Data Centric Research at The University of Queensland

Xiaofang Zhou Shazia Sadiq
School of Information Technology and Electrical Engineering
The University of Queensland, Australia
{zxf,shazia}@uq.edu.au

1. INTRODUCTION

The University of Queensland (UQ) is in the top 100 universities worldwide measured through a number of major rankings. In 2012, the University had more than 46,000 students including 11,000 international students from 134 nations. It has one of Australia's largest PhD enrolments, with more than 12,600 postgraduate students. The University's outstanding 200,000-plus alumni include a Nobel laureate, the CEO of a Fortune 500 company, an Academy Award winner, and leaders in government, law, science, public service and the arts.

UQ hosts a cross-disciplinary group of researchers working on data centric approaches, spanning the engineering, business and science faculties. The group's strength lies in its diversity and ability to cater for the entire data pipeline. Members of the group have, and continue to contribute to all aspects of data-centric research including data modelling, indexing, querying, cleansing, analyzing, visualizing, as well as use and adoption over a number of application areas. Currently the group has 26 full time staff and over 50 PhD students. The group was awarded the highest ranking under the Australian Research Council (ARC) Excellence of Research in Australia scheme in 2012 (www.arc.gov.au/era/).

From 2005-2010, we managed the ARC Research Network in Enterprise Information Infrastructure (eii.edu.au). The aims of EII were to provide focus for research exchange via networking and collaboration and to improve the quality, impact and visibility of Australian ICT research. This was a unique initiative of the ARC, funded by the ARC in conjunction with financial support provided by 21 Australian and international institutions. Over the course of 5 years, and led by Xiaofang Zhou as convenor of the network, EII made a profound impact on the Australian research community, University sector and IT industry in terms of research collaboration, research training (especially through the annual PhD school), and research quality.

Over the years, we have also organized, or contributed to the organization of a number of major conferences in Brisbane including VLDB 1990, WWW 1998, WISE 2004, BPM 2007, DASFAA 2009, ACIS 2010 and ICDE 2013. ACM Multimedia will also be hosted by us in 2015.

In this article we first present a short chronology outlining the major topics and achievements of the group in the last three decades. We then present details on a selected number of current topics that we would like to share our viewpoint on, in terms of our focus, contributions and way forward.

1.1 Research Chronology

Late 1980s - Early 1990s:

Conceptual Modelling: One of the earliest strengths of the group was in the area of conceptual modelling[29]. In the late eighties and early nineties, Sjir Nijssen, considered as a founder of verbalization in computer science, developed a fact-based modelling methodology for information analysis, known as NIAM, subsequently extended as ORM (Object Role Modelling) together with Terry Halpin. This work was instrumental in promoting the benefits of conceptual modelling to many businesses, and over two decades later, the ORM approach continues to be used widely in teaching and research (www.ormfoundation.org).

Early - Mid 1990s:

Theory of Representation: At the same time as the influential works on conceptual modelling were being developed, UQ researchers were developing their theory of representation[42], also known as BWW model. The models that made up this theory provided a basis for work over the ensuing 25 years to date that has led to increased understanding of conceptual modelling, and the strengths and weaknesses of languages and tools that have been commonly used in conceptual modelling.

Mid - Late1990s:

Distributed Systems: During the nineties, the group was heavily involved with the Distributed Systems Technology Centre (DSTC), a centre of excellence that gained an international reputation for one of the most influential IT research organizations in Australia. During this period, the group's focus was on advanced distributed systems including data integration, semantics and advanced database applications.

Late 1990s - Mid 2000s:

Workflow Technology: Led by Maria Orlowska and the DSTC team, a number of industry strength applications in the area of workflow systems were developed, including FlowMake, a business process modelling and verification tool [36]; Chameleon, a high performance workflow engine for non-traditional workflow applications; and FlexeL, an eLearning platform based on workflow technology. Members of the group were one of the first to introduce a declarative constraint based approach to business process modelling [34], which influenced a shift in research on flexible business processes.

Bioinformatics: From 2004 - 2010, we were part of ARC Centre of Excellence in Bioinformatics, and worked on data management issues in bioinformatics, including 3D protein structure search and fuzzy biomedical image search [18, 50].

Current:

Spatial and Multimedia Databases: Led by Xiaofang Zhou and Heng Tao Shen, the group has made wide ranging contributions to the topics of spatial and multimedia databases. These are further detailed in sections 2 and 3 below.

