Bibliography

[1] E. F. Codd. "A Relational Model of Data for Large Shared Data Banks". In: *Commun. ACM* 13.6 (June 1970), pp. 377–387.

[2] David A Bearman and Richard H Lytle. "The power of the principle of provenance". In: *Archivaria* (1985), pp. 14–27.

[3] Gautam Bhargava and Shashi K Gadia. "Relational database systems with zero information loss". In: *IEEE Transactions on Knowledge and Data engineering* 5.1 (1993), pp. 76–87.

[4] Avi Silberschatz, Henry F Korth, and S Sudarshan. "Data models". In: *ACM Computing Surveys (CSUR)* 28.1 (1996), pp. 105–108.

[5] Wilburt J Labio et al. "The WHIPS prototype for data warehouse creation and maintenance". In: *Proceedings of ACM SIGMOD International conference on Management of data*. 1997, pp. 557–559.

[6] Allison Woodruff and Michael Stonebraker. "Supporting fine-grained data lineage in a database visualization environment". In: *Proceedings 13th International Conference on Data Engineering*. IEEE. 1997, pp. 91–102.

[7] Yingwei Cui and Jennifer Widom. "Storing auxiliary data for efficient maintenance and lineage tracing of complex views". In: *Proceedings of 2nd International Workshop on Design and Management of Data Warehouses (DMDW)*. Stanford InfoLab. 1999.

[8] John R Mashey. "Big data and the next wave of InfraStress problems, solutions, opportunities". In: {*USENIX*} *Annual Technical Conference*. 1999.

[9] Peter Buneman, Sanjeev Khanna, and Wang-Chiew Tan. "Data provenance: Some basic issues". In: *Proceedings of International Conference on Foundations of Software Technology and Theoretical Computer Science*. Springer. 2000, pp. 87–93.

[10] Yingwei Cui and Jennifer Widom. "Lineage tracing in a data warehousing system". In: *Proceedings of 16th International Conference on Data Engineering*. IEEE. 2000, pp. 683–684.

[11] Yingwei Cui and Jennifer Widom. "Practical lineage tracing in data warehouses". In: *Proceedings of 16th International Conference on Data Engineering (Cat. No. 00CB37073)*. IEEE. 2000, pp. 367–378.

[12] Yingwei Cui, Jennifer Widom, and Janet L Wiener. "Tracing the lineage of view data in a warehousing environment". In: *ACM Transactions on Database Systems (TODS)* 25.2 (2000), pp. 179–227.

[13] Peter Buneman, Sanjeev Khanna, and Tan Wang-Chiew. "Why and where: A characterization of data provenance". In: *Proceedings of International conference on database theory*. Springer. 2001, pp. 316–330.

[14] Yingwei Cui. "Lineage tracing in data warehouses". PhD thesis. Stanford InfoLab, 2001.

[15] Yingwei Cui and Jennifer Widom. *Run-time translation of view tuple deletions using data lineage*. Tech. rep. Stanford, 2001.

[16] Peter Buneman, Sanjeev Khanna, and Wang-Chiew Tan. "On propagation of deletions and annotations through views". In: *Proceedings of 21st ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 2002, pp. 150–158.

[17] M Greenwood et al. "Provenance of e-Science Experiments-Experience from Bioinformatics". In: *Proceedings of UK e-Science Programme All Hands Conference*. Engineering and Physical Sciences Research Council. 2003, pp. 223–226.

[18] James D Myers et al. *Multi-scale science: supporting emerging practice with semantically derived provenance*. Tech. rep. Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2003.

[19] Robert D Stevens, Alan J Robinson, and Carole A Goble. "myGrid: personalised bioinformatics on the information grid". In: *Bioinformatics* 19.suppl_1 (2003), pp. i302–i304.

[20] Wang-Chiew Tan. "Containment of relational queries with annotation propagation". In: *Proceedings of International Workshop on Database Programming Languages*. Springer. 2003, pp. 37–53.

[21] Jun Zhao et al. "Annotating, linking and browsing provenance logs for e-science". In: *Proceedings of Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*. Vol. 176. Citeseer. 2003.

[22] Peter Buneman et al. "Archiving scientific data". In: *ACM Transactions on Database Systems (TODS)* 29.1 (2004), pp. 2–42.

