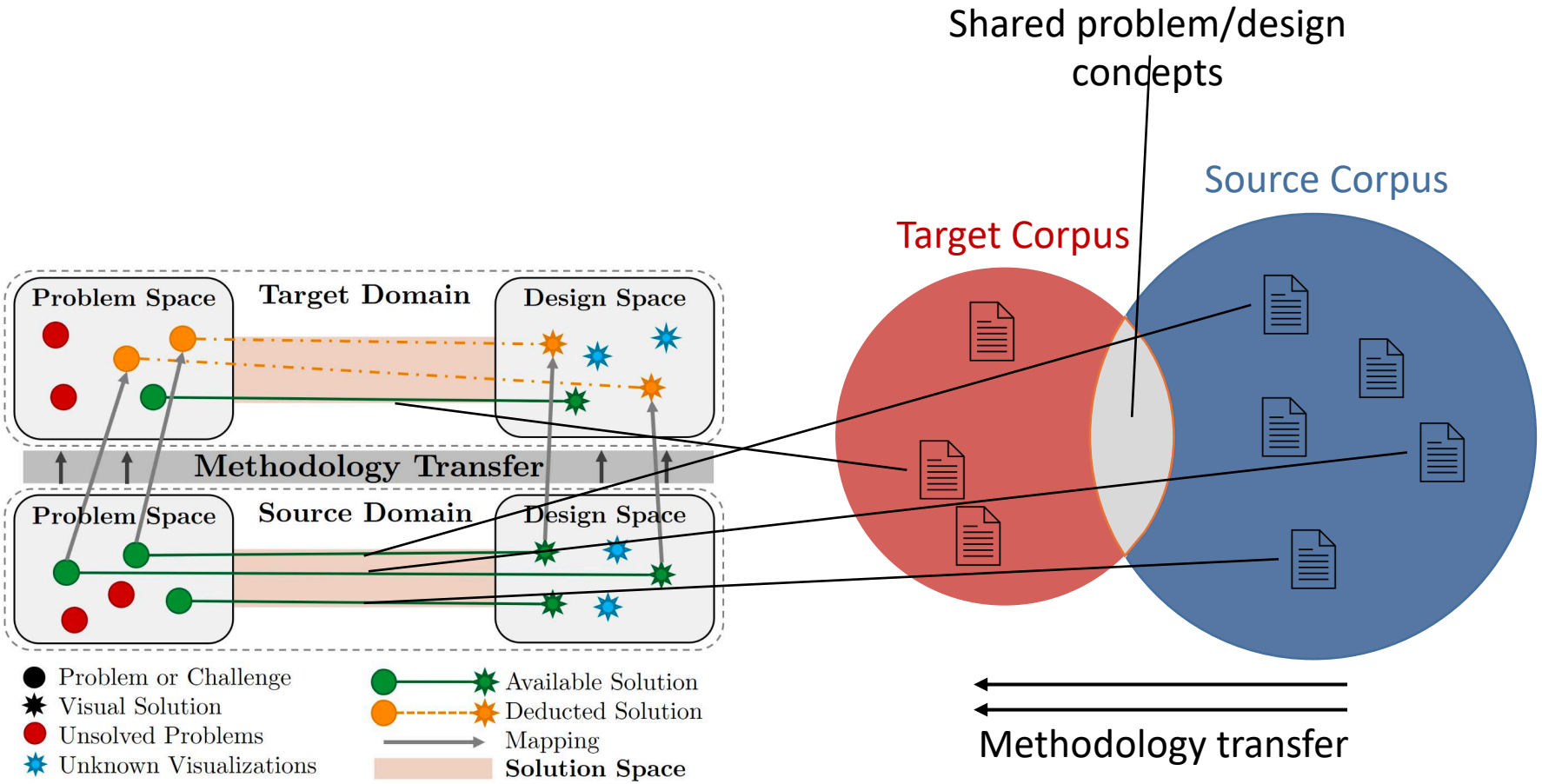
- To understand current **challenges** in **interdisciplinary visualization research**.
- To develop a **methodology** to **detect, frame and map** novel expressions of **PDVR**.
- To study the **properties** of the **language** defined by author-assigned **keywords**.
- To accelerate **knowledge discovery** in the **document exploration** task when performed in an interdisciplinary research context.

Main hypotheses

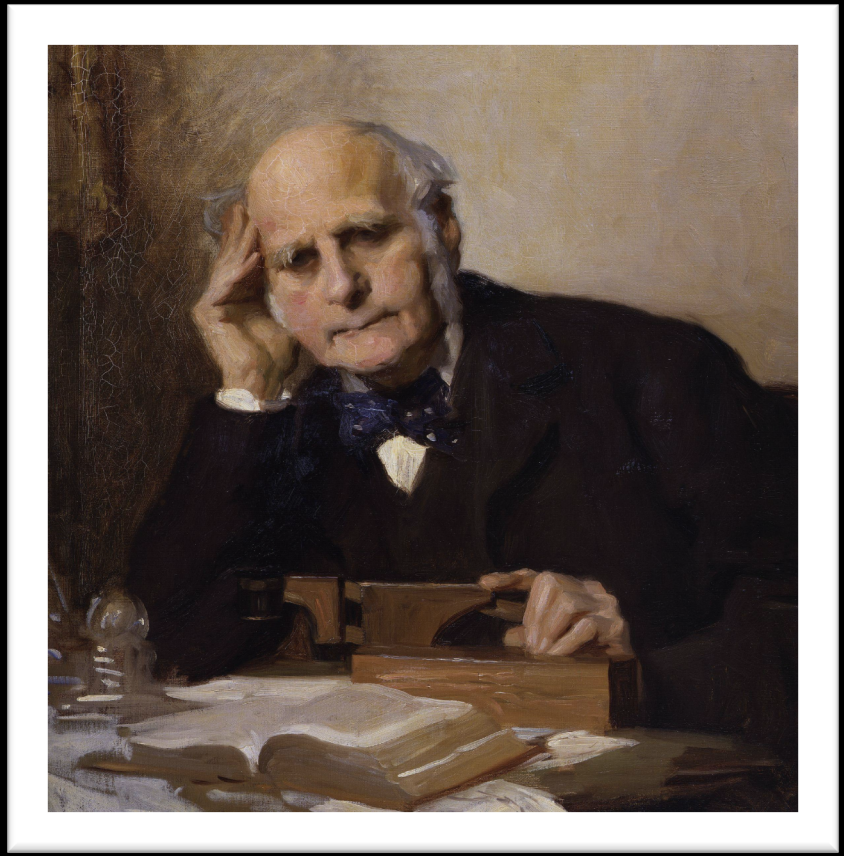
- It is possible to apply **literature-based discovery** and **linguistic analysis** techniques to support, by computational means, the **transfer of methodologies** between application areas in visualization research.

- It is possible to build **interactive systems** that mimic the **sensemaking model** based on **methodology transfer** adopted by interdisciplinary researchers in the **document exploration task**.

Main hypotheses



Data Collection



Furse, C.W. (1903) Sir Francis Galton [Oil on canvas]

Data collection

- Aimed at obtaining a representative dataset of research interests in a key PDVR discipline like visualization for the digital humanities (VIS4DH).
- DH are hard to define
  - 21 definitions in the 2012 edition of Debates in the Digital Humanities.
  - There is not a canonical definition.
- How could we **survey a topic that could not be defined**?
  - Paradox: knowing the input query to obtain relevant documents implied knowing the results of the study.
  - Many DH publications are not indexed in the typical scientific paper databases.
  - DH and VIS authors publish their results in different venues, journals and formats.

Sampling relevant publications

- We obtained a sample of publications made by two types of researchers:
  - A) Visualization researchers with a “declared” interest in DH (VIS4DH)
  - B) Digital humanists with a “declared” interest in visualization (DH4VIS)
- Two main venues:
  - A) Workshop on Visualization for the Digital Humanities (VIS4DH Workshop).
  - B) Alliance of Digital Humanities Organizations
    - DH Conference
    - Digital Humanities Quarterly (DHQ)

1st Workshop on Visualization for the Digital Humanities

Monday, 24 October 2016 - Baltimore, Maryland, USA

2nd Workshop on Visualization for the Digital Humanities

Monday, 02 October 2017 - Phoenix, Arizona, USA

3rd Workshop on Visualization for the Digital Humanities

Sunday, 21 October 2018 - Berlin, Germany

4th Workshop on Visualization for the Digital Humanities

Sunday, 20 October 2019 - Vancouver, Canada

TEI files matching the regular expression [Vv]isua* in their (1) title, (2) list of keywords, (3) list of ADHO topics (pre-defined taxonomy).

(1)

```
<title type="article">Language DNA: Visualizing a Language Decomposition </title>
```

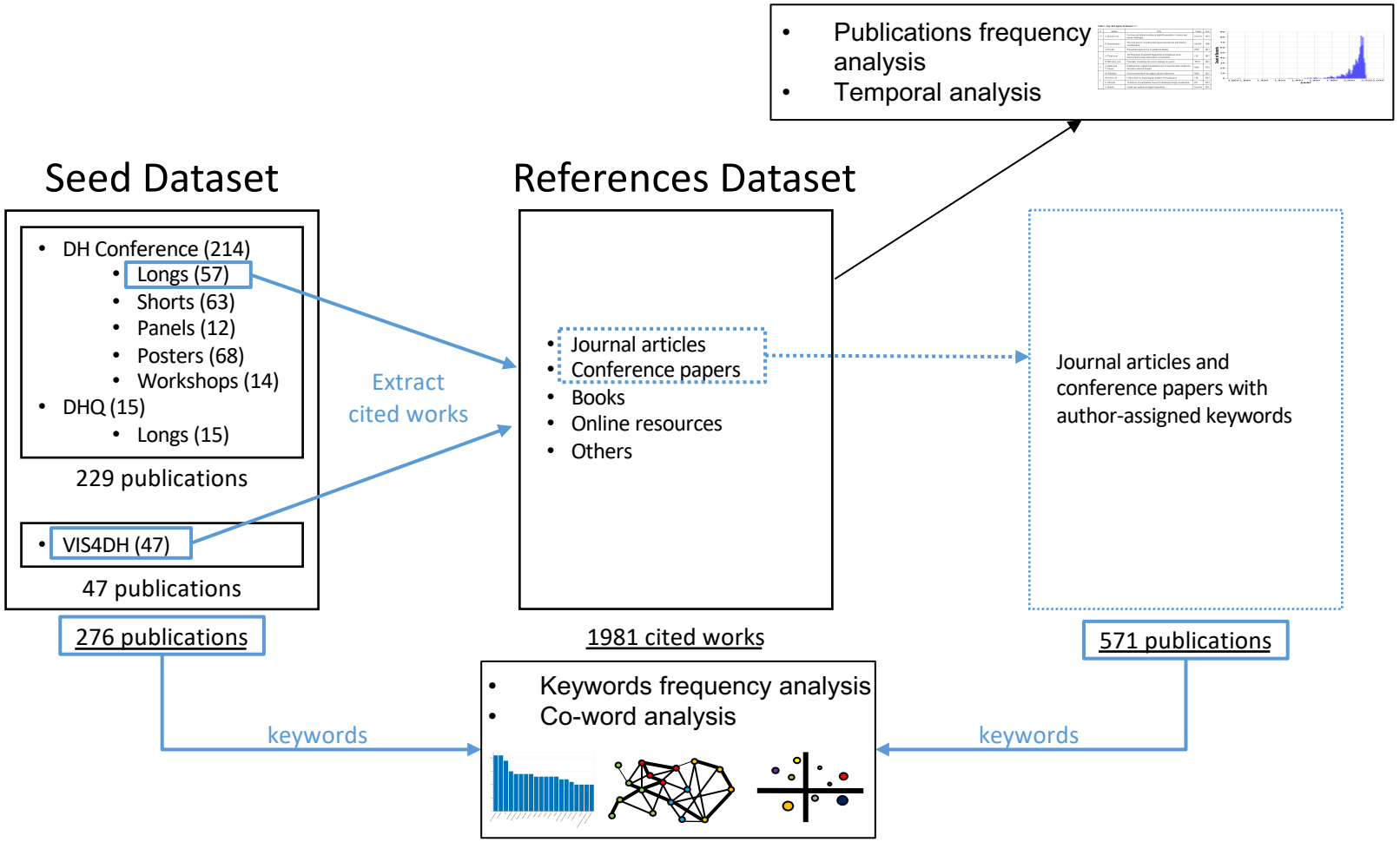
(2)

```
<keywords scheme="ConfTool" n="keywords">
  <term>Topic Characterization; Data Mining; Visual Analytics; Blogging;</term>
</keywords>
```

(3)

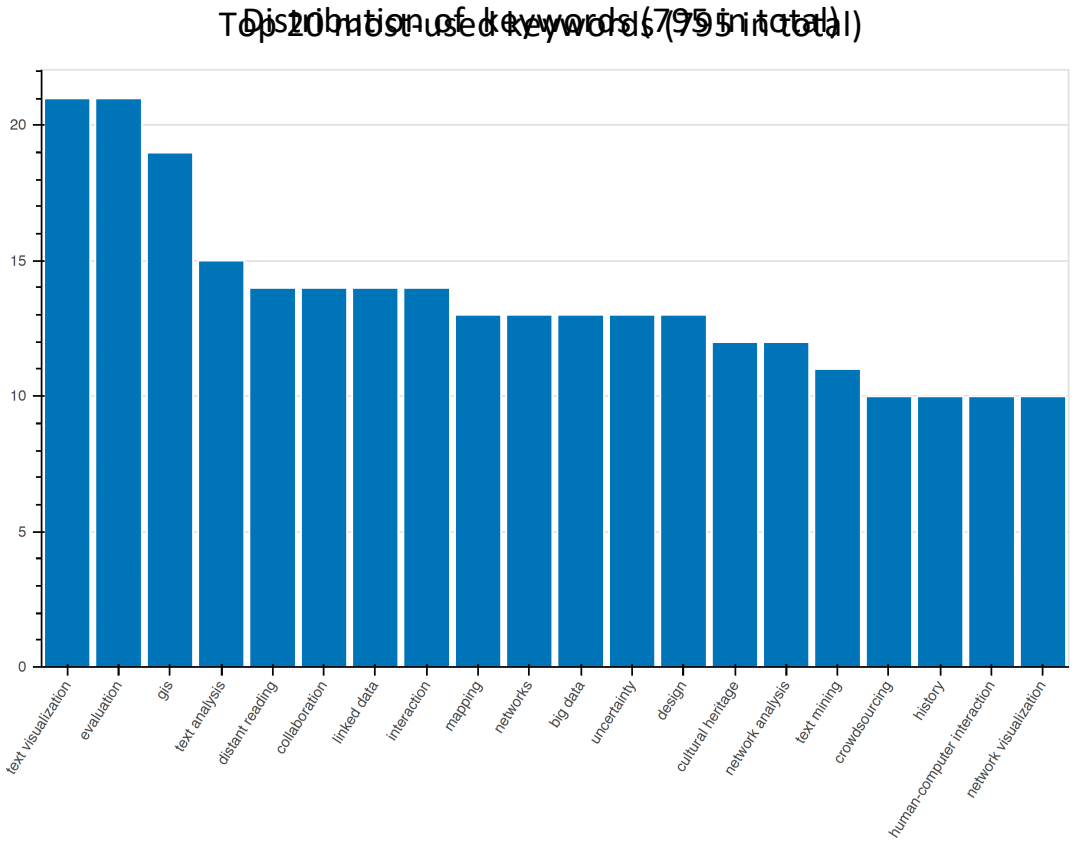
```
<keywords scheme="ConfTool" n="topics">
  <term>interface and user experience design</term>
  <term>literary studies</term>
  <term>text analysis</term>
  <term>knowledge representation</term>
  <term>visualisation</term>
  <term>linking and annotation</term>
  <term>English</term>
```

Analysis methodology



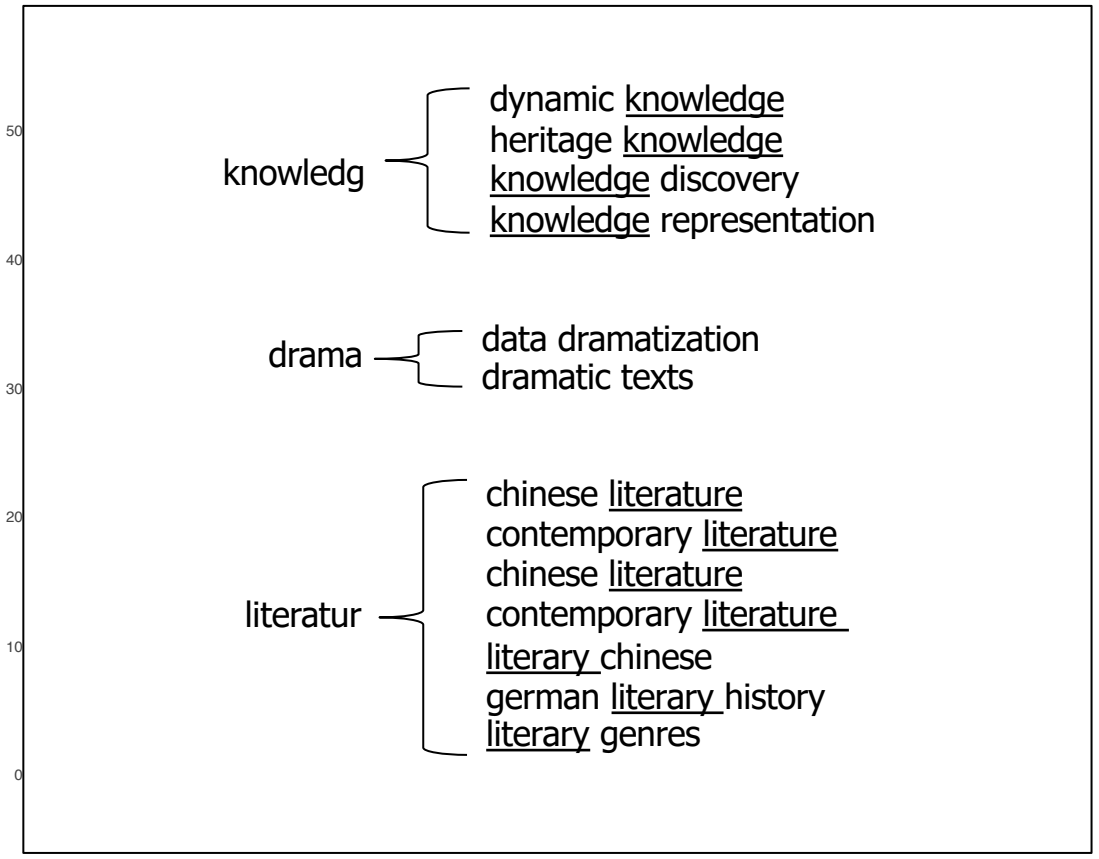
Keyword frequency analysis

- The major DH themes are represented:
  - Text visualization
  - Distant reading
  - GIS, Mapping
  - Networks
- Two **main concerns** of VIS4DH also appeared:
  - Evaluation
  - Collaboration
- Uncertainty?



