

Visualizing, Exploring and Analyzing Big Data: A 6-Year Story

Report on the Big Data Visual Exploration and Analytics Workshops 2018-2023

Nikos Bikakis¹ George Papastefanatos² Panos K. Chrysanthis³ Olga Papemmanuil⁴ David Auber⁵
Steffen Frey⁶ Issei Fujishiro⁷ Hanna Hauptmann⁸ Shixia Liu⁹ Kwan-Liu Ma¹⁰
Tobias Schreck¹¹ Michael Sedlmair¹² Mohamed A. Sharaf¹³

1. INTRODUCTION

Information Visualization has been one of the cornerstones of Data Science, turning the abundance of Big Data being produced through modern systems into actionable knowledge. Indeed, the Big Data era has realized the availability of voluminous datasets that are dynamic, multidimensional, noisy and heterogeneous in nature. Transforming a data-curious user into someone who can access and analyze that data is even more burdensome now for a great number of users with little or no support and expertise on the data processing part.

In this context, several traditional problems from *Data Management & Mining*, *Information Visualization*, and *Human-computer Interaction* communities, such as efficient data storage, querying, and indexing to enable visual analytics, as well as new ways and AI techniques for visual presentation of massive data, need to be revisited. For instance, techniques that provide mechanisms for information abstraction, prefetching, sampling, progressive data visualization, and summarization to address problems related to visual information overplotting. Furthermore, it is essential to develop new methods that enhance user comprehension by providing customization options tailored to various user-defined exploration scenarios and preferences.

This article presents the main outcomes of the *International Workshop on Big Data Visual Exploration and Analytics* (BigVis). BigVis has been held annually from 2018 to 2023 in conjunction with

the International Conference on Extending Database Technology (EDBT). The series of BigVis workshops has offered a forum for scientists and engineers from different research areas to discuss, exchange and disseminate their work, as well as to highlight challenges that bridge together these communities.

Over the past 6 years (2018–2023), more than 250 participants, 43 published papers, 147 papers' authors, 103 PC members, 3 Journal Special Issues, and 6 Keynotes have contributed to the success of the BigVis workshops. This report summarizes six years of organizing the BigVis workshop, presenting the main findings and outcomes.

2. HIGHLIGHTS

This section provides an overview of the key highlights, including keynotes, special issues, reports, and blog posts.

Keynotes

- **2023:** Daniel Keim: "*The Role of Interactive Visualization in Human-Centered AI*"
- **2022:** Danyel Fisher: "*Co-Designing the Data Structure and the User Experience*"
- **2022:** Steffen Frey: "*Visual Mapping, Comparison and Exploration of Large Multifield Data*"
- **2021:** Georgia Koutrika: "*The Rise of Intelligent Data Assistants: Democratizing Data Access*"
- **2021:** Michael Sedlmair: "*Machine Learning meets Visualization*"
- **2018:** Bill Howe: "*Viziometrics: Mining Visualizations in the Scientific Literature*"

¹Hellenic Mediterranean University, Greece ²ATHENA Research Center, Greece ³University of Pittsburgh, USA
⁴Brandeis University, USA ⁵University Bordeaux, France ⁶University of Groningen, Netherlands ⁷Keio University, Japan
⁸Utrecht University, Netherlands ⁹University of Groningen, Netherlands ¹⁰University of California-Davis, USA
¹¹Graz University of Technology, Austria ¹²University of Stuttgart, Germany ¹³United Arab Emirates University, United Arab Emirates

Table 1. Published Papers Categorization based on CFP Topics & Years

Topic	2018	2019	2020	2021	2022	2023
Graph Data	[40][43]		[24] [27]	[17]		
Knowledge Graph & Linked Data	[42]	[35][36]	[21][26]			[5]
Spatiotemporal	[41]	[34]	[20][23]	[13]	[9]	[1]
Data Mining, Knowledge & ML		[32][36]	[22][28]		[8][10]	[3]
Personalization, Recommendations & Assistance		[30]	[25]			
Data representation & Novel interface	[39]		[29]	[14]		[4]
Data modeling, Storage, Indexing & Query processing		[37]		[16]	[11]	
Domain-specific approaches		[31][33]		[15][18]	[7][12]	[5][2]

Journal Special Issues

- **2022:** *Machine Learning Approaches in Big Data Visualization*, IEEE Computer Graphics and Applications, IEEE. The article [51] is an extended version of an invited paper that appears in BigVis 2021.
- **2020:** *Interactive Big Data Visualization and Analytics*, Big Data Research Journal, Elsevier. The articles [46][47][48][49][50] are extended versions of invited papers that appear in BigVis 2020.
- **2018:** *Big Data Exploration*, Visualization and Analytics, Big Data Research Journal, Elsevier. The articles [44][45] are extended versions of invited papers that appear in BigVis 2018.