[23] Dennis P Groth. "Information provenance and the knowledge rediscovery problem". In: *Proceedings of 8th International Conference on Information Visualisation*. IEEE. 2004, pp. 345–351.

[24] Jun Zhao et al. "Semantically linking and browsing provenance logs for e-science". In: *Proceedings of International Conference on Semantics for the Networked World*. Springer. 2004, pp. 158–176.

[25] Jun Zhao et al. "Using semantic web technologies for representing e-science provenance". In: *Proceedings of International Semantic Web Conference*. Springer. 2004, pp. 92–106.

[26] Omar Benjelloun et al. *ULDBs: Databases with uncertainty and lineage*. Tech. rep. Stanford, 2005.

[27] Deepavali Bhagwat et al. "An annotation management system for relational databases". In: *The VLDB Journal* 14.4 (2005), pp. 373–396.

[28] Laura Chiticariu, Wang-Chiew Tan, and Gaurav Vijayvargiya. "DBNotes: a post-it system for relational databases based on provenance". In: *Proceedings of ACM SIGMOD International conference on Management of data*. 2005, pp. 942–944.

[29] Zachary G Ives et al. "ORCHESTRA: Rapid, Collaborative Sharing of Dynamic Data." In: *Proceedings of CIDR'05*. 2005, pp. 107–118.

[30] Yogesh L Simmhan, Beth Plale, and Dennis Gannon. "A survey of data provenance techniques". In: *Computer Science Department, Indiana University, Bloomington IN* 47405 (2005), p. 69.

[31] J Widom Trio. "A system for integrated management of data, accuracy, and lineage". In: *Proceedings of CIDR* (2005).

[32] Yannis Velegrakis, Renee J Miller, and John Mylopoulos. "Representing and querying data transformations". In: *Proceedings of 21st International Conference on Data Engineering (ICDE'05)*. IEEE. 2005, pp. 81–92.

[33] Parag Agrawal et al. "Trio: a system for data, uncertainty, and lineage". In: *Proceedings of 32nd International conference on Very large data bases*. 2006, pp. 1151–1154.

[34] Ilkay Altintas, Oscar Barney, and Efrat Jaeger-Frank. "Provenance collection support in the kepler scientific workflow system". In: *Proceedings of International Provenance and Annotation Workshop, IPAW*. Springer. 2006, pp. 118–132.

[35] Omar Benjelloun et al. *An introduction to ULDBs and the Trio system*. Tech. rep. Stanford InfoLab, 2006.

[36] Omar Benjelloun et al. "ULDBs: Databases with Uncertainty and Lineage". In: *Proceedings of 32nd International Conference on Very Large Data Bases*. VLDB '06. Seoul, Korea: VLDB Endowment, 2006, pp. 953–964.

[37] Shawn Bowers et al. "A model for user-oriented data provenance in pipelined scientific workflows". In: *Proceedings of International Provenance and Annotation Workshop, IPAW*. Springer. 2006, pp. 133–147.

[38] Peter Buneman et al. "A provenance model for manually curated data". In: *Proceedings of International Provenance and Annotation Workshop, IPAW*. Springer. 2006, pp. 162–170.

[39] Steven P Callahan et al. "VisTrails: visualization meets data management". In: *Proceedings of ACM SIGMOD International conference on Management of data*. 2006, pp. 745–747.

[40] Mohamed Y Eltabakh, Mourad Ouzzani, and Walid G Aref. "BDBMS–a database management system for biological data". In: *arXiv preprint cs/0612127* (2006).

[41] Floris Geerts, Anastasios Kementsietsidis, and Diego Milano. "Mondrian: Annotating and querying databases through colors and blocks". In: *Proceedings of International Conference on Data Engineering (ICDE'06)*. IEEE. 2006, pp. 82–82.

[42] Bertram Ludäscher et al. "Scientific workflow management and the Kepler system". In: *Concurrency and computation: Practice and experience* 18.10 (2006), pp. 1039–1065.

[43] Anish Das Sarma et al. "Working models for uncertain data". In: *Proceedings of 22nd International Conference on Data Engineering (ICDE'06)*. IEEE. 2006, pp. 7–7.

[44] Parag Agrawal and Jennifer Widom. "Confidence-aware joins in large uncertain databases". In: *Stanford University Technical Report* (2007).