Keyword analysis

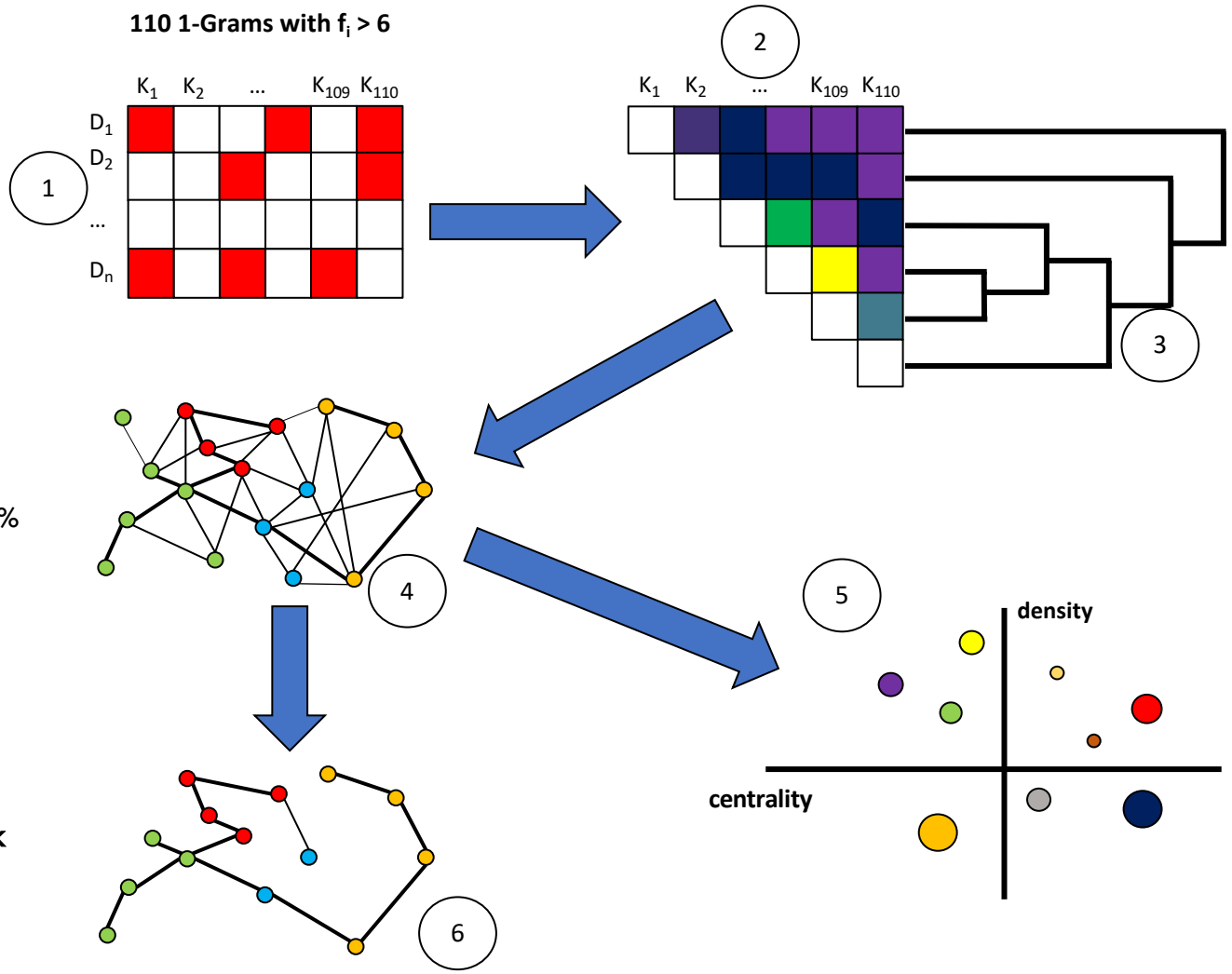
- Tokenization & Stemming
  - Author keywords were tokenized and stemmed using the Porter algorithm.
  - This had the positive effect of compressing the input data and relating n-grams with coincident parts.
  - Stems are similarly distributed to keywords
- Allowed us to avoid an intensive manual coding and classification of the keywords as done in previous studies [12].



[12] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller, 'Visualization as Seen through its Research Paper Keywords', *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 771–780, Jan. 2017

Co-occurrence analysis

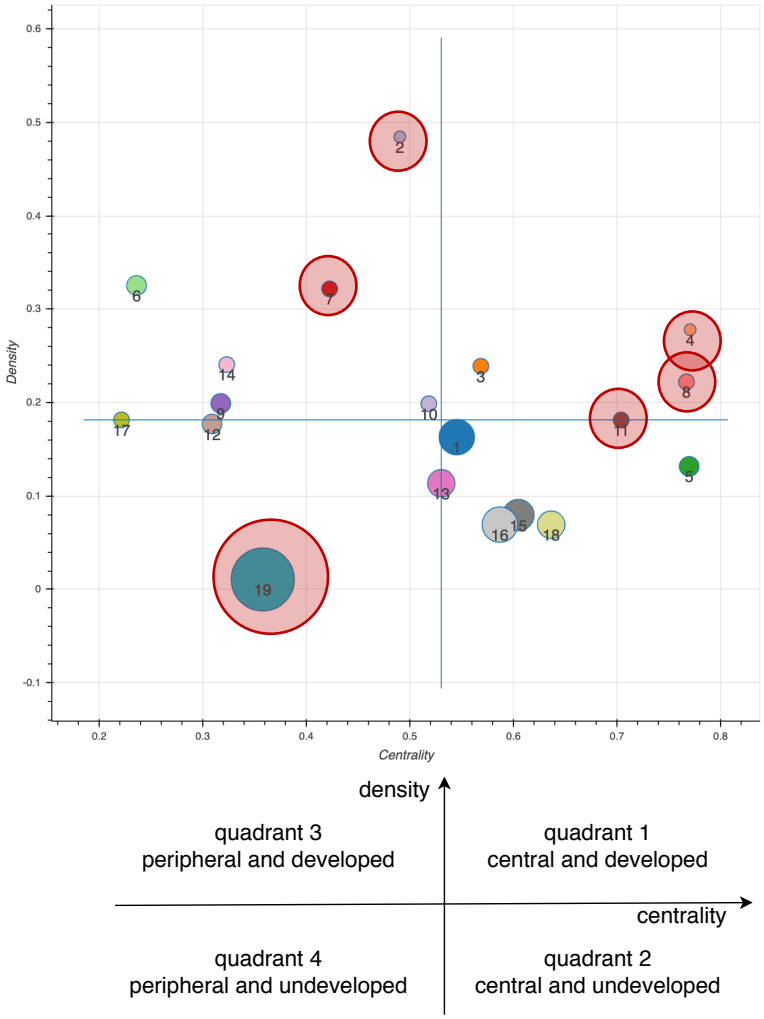
1. **Term-doc matrix**
   - 2511 unique 1-gram keywords
   - 110 1-grams appearing more than a threshold frequency (6).
   - 375 documents.
2. **Keyword correlation matrix**
   - Scipy's corrcoef function.
3. **Hierarchical clustering**
   - Ward's method.
   - Sq. Euclidean distance.
   - Max distance criterion (<95% max distance between any two pairs)
4. **Keywords network**
   - Edges show positive correlations.
5. **Strategic Diagram**
   1. Density
   2. Centrality
6. **Pruned keywords network**
   - R ≥ 0.20



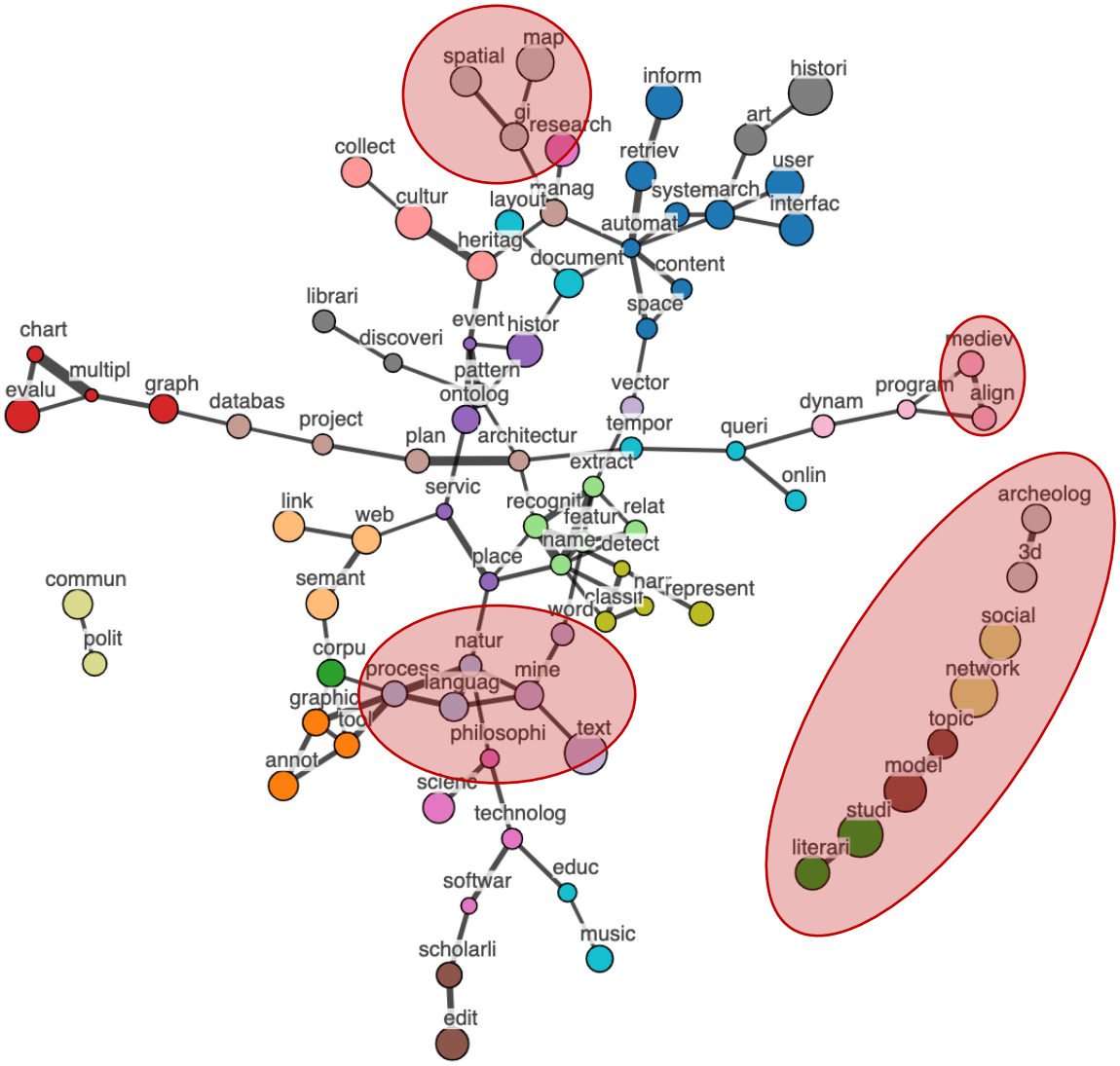
[12] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller, 'Visualization as Seen through its Research Paper Keywords', *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 771–780, Jan. 2017

Strategic diagram

ID	Members (by frequency)
1	**user, inform**, interfac, retriev, search, system, content, space, automat
2	**languag, process**, natur
3	**imag, annot**, graphic, tool
4	**semant, link**, web
5	**studi, literari**, literatur, linguist, corpu
6	**recognit, relat**, extract, featur, name
7	**evalu, graph**, chart, multipl
8	**cultur, collect**, heritag, explor
9	**histor, ontolog**, place, servic, event
10	**text, mine**, vector, word
11	**model, edit**, topic, scholarli
12	**manag, databas**, plan, architectur, project
13	**design, research**, scienc, knowledg, technolog, philosophi, softwar
14	**mediev, align**, dynam, program
15	**histori, collabor**, art, archiv, learn, pattern, librari, discoveri
16	**map, media**, spatial, 3d, archeolog, gi, virtual, mobil, discours
17	**represent, classif**, narr, detect
18	**network, social**, commun, critic, cartographi, polit, theori
19	**interact, video**, uncertainti, document, method, layout, music, tempor, geograph, onlin, structur, concept, educ, queri, cluster, intellig



Keywords network



Results

- A **normalized dataset** of VIS4DH/DH4VIS publications and related works metadata.
  - Keywords
  - Authors
  - Publication venues and years
  - Titles
- **Bibliographic analysis**
  - A quantitative analysis of citation metadata
    - Titles
    - Publication Years
  - An analysis of author-assigned keywords
    - Quantitative analysis of N-grams and 1-Grams (stems)
    - Co-occurrence analysis of 1-Grams (co-word analysis)

Conclusions

- The **study of keywords** revealed several interesting **structural patterns** that characterize the novel field of **visualization for the digital humanities**.
- The **selection of keywords** found in a **document** is a **random draw** from an empirical distribution that **resembles a power law (Zipf´s Law)**.
- **Tokenization and stemming**, when applied to keywords, effectively compress the data **relating multi-word** phrases with **coincident parts**.
  - However, it also introduces some errors that are difficult to solve by automatic means.
  - But it makes the analysis process more reproducible (does not require manual curation/classification as in previous studies [12]).

[12] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller, 'Visualization as Seen through its Research Paper Keywords', *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 771–780, Jan. 2017

Publications

- A. Benito-Santos and R. Therón Sánchez, 'A Data-Driven Introduction to Authors, Readings and Techniques in Visualization for the Digital Humanities', IEEE Computer Graphics and Applications, 2020.
  - Status: **Published**
  - DOI: 10.1109/MCG.2020.2973945
  - Impact Factor (JCR 2019): **1.627**
  - Subject Category: Computer Science, Software Engineering
  - Quartile: (51/108) **Q2**
  - Presented at IEEEVIS 2020. Conference Rating **A+** (GII-GRIN-SCIE 2018)
  - Indexed in IEEE Xplore.

- A. Benito-Santos and R. Therón, 'Pilaster: A Collection of Citation Metadata Extracted From Publications on Visualization for the Digital Humanities', presented at the 5th Workshop on Visualization for the Digital Humanities, collocated to IEEEVIS 2020, Oct. 2020.
  - Status: **Published**
  - Indexed in IEEE Xplore.