Community Report & Blog Posts

- **2020:** *Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications* [19]
- **2020:** Two ACM SIGMOD Blog posts¹

¹ Part1: <https://wp.sigmod.org/?p=3037>
Part2: <https://wp.sigmod.org/?p=3123>

3. PUBLISHED WORKS

The BigVis workshops have hosted several types of submissions, including novel research works, completed or in-progress work, vision, and system or demonstration papers. These works covered a wide spectrum of research topics related to the types of visualized data and the methods used to prepare data for visualization, or the techniques used to visualize, analyze, and facilitate user interaction. Table 1 summarizes the presented contributions, organizing them based on their main topic and year of presentation.

Five papers referred to the visualization of *graph data*, mainly focusing on the effective visualization of complex networks and big data graphs, such as summarized or 3D visualizations of graph networks. In addition, six focused on the representation of the *semantics of graph data*, mainly working on ways to visualize *knowledge graphs* and *Linked Data*. Another line of presented works (7 papers) referred to *spatiotemporal* data, i.e., methods for visualizing spatial objects, traces, trajectories, timeseries and novel applications dealing with such data. Next, the use of *data mining and machine learning* techniques has been explored by seven works that aim at facilitating the

interactive exploration, and visual analysis of different tasks, such as anomaly detection, opinion analysis, or popularity bias. Other works aimed at offering a *user-centric approach to visual analytics*, offering ways for the personalization of visual results, visual recommendations, and user assistance in accomplishing a visual task. Yet other works proposed new methods for *data representation and user interaction*, combining visual and chat-based interfaces, whereas a few works dealt with topics related to the *efficiency of producing visual results*, focusing on issues related to data modeling, caching, indexing and query processing. Finally, eight works presented *domain-specific* visual applications and demonstrations from different sectors, such as agriculture, energy, media industry, cultural heritage, and sports.

4. CHALLENGES & EMERGING APPLICATIONS

Another highlight in the context of the BigVis 2020, concerned the joint report prepared by *fourteen distinguished scientists*, from different communities, who were invited to provide their insights regarding promising research challenges and applications related to *Big Data visualization and analytics* [19].

The report identified that modern systems *face significant user-centric challenges*. They must comprehend users' needs to help them solve problems and offer guidance ("Show the Data not Seen by Humans"). In this context, fundamental challenges arise, such as: (a) recommending data views that the users might want to analyze; (b) identifying the data segments that will be useful for specific tasks; (c) creating data stories and explanations; (d) designing novel interfaces that assist users to understand data types and properties; (e) integrating human factors related to vision and perception into the data analysis pipeline, enabling users to supervise or provide feedback to the systems.

Another key challenge pertains *to the scalability and efficiency* of these systems. This is to enable visualization systems to efficiently handle billion objects datasets, while ensuring that the response time is limited to just a few milliseconds. Addressing this challenge involves developing tools capable of performing interactive operations and complex analytics over massive sets of data. In that respect, there is the need to develop novel approaches (e.g., progressive data processing & indexing) that can handle large streaming, sampled, uncertain, high-dimensional, and noisy data.

Next, within the context of visual analysis, addressing data management problems reveals "new" challenges for *data-intensive visual applications*, such as visualization-centric algebras, design of visualization operators, optimization techniques, and effective storage and indexing scheme for visual analysis.

Finally, building *interactive tools* and enabling visual analysis *for Machine Learning* applications presents a significant challenge. For example, developing visual methods for interpreting and techniques for interacting with ML models; implementing visualization systems that facilitate model troubleshooting, debugging, and comparison.