[45] Giuseppe DeCandia et al. "Dynamo: amazon's highly available key-value store". In: *ACM SIGOPS operating systems review* 41.6 (2007), pp. 205–220.

[46] Todd J Green, Grigoris Karvounarakis, and Val Tannen. "Provenance semirings". In: *Proceedings of 26th ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 2007, pp. 31–40.

[47] Todd J Green et al. "ORCHESTRA: facilitating collaborative data sharing". In: *Proceedings of ACM SIGMOD International conference on Management of data*. 2007, pp. 1131–1133.

[48] Todd J Green et al. "Update exchange with mappings and provenance". In: *VLDB* (2007).

[49] Michi Mutsuzaki et al. "Trio-One: Layering uncertainty and lineage on a conventional DBMS". In: *Proceedings of 3rd Biennial Conference on Innovative Data Systems Research*. 2007, pp. 269–274.

[50] Anish Das Sarma, Jeffrey Ullman, and Jennifer Widom. "Functional dependencies for uncertain relations". In: *ICDE 2008* (2007).

[51] Anish Das Sarma, Jeffrey Ullman, and Jennifer Widom. "Schema design for uncertain databases". In: *Proceedings of Foundations of Data Management*. Stanford InfoLab, 2007.

[52] Claudio T Silva, Juliana Freire, and Steven P Callahan. "Provenance for visualizations: Reproducibility and beyond". In: *Computing in Science & Engineering* 9.5 (2007), pp. 82–89.

[53] Divesh Srivastava and Yannis Velegrakis. "Intensional associations between data and metadata". In: *Proceedings of ACM SIGMOD international conference on Management of data*. 2007, pp. 401–412.

[54] Divesh Srivastava and Yannis Velegrakis. "MMS: using queries as data values for metadata management". In: *Proceedings of 23rd International Conference on Data Engineering*. IEEE. 2007, pp. 1481–1482.

[55] Divesh Srivastava and Yannis Velegrakis. "Using queries to associate metadata with data". In: *Proceedings of 23rd International Conference on Data Engineering*. IEEE. 2007, pp. 1451–1453.

[56] Stijn Vansummeren and James Cheney. "Recording Provenance for SQL Queries and Updates." In: *IEEE Data Eng. Bull.* 30.4 (2007), pp. 29–37.

[57] Renzo Angles and Claudio Gutierrez. "Survey of graph database models". In: *ACM Computing Surveys (CSUR)* 40.1 (2008), pp. 1–39.

[58] Shawn Bowers, Timothy M McPhillips, and Bertram Ludäscher. "Provenance in collection-oriented scientific workflows". In: *Concurrency and Computation: Practice and Experience* 20.5 (2008), pp. 519–529.

[59] Fay Chang et al. "Bigtable: A distributed storage system for structured data". In: *ACM Transactions on Computer Systems (TOCS)* 26.2 (2008), pp. 1–26.

[60] Susan B Davidson and Juliana Freire. "Provenance and scientific workflows: challenges and opportunities". In: *Proceedings of ACM SIGMOD International conference on Management of data*. 2008, pp. 1345–1350.

[61] Tommy Ellkvist et al. "Using provenance to support real-time collaborative design of workflows". In: *Proceedings of International Provenance and Annotation Workshop, IPAW*. Springer. 2008, pp. 266–279.

[62] Mohamed Y Eltabakh et al. "Managing biological data using bdbms". In: *Proceedings of 24th International Conference on Data Engineering*. IEEE. 2008, pp. 1600–1603.

[63] Bill Howe et al. "End-to-end escience: Integrating workflow, query, visualization, and provenance at an ocean observatory". In: *Proceedings of 4th International Conference on eScience*. IEEE. 2008, pp. 127–134.

[64] Zachary G Ives et al. "The orchestra collaborative data sharing system". In: *ACM Sigmod Record* 37.3 (2008), pp. 26–32.

[65] Bertram Ludäscher et al. "From computation models to models of provenance: the RWS approach". In: *Concurrency and Computation: Practice and Experience* 20.5 (2008), pp. 507–518.

[66] Anish Das Sarma, Martin Theobald, and Jennifer Widom. *Data modifications and versioning in Trio*. Tech. rep. Stanford InfoLab, 2008.

