

Apresentação Paper #1

SCC 5836 – Visualização Computacional

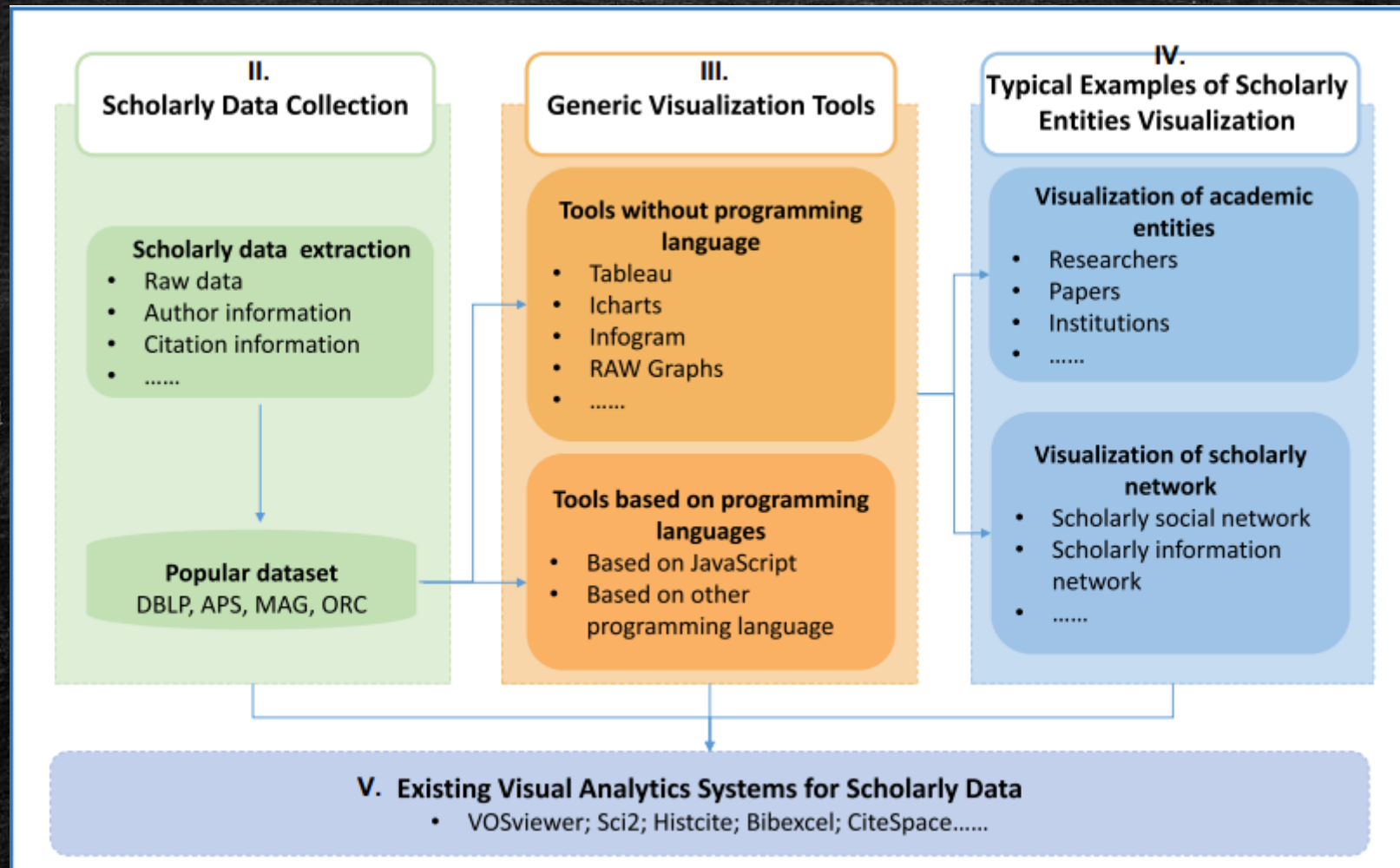
Luiz Alberto Hiroshi Horita

A Survey of Scholarly Data Visualization

JIAYING LIU, TAO TANG, WEI WANG, BO XU, XIANGJIE KONG, AND FENG XIA.

Publicado no Open Access Journal IEEE em 23 de Abril de 2018.