Visualization & Visual Analysis Techniques

Pignatta, G. (1712) Portrait of a Cartographer [Oil on canvas]

Design goals

- **D.G.1: Motivate a personalized exploration of scientific corpora that is tailored to the user's research aims.**
  - *What kind of knowledge does the user want to extract from a dataset?*
- **D.G.2: Potentiate the discovery of methodologies that could potentially be transferred from other existing design spaces to the source domain.**
  - *How can we measure the degree of transferability of solutions conceived in other knowledge domains?*

Design goals

- **D.G.3: Accelerate sensemaking and language acquisition in the context of PDVR.**
  - *What are the terms that best describe a dataset according to the user's level of expertise and grounded knowledge?*
  - *What themes are especially interesting for the user?*
- **D.G.4: Provide a reading order for discovered documents.**
  - *What documents are the most important for the user?*

Distributional similarity

Academic papers
(defined by keywords)

Target Corpus
(known to the user)

Paper #T1: B-concept$_1$, B-concept$_2$, **A-concept**, B-concept$_4$, B-concept$_5$

More similar

Source Corpus
(to be explored by the user)

Paper S1: B-concept$_1$, B-concept$_2$, C-concept$_1$, B-concept$_4$, B-concept$_5$
(4 matches with T1)

Paper S2: B-concept$_1$, C-concept$_2$, C-concept$_1$, B-concept$_4$, B-concept$_5$
(3 matches with T1)

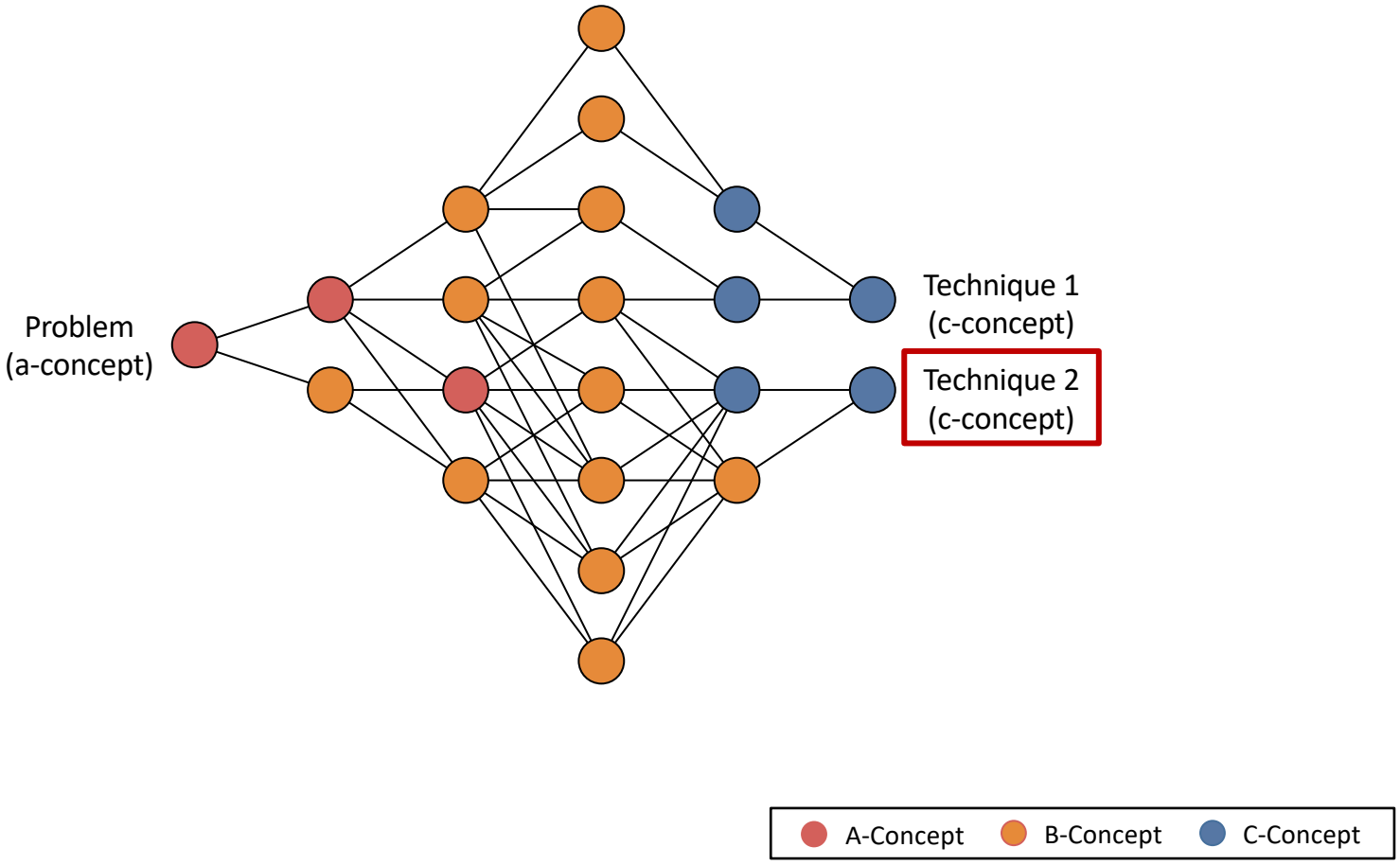
Paper S3: B-concept$_1$, C-concept$_2$, C-concept$_1$, C-concept$_3$, B-concept$_5$
(2 matches with T1)

Paper S4: C-concept$_4$, C-concept$_2$, C-concept$_1$, C-concept$_3$, B-concept$_5$
(1 match with T1)

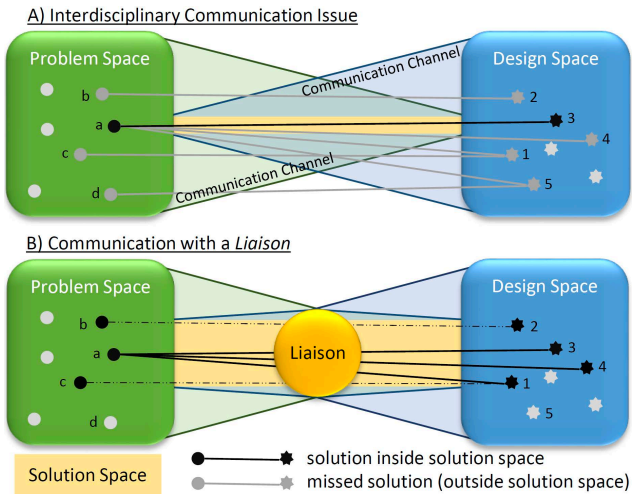
Less similar

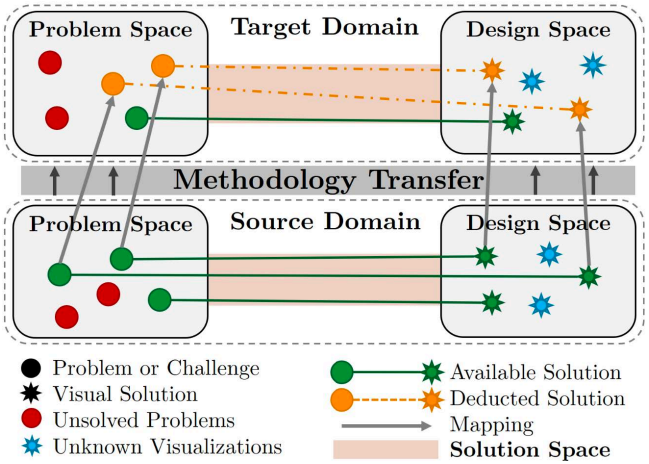
DistribSim (A-concept, C-concept$_1$) ≈ 1

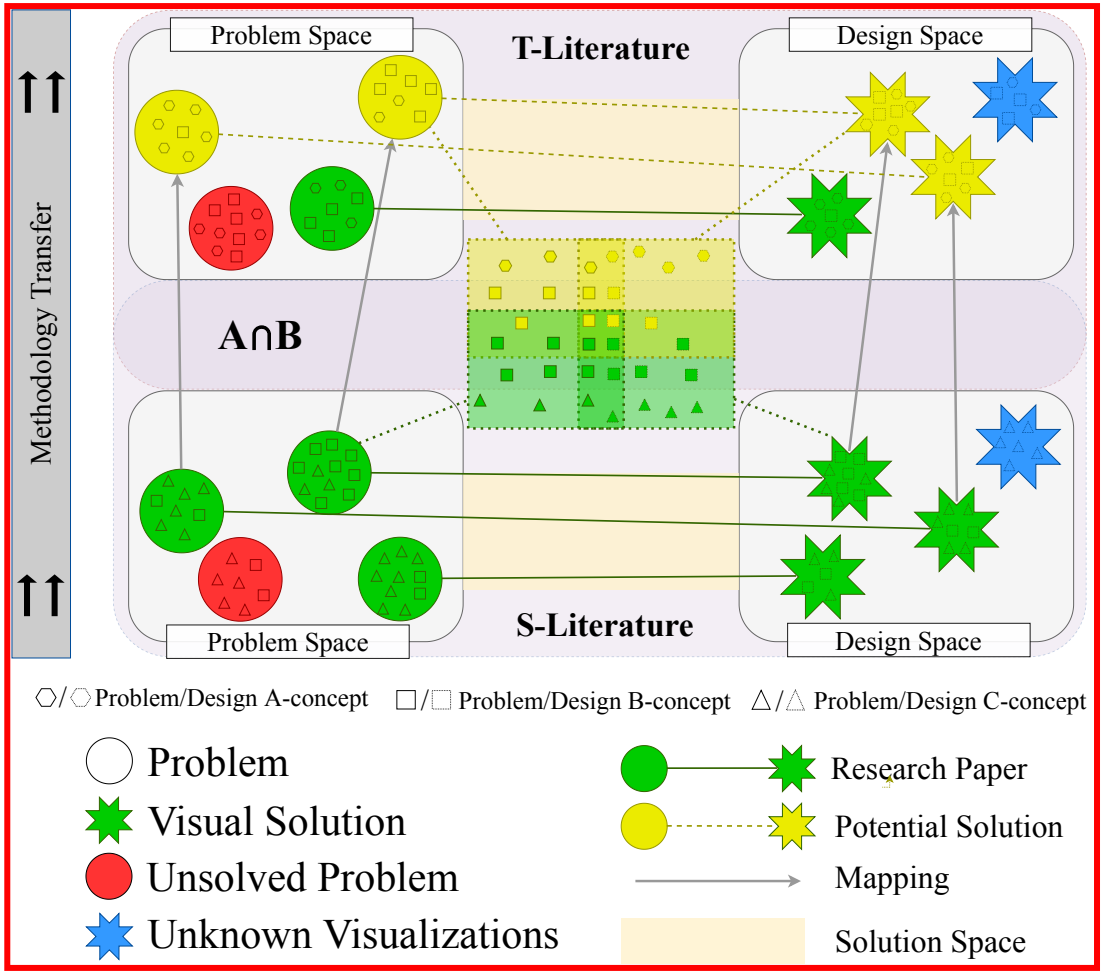
High-order co-occurrence



Exploration model







Previous models

Our model

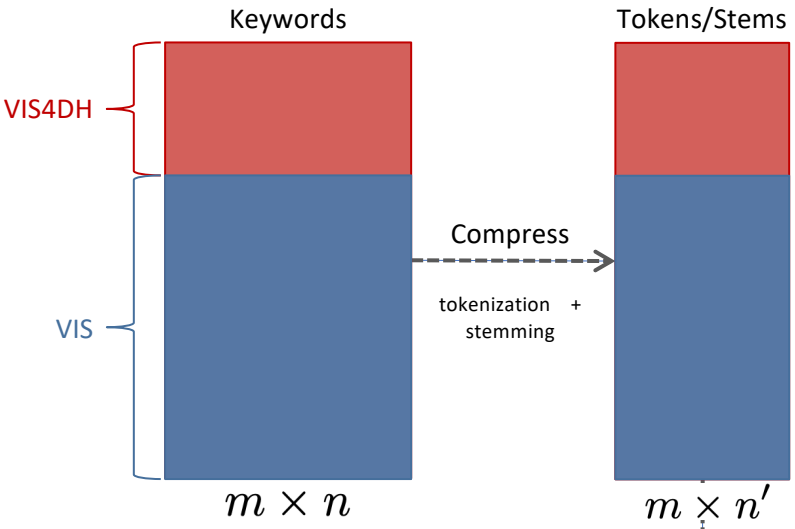
Keyword embeddings

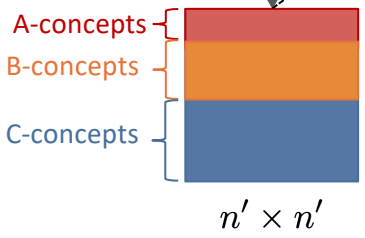
[12] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller, 'Visualization as Seen through its Research Paper Keywords', *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 771–780, Jan. 2017

[23] P. Isenberg et al., 'Vispubdata.org: A Metadata Collection About IEEE Visualization (VIS) Publications', IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 9, pp. 2199–2206, Sep. 2017, doi: 10.1109/TVCG.2016.2615308.

[24] O. Levy, Y. Goldberg, and I. Dagan, 'Improving Distributional Similarity with Lessons Learned from Word Embeddings', Transactions of the Association for Computational Linguistics, vol. 3, no. 0, pp. 211–225, May 2015.

[25] O. Levy and Y. Goldberg, 'Neural Word Embedding as Implicit Matrix Factorization', in Advances in Neural Information Processing Systems 27, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2177–2185.



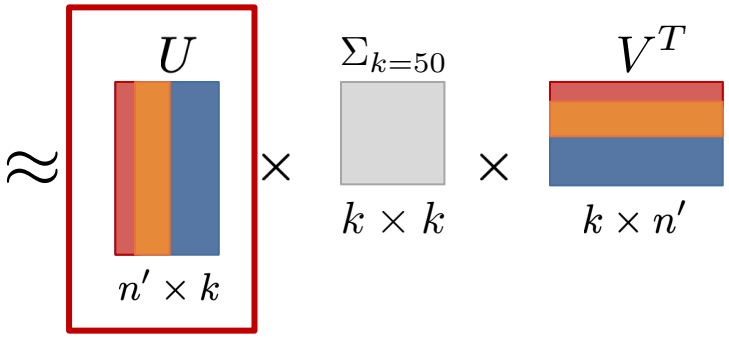


$$PMI(w,c) = \log \frac{P(w,c)}{P(w)P(c)} = \log \frac{\#(w,c) \cdot |D|}{\#(w) \cdot \#(c)} \quad (1)$$

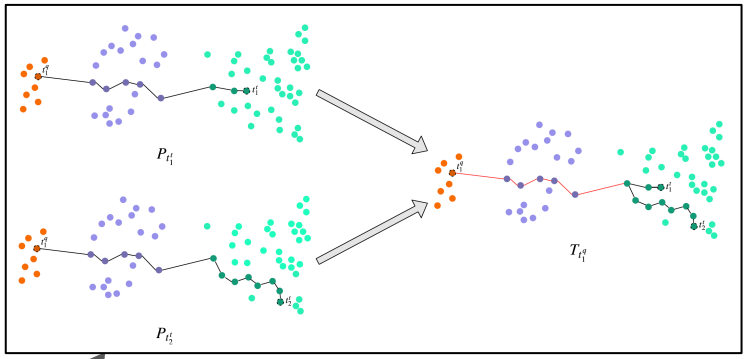
$$PPMI(w,c) = max(PMI(w,c), 0) \quad (2)$$

$$SPPMI(w,c) = \log \frac{\hat{P}(w,c)}{\hat{P}(w)\hat{P}_\alpha(c)} \quad (3)$$

$$\hat{P}_\alpha(c) = \frac{\#(c)^\alpha}{\sum_c \#(c)^\alpha} \quad (4)$$



Similarity matrix



1. Significant paths

U

$n' \times k$

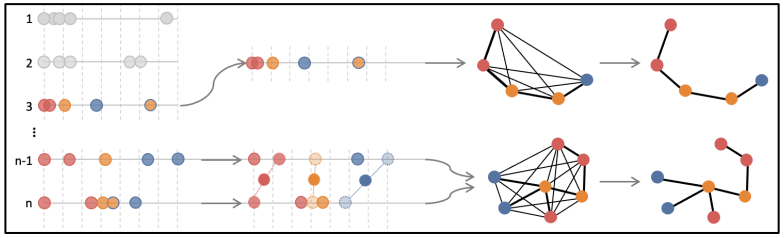
Pairwise Distance
Synonym detection

S

$n'' \times n''$
$n'' < n'$

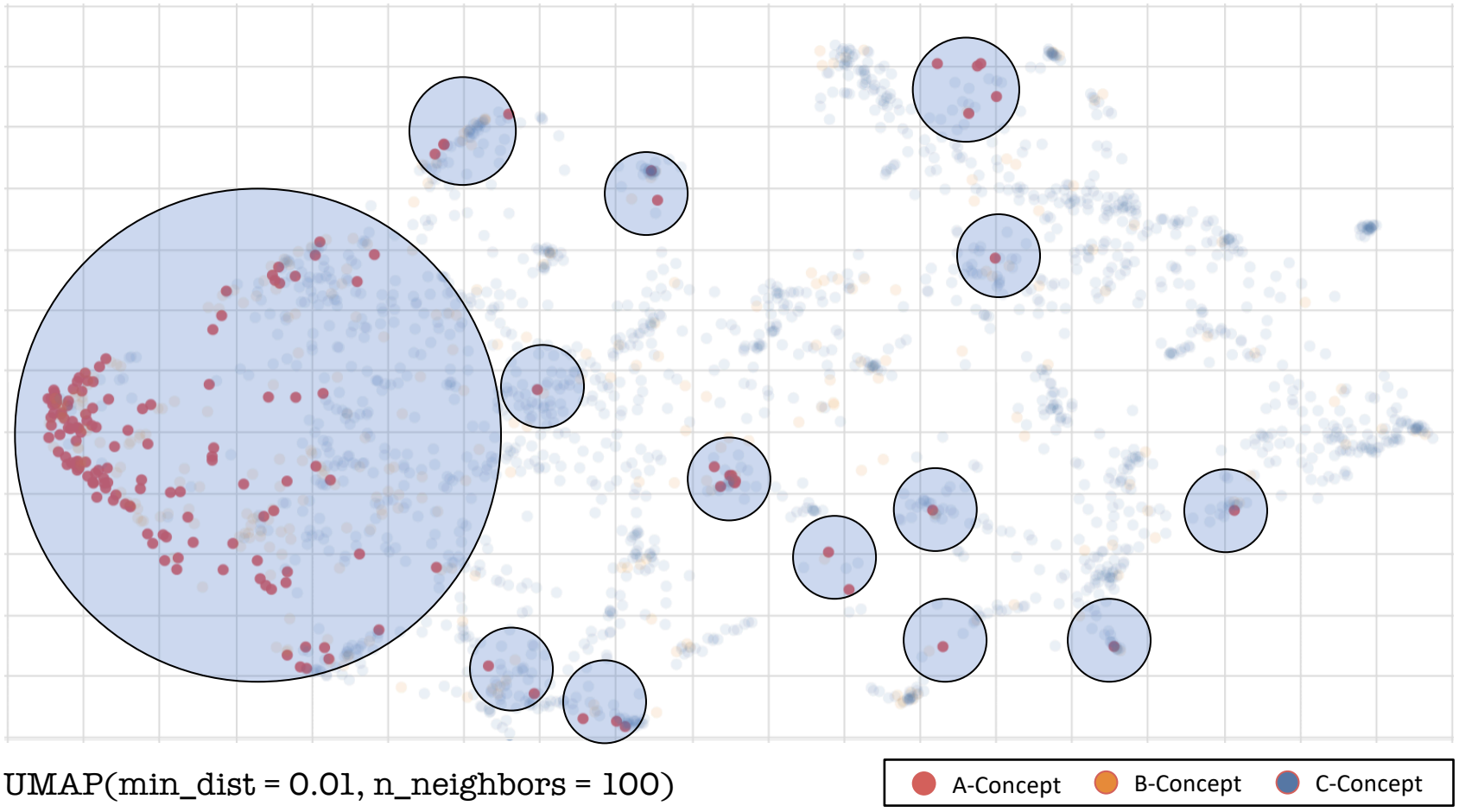
$$D(x,y) = cos(x,y) = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2}}$$

$$S(x,y) = 1 - D(x,y)$$



2. Quality neighborhoods

UMAP 2D projection



UMAP(min_dist = 0.01, n_neighbors = 100)

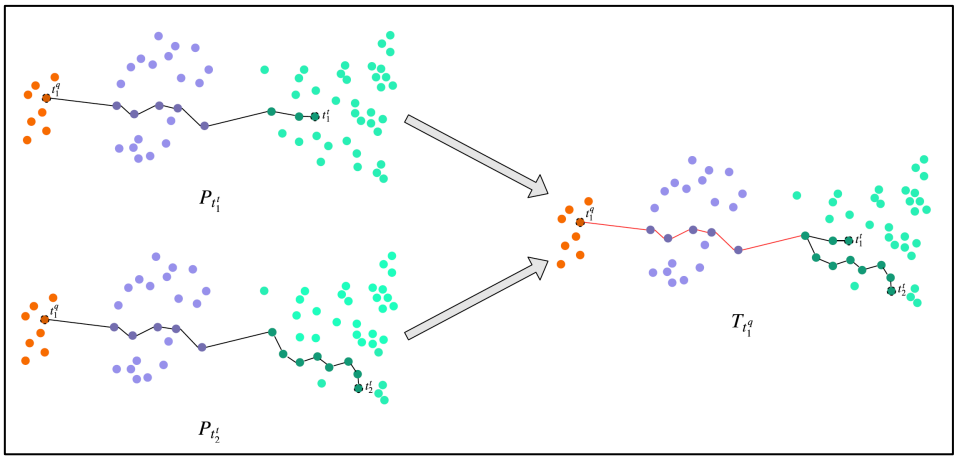
[26] L. McInnes, J. Healy, N. Saul, and L. Grossberger, 'UMAP: Uniform manifold approximation and projection', The Journal of Open Source Software, vol. 3, no. 29, p. 861, 2018.