[67] Anish Das Sarma, Martin Theobald, and Jennifer Widom. "Exploiting lineage for confidence computation in uncertain and probabilistic databases". In: *Proceedings of 24th International Conference on Data Engineering*. IEEE. 2008, pp. 1023–1032.

[68] Anish Das Sarma et al. "Towards Special-Purpose Indexes and Statistics for Uncertain Data". In: *Proceedings of Workshop on Management of Uncertain Data*. 2008, p. 57.

[69] Carlos Scheidegger et al. "Tackling the provenance challenge one layer at a time". In: *Concurrency and Computation: Practice and Experience* 20.5 (2008), pp. 473–483.

[70] Carlos E Scheidegger et al. "Querying and re-using workflows with VsTrails". In: *Proceedings of ACM SIGMOD International conference on Management of data*. 2008, pp. 1251–1254.

[71] Yogesh L Simmhan, Beth Plale, and Dennis Gannon. "Karma2: Provenance management for data-driven workflows". In: *International Journal of Web Services Research (IJWSR)* 5.2 (2008), pp. 1–22.

[72] Yogesh L Simmhan, Beth Plale, and Dennis Gannon. "Query capabilities of the Karma provenance framework". In: *Concurrency and Computation: Practice and Experience* 20.5 (2008), pp. 441–451.

[73] Jennifer Widom et al. "Trio: A System for Integrated Management of Data". In: *Uncertainty, and Lineage* (2008).

[74] Charu C Aggarwal. "Trio a system for data uncertainty and lineage". In: *Managing and Mining Uncertain Data*. Springer, 2009, pp. 1–35.

[75] Manish Kumar Anand et al. "Efficient provenance storage over nested data collections". In: *Proceedings of 12th International Conference on Extending Database Technology: Advances in Database Technology*. 2009, pp. 958–969.

[76] Manish Kumar Anand et al. "Exploring scientific workflow provenance using hybrid queries over nested data and lineage graphs". In: *Proceedings of International Conference on Scientific and Statistical Database Management*. Springer. 2009, pp. 237–254.

[77] Bin Cao et al. "Provenance information model of karma version 3". In: *Proceedings of Congress on Services-I*. IEEE. 2009, pp. 348–351.

[78] James Cheney et al. "Provenance: a future history". In: *Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications*. 2009, pp. 957–964.

[79] Mohamed Y Eltabakh et al. "Supporting annotations on relations". In: *Proceedings of 12th International Conference on Extending Database Technology: Advances in Database Technology*. 2009, pp. 379–390.

[80] B. Glavic and G. Alonso. "Perm: Processing Provenance and Data on the Same Data Model through Query Rewriting". In: *Proceedings of 25th International Conference on Data Engineering*. 2009, pp. 174–185.

[81] Robert Ikeda and Jennifer Widom. *Data lineage: A survey*. Tech. rep. Stanford InfoLab, 2009.

[82] Timothy McPhillips et al. "Scientific workflow design for mere mortals". In: *Future Generation Computer Systems* 25.5 (2009), pp. 541–551.

[83] Peter Buneman and Susan B Davidson. "Data provenance–the foundation of data quality". In: *Proceedings of Workshop: Issues and Opportunities for Improving the Quality and Use of Data*. 2010.

[84] E Deelman et al. "Chapter 12: Metadata and provenance management". In: *Scientific Data Management: Challenges, Existing Technology, and Deployment. CRC Press. Available at: http://arxiv. org/ftp/arxiv/papers/1005/1005.2643. pdf* (2010).

[85] Dietrich Featherston. "Cassandra: Principles and application". In: *Department of computer science university of illinois at Urbana-champaign* (2010).

[86] Boris Glavic. "Perm: efficient provenance support for relational databases". PhD thesis. University of Zurich, 2010.

[87] Boris Glavic et al. "TRAMP: understanding the behavior of schema mappings through provenance". In: *Proceedings of the VLDB Endowment* 3.1-2 (2010), pp. 1314–1325.

[88] Todd Green et al. "Provenance in ORCHESTRA". In: *IEEE Data Eng. Bull.* 33 (Oct. 2010), pp. 9–16.

[89] Andreas M Kaplan and Michael Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media". In: *Business horizons* 53.1 (2010), pp. 59–68.