I. Introdução



II. Coleta de dados acadêmicos

A. Dados acadêmicos

1) Dados brutos (ou Metadata)

- Autores, títulos, abstract, palavras-chave, número de página, etc.

2) Informação de autor

- Nome, afiliação, interesse, instituição, e-mail, hoepages.

3) Informação de citação

- Referências, Bibliografia ou Fontes.

4) Outras informações

- Instituições, venues (nome, ano, local, etc), conteúdo (algoritmos, figuras, tabelas, etc), reconhecimentos.

II. Coleta de dados acadêmicos

B. Datasets populares

1) Digital Bibliography & Library Project (DBLP)

- Bibliografias na área de Ciência da Computação.

2) American Physical Society (APS)

- Bibliografias na área de Física.

3) Microsoft Academic Graph (MAG)

- Bibliografias multidisciplinares.

4) Open Research Corpus (ORC)

- Bibliografias nas áreas de Ciência da Computação e Neurociência.

III. Ferramentas Comuns de Visualização

A. Ferramentas sem linguagem de programação

Ferramenta	Desktop / web-based	Input format	Interface com banco de dados	Recursos
TABLEAU	Desktop	Txt, xlsx, csv	Oracle, MySQL	Intuitivo, visualização flexível, interativo.
ICHARTS	Web-based	–	NetSuite, Salesforce, Google Cloud Platform	Atualização tempo real com banco de dados, análise visual, relatório de usuário, dashboard interativos.
INFOGRAM	Web-based	Xls, csv, xlsx, json	GoogleDrive, Dropbox, OneDrive	Interface amigável, compartilhamento de charts, processamento em tempo real dos dados.
RAW GRAPHS	Web-based	Tsv, dsv, csv, json, xls	Online data with a public API	Intuitivo, dado sempre seguro (processamento no browser), possível criar novos charts com D3.js.
VISUALIZE FREE	Web-based	Xls, xlsx, txt, csv	–	Intuitivo, adequado para pequenas quantidades de dados.

III. Ferramentas Comuns de Visualização

B. Ferramentas baseadas em uma linguagem de programação

1) Baseadas em JavaScript

Framework name	Input data format	Rendered charts by	Charts and maps type	License from
D3.js	JSON, CSV, and XML	HTML5 canvas, SVG and CSS	A powerful D3 gallery with multiple charts, graphs, and maps including the world map and the US map.	BSD-3
Chart.js	JavaScript API	Only HTML5 Canvas	8 chart types, including over 23 charts and graphs.	MIT LICENSE
FusionCharts	JSON, XML	SVG, VML	90+ charts and graphs, 1000+ maps including all continents, major countries, and all US states.	Free for basic edition and advanced commercial edition.
Flot Chart	JavaScript API	Only HTML5 Canvas	The charts of lines, points, filled areas, bars and any combinations of these charts. Doesn't support maps.	Free
ZingChart	JavaScript API	HTML5 Canvas, SVG, and VML	Plenty of chart and graph types in its ZingChart gallery. Support almost every country and area.	Free for basic edition and advanced commercial edition.

III. Ferramentas Comuns de Visualização

B. Ferramentas baseadas em uma linguagem de programação

2) Baseadas em outras linguagens

Tools	Input data format	Language-based	Features	License from
Gephi	CSV, Excel files,	Java, OpenGL	Powered by OpenGL engine. Force-based layout algorithms shape the graph.	GUN, GPL
Nodebox 3	CSV	Python and Clojure	Integrate all the functional parts in nodes.	GPL
Ggplot2	R, API	R	Plotting based on layers. Graphs composed of layers.	GUN, GPL, and V2
Processing	Multiples of formats are available in its library	Java, plugins for Python and JavaScript	Integrate the OpenGL engine. Over 100+ libraries offered to expand its usage.	GPL, LGPL
JpGraph	CSV, From databases such as MySQL	PHP	Tiny size of generating images. Anti-spam images is supported. 3D effects is also supported.	Free QPL, paid for commercial