5. REFERENCES

- [1] N. Bikakis, G. Giannopoulos, N. Sidiropoulos, C. Flouda, A. Doupas, V. Giouli, P. Karioris, P. Botini, A. Vacalopoulou, G. Stainhaouer: Exposing Geospatial Cultural Heritage Content in Map-based Applications. BigVis 2023
- [2] I. Mavroudopoulos, T. Toliopoulos, A. Gounaris, G. Kynigopoulos, A. Andreadis, I. Kalfas: HARVEST: a complete solution for smart agriculture monitoring. BigVis 2023
- [3] I. Fujishiro, N. Sawada, M. Uemura: Clustering, Universality, and Evolutionary Schema Design. BigVis 2023
- [4] G. Katsogiannis-Meimarakis, C. Tsapelas, G. Koutrika: Natural Language Data Interfaces: From Keyword Search to ChatGPT, are we there yet? BigVis 2023
- [5] V. Stamatopoulos, S. Maroulis, K. Kozanis, I. Psarros, G. Papastefanatos, G. Giannopoulos, M. Terrovitis: A Tool for Visual Exploration and Analysis of Solar Photovoltaic Module Data. BigVis 2023
- [6] J. Abello, H. Zhang: Graph Peeling Semantics. BigVis 2023
- [7] B. Wang, N. Adachi, I. Fujishiro: FORSETI: A Provenance-aware Visual Analysis Environment for the Lifecycle Management of E-autopsy Reports. BigVis 2022
- [8] A. Garg, N. Tandon, A.S. Varde: CSK-SNIFFER: Commonsense Knowledge for Sniffing Object Detection Errors. BigVis 2022
- [9] M. Schoemans, A.A. Sakr, E. Zimányi: MOVE: Interactive Visual Exploration of Moving Objects. BigVis 2022
- [10] G. Papastefanatos, G. Alexiou, N. Bikakis, S. Maroulis, V. Stamatopoulos: VisualFacts: A Platform for In-Situ Visual Exploration and Real-Time Entity Resolution. BigVis 2022
- [11] Q. Bai, S. Alsudais, C. Li: Demo of VisBooster: Accelerating Tableau Live Mode Queries Up to 100 Times Faster. BigVis 2022
- [12] X. Zhang, J. Xu, M. Keskin, M. Colaresi, V. Zadorozhny: HR-VEAW: A Human Rights Violation Exploration Analytics, and Warning System. BigVis 2022
- [13] P. Mitropoulos, K. Patroumpas, D. Skoutas, T. Vakkas, S. Athanasiou: BigDataVoyant: Automated Profiling of Large Geospatial Data. BigVis 2021
- [14] T. Cross, R. Bovo, T. Heinis: Browsing Internet Content in Multiple Dimensions. BigVis 2021