[90] Grigoris Karvounarakis, Zachary G Ives, and Val Tannen. "Querying data provenance". In: *Proceedings of ACM SIGMOD International Conference on Management of data*. 2010, pp. 951–962.

[91] Avinash Lakshman and Prashant Malik. "Cassandra: a decentralized structured storage system". In: *ACM SIGOPS Operating Systems Review* 44.2 (2010), pp. 35–40.

[92] Raghotham Murthy, Robert Ikeda, and Jennifer Widom. "Making aggregation work in uncertain and probabilistic databases". In: *IEEE Transactions on knowledge and data engineering* 23.8 (2010), pp. 1261–1273.

[93] Chad Vicknair et al. "A comparison of a graph database and a relational database: a data provenance perspective". In: *Proceedings of 48th annual Southeast regional conference*. 2010, pp. 1–6.

[94] Geoffrey Barbier and Huan Liu. "Information provenance in social media". In: *Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer. 2011, pp. 276–283.

[95] Daniel Crawl, Jianwu Wang, and Ilkay Altintas. "Provenance for mapreduce-based data-intensive workflows". In: *Proceedings of 6th workshop on Workflows in support of large-scale science*. 2011, pp. 21–30.

[96] Boris Glavic and Renée J Miller. "Reexamining Some Holy Grails of Data Provenance." In: *Proceedings of 3rd {USENIX} Workshop on the Theory and Practice of Provenance, TaPP '11*. 2011.

[97] Todd J Green. "Containment of conjunctive queries on annotated relations". In: *Theory of Computing Systems* 49.2 (2011), pp. 429–459.

[98] Robert Ikeda, Hyunjung Park, and Jennifer Widom. "Provenance for generalized map and reduce workflows". In: *Proceedings of 5th Biennial Conference on Innovative Data Systems Research (CIDR '11)*. Stanford InfoLab, 2011, pp. 273–283.

[99] Tim O'Reilly and Sarah Milstein. *The twitter book*. " O'Reilly Media, Inc.", 2011.

[100] Hyunjung Park, Robert Ikeda, and Jennifer Widom. "Ramp: A system for capturing and tracing provenance in mapreduce workflows". In: *Proceedings of the VLDB Endowment* 4.12 (2011), pp. 1351–1354.

[101] Divyakant Agrawal et al. "Challenges and Opportunities with Big Data: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association". In: *Computing Research Association* (2012).

[102] Yuan Cheng et al. "Towards provenance and risk-awareness in social computing". In: *Proceedings of First International Workshop on Secure and Resilient Architectures and Systems*. 2012, pp. 25–30.

[103] Boris Glavic. "Big data provenance: Challenges and implications for benchmarking". In: *Specifying big data benchmarks*. Springer, 2012, pp. 72–80.

[104] Grigoris Karvounarakis and Todd J Green. "Semiring-annotated data: queries and provenance?" In: *ACM SIGMOD Record* 41.3 (2012), pp. 5–14.

[105] Guoxi Wang and Jianfeng Tang. "The nosql principles and basic application of cassandra model". In: *Proceedings of International conference on computer science and service system*. IEEE. 2012, pp. 1332–1335.

[106] Sherif Akoush, Ripduman Sohan, and Andy Hopper. "Hadoopprov: Towards provenance as a first class citizen in mapreduce". In: *Proceedings of 5th {USENIX} Workshop on the Theory and Practice of Provenance, TaPP '13*. 2013.

[107] Ciro Cattuto et al. "Time-varying social networks in a graph database: a Neo4j use case". In: *Proceedings of First International workshop on graph data management experiences and systems*. 2013, pp. 1–6.

[108] Dunren Che, Mejdl Safran, and Zhiyong Peng. "From big data to big data mining: challenges, issues, and opportunities". In: *Proceedings of International conference on database systems for advanced applications*. Springer. 2013, pp. 1–15.

[109] You-Wei Cheah et al. "Milieu: Lightweight and configurable big data provenance for science". In: *Proceedings of IEEE International Congress on Big Data*. IEEE. 2013, pp. 46–53.

[110] David DeBoer, Wenchao Zhou, and Lisa Singh. "Using substructure mining to identify misbehavior in network provenance graphs". In: *Proceedings of First International Workshop on Graph Data Management Experiences and Systems*. 2013, pp. 1–6.