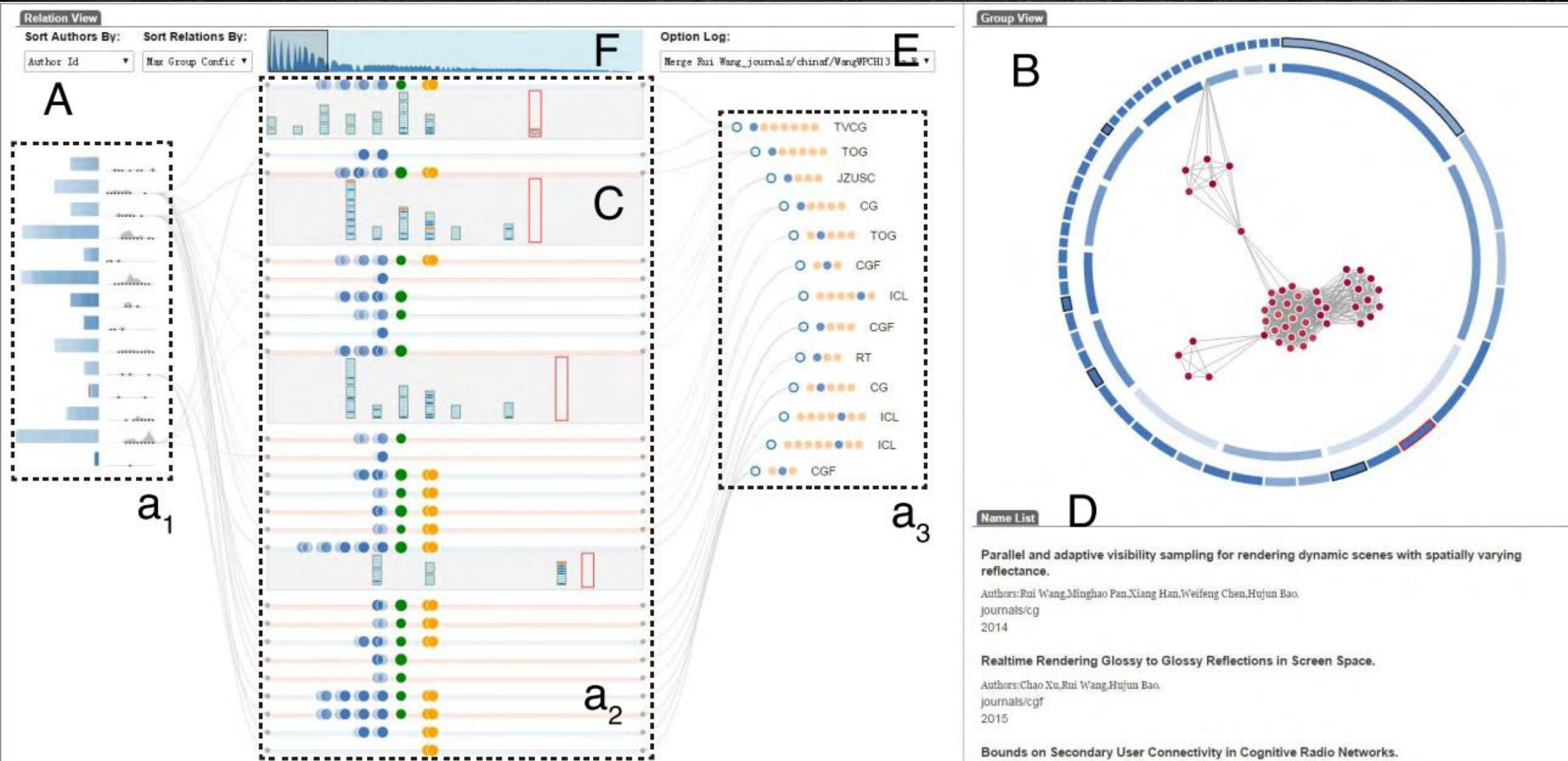
IV. Visualização de dados acadêmicos

A. Entidades acadêmicas

1) Visualização de Pesquisadores

- Auxilia a análise da **formação de grupos científicos** e compreensão sobre as **interações de pesquisa**.
- Ex.: Resolver problemas de ambiguidade de nomes com **NameClarifier** (Shen et. al., Jan. 2017).