Data Quality Management: In section 4, we outline our work on data quality management, led by Shazia Sadiq and Marta Indulska, highlighting specifically the initiatives for improved cross-fertilization between research, industry and user groups in the area of data quality management.

Information Use and Adoption: Led by researchers from the UQ business school, Peter Green and Andrew Burton-Jones, a range of topics relating to theory of representation, effective use of information and IT governance and compliance have been studied. The contributions of this research which is predominantly empirical in nature, are further detailed in section 5.

Application Driven Research: Although all aspects of our theoretical, experimental and empirical research are positioned in the context of meaningful use cases with strong links to application scenarios and often developed together with industry partners and domain experts, the eResearch group at UQ, led by Jane Hunter is an Australian hub of activities on data centric applications in science and humanities. In section 6, we will present a summary listing of some of the current cross-disciplinary research projects.

Next we will briefly present some of our recent work on these current topics.

2. TRAJECTORY COMPUTING

Location and time are two ubiquitous attributes of useful data. With the increasing ease of capturing spatiotemporal data, it becomes an important aspect of modern data management to efficiently and effectively manage spatial and spatiotemporal data, and to discover rich information about spatial and spatiotemporal patterns and trends.

Our research in this area is rooted in the development and deployment of an early spatial database management system SIRODBMS [3], especially for a set of efficient parallel spatial join algorithms [53] and spatial data warehousing algorithms [54]. When spatial DBMS evolved from a highly specialized tool used by well-trained professionals to a necessity in our daily life used by everyday people from the Internet and via mobile devices, our focus moved to provide light weight access to cater for a wide range of users and devices with variable resolutions and to minimize IO costs and communication bandwidth demand. To that end we proposed multi-resolution map data access methods and spatial query processing algorithms in both the Euclidean space and some constrained spaces such as road-networks and terrain surfaces [45, 9, 43].

With the prevalence of GPS devices and smartphones, an increasing amount of spatiotemporal data has become available. New applications such as environmental resource management, intelligent transport systems, urban planning, fleet management, location-based services and marketing, and location-based social networks can benefit from analyzing large amounts of trajectory data. We developed data mining algorithms to identify object movement patterns for future location predication [22] and object "convoy" identification [23], various trajectory search algorithms [7], location-based search combined with text keywords and multimedia tags [51, 27], and handling of low-sampling trajectory data that comes with location uncertainty [52].

In a recent work [41], we discovered that raw trajectory data cannot always be used for similarity comparison directly for a large range of proposed trajectory similarity measures, especially for low sampling data. Based on that we introduced trajectory calibration as a fundamental operation for similarity-based trajectory analytics. This work further opens new opportunities in trajectory computing, such as incorporating the temporal information in calibration, novel indexing and query processing methods with calibrated trajectories.

One new direction for our group on trajectory computing is to develop new spatiotemporal database systems for real-time trajectory analytics, based on in-memory technologies and multi-core parallelism. That approach is expected to bring an order of magnitude processing performance improvement.

3. MULTIMEDIA SEARCH

Driven by the rapid advances in hardware and multimedia technologies, we have witnessed unprecedented interest in large-scale multimedia retrieval in many computer science areas, including database management, multimedia analysis, information retrieval, computer vision, machine learning, image processing and distributed computing. It emerges as a key technology for a broad range of critical application domains, including security, health management, environment monitoring and astronomy.

Lack of indexing support for Web-scale multimedia databases has been recognized as one of the bottlenecks for fast multimedia search. In the last few years, we have made significant contributions to effective and efficient multimedia search. To enable real-time content-based search over large-scale multimedia databases, we have proposed various novel indexing approaches, including optimal one-dimensional indexing method [38], dimensionality reduction [39], data co-reduction [16] and hashing [40]. In particular, we have developed the first real-time near-duplicate video retrieval system over millions of video clips by introducing compact video signatures and effective indexing structures [17]. The system has been demonstrated in several major conferences such as VLDB, ICDE, ACM Multimedia, and attracted interest from industry with three filed patents. Our recent survey paper summarizes the state of the art on near-duplicate video retrieval and sheds light on its future research trends [28].

Complementary to content-based search, semantic-based search provides a convenient way for users to find relevant multimedia data with textual keywords. However, one major limitation of the traditional multimedia tagging framework is that tags are only assigned or propagated at the global level of objects. We have further advanced existing methods by allocating tags at the local level, i.e., regions in images [47] and shots in videos [55], to provide finer semantic indexing for multimedia objects. Web-based image tagging tools have also been developed to harness the richness of Web data [46].