[111] Devarshi Ghoshal and Beth Plale. "Provenance from log files: a BigData problem". In: *Proceedings of Workshops of the EDBT/ICDT 2013 Joint Conference*. 2013, pp. 290–297.

[112] Boris Glavic, Renée J Miller, and Gustavo Alonso. "Using SQL for efficient generation and querying of provenance information". In: *In search of elegance in the theory and practice of computation*. Springer, 2013, pp. 291–320.

[113] Pritam Gundecha, Zhuo Feng, and Huan Liu. "Seeking provenance of information using social media". In: *Proceedings of 22nd ACM International conference on Information & Knowledge Management*. 2013, pp. 1691–1696.

[114] Pritam Gundecha et al. "A tool for collecting provenance data in social media". In: *Proceedings of 19th ACM SIGKDD International conference on Knowledge discovery and data mining*. 2013, pp. 1462–1465.

[115] Devdatta Kulkarni. "A fine-grained access control model for key-value systems". In: *Proceedings of third ACM conference on Data and application security and privacy*. 2013, pp. 161–164.

[116] Devdatta Kulkarni. "A provenance model for key-value systems". In: *Proceedings of 5th{USENIX} Workshop on the Theory and Practice of Provenance, TaPP '13*. 2013.

[117] Milan Markovic, Peter Edwards, and David Corsar. "A role for provenance in social computation". In: *Proceedings of First International Workshop on Crowdsourcing the Semantic Web-CrowdSem 2013*. CEUR-WS. 2013.

[118] Suhas Ranganath, Pritam Gundecha, and Huan Liu. "A tool for assisting provenance search in social media". In: *Proceedings of 22nd ACM International conference on Information & Knowledge Management*. 2013, pp. 2517–2520.

[119] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph databases*. " O'Reilly Media, Inc.", 2013.

[120] Rajeev Agrawal et al. "A layer based architecture for provenance in big data". In: *Proceedings of IEEE International Conference on Big Data (Big Data)*. IEEE. 2014, pp. 1–7.

[121] Bahareh Arab et al. "A generic provenance middleware for queries, updates, and transactions". In: *Proceedings of 6th {USENIX} Workshop on the Theory and Practice of Provenance, TaPP '14*. 2014.

[122] Melyssa Barata, Jorge Bernardino, and Pedro Furtado. "Ycsb and tpc-h: Big data and decision support benchmarks". In: *Proceedings of International Congress on Big Data*. IEEE. 2014, pp. 800–801.

[123] Graham Kirby et al. "Comparing relational and graph databases for pedigree data sets". In: *Proceedings of Workshop on Population Reconstruction, Amsterdam, Netherlands*. 2014.

[124] Vladimir Korolev, Anupam Joshi, et al. "PROB: A tool for tracking provenance and reproducibility of big data experiments". In: *Reproduce'14. HPCA 2014* (2014).

[125] David Bermbach et al. "Informed schema design for column store-based database services". In: *Proceedings of 8th International Conference on Service-Oriented Computing and Applications (SOCA)*. IEEE. 2015, pp. 163–172.

[126] Amit Chavan et al. "Towards a unified query language for provenance and versioning". In: *Proceedings of 7th {USENIX} Workshop on the Theory and Practice of Provenance, TaPP '15*. 2015.

[127] Artem Chebotko, Andrey Kashlev, and Shiyong Lu. "A big data modeling methodology for Apache Cassandra". In: *Proceedings of International Congress on Big Data*. IEEE. 2015, pp. 238–245.

[128] Alfredo Cuzzocrea. "Provenance research issues and challenges in the big data era". In: *Proceedings of 39th Annual Computer Software and Applications Conference*. Vol. 3. IEEE. 2015, pp. 684–686.

[129] Tom De Nies et al. "Towards multi-level provenance reconstruction of information diffusion on social media". In: *Proceedings of 24th ACM International on Conference on Information and Knowledge Management, CIKM*. 2015, pp. 1823–1826.

[130] Rosa Filgueira et al. "dispel4py: An agile framework for data-intensive escience". In: *Proceedings of 11th International Conference on e-Science*. IEEE. 2015, pp. 454–464.

[131] Roger Hernandez et al. "Automatic query driven data modelling in Cassandra". In: *Procedia Computer Science* 51 (2015), pp. 2822–2826.