NameClarifier: A visual analytics system for author name disambiguation



IV. Visualização de dados acadêmicos

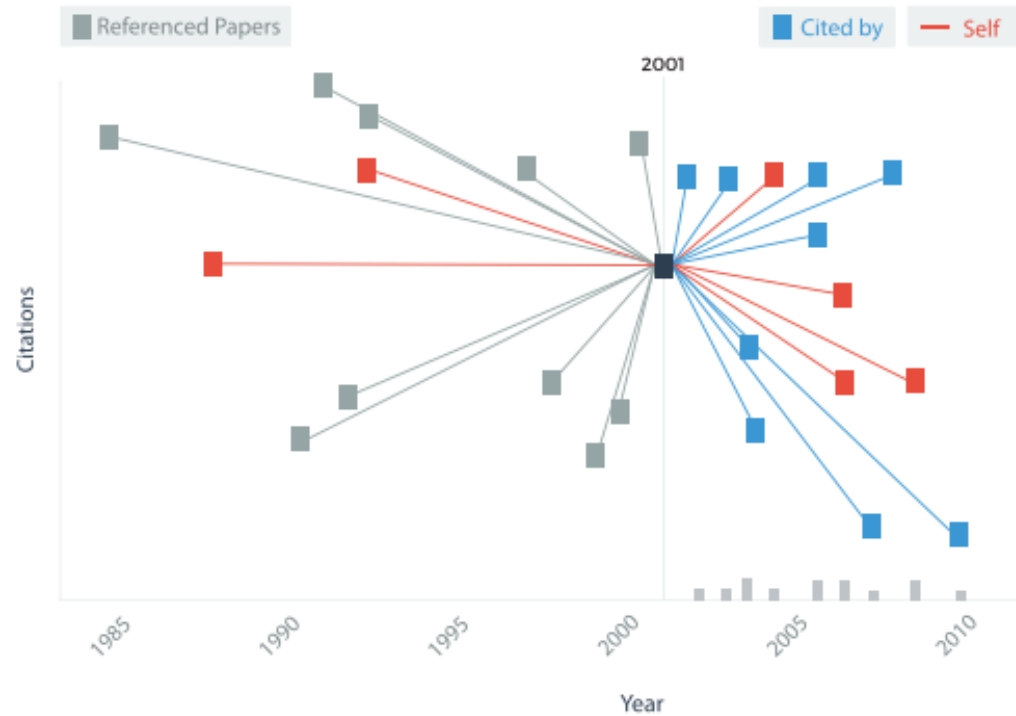
A. Entidades acadêmicas

2) Visualização de Papers

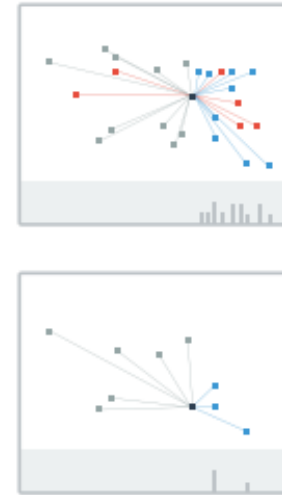
- Propicia abrangente cognição sobre **campos de pesquisa** referentes aos papers científicos.
- Auxilia o entendimento das **tendências e impactos da pesquisa**, a descoberta de **conexões e padrões** entre documentos científicos.
- Ex.: descobrir relações explícitas ou implícitas entre papers com **Visualization of Publication Impact** (Maguire et al., 2016).
- Ex.2: relações, representação da estrutura hierárquica e similaridades entre tópicos com **A text visualization method for cross-domain research topic mining** (Jiang et al., 2016).