Significant paths

- Significant path: a path that maximizes the similarity score of all possible paths that connect every c-concept to a given fixed a-concept.
- The process is repeated for every a-concept until all a-concepts are assigned a unique path.
- In the end, every a-concept will be related to at least one c-concept

$$sim(t_j^t, t_i^q) = 1 - \min_{P \in P'_j} \{ \sum_{k=1}^{l} dist(t_k, t_{k+1}) \mid (t_k, t_{k+1}) \in P \}$$

Significant paths



co-retweet
co-citation
controversy
duplications
geography-based
interreflections
mood
paper-reference
pathfinder
socialnetsense
tree
tree-based
twitter

faces
adversarial
anonymization/anonymity
applied
artificial
active/activity
attention/attentionally
awareness/aware
blue
call
chromosome
debugging
experimental/experimentation
explainability/explainable
explanatory
forests/forest
graphlet
healthcare
inference
large-scale
lstm
mixed-initiative
peer
privacy
q-network
random
recommendation/recommender
retweeting
rule
sampling/sample/sampled
situation/situational/situated
user-steerable
vis4ml
visual-interactive

temporospatial
integration / integral / integrating / integrated
photogrammetry
astronomy
boids
context-aware
cosmology
exponents
ftle
fusion
finite-time
hyperspectral
inertial
lagrangian
oil
particles/particle
photogrammetry
texture/texturing/textured
tunneling/tunnel
vectorfield
vrml
wind

encyclopedia
adjacency
attack
automotive
betweenness
centrality/centralized
concentration
confluent
dataflow
financial
grand
intrusion
lenses
market
modular
paradigm
play
readability
security
triage
vulnerability
world-in-miniature
wormhole

bibliography
close/closed
allocation
atomistic
body
box
cartesiangrid
centered/centering
cubic
evenly-spaced
face-centered
latent
lattice
propagation
strategy
topic/topics

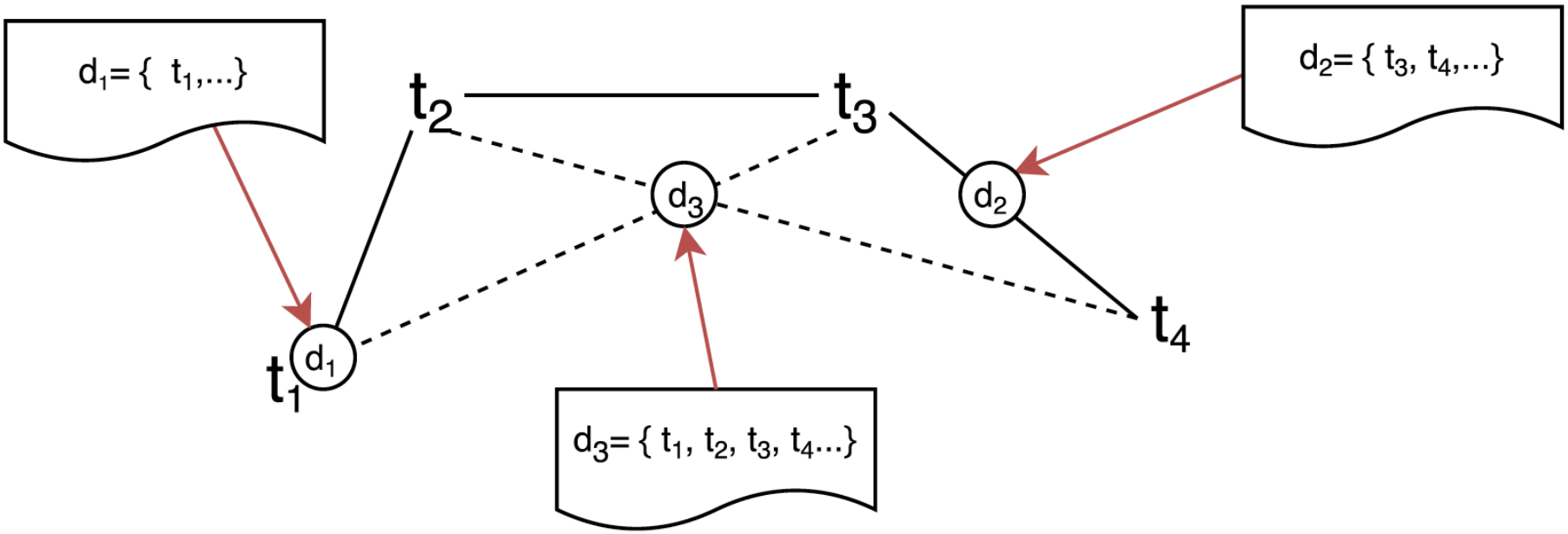
editor
concept
ajax
color-contrast
common
confusion
contact
cue/cues
deficiency
embedder
exemplar
expected
hasse
incomplete
interdomain
matrices
mds
node-link
photos
proximity
recoloring
results/result
seeking
spring
ui

hypergraph
2d1/2
directx
egocentric
essential
faceted
k-d
multi-focus+context
recall
size
slider
streamgraph
timeline/timelines
updating
undo

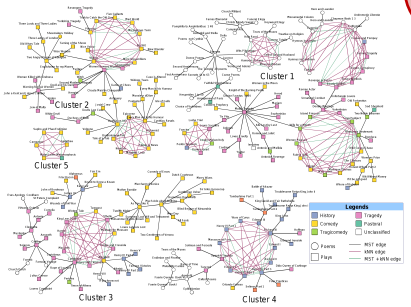
audiovisual
collection/collective
co-located
entity-based
foraging
intelligent/intelligence
meta-visualization
note
tabletop
treatment
workspace

- Paths with coincident intermediate nodes are merged under their common minimum spanning tree [1].
- Produces a variant of spectral clustering.

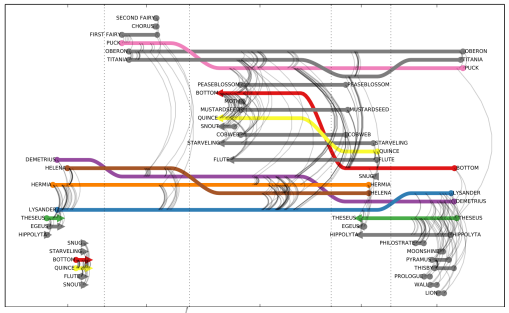
Significant paths



- Each path is projected into the plane using the Kamada-Kawai force-directed graph layout algorithm
  - Distances in the plane represent distributional similarity.
- After keywords have been placed, documents in the same subspace are projected accordingly by triangulating their position (i.e., geometrical centroid in Euclidean space).



A. S. Arefin, R. Vimieiro, C. Riveros, H. Craig, and P. Moscato, 'An Information Theoretic Clustering Approach for Unveiling Authorship Affinities in Shakespearean Era Plays and Poems', PLOS ONE, vol. 9, no. 10, p. e111445, Oct. 2014, doi: 10.1371/journal.pone.0111445.



D. L. Arendt and L. M. Blaha, 'SVEN: Informative Visual Representation of Complex Dynamic Structure', *arXiv:1412.6706 [physics]*, Dec. 2014, Accessed: Nov. 26, 2020. [Online]. Available: http://arxiv.org/abs/1412.6706.

shakespeare
2 documents
(DH.1,DH.2)

spine
2 documents

cliques

'Shall I compare thee to a network?': Visualizing the Topological Structure of Shakespeare's Plays
(DH.3)

reeb

Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees

Clique Community Persistence: A Topological Visual Analysis Approach for Complex Networks
(VIS.1)

Augmented Topological Descriptors of Pore Networks for Material Science
(VIS.2)

homology

arrays
Fixed-Rate Compressed Floating-Point Arrays

voxelization
3 documents

Quasi-static approach approximation for 6 degrees-of-freedom haptic rendering

approximation
13 documents

oct-trees

plant
2 documents

Computing Robustness and Persistence for Images

arc
Arc diagrams: visualizing structure in strings
(VIS.3)

sets
23 documents

numbers
An Extension of Wilkinson's Algorithm for Positioning Tick Labels on Axes

Efficient computation of the topology of level sets
9 documents

betti

Counting cases in marching cubes: toward a generic algorithm for producing substitopes

level
19 documents

separating
2 documents

Significant paths: problems

- Although the significant paths technique produced interesting results, it **could not operate in interaction times**.

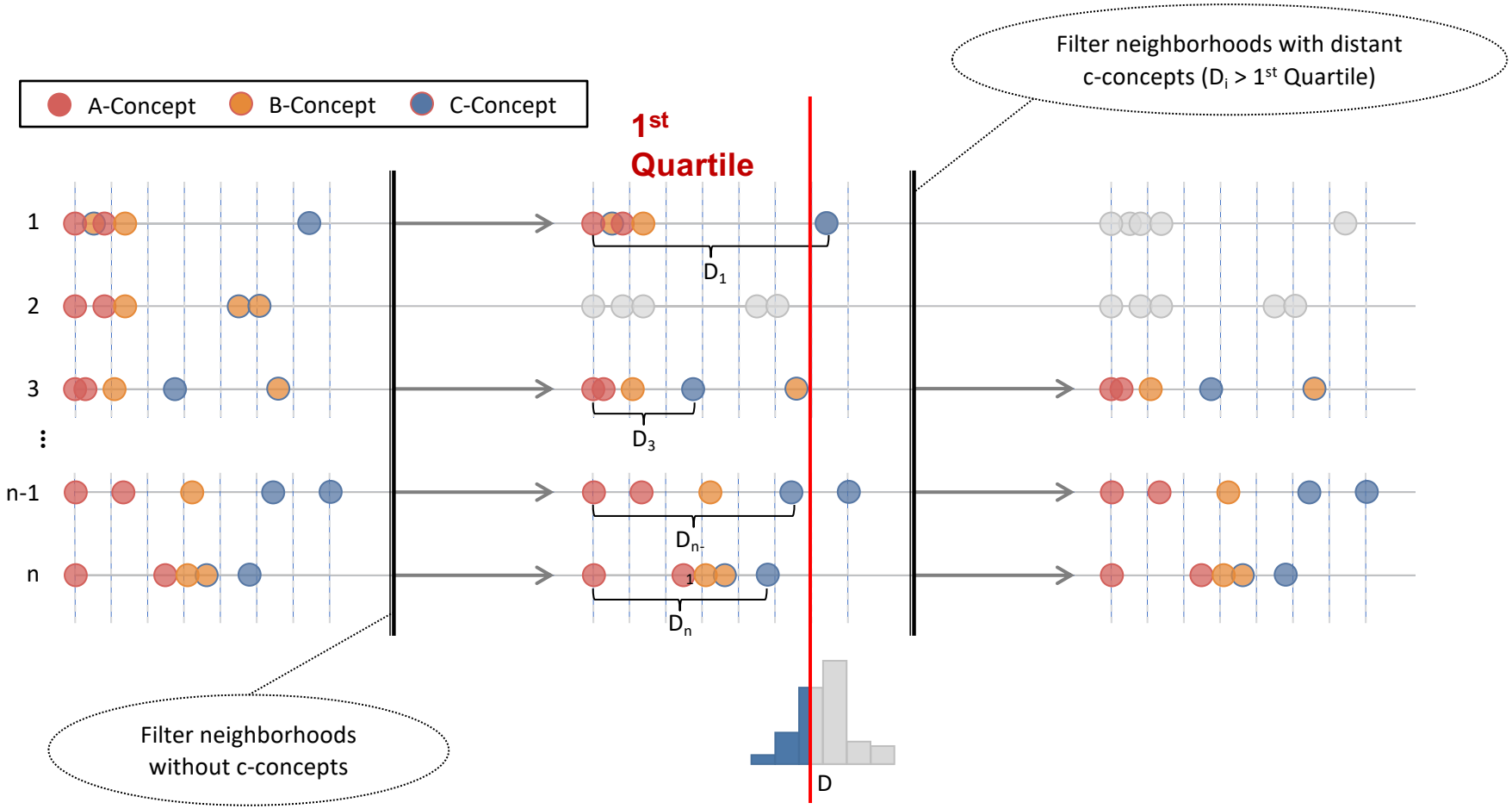
- The visualization was **limited** in interactivity and **it did not allow** to see the contexts in which the represented keywords and documents appeared in the corpora.
  - This synoptic task is key to get a full understanding of the relationship between the two collections [1].

[27] P. Federico, F. Heimerl, S. Koch, and S. Miksch, 'A Survey on Visual Approaches for Analyzing Scientific Literature and Patents', IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 9, pp. 2179–2198, Sep. 2017, doi: 10.1109/TVCG.2016.2610422.

VTA prototype GlassViz

- Interactive VTA application
  - Supports the inspection of cohesive local A-concept **quality neighborhoods** called "**entry points**."
- Linked-Views system:
  - Documents and contexts
    - Rank-frequency lists.
  - Entry points: semantically-cohesive keyword groups
    - Each entry point is shown in a designated area of the main view.
    - Displayed using force-directed layout (Fruchterman & Reingold).