[132] Fangfang Li et al. "Coupling Analysis Between Twitter and Call Centre". In: *arXiv preprint arXiv:1509.02238* (2015).

[133] Felipe Mathias Schmidt et al. "Change data capture in NoSQL databases: A functional and performance comparison". In: *Proceedings of IEEE Symposium on Computers and Communication (ISCC)*. IEEE. 2015, pp. 562–567.

[134] Sugam Sharma. "An extended classification and comparison of nosql big data models". In: *arXiv preprint arXiv:1509.08035* (2015).

[135] Io Taxidou et al. "Modeling information diffusion in social media as provenance with W3C PROV". In: *Proceedings of the 24th International Conference on World Wide Web*. 2015, pp. 819–824.

[136] Jianwu Wang et al. "Big data provenance: Challenges, state of the art and opportunities". In: *Proceedings of IEEE International Conference on Big Data (Big Data)*. IEEE. 2015, pp. 2509–2516.

[137] Teddy Aryono. "Modelling Social Media Semi-structured Data with Graph Database". In: *Proceedings of International Conference on Innovation, Entrepreneurship and Technology (ICONIET)*. 2016.

[138] David Corsar, Milan Markovic, and Peter Edwards. "Social media data in research: provenance challenges". In: *Proceedings of International Provenance and Annotation Workshop, IPAW*. Springer. 2016, pp. 195–198.

[139] Alfredo Massimiliano Cuzzocrea. "Big data provenance: State-of-the-art analysis and emerging research challenges". In: *Proceedings of Workshops of the EDBT/ICDT 2016 Joint Conference*. Vol. 1558. CEUR-WS. 2016.

[140] Khalid Mahmood. "Performance Comparison of NOSQL Database Cassandra and SQL Server for Large Databases". In: *Journal of Independent Studies and Research (JISR)* 14.2 (2016).

[141] Dharavath Ramesh, Ashay Sinha, and Suraj Singh. "Data modelling for discrete time series data using Cassandra and MongoDB". In: *Proceedings of 3rd international conference on recent advances in information technology (RAIT)*. IEEE. 2016, pp. 598–601.

[142] Yucel Tas, Mohamed Jehad Baeth, and Mehmet S Aktas. "An approach to standalone provenance systems for big social provenance data". In: *Proceedings of 12th International Conference on Semantics, Knowledge and Grids (SKG)*. IEEE. 2016, pp. 9–16.

[143] Liang Zhao et al. "A topic-focused trust model for Twitter". In: *Computer Communications* 76 (2016), pp. 1–11.

[144] Mohamed Jehad Baeth and Mehmet S Aktas. "A large scale synthetic social provenance database". In: *Proceedings of Ninth International Conference on Advances in Databases, Knowledge, and Data Applications*. 2017, pp. 16–22.

[145] Roberto Boselli et al. "A Pipeline for Multimedia Twitter Analysis through Graph Databases: Preliminary Results." In: *Proceedings of 6th International Conference on Data Science, Technology and Applications, DATA*. 2017, pp. 343–349.

[146] Anu Chacko and SD Madhu Kumar. "Big data provenance research directions". In: *Proceedings of IEEE Region 10 Conference, TENCON*. IEEE. 2017, pp. 651–656.

[147] Chi Thang Duong et al. "Provenance-based rumor detection". In: *Proceedings of Australasian Database Conference, ADC*. Springer. 2017, pp. 125–137.

[148] Gabriel Campero Durand et al. "Backlogs and Interval Timestamps: Building Blocks for Supporting Temporal Queries in Graph Databases." In: *Proceedings of EDBT/ICDT Workshops*. 2017.

[149] Melanie Herschel, Ralf Diestelkämper, and Houssem Ben Lahmar. "A survey on provenance: What for? What form? What from?" In: *The VLDB Journal* 26.6 (2017), pp. 881–906.

[150] Fernanda Hondo et al. "Data provenance management for bioinformatics workflows using NoSQL database systems in a cloud computing environment". In: *Proceedings of International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2017, pp. 1929–1934.

[151] David Reinsel, John Gantz, and John Rydning. "Data age 2025: The evolution of data to life-critical". In: *Don't Focus on Big Data* (2) (2017).