Visualization of Publication Impact

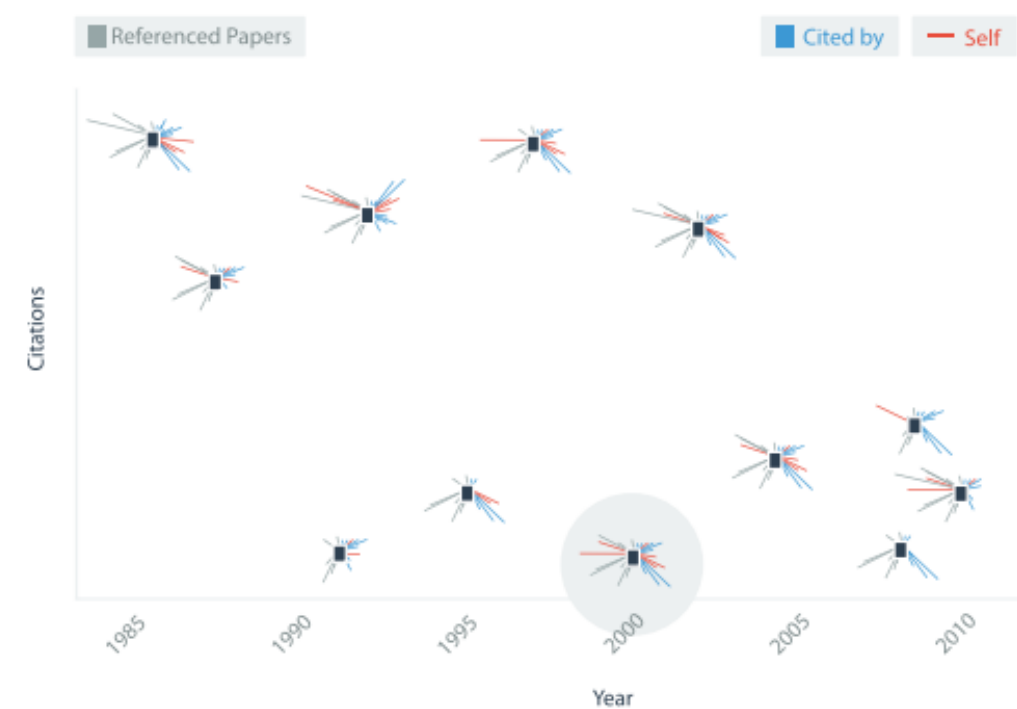
A) Publication Impact Graph



B) Publication Impact Glyphs

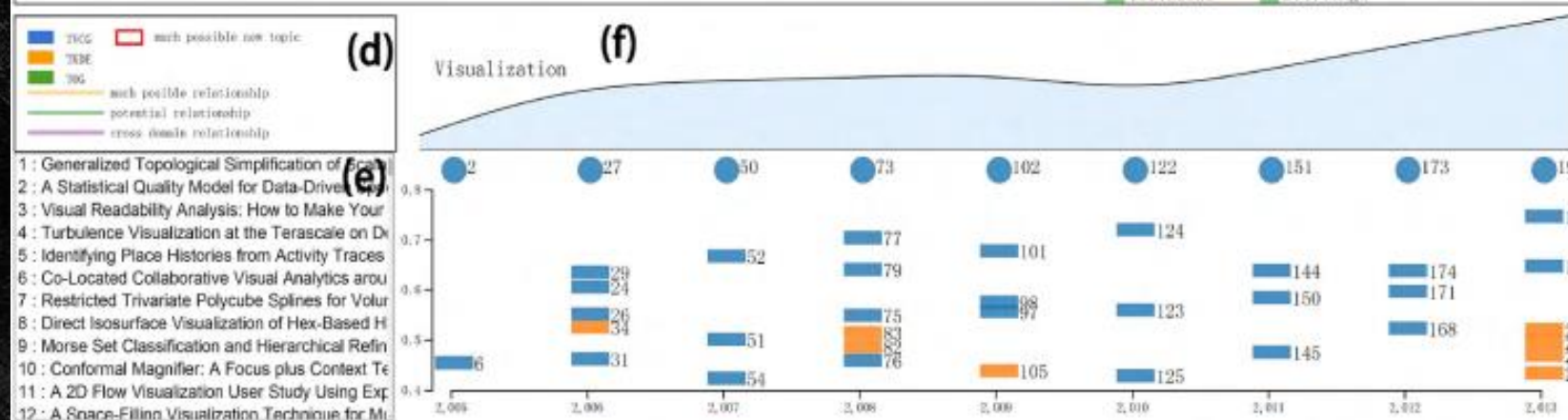
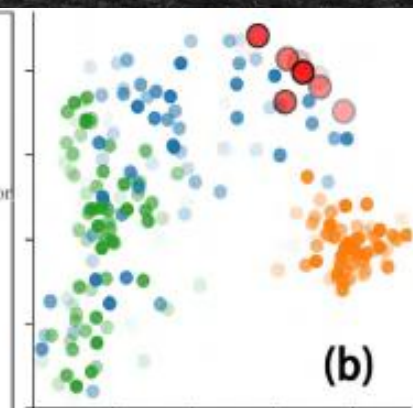
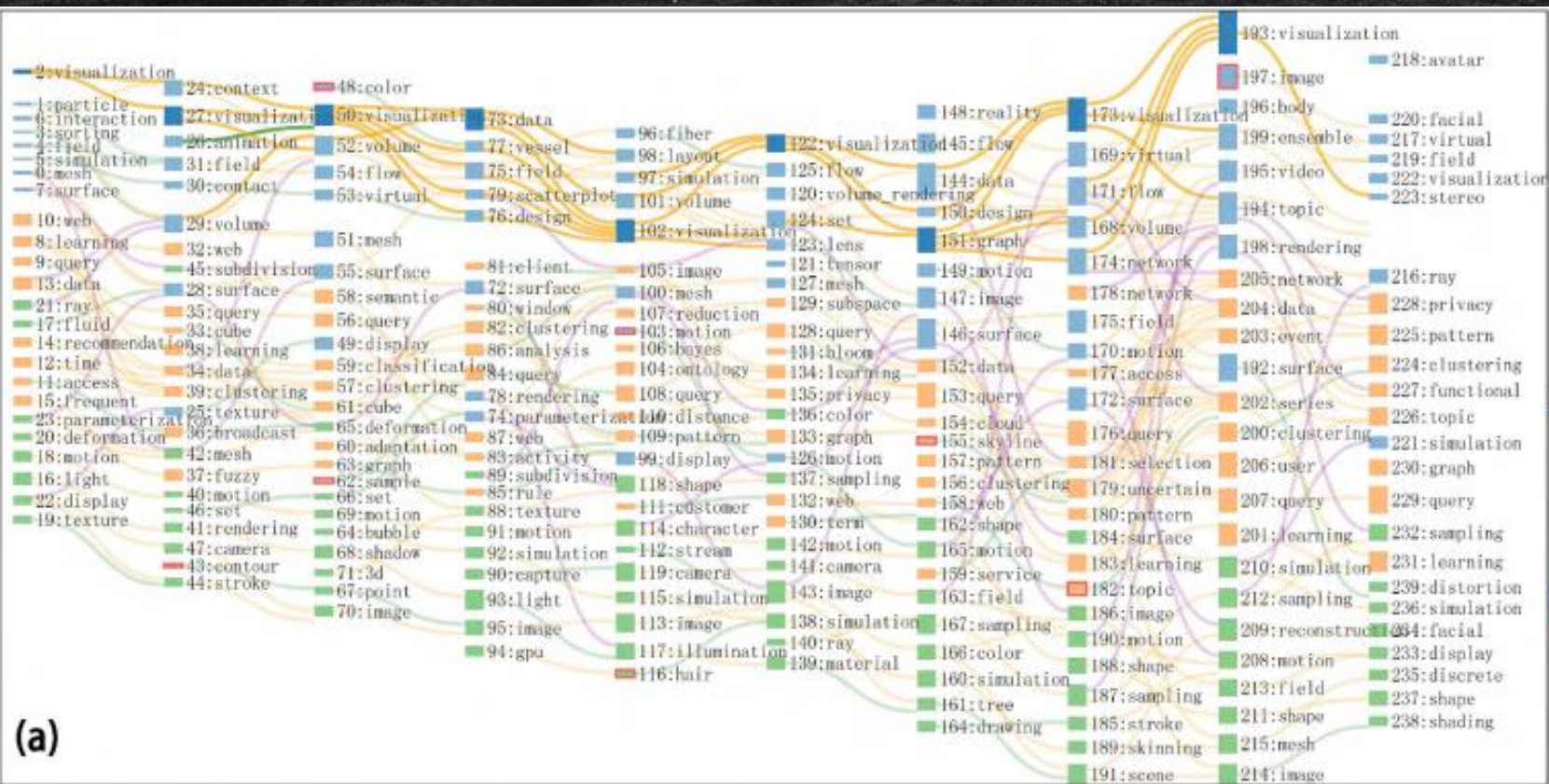


C) Publication Impact Collection



- A) Impact graphs: apresenta a informação específica do paper incluindo referências e citações.
- B) Impact glyphs: compacta o impact graphs para mostrar a importância do paper.
- C) Impact overview: posiciona o impact graphs e impact glyphs para as informações relacionadas.