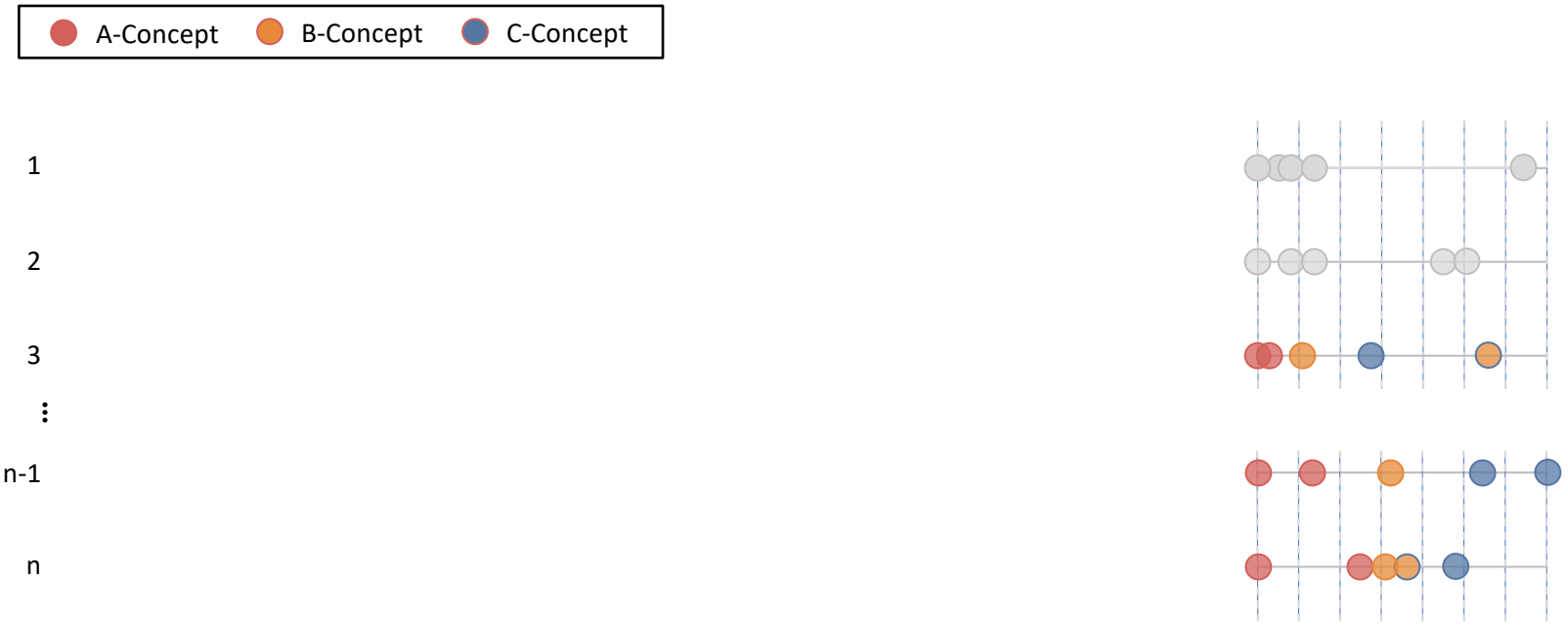
Quality neighborhoods



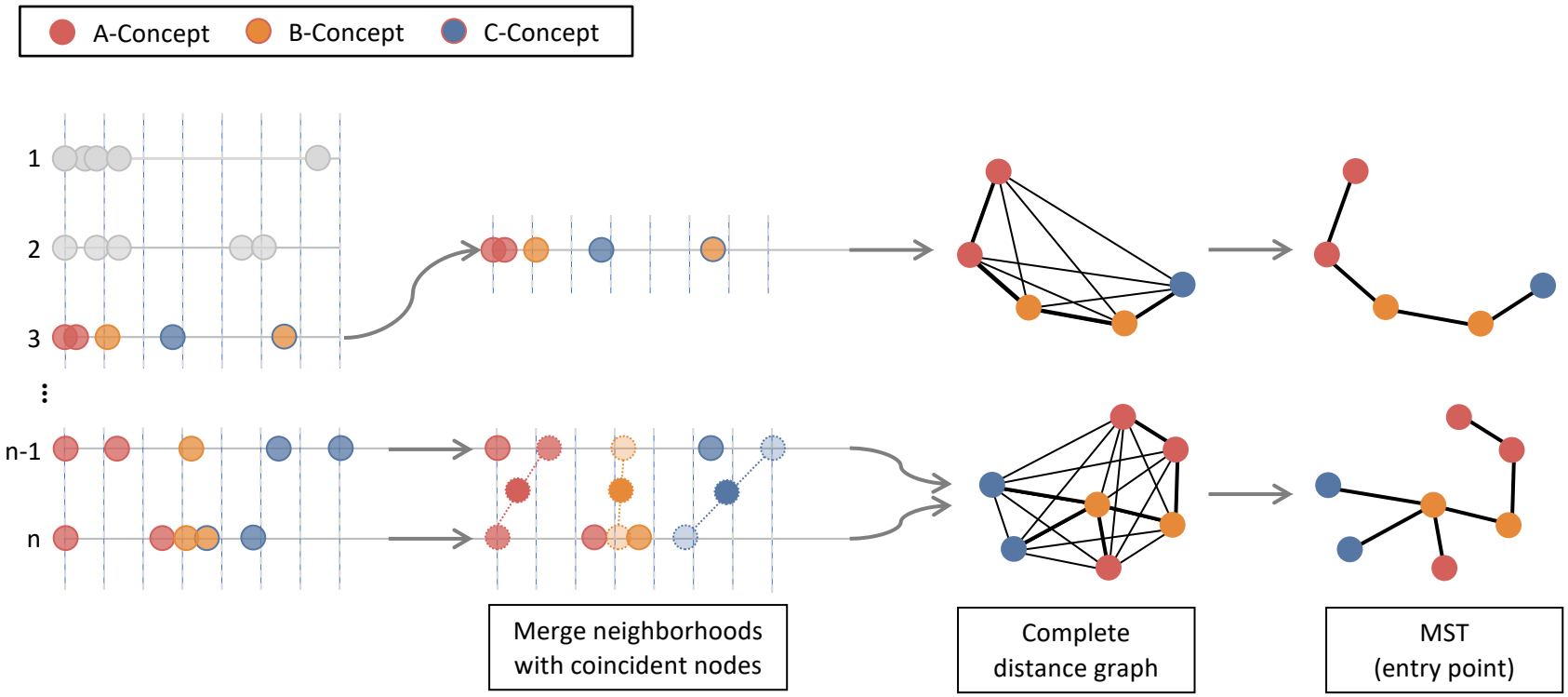
[28] E. Alexander and M. Gleicher, 'Task-Driven Comparison of Topic Models', IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 320–329, Jan. 2016

[29] F. Heimerl and M. Gleicher, 'Interactive Analysis of Word Vector Embeddings', Computer Graphics Forum, vol. 37, no. 3, pp. 253–265, Jun. 2018

Quality Neighborhoods

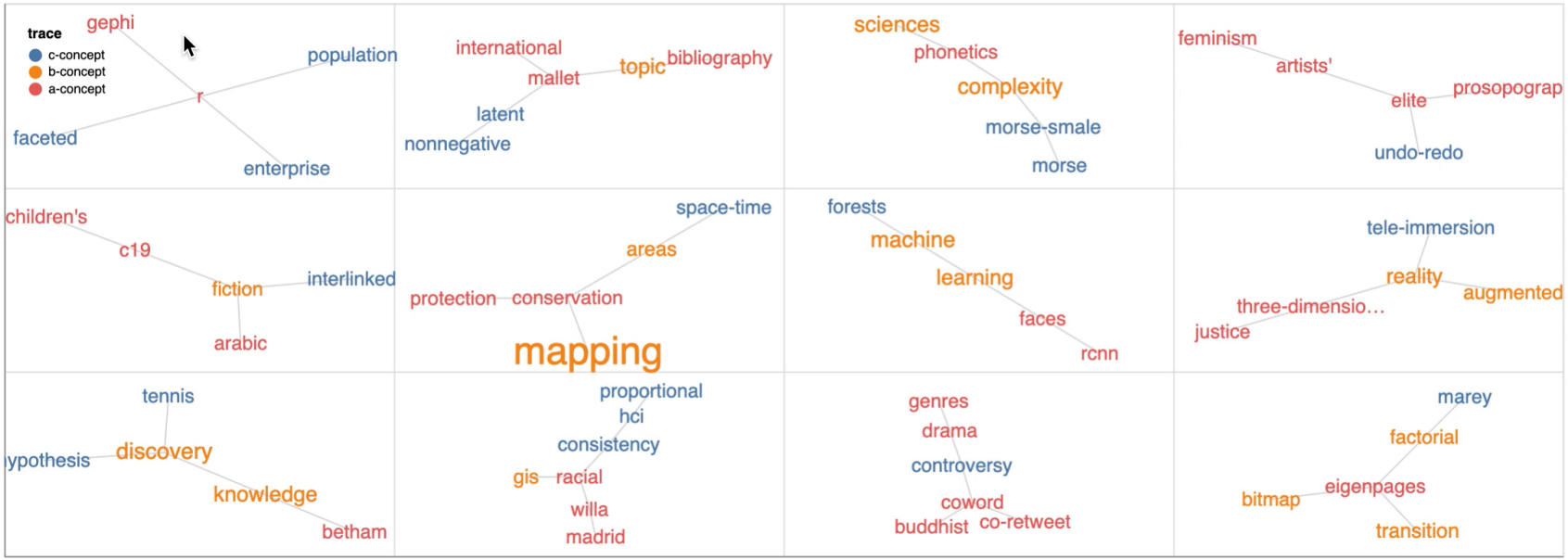


Entry Points



Entry Points

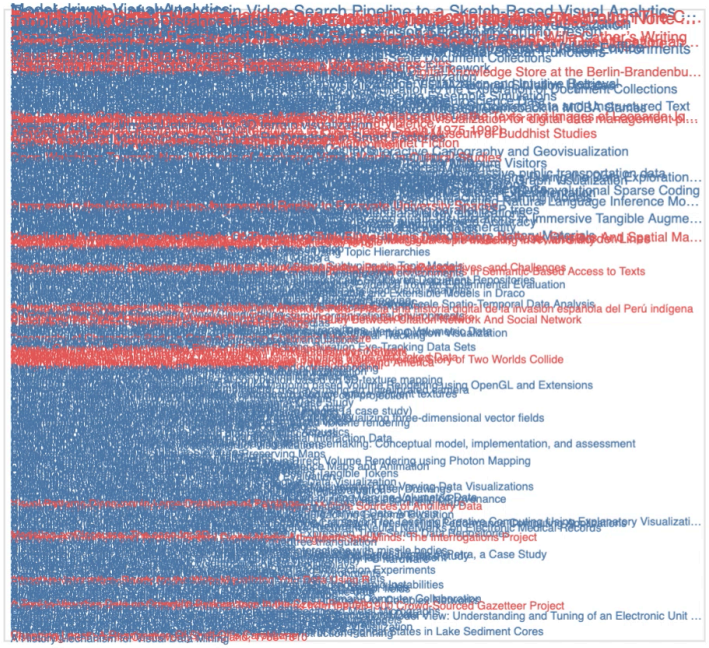
Click and drag while pressing shift to pan & zoom. Hold alt/option(⌥) to brush.



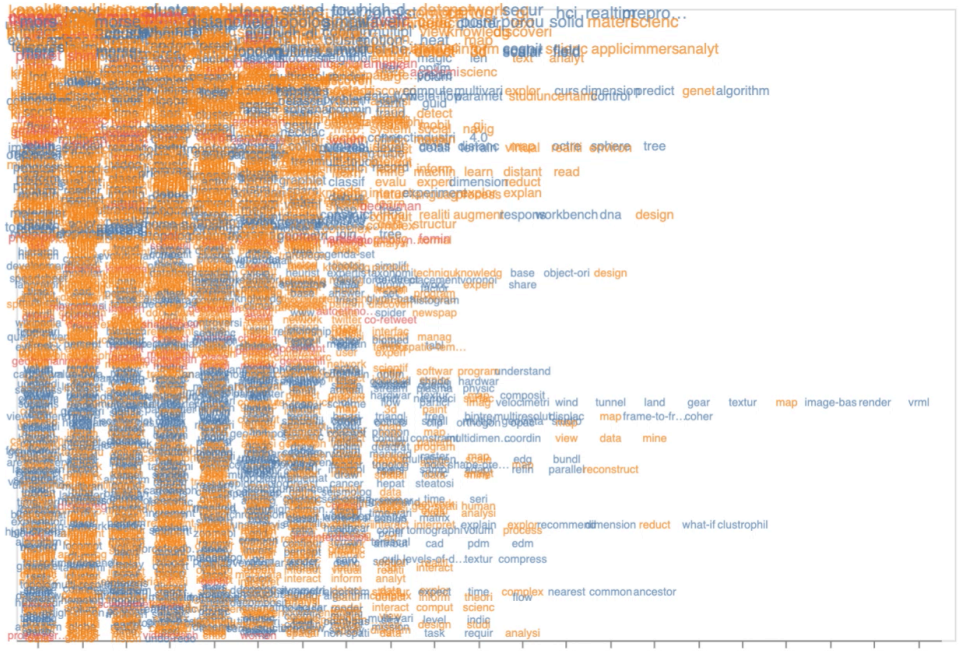
Contexts (by rank)

map
data
analyt
knowledg
learn
analysi
discoveri
machin
interact
realiti
model
inform
scienc
virtual
network
complex
studi
topic
volum
render
topolog
comput
cluster
represent
theori
augment
morse-smal

Papers (by rank)



Keyword Tokens



tele-immersion

reality

augmented

three-dimensio…

justice

Entry point

DH in 3D: Multidimensional Research and Education in the Digital Humanities

Toward a compelling sensation of telepresence: demonstrating a portal to a distant (static) office

Augmenting the University: Using Augmented Reality to Excavate University Spaces

XR in DH: Extended Reality in the Digital Humanities

Real-time incremental visualization of dynamic ultrasound volumes using parallel BSP trees

Collaborative augmented reali

Documents

three-… virtual reali

telepres tele-im… virtua

augment realiti institu

virtual augment realit

augment realiti ultrasou

virtual environ scient

Tokenized Document Keywords

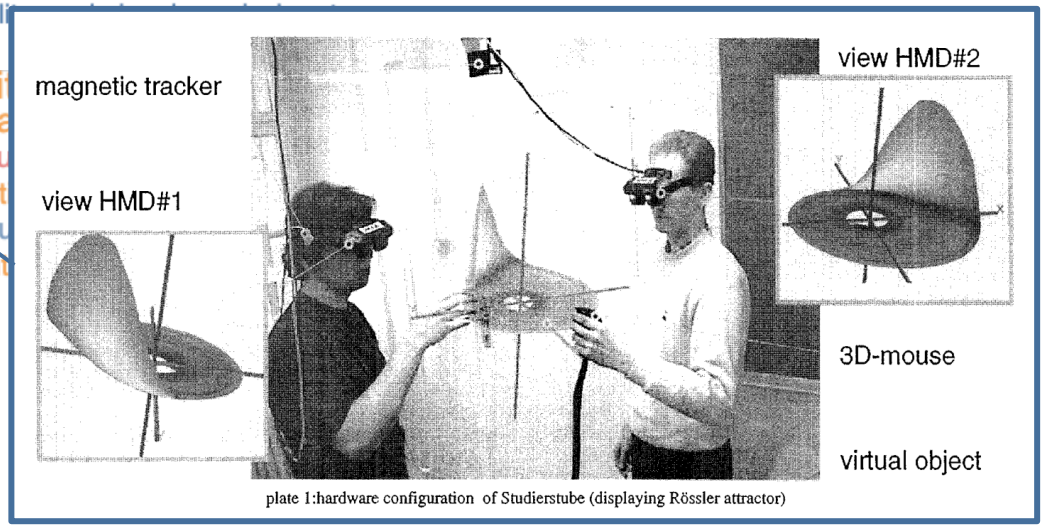
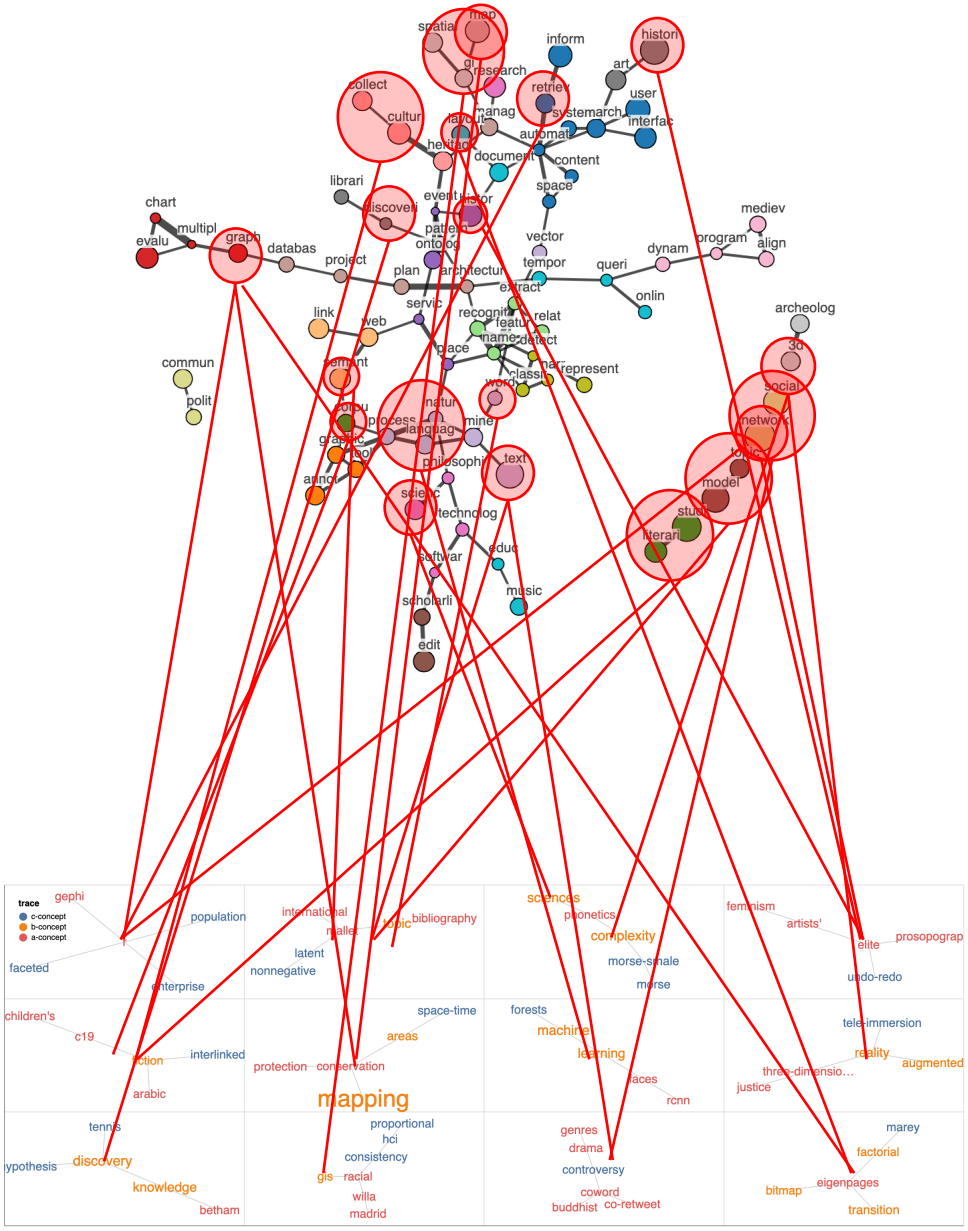


plate 1:hardware configuration of Studierstube (displaying Rössler attractor)

Show the interesting

- **Entry points** display the **most cohesive** inter-collection **keyword associations** as found by the model.
  - They generally connect **domain problems** to potential visualization solutions and problems in other **domains**.
- They also capture many **significant themes** of the domain the user is familiar with.