[152] Mirela Riveni et al. "Provenance in Social Computing: A Case Study". In: *Proceedings of 13th International Conference on Semantics, Knowledge and Grids (SKG)*. IEEE. 2017, pp. 77–84.

[153] Ilkay Melek Yazici, Mehmet S Aktas, and Mehmet Gokturk. "A Novel Approach to User Involved Big Data Provenance Visualization". In: *DBKDA 2017* (2017), p. 19.

[154] Enzhi Zhang, Jinan Fiaidhi, and Sabah Mohammed. "Social Recommendation using Graph Database Neo4j: Mini Blog, Twitter Social Network Graph Case Study". In: *International Journal of Future Generation Communication and Networking* 10.2 (2017), pp. 9–20.

[155] Renzo Angles and Claudio Gutierrez. "An introduction to graph data management". In: *Graph Data Management*. Springer, 2018, pp. 1–32.

[156] Zhuo Feng, Pritam Gundecha, and Huan Liu. "Social Provenance". In: *Encyclopedia of Social Network Analysis and Mining*. Springer New York, 2018, pp. 2768–2772.

[157] Diogo Fernandes and Jorge Bernardino. "Graph Databases Comparison: AllegroGraph, ArangoDB, InfiniteGraph, Neo4J, and OrientDB." In: *Proceedings of the 7th International Conference on Data Science, Technology and Applications,DATA*. 2018, pp. 373–380.

[158] Kaggle. *twitter data set*. 2018. URL: `https://www.kaggle.com/umarhabib/pulwama-killing-twitter-data`.

[159] Dharavath Ramesh and Anand Kumar. "Query Driven implementation of Twitter base using Cassandra". In: *Proceedings of International Conference on Current Trends towards Converging Technologies (ICCTCT)*. IEEE. 2018, pp. 1–4.

[160] Yann Ramusat, Silviu Maniu, and Pierre Senellart. "Semiring provenance over graph databases". In: *Proceedings of 10th {USENIX} Workshop on the Theory and Practice of Provenance, TaPP '18*. 2018.

[161] Anisha P Rodrigues and Niranjan N Chiplunkar. "Real-time Twitter data analysis using Hadoop ecosystem". In: *Cogent Engineering* 5.1 (2018), p. 1534519.

[162] Axel J. Soto et al. "Data Quality Challenges in Twitter Content Analysis for Informing Policy Making in Health Care". In: *Proceedings of 51st Hawaii International Conference on System Sciences, HICSS, Hilton Waikoloa Village, Hawaii, USA, January 3-6*. 2018, pp. 1–10.

[163] Io Taxidou et al. "Web-scale provenance reconstruction of implicit information diffusion on social media". In: *Distributed and Parallel Databases* 36.1 (2018), pp. 47–79.

[164] Zhihao Yuan et al. "Utilizing provenance in reusable research objects". In: *Informatics*. Vol. 5. 1. Multidisciplinary Digital Publishing Institute. 2018, p. 14.

[165] David Allen et al. "Understanding trolls with efficient analytics of large graphs in neo4j". In: *BTW 2019* (2019).

[166] Peter Buneman and Wang-Chiew Tan. "Data provenance: What next?" In: *ACM SIGMOD Record* 47.3 (2019), pp. 5–16.

[167] Dan Kerchner et al. *The provenance of a tweet*. 2019.

[168] Mohammad Hossein Namaki et al. "Answering why-questions by exemplars in attributed graphs". In: *Proceedings of International Conference on Management of Data*. 2019, pp. 1481–1498.

[169] Vicky Papavasileiou, Ken Yocum, and Alin Deutsch. "Ariadne: Online Provenance for Big Graph Analytics". In: *Proceedings of International Conference on Management of Data*. 2019, pp. 521–536.

[170] Pierre Senellart. "Provenance in Databases: Principles and Applications". In: *Reasoning Web. Explainable Artificial Intelligence*. Springer, 2019, pp. 104–109.

[171] Disha Soni et al. "Leveraging Twitter and Neo4j to Study the Public Use of Opioids in the USA". In: *Proceedings of 2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA)*. 2019, pp. 1–5.

[172] Jingchao Yang et al. "A twitter data credibility framework—Hurricane harvey as a use case". In: *ISPRS International Journal of Geo-Information* 8.3 (2019), p. 111.