A text visualization method for cross-domain research topic mining



IV. Visualização de dados acadêmicos

A. Entidades acadêmicas

3) Visualização de Instituições

- Possibilita visualizar a **rede de colaboração entre instituições** através de seus membros.
- Ex.: visualização de filiações sobre o mapa com **Acemap** (<http://acemap.sjtu.edu.cn/app/affiliationMap/index.html>).

Acemap



Author/organization/conference/topic/paper Search

Harvard University ^

Author Map Total number of papers: 58788 Total number of authors: 34316

COLORING TOP 5 FIELD FILTER

Field filter

Field Of Study:

- ☐ Bioinformatics
- ☐ Genetics
- ☐ Thermodynamics
- ☐ Quantum Mechanics
- ☐ Nanotechnology
- ☐ Particle Physics
- ☐ Remote Sensing
- ☐ Neuroscience
- ☐ Artificial Intelligence
- ☐ Algorithm
- ☐ Computational Biology
- ☐ Computer Network
- ☐ Computer Hardware
- ☐ Real-time Computing



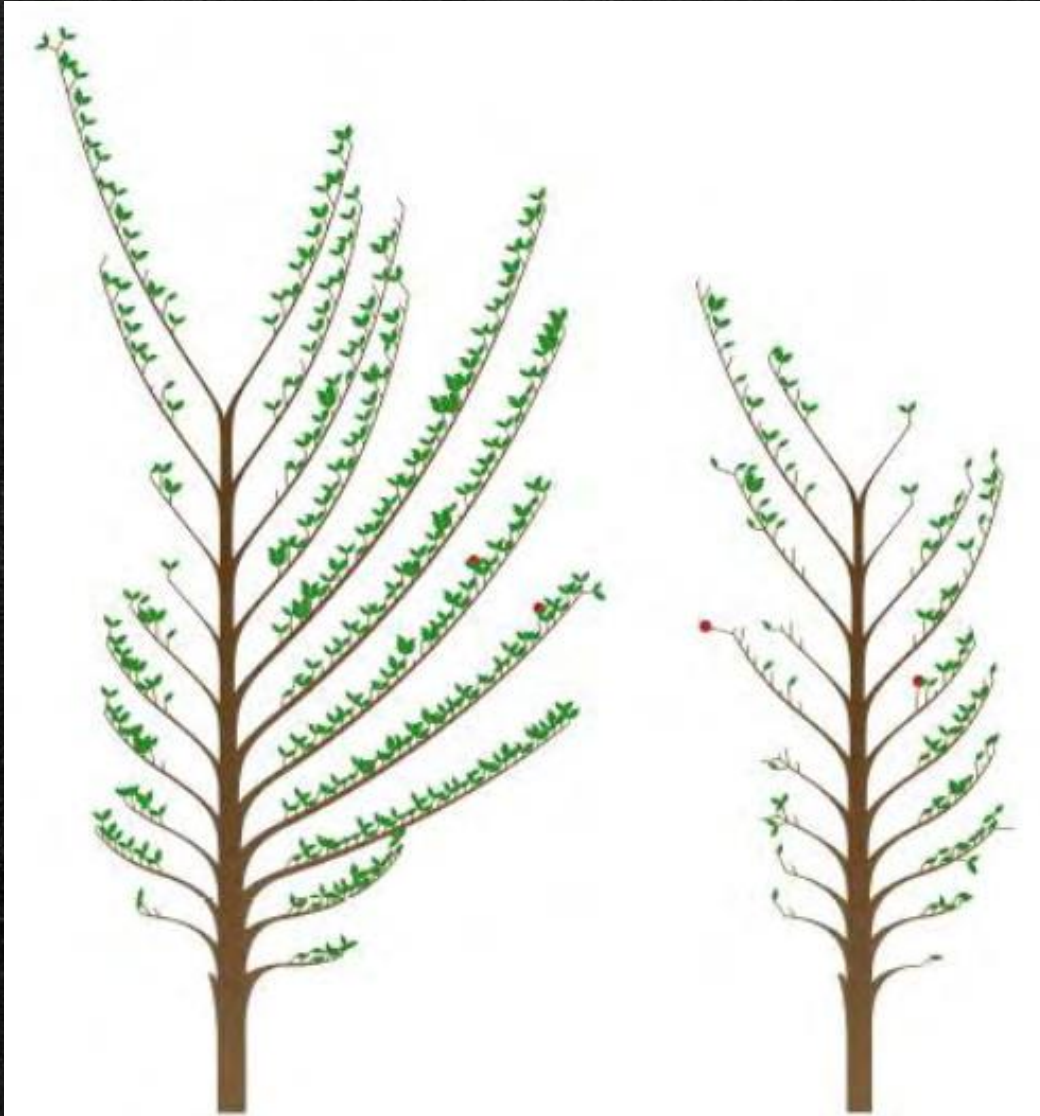
IV. Visualização de dados acadêmicos

B. Rede acadêmica

1) Visualização de rede social acadêmica (rede de colaboração)

- Propicia a compreensão sobre o acúmulo das interações da pesquisa.
- Auxilia o cientista a **examinar sua carreira pessoal e promover seu desenvolvimento.**
- Ex.: **Visual characterization of personal bibliographic data using a botanical tree design** (Fung and Ma, 2015).

Visual characterization of personal bibliographic data using a botanical tree design.



Scholarly tree de dois pesquisadores ativos.

Cada ramo representa publicações em um ano.

Árvore da esquerda: 1993-2016

Árvore da direita: 1995-2012

V. Sistemas analíticos visuais para dados acadêmicos existentes

System name	Main function	Supported data file format	Features	Operating environment	Limitation and the key difference
VOSViewer	Taking citation analysis.	Data from WOS, Scopus, PubMed, RIS, Pajek, and GML.	Density and overlay visualizations. Create bibliometric networks based on co-authorship, bibliographic coupling, and co-citation networks, etc. Natural language processing techniques are available for creating term co-occurrence networks.	Windows, Mac OS X, Other systems with the support of Java 6 or later updates, the web client based on Java installed.	Only support node network diagram and corresponding heat map. Strong graphic display ability, suitable for large-scale scholarly data.