Looking *through*

Publications

- A. Benito-Santos and R. Therón Sánchez, 'Cross-domain Visual Exploration of Academic Corpora via the Latent Meaning of User-Authored Keywords', IEEE Access, vol. 7, pp. 98144–98160, 2019.
  - Status: **Published**
  - DOI: 10.1109/ACCESS.2019.2929754
  - Impact Factor (JCR 2019): **3.745**
  - Subject Category: Computer Science, Information Systems
  - Quartile: (35/156) **Q1**

- A. Benito-Santos and R. Therón, 'GlassViz: Visualizing Automatically-Extracted Entry Points for Exploring Scientific Corpora in Problem-Driven Visualization Research', presented at the 2020 IEEE Visualization Conference (VIS), Oct. 2020.
  - Status: **To be published** in IEEE Xplore (Conf. proceedings)
  - Conference Rating **A+** (GII-GRIN-SCIE 2018)

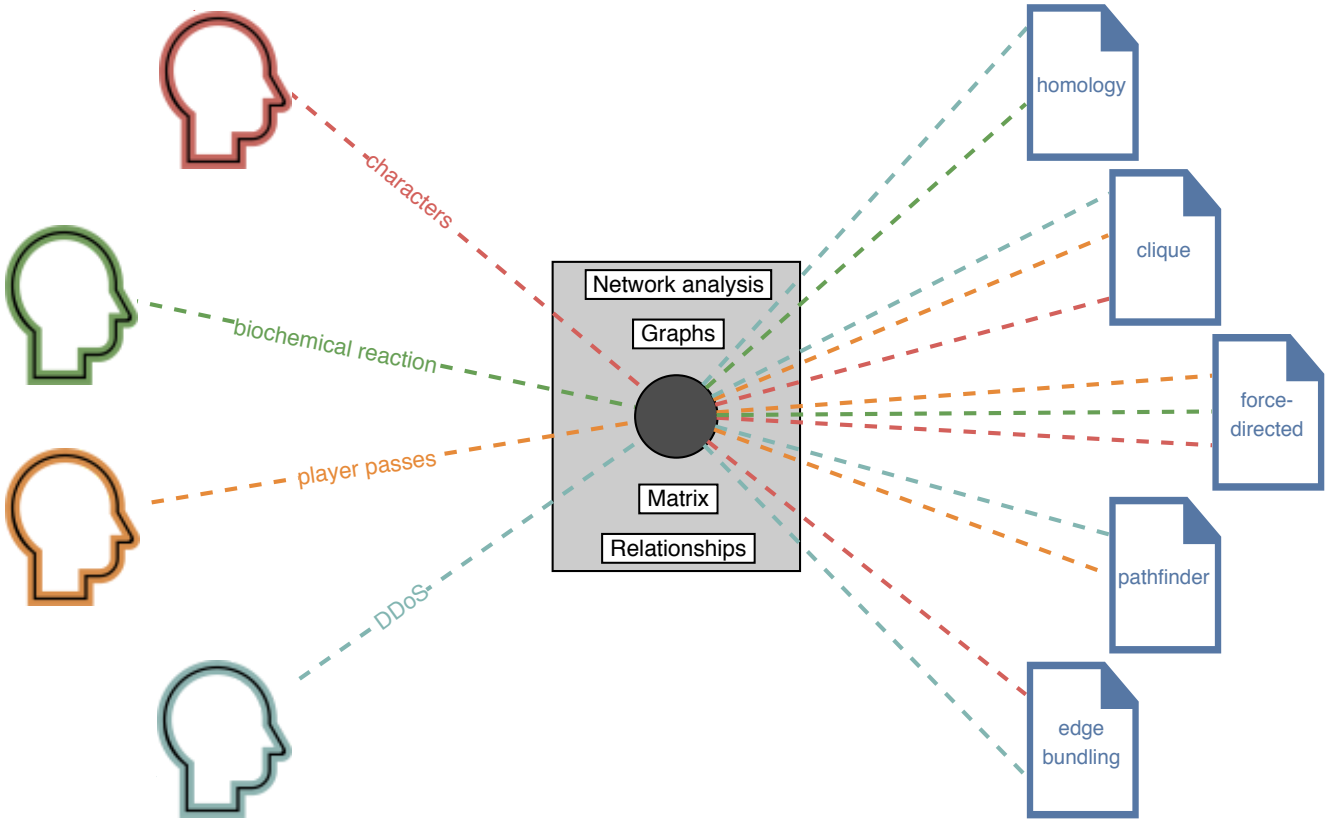
Applicability of the findings to knowledge defragmentation

Llull, R. (ca. 1295), Arbre de ciència (The Tree of Science).
Image from a 1505 edition of the book.

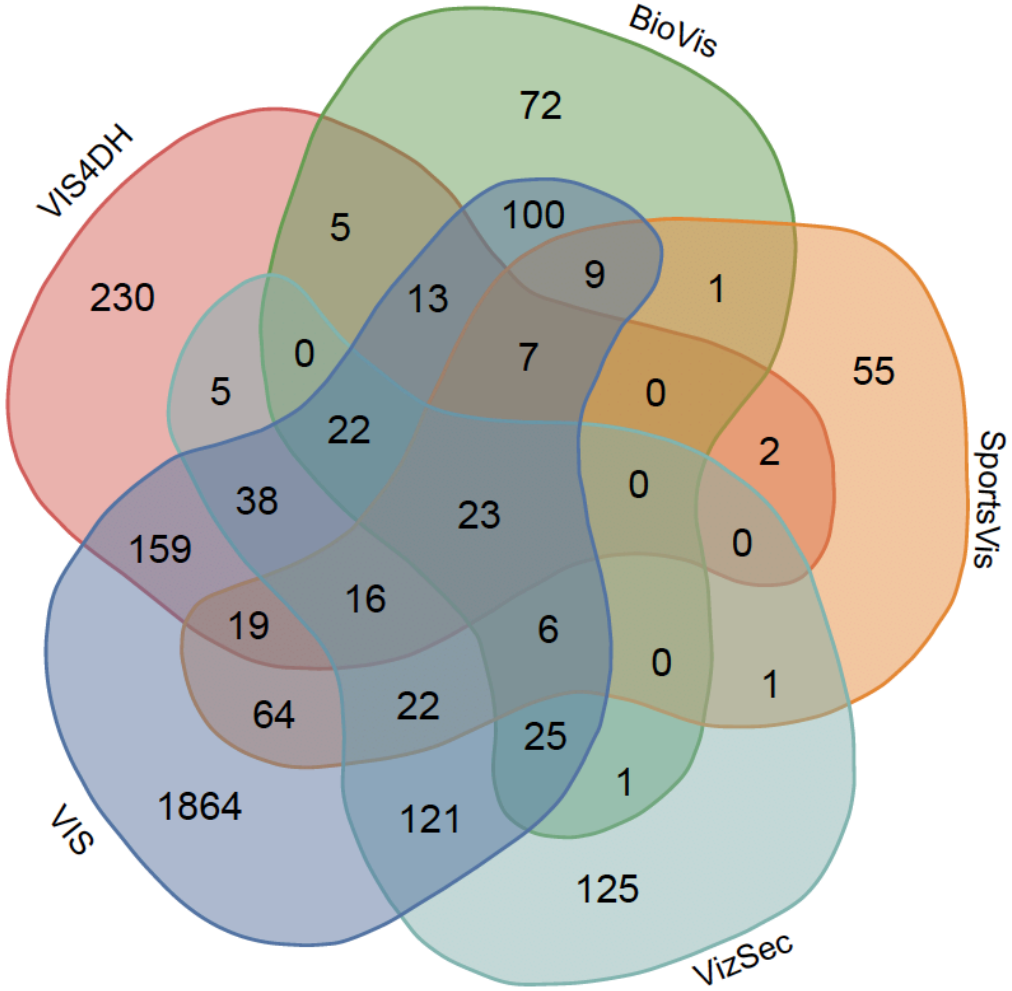
Fragmentation

- The **increasing specialization** of visualization research is motivating the **fragmentation** of the field into **isolated communities of practice**.
- **Potentially redundant** solutions for **generic tasks** are being created, leading to a **waste of time and human resources**.
- In a new study, we provided evidence that our findings are a valid alternative to identify **commonalities and differences** between such communities with the aim of bringing them together.
  - We adapted GlassViz to support this task.
  - We collected 3 more datasets representative of different areas of PDVR (BioVis, SportsVis, VizSec).

Reusing the exploration model

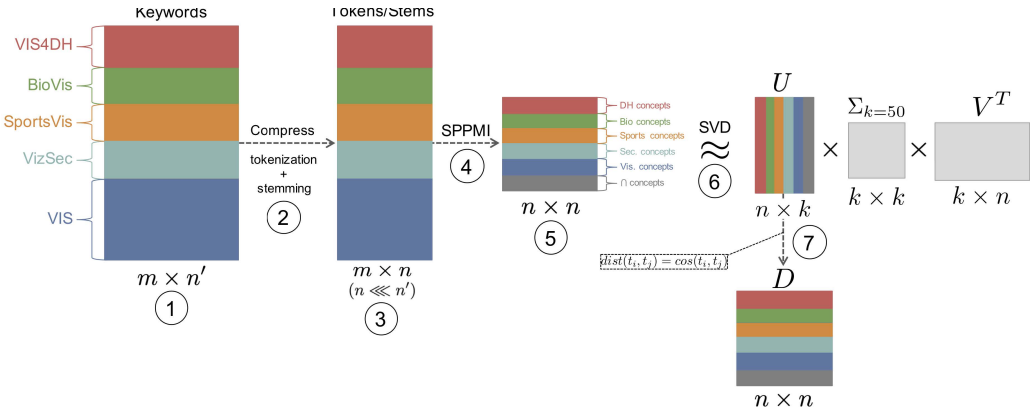


Dataset	# Documents	# Unique Tokens	Avg. Keyword Tokens per Doc.	# Exclusive Tokens
VIS4DH	221	539	4.47 ± 0.99	230 (42.7%)
BioVis	69	284	4.57 ± 1.85	72 (25.4%)
SportsVis	59	225	4.73 ± 1.55	55 (24.4%)
VizSec	175	405	4.63 ± 1.75	125 (30.9%)
VIS	2253	2508	4.66 ± 1.61	1864 (74.3%)

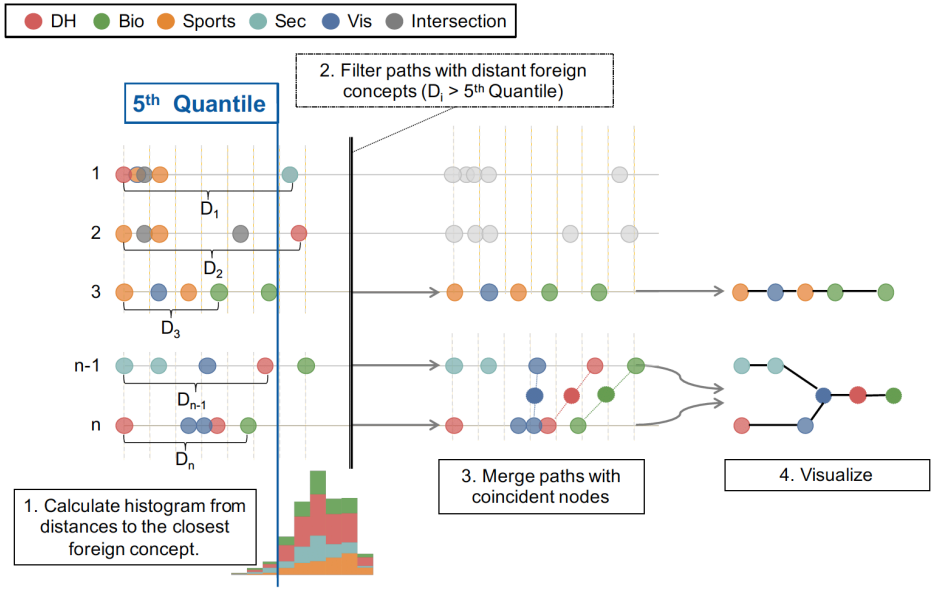


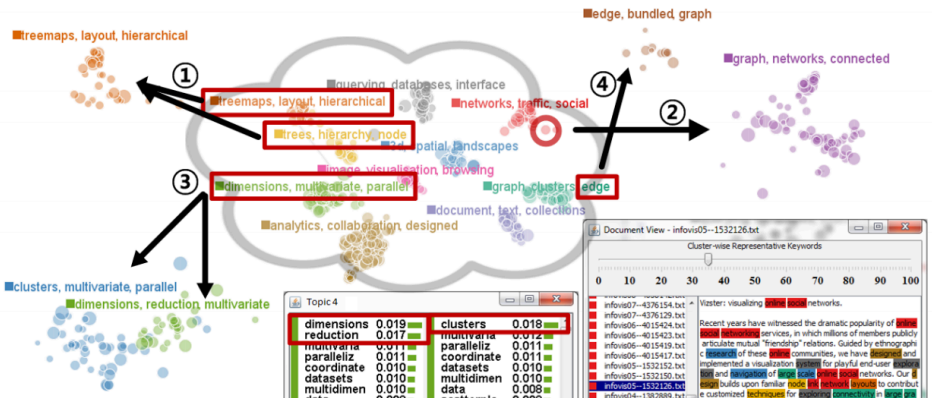
Main modifications

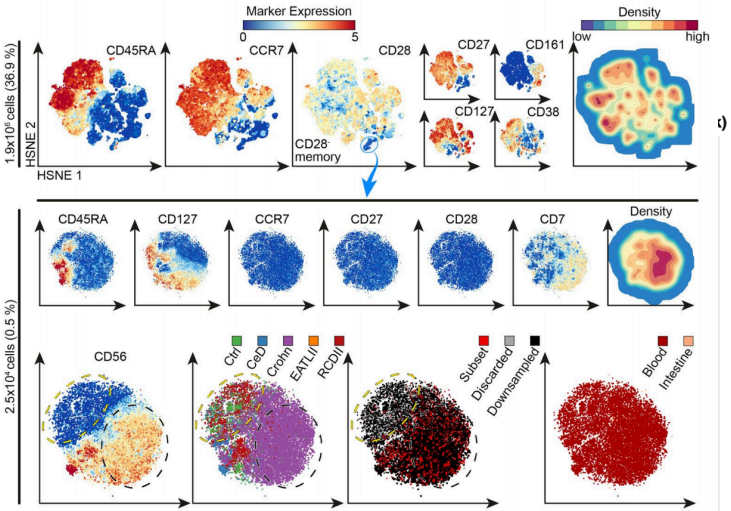
- Distributional similarity model:
  - Each **keyword** was **indexed** according to its **provenance**.
  - We obtained **intersection sets** between keywords in the different datasets.
  - The distance and similarity matrices were also obtained in this case using a cosine metric.

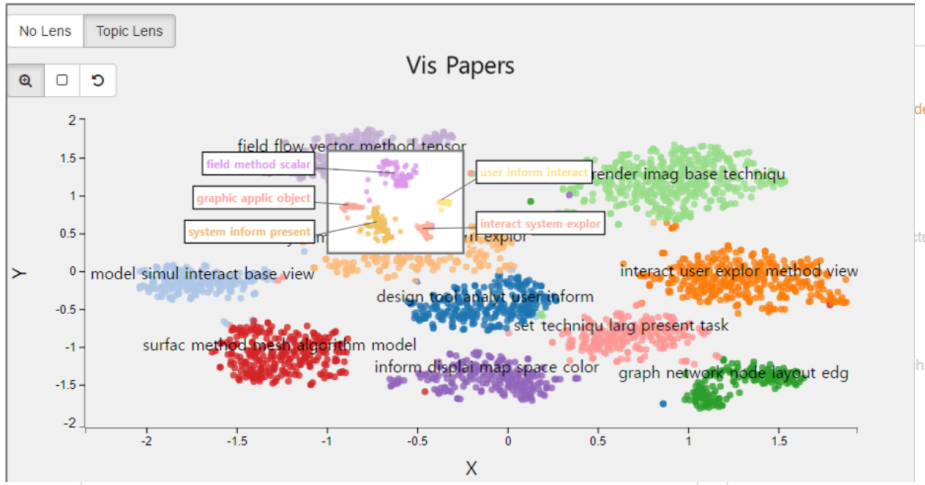


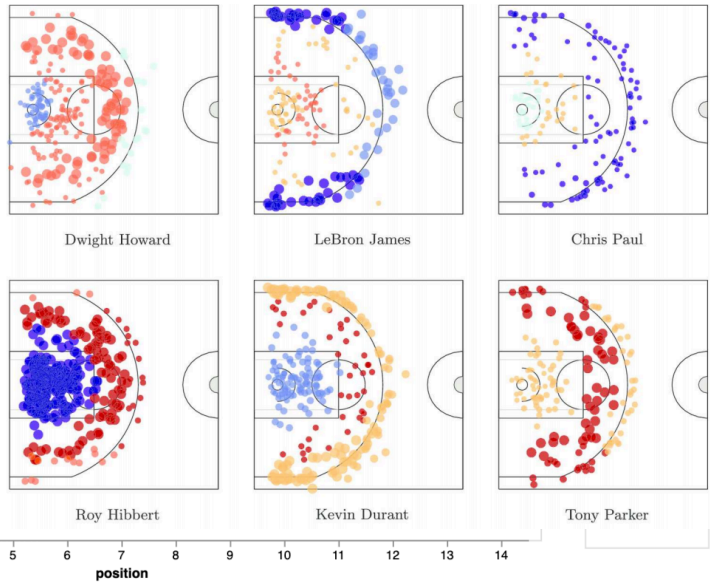
- Mixed approach to explore the similarity matrix:
  - Significant inter-collection paths were obtained for every token appearing exclusively in one of the collections.
  - We applied a distance cutoff at the 5th quartile to filter out less relevant connections.
  - The resulting graphs were merged and pruned using MSTs.