- [15] F. Rollo, L. Po: SenseBoard: Sensor Monitoring for Air Quality Experts. *BigVis* 2021
- [16] L. Caruccio, S. Cirillo, V. Deufemia, G. Polese: Real-time visualization of profiling metadata upon data insertions. *BigVis* 2021
- [17] J. Abello, D. Nakhimovich, C. Han, M. Aanjaneya: Graph Cities: Their Buildings, Waves, and Fragments. *BigVis* 2021
- [18] I. Ivkovic-Kihic, D. Seebacher, M. Stein, T. Schreck, R. Preiner: Visual Analysis of Player Interactions in Soccer Games. *BigVis* 2021
- [19] G.L. Andrienko, N.V. Andrienko, S.M. Drucker, J-D. Fekete, D. Fisher, S. Idreos, T. Kraska, G. Li, K-L. Ma, J.D. Mackinlay, A. Oulasvirta, T. Schreck, H. Schumann, M. Stonebraker, D. Auber, N. Bikakis, P.K. Chrysanthis, G. Papastefanatos, M.A. Sharaf: Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications. *BigVis* 2020
- [20] J. Suschnigg, B. Mutlu, A.K. Fuchs, V. Sabol, S. Thalmann, T. Schreck: Exploration of Anomalies in Cyclic Multivariate Industrial Time Series Data for Condition Monitoring. *BigVis* 2020
- [21] K. Li, H. Naacke, B. Amann: Exploring the Evolution of Science with Pivot Topic Graphs. *BigVis* 2020
- [22] B. Breve, L. Caruccio, S. Cirillo, V. Deufemia, G. Polese: Visualizing Dependencies during Incremental Discovery Processes. *BigVis* 2020
- [23] P. Quirós, B. Rosario C. Álvarez: The new Spanish political scenario: Twitter graph and opinion analysis with an interactive visualisation. *BigVis* 2020
- [24] J. Abello, D. Nakhimovich: Graph Waves. *BigVis* 2020
- [25] W. Cui, G. Strazdins, H. Wang: Web-based Scalable Visual Exploration of Large Multidimensional Data Using Human-in-the-Loop Edge Bundling in Parallel Coordinates. *BigVis* 2020
- [26] F. Desimoni, L. Po: Providing Effective Visualizations over Big Linked Data. *BigVis* 2020
- [27] P. Camarillo-Ramírez, F. Cervantes-Alvarez, L. F. Gutiérrez-Preciado: A task-based evaluation methodology for visual representation of dynamic networks. *BigVis* 2020
- [28] R. Borges, K. Stefanidis: On Measuring Popularity Bias in Collaborative Filtering Data. *BigVis* 2020
- [29] X. Zhang, X. Ge, P.K. Chrysanthis: A GUI Design for Composition Discovery of View Interestingness. *BigVis* 2020
- [30] X. Zhang, X. Ge, P.K. Chrysanthis, M.A. Sharaf: ViewSeeker: An Interactive View Recommendation Tool. *BigVis* 2019
- [31] T. Cerquitelli, E. Di Corso, S. Proto, A. Capozzoli, F. Bellotti, M.G. Cassese, E. Baralis, M. Mellia, S. Casagrande, M. Tamburini: Exploring energy performance certificates through visualization. *BigVis* 2019
- [32] J.S. Andersen: A User Centric Visual Analytics Framework for News Discussions. *BigVis* 2019
- [33] P. Bouros, I. Assent, J. Jeppesen, T.S. Toftgaard: Processing Satellite Imagery for Decision-Making in Precision Agriculture. *BigVis* 2019
- [34] A. Gogolou, T. Tsandilas, T. Palpanas, A. Bezerianos: Progressive Similarity Search on Time Series Data. *BigVis* 2019
- [35] I. Manolescu, M. Mazuran: Speeding up RDF Aggregate Discovery through Sampling. *BigVis* 2019
- [36] M. Meimaris: Visualizing Implicit RDF Schema with the Help of Embeddings. *BigVis* 2019
- [37] A. Watson, S. Bateman, S. Ray: PySnippet: Accelerating Exploratory Data Analysis in Jupyter Notebook through Facilitated Access to Example Code. *BigVis* 2019
- [38] C. Athinaiou, H. Kondylakis: VESEL: Visual Exploration of Schema Evolution using Provenance Queries. *BigVis* 2019
- [39] G. Richer, J. Sansen, F. Lalanne, D. Auber, Romain Bourqui: Enabling Hierarchical Exploration for Large-Scale Multidimensional Data with Abstract Parallel Coordinates. *BigVis* 2018
- [40] D. Bleco, Y. Kotidis: Finding the Needle in a Haystack: Entropy Guided Exploration of Very Large Graph Cubes. *BigVis* 2018
- [41] G. Chatzigeorgakidis, D. Skoutas, K. Patroumpas, S. Athanasiou, S. Skiadopoulos: Map-Based Visual Exploration of Geolocated Time Series. *BigVis* 2018
- [42] M-E Papadaki, P. Papadacos, M. Mountantonakis, Y. Tzitzikas: An Interactive 3D Visualization for the LOD Cloud. *BigVis* 2018
- [43] T. Repke, R. Krestel: Topic-aware Network Visualisation to Explore Large Email Corpora. *BigVis* 2018
- [44] G. Richer, J. Sansen, F. Lalanne, D. Auber, R. Bourqui: HiePaCo: Scalable Hierarchical Exploration in Abstract Parallel Coordinates Under Budget Constraints. *Big Data Res.* 17, 2019
- [45] G. Chatzigeorgakidis, K. Patroumpas, D. imitrios Skoutas, S. Athanasiou, S. Skiadopoulos: Visual Exploration of Geolocated Time Series with Hybrid Indexing. *Big Data Res.* 15, 2019
- [46] X. Zhang, X. Ge, P.K. Chrysanthis, M.A. Sharaf: ViewSeeker: An Interactive View Recommendation Framework. *Big Data Res.* 25, 2021
- [47] B. Breve, L. Caruccio, S. Cirillo, V. Deufemia, G. Polese: Dependency Visualization in Data Stream Profiling. *Big Data Res.* 25, 2021
- [48] J. Abello, D. Nakhimovich: Graph Waves. *Big Data Res.* 29, 2022
- [49] K. Li, H. Naacke, B. Amann: An Analytic Graph Data Model and Query Language for Exploring the Evolution of Science. *Big Data Res.* 26, 2021
- [50] J. Suschnigg, B. Mutlu, G. Koutroulis, V. Sabol, S. Thalmann, T. Schreck: Visual Exploration of Anomalies in Cyclic Time Series Data with Matrix and Glyph Representations. *Big Data Res.* 26, 2021
- [51] J. Abello, H. Zhang, D. Nakhimovich, C. Han, M. Aanjaneya: Giga Graph Cities: Their Buckets, Buildings, Waves, and Fragments. *IEEE Computer Graphics and Applications* 42(3), 2022