Sci2	Taking network analysis, especially directed network analysis.	TXT, CSV, Network data (in-memory graph/network object or network files saved as Graph/ML, XGMML, NWB, Pajek .net or Edge list format), Matrix data (Pajek.mat), In-memory database, Tree data (TreeML)	Visualize the scholarly datasets into kinds of networks. Perform different types of analysis with the most effective algorithms available. Access science datasets online or load their own.	Mac OS, Windows, Linux	High memory footprint while processing large datasets. Temporal bar graph, Choropleth map, UCSD science map and Bimodal network visualization are supported.
Histcite	Taking statistic analysis.	Data from WOS	Scholarly data visualization and various types of bibliometric analysis	Windows, based on the IE browser	No longer in official supported. Only support WOS data. Diagrams the development of one scientific field based on the timeline.
Bibexcel	Processing scholarly database.	Plain text data from WOS, SCIE, DII (Derwent Innovations Index), Medline	Able to do various types of bibliometric analysis. Export its processed data into other visualization tools (Gephi, Pajek, VOSviewer, etc.) that can take a comprehensive visual analysis.	Windows	Flexible in processing scholarly data but not easy to use without its help document.
CiteSpace	Taking co-citation analysis.	Data from WOS, Scopus, PubMed, ADS, arXiv, CNKI, CSSCI, Derwent, NSF, Project DX, ADS, and CiteSpace Built-in database.	Visualize and analyze the patterns and trends in scientific literature	Windows, Require Java 8	Can not delete the irrelevant node within the generated node network.

VI. Questões abertas e panorama

- Como integrar informação de complexos dados acadêmicos de diferentes fontes.
- Como combinar diferentes técnicas de visualização com o processo de análise de forma mais adequada.

Crítica do trabalho

Positivo

Abrangeu bem as técnicas e ferramentas.

Teve boa profundidade na exploração de cada tópico, citando a importância e seus principais recursos .

Negativo

Faltou uma padronização de como exibir informações sobre as ferramentas e exemplos.

Itens disponibilizados



dados



Código

Referências

- Google Scholar: citado em 2 artigos.
- As referências são na área de aplicação.

Perguntas

- 1) O que são os dados acadêmicos?
- 2) Por que é importante a visualização dos dados acadêmicos?
- 3) A visualização de dados acadêmicos é a combinação de três teorias. Quais e do que se tratam?

OBRIGADO!