treemaps, layout, hierarchical
querying, databases, interface
networks, traffic, social
trees, hierarchy, node
spatial, landscapes
image, visualization, browsing
dimensions, multivariate, parallel
graph, clusters, edge
document, text, collections
analytics, collaboration, designed
edge, bundled, graph
graph, networks, connected
clusters, multivariate, parallel
dimensions, reduction, multivariate
Topic 4
dimensions 0.019
reduction 0.017
clusters 0.018
parallelliz 0.011
coordinate 0.010
datasets 0.010
multidimen 0.010
data 0.008
scatterplo 0.008
variable 0.008
dimensiona 0.007
number 0.007
Remove
Remove completely
OK
Document View - infovis05--1532126.txt
Cluster-wise Representative Keywords
0 10 20 30 40 50 60 70 80 90 100
Vizster: visualizing online social networks.
Recent years have witnessed the dramatic popularity of online social networking services, in which millions of members publicly articulate mutual "friendship" relations. Guided by ethnographi
research of these online communities, we have designed and implemented a visualization system for playful end-user explora
tion and manipulation of large scale online social networks. Our
design builds upon familiar node link network layouts to contribut
e customized techniques for exploring connectivity in large so
cial structures, supporting visual search and analysis, and aute
matically identifying and visualizing community structures. Both
public installation and controlled studies of the system provide
evidence of the system's usability, capacity for facilitating disco
very, and potential for fun and engaged social activity
rare, sne, cytometry, hsne, sin...
shakespeare, drama
coword
neighbor, t-distributed
exertion, helmet, body-worn, bi...
text
basketbal
bayesian


Marker Expression
0
5
Density
low
high
Level 3
1.9x10⁶ cells (36.9 %)
HSNE 2
HSNE 1
CD45RA
CCR7
CD28
CD27
CD161
CD127
CD38
CD28- memory
Density
CD45RA
CD127
CCR7
CD27
CD28
CD7
Density
Data Level
2.5x10⁴ cells (0.5 %)
CD56
Ctrl
CeD
Crohn
EATL/I
RCD/I
Subset
Discarded
Downsampled
Blood
Intestine


No Lens
Topic Lens
Vis Papers
field flow vector method tensor
field method scalar
graphic applic object
system inform present
user inform interact
render imag base techniqu
interact system explor
model simul interact base view
interact user explor method view
design tool analyt user inform
set techniqu larg present task
surfac method mesh algorithm model
inform displai map space color
graph network node layout edg
X
Y
UTOPIAN: User-Driven Topic Modeling Based on Interactive Nonnegative Matrix Factorization


Dwight Howard
LeBron James
Chris Paul
Roy Hibbert
Kevin Durant
Tony Parker
position

Publication

- A. Benito-Santos and R. Therón Sánchez, 'Defragmenting Research Areas with Knowledge Visualization and Visual Text Analytics', Applied Sciences, vol. 10, no. 20, Art. no. 20, Jan. 2020.
  - Status: **Published**
  - DOI: 10.3390/app10207248
  - Impact Factor (JCR 2019): **2.474**
  - Subject Category: Engineering, Multidisciplinary
  - Quartile: (32/91) **Q2**.

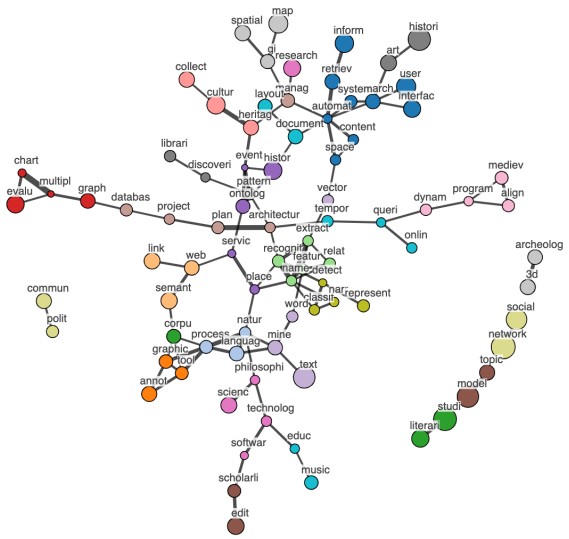
Conclusions

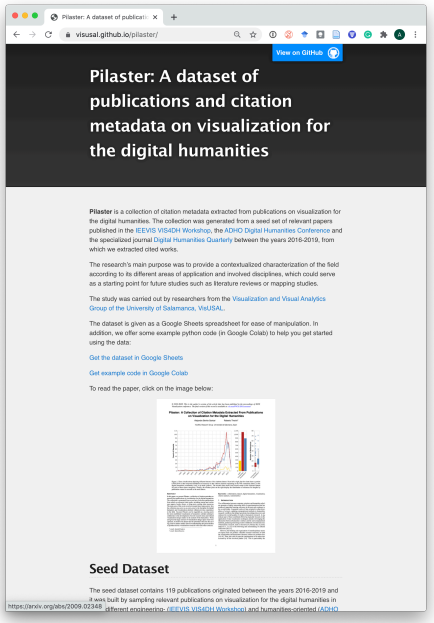
- Main contributions
- Future work

Goya, F. (ca. 1823) The Dog [Oil mural on plaster, transferred to canvas]

Main contributions

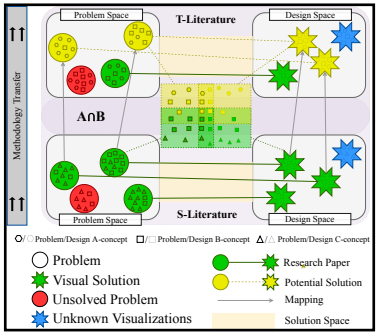
- A <u>community-centered methodology</u> and a set of unsupervised methods to <u>map novel and diffuse interdisciplinary</u> research areas.
- A <u>data-driven characterization</u> of the state-of-the-art in VIS4DH.
- A <u>curated dataset</u> of ~2,000 core VIS4DH works.

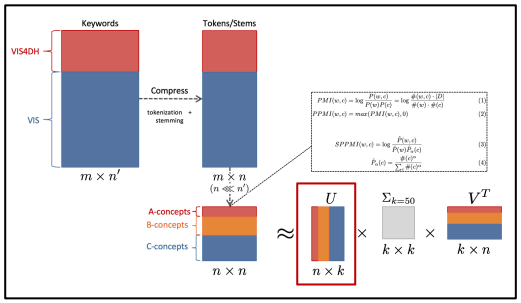


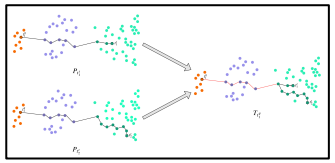


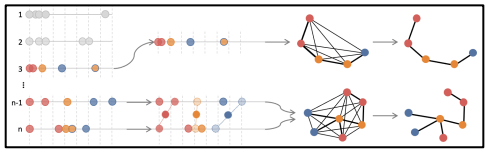
https://visual.github.io/pilaster/

Main contributions









- A <u>revised communication model</u> for PDVR that employs Swanson's <u>ABC model</u> to describe the channel shared by researchers in the two sides of a <u>PDVR collaboration</u>.
- An <u>automatic, unsupervised computational method</u> that relies on <u>distributional similarity found between paper keywords</u> to detect and display potential <u>methodology transfers</u> that are available in the literature.
- <u>Two psychometric scaling and visualization techniques</u> for extracting and representing interdomain knowledge from **proximity data.**

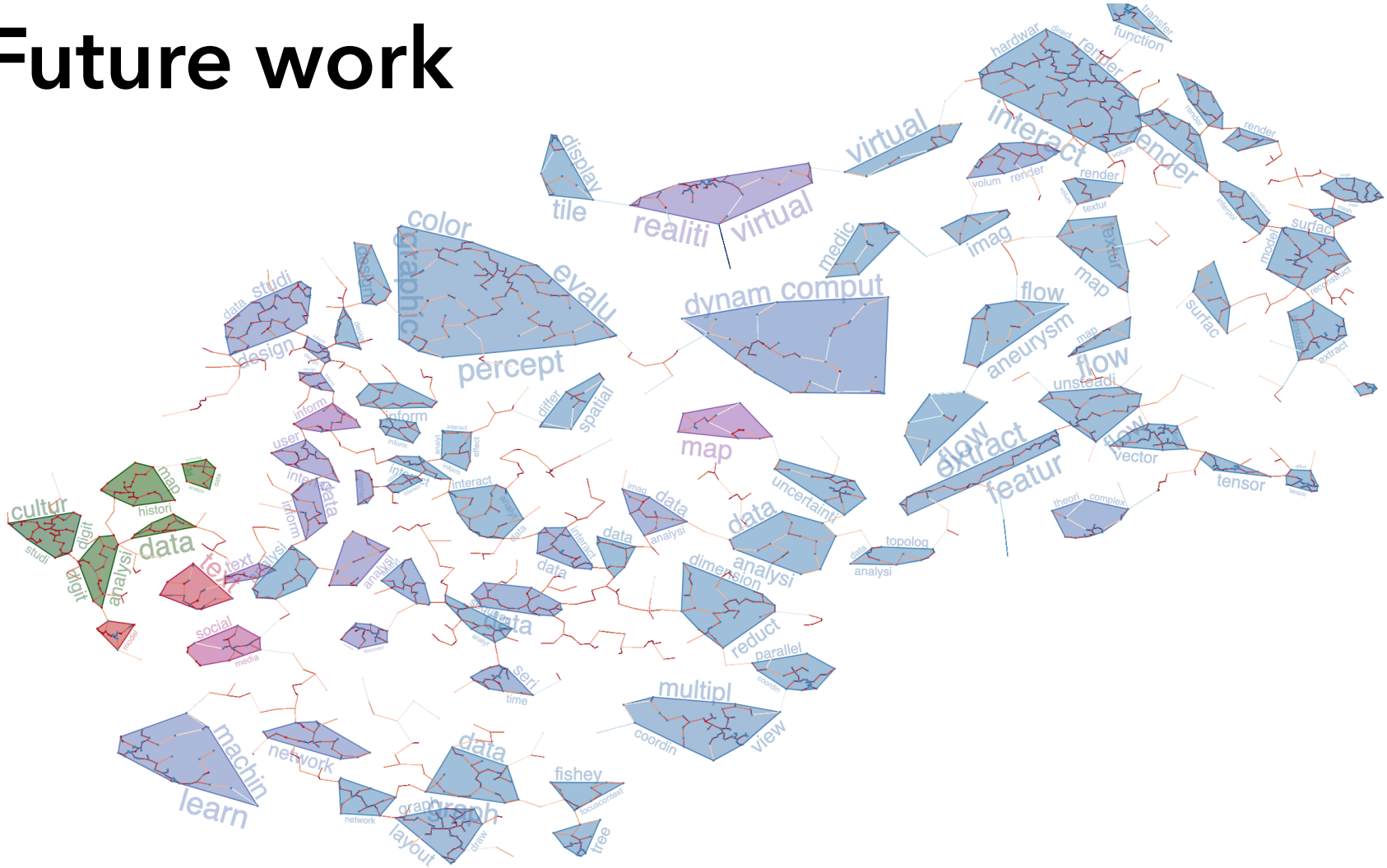
Main contributions

- A fully-functional **visual text analytics interface prototype**, GlassViz, for document:
  - Implements the **sensemaking model** of methodology transfer typically adopted by interdisciplinary researchers.
  - Takes advantage of the **communicative power** and **expert knowledge** encoded in **author-assigned keywords**.
  - Addresses the **cold-start problem** by skipping input a specific query string to start the exploration.
  - Motivates a progressive learning of interesting vocabulary that goes from the known to the unknown.

Main contributions

- A modification of GlassViz to assess fragmentation in VIS research.
  - Facilitates the detection of related studies employing similar algorithms and visualization techniques.
  - Can be used to

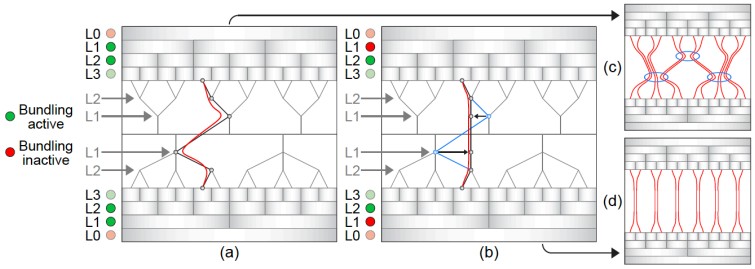
Future work



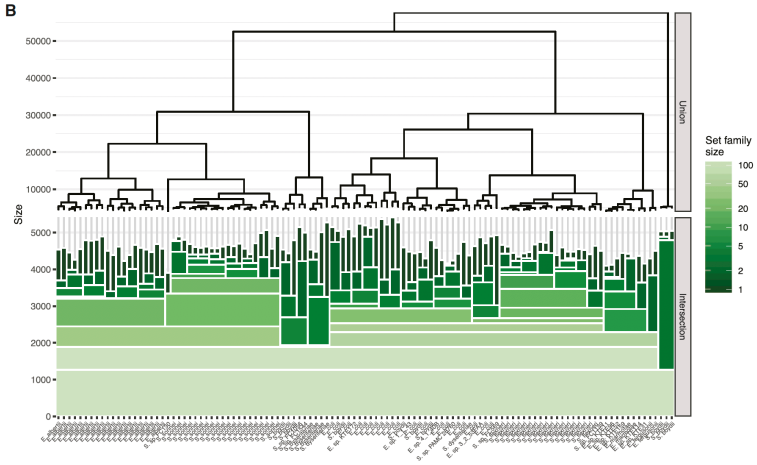
Joint projection of keywords and documents
(UMAP + HDBSCAN)

[26] L. McInnes, J. Healy, N. Saul, and L. Grossberger, 'UMAP: Uniform manifold approximation and projection', The Journal of Open Source Software, vol. 3, no. 29, p. 861, 2018.

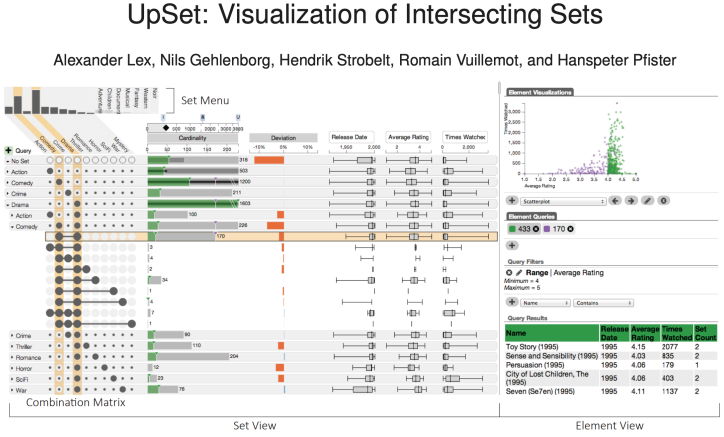
[30] L. McInnes, J. Healy, and S. Astels, 'hdbscan: Hierarchical density based clustering', The Journal of Open Source Software, vol. 2, no. 11, Mar. 2017, doi: 10.21105/joss.00205.



[31] D. Holten and J. J. V. Wijk, 'Visual Comparison of Hierarchically Organized Data', Computer Graphics Forum, vol. 27, no. 3, pp. 759–766, 2008, doi: 10.1111/j.1467-8659.2008.01205.x.

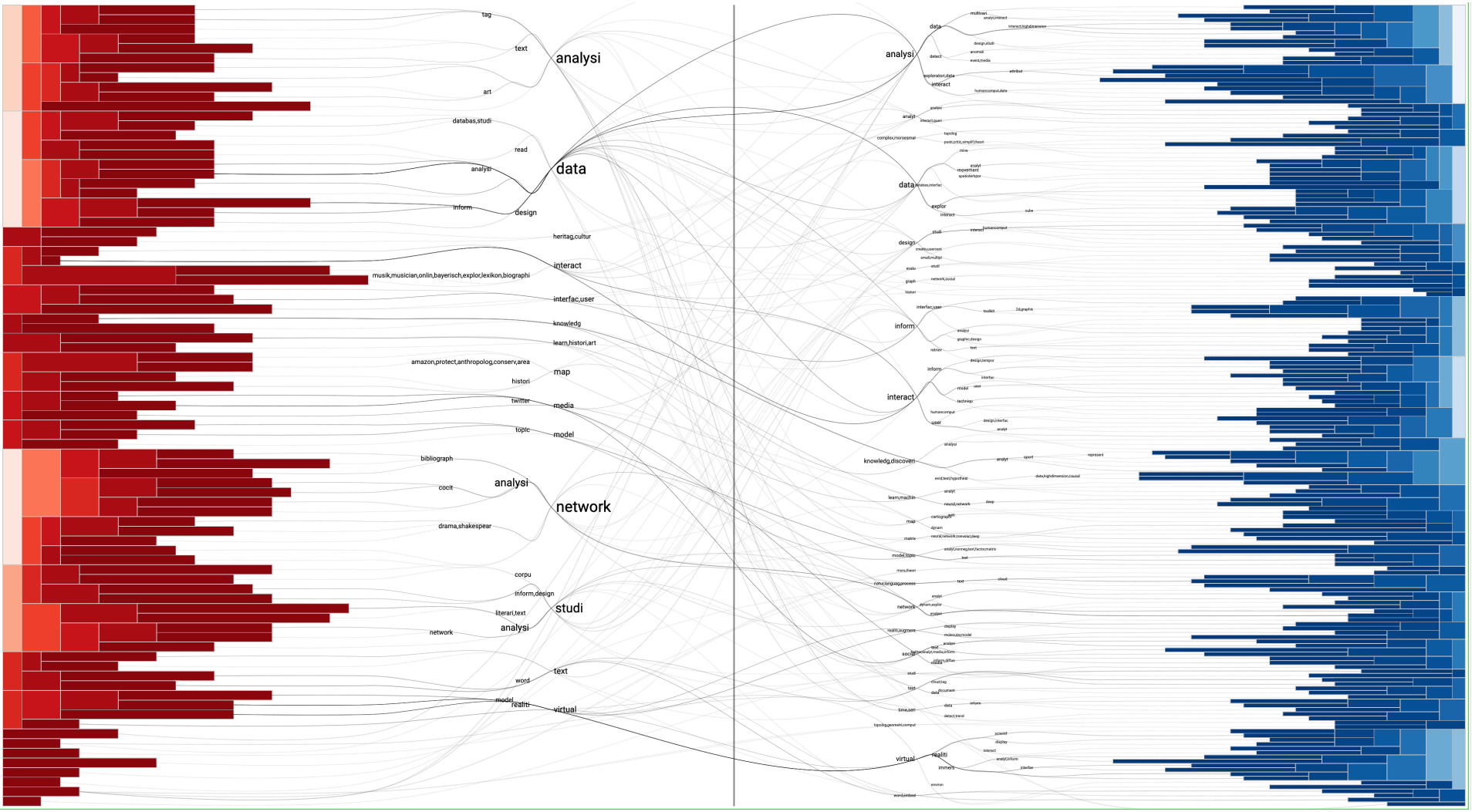


[32] T. L. Pedersen, 'Hierarchical sets: analyzing pangenome structure through scalable set visualizations', Bioinformatics, vol. 33, no. 11, pp. 1604–1612, Jun. 2017, doi: 10.1093/bioinformatics/btx034.



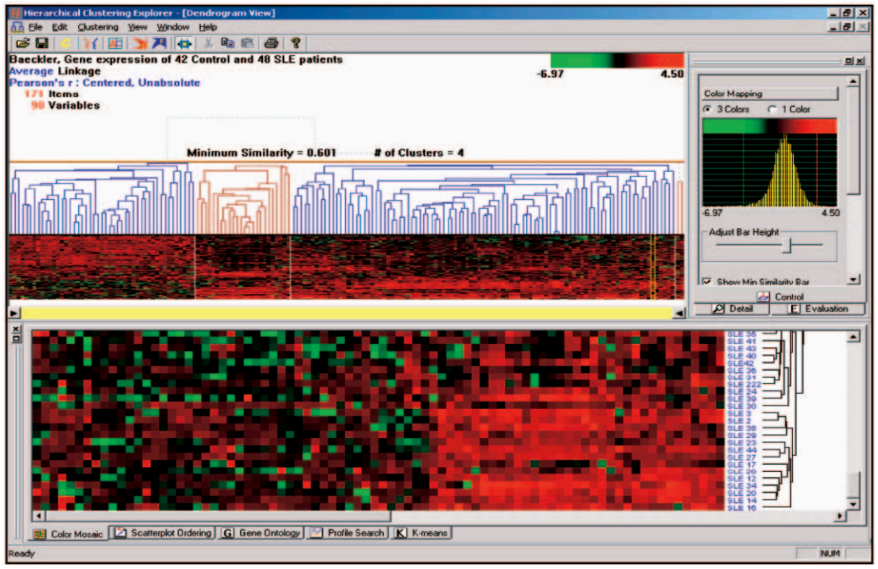
[33] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, 'UpSet: Visualization of Intersecting Sets', IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1983–1992, Dec. 2014, doi: 10.1109/TVCG.2014.2346248.

[34] J. R. Conway, A. Lex, and N. Gehlenborg, 'UpSetR: an R package for the visualization of intersecting sets and their properties', Bioinformatics, vol. 33, no. 18, pp. 2938–2940, Sep. 2017, doi: 10.1093/bioinformatics/btx364.

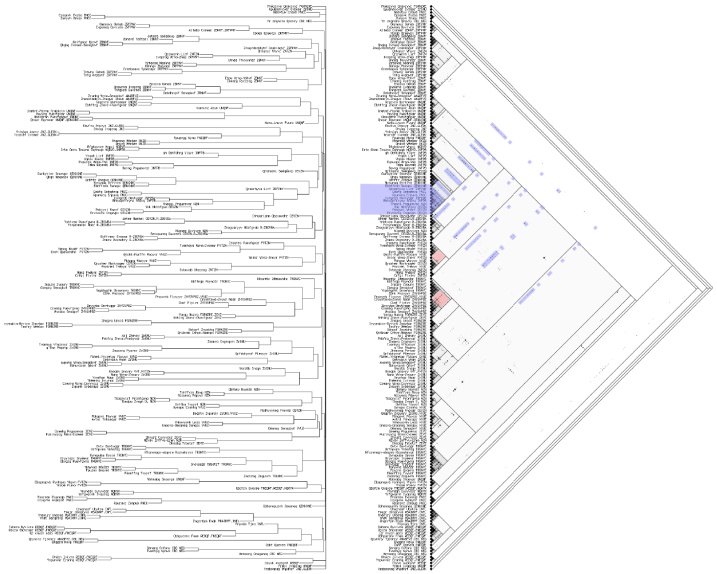


Hierarchical visualization of keywords, documents and inter-collection relationships

Insight-Based evaluation



[35] P. Saraiya, C. North, and K. Duca, 'An insight-based methodology for evaluating bioinformatics visualizations', IEEE Transactions on Visualization and Computer Graphics, vol. 11, no. 4, pp. 443–456, Jul. 2005, doi: 10.1109/TVCG.2005.53.

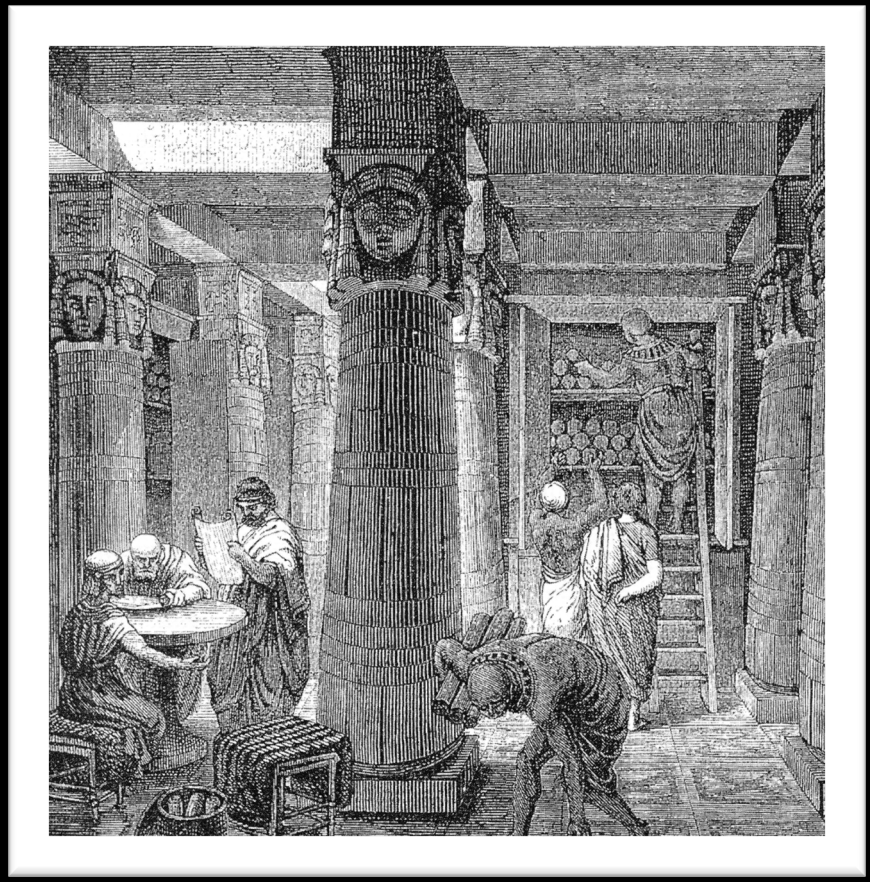


[36] [1]R. Blanch, R. Dautriche, and G. Bisson, 'Dendrogramix: A hybrid tree-matrix visualization technique to support interactive exploration of dendrograms', in 2015 IEEE Pacific Visualization Symposium (PacificVis), Apr. 2015, pp. 31–38, doi: 10.1109/PACIFICVIS.2015.7156353.

Questions

Rafael (1511), The School of Athens [Fresco]

Bibliography



Von Corven, O. (19th century), The Great Library of Alexandria [Engraving]

- [1] A. G. Losada, R. Therón, and A. Benito, ‘BKViz: A Basketball Visual Analysis Tool’, IEEE Computer Graphics and Applications, vol. 36, no. 6, pp. 58–68, 2016.
- [2] A. Benito-Santos, R. Theron, A. Losada, J. E. Sampaio, and C. Lago-Peñas, ‘Data-Driven Visual Performance Analysis in Soccer: An Exploratory Prototype’, Front. Psychol., vol. 9, 2018, doi: 10.3389/fpsyg.2018.02416
- [3] R. Therón Sánchez, A. Benito-Santos, R. S. Santamaría Vicente, and A. Losada Gómez, ‘Towards an Uncertainty-Aware Visualization in the Digital Humanities’, Informatics, vol. 6, no. 3, p. 31, Sep. 2019, doi: 10.3390/informatics6030031.
- [4] A. Dorn et al., ‘A project review under the focus of complexities on the example of ExploreAT!’, presented at the Digital Humanities 2019, Utrecht, 2019.
- [5] A. Benito, A. Dorn, R. Therón, E. Wandl-Vogt, and A. Losada, ‘Shedding Light on Indigenous Knowledge Concepts and World Perception through Visual Analysis’, in Digital Humanities 2018 Book of Abstracts, Mexico City, Mexico, 2018, pp. 537–538.
- [6] A. Benito, R. Therón, A. Losada, E. Wandl-Vogt, and A. Dorn, ‘Exploring Lemma Interconnections in Historical Dictionaries’, presented at the 2nd Workshop on Visualization for the Digital Humanities (VIS4DH), 2017.
- [7] A. Benito Santos and R. Therón Sánchez, ‘Visualización de Datos en Humanidades Digitales’, Tesis de Master, Universidad de Salamanca, España, 2016.
- [8] Bracken, L.J., Oughton, E.A.: `What do you mean?' The importance of language in developing interdisciplinary research. Transactions of the Institute of British Geographers 31(3), 371{382 (Jul 2006). https://doi.org/10.1111/j.1475-5661.2006.00218.x
- [9] ‘Global scientific output doubles every nine years : News blog’. http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html (accessed Nov. 25, 2020).
- [10] J. Chuang, C. D. Manning, and J. Heer, ‘“Without the Clutter of Unimportant Words”: Descriptive Keyphrases for Text Visualization’, ACM Trans. Comput.-Hum. Interact., vol. 19, no. 3, p. 19:1–19:29, Oct. 2012, doi: 10.1145/2362364.2362367.
- [11] K. El-Arini and C. Guestrin, ‘Beyond Keyword Search: Discovering Relevant Scientific Literature’, in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2011, pp. 439–447, doi: 10.1145/2020408.2020479.
- [12] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller, ‘Visualization as Seen through its Research Paper Keywords’, IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 1, pp. 771–780, Jan. 2017

- [13] S. Simon, S. Mittelstädt, D. A. Keim, and M. Sedlmair, ‘Bridging the gap of domain and visualization experts with a Liaison’, Eurographics Conference on Visualization (EuroVis 2015, Short Paper), Cagliari, Italy, 2015, vol. 2015.
- [14] M. Wattenberg, ‘Arc diagrams: Visualizing Structure in Strings’, in IEEE Symposium on Information Visualization, 2002. INFOVIS 2002., Oct. 2002, pp. 110–116, doi: 10.1109/INFVIS.2002.1173155.
- [15] M. Miller, H. Schäfer, M. Kraus, M. Leman, D. A. Keim, and M. El-Assady, ‘Framing Visual Musicology through Methodology Transfer’, Proceedings of the Workshop on Visualization for the Digital Humanities (VIS4DH) at IEEE VIS 2019, Oct. 2019.
- [16] D. R. Swanson, ‘Fish Oil, Raynaud’s Syndrome, and Undiscovered Public Knowledge’, Perspectives in Biology and Medicine, vol. 30, no. 1, pp. 7–18, 1986, doi: 10.1353/pbm.1986.0087
- [17] M. Thilakaratne, K. Falkner, and T. Atapattu, ‘A Systematic Review on Literature-based Discovery’, ACM Computing Surveys (CSUR), Dec. 2019
- [18] M. Thilakaratne, K. Falkner, and T. Atapattu, ‘Automatic Detection of Cross-Disciplinary Knowledge Associations’, Jul. 2018, pp. 45–51,
- [19] F. Heimerl, M. John, Qi Han, S. Koch, and T. Ertl. DocuCompass: Effective exploration of document landscapes. In 2016 IEEE Conference on Visual Analytics Science and Technology (VAST), pages 11–20, October 2016.
- [20] J. He, Q. Ping, W. Lou, and C. Chen, ‘PaperPoles: Facilitating adaptive visual exploration of scientific publications by citation links’, Journal of the Association for Information Science and Technology, vol. 70, no. 8, pp. 843–857, 2019, doi: 10.1002/asi.24171.
- [21] E. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher, ‘Serendip: Topic model-driven visual exploration of text corpora’, in 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), Paris, France, Oct. 2014, pp. 173–182, doi: 10.1109/VAST.2014.7042493.
- [22] A. J. Bradley et al., ‘Visualization and the Digital Humanities: Moving Toward Stronger Collaborations’, IEEE Computer Graphics and Applications, vol. 38, no. 6, pp. 26–38, Nov. 2018
- [23] P. Isenberg et al., ‘Vispubdata.org: A Metadata Collection About IEEE Visualization (VIS) Publications’, IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 9, pp. 2199–2206, Sep. 2017, doi: 10.1109/TVCG.2016.2615308.
- [24] O. Levy, Y. Goldberg, and I. Dagan, ‘Improving Distributional Similarity with Lessons Learned from Word Embeddings’, Transactions of the Association for Computational Linguistics, vol. 3, no. 0, pp. 211–225, May 2015.

- [25] O. Levy and Y. Goldberg, ‘Neural Word Embedding as Implicit Matrix Factorization’, in Advances in Neural Information Processing Systems 27, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2177–2185.
- [26] L. McInnes, J. Healy, N. Saul, and L. Grossberger, ‘UMAP: Uniform manifold approximation and projection’, The Journal of Open Source Software, vol. 3, no. 29, p. 861, 2018.
- [27] P. Federico, F. Heimerl, S. Koch, and S. Miksch, ‘A Survey on Visual Approaches for Analyzing Scientific Literature and Patents’, IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 9, pp. 2179–2198, Sep. 2017, doi: 10.1109/TVCG.2016.2610422.
- [28] E. Alexander and M. Gleicher, ‘Task-Driven Comparison of Topic Models’, IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 320–329, Jan. 2016
- [29] F. Heimerl and M. Gleicher, ‘Interactive Analysis of Word Vector Embeddings’, Computer Graphics Forum, vol. 37, no. 3, pp. 253–265, Jun. 2018
- [30] L. McInnes, J. Healy, and S. Astels, ‘hdbscan: Hierarchical density based clustering’, The Journal of Open Source Software, vol. 2, no. 11, Mar. 2017, doi: 10.21105/joss.00205.
- [31] D. Holten and J. J. V. Wijk, ‘Visual Comparison of Hierarchically Organized Data’, Computer Graphics Forum, vol. 27, no. 3, pp. 759–766, 2008, doi: 10.1111/j.1467-8659.2008.01205.x.
- [32] T. L. Pedersen, ‘Hierarchical sets: analyzing pangenome structure through scalable set visualizations’, Bioinformatics, vol. 33, no. 11, pp. 1604–1612, Jun. 2017, doi: 10.1093/bioinformatics/btx034.
- [33] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, ‘UpSet: Visualization of Intersecting Sets’, IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1983–1992, Dec. 2014, doi: 10.1109/TVCG.2014.2346248.
- [34] J. R. Conway, A. Lex, and N. Gehlenborg, ‘UpSetR: an R package for the visualization of intersecting sets and their properties’, Bioinformatics, vol. 33, no. 18, pp. 2938–2940, Sep. 2017, doi: 10.1093/bioinformatics/btx364.
- [35] P. Saraiya, C. North, and K. Duca, ‘An insight-based methodology for evaluating bioinformatics visualizations’, IEEE Transactions on Visualization and Computer Graphics, vol. 11, no. 4, pp. 443–456, Jul. 2005, doi: 10.1109/TVCG.2005.53.
- [36] [1]R. Blanch, R. Dautriche, and G. Bisson, ‘Dendrogramix: A hybrid tree-matrix visualization technique to support interactive exploration of dendrograms’, in 2015 IEEE Pacific Visualization Symposium (PacificVis), Apr. 2015, pp. 31–38, doi: 10.1109/PACIFICVIS.2015.7156353.

Supporting Methodology Transfer in Visualization Research with Literature-Based Discovery and Visual Text Analytics

Doctoral Dissertation to obtain the

Degree of Doctor of Philosophy in Computer Engineering

by Alejandro Benito-Santos

supervised by **Roberto Therón Sánchez, PhD**

27th November 2020

Visual Analytics and Visualization Information Group

Universidad de Salamanca

Research Group Interaction and e-Learning