Using new cost-effective depth cameras, we have investigated effective feature extraction methods for improved human gesture recognition from motion data streams [49]. Human body gesture recognition has many valuable applications in video surveillance, patient monitoring, smart homes and entertainment.

Clearly, the emergence of big multimedia data has presented many new opportunities across different research disciplines. One future trend for multimedia search is to fuse heterogeneous data sources with multi-modal representations to greatly improve the quality of the results. Additionally, the wide availability of growing social multimedia data has inspired us to leverage crowd wisdom to reshape the landscape of multimedia search, by discovering and utilizing the synergy between multimedia data and social data. With the application shift from desktops to mobiles, another future trend is to propose mobile solutions for novel applications, where mobile context, user profiles, device configurations and user experiences all play a part in the search process. We envisage that our future work related to multimedia search will be mainly focussed on main memory indexing of hash codes generated from scalable learning methods on heterogeneous data, utilizing social knowledge, and developing mobile platforms for multimedia retrieval.

4. DATA QUALITY

Data quality can be broadly interpreted as the state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use. The issue of data quality is as old as data itself. However due to the proliferation of large scale data in every walk of life, it is now exposed at a much wider and critical level, increasing manifold the stakes involved for corporations, government agencies and individuals. Due to the changing nature of data management, traditional approaches and solutions to data quality control are challenged, and there is an evident need to incorporate data quality considerations into the whole data cycle.

The topic of data quality has been pervasive in several aspects of our work. These include data consistency, linkage and matching [8, 25], data un-

certainty [52], data completeness [26], data diversity [24], and data freshness [37, 44]. Although historically issues of data integrity control have been addressed widely from the database research community, we identified an evident gap between research and practice [35], which has resulted in relative isolation between the organizational and computational aspects of data quality management and poor uptake of significant technical contributions from research. We initiated a number of initiatives for improved cross-fertilization between research and industry, which includes a Handbook of Data Quality [33], formation of a community of practice in data quality in Asia Pacific (apac.iaidq.org), as well as more locally establishing a data quality roundtable that currently has over 20 members from 7 different organizations.

We observe that current data quality management tools and practices cannot cope with the phenomenon of big data, where an overwhelmingly large amount of diverse and dynamic data needs to be processed and used for real-time decision-making. There are two intrinsic issues that make flexible, automated, and efficient support of data quality management highly challenging: Firstly, the quality of data is fundamentally perceived through the fitness for use lens, wherein the application (user and purpose) perspective makes generic data quality control very difficult. Secondly, the prevalence of large, multi-source data sets makes it very difficult for data owners and users to have a comprehensive understanding of the data characteristics and properties including data quality levels and targets. In light of these challenges, our current research aims at improving quality awareness [48] in data management systems, especially developing extensions in data management systems that link together data quality specification, measurement, improvement and tracking capabilities.

5. EMPIRICAL RESEARCH

Current members of the group, led by Peter Green, at the UQ Business School have continued to contribute to the body of knowledge on conceptual modelling as well as making significant contributions to various empirical aspects of data management. The theory of representation has been used and extended into many areas including business process modelling [12, 31], software maintenance [15], activity-based costing in management costing systems [32], and process interoperability languages [13]. Most recently, Andrew Burton-Jones has returned to his alma mater from Georgia State University and then UBC, Vancouver. He has continued the strong research into conceptual modelling by theorizing how representation is at the heart of effective use of information systems by users in organizations [6].

As organizations generate, collect and store continually increasing amounts of data, the drastic and ongoing increase has spawned the need not only for new research on data processing and retrieval techniques and data quality management approaches, but also for auditing of information systems use practices, IT governance approaches and structures to ensure accountability for the use of systems and data, and compliance management to ensure processes that use and generate the data are in line with legal and regulatory requirements.

Our research in the broad domain of compliance management has identified key challenges for the Information Systems community [1], and addressed the specific problem of lack of shared understanding in the compliance management environment within an organization by developing a validated ontology of compliance management, the first of its kind, to facilitate the required common conceptual model of the phenomenon. The Compliance Management Ontology (CoMOn) [2] has been validated with several organizations and has received significant attention from industry with respect to use in training and professional development of compliance professionals.

In the area of IT governance which pertains to the establishment of structures and practices that indicate accountability for the use of data and systems in organizations and encourage desirable behaviour in the use of IT, our research includes a study of factors that facilitate top management to increase IT governance absorptive capacity [5] and identification of IT governance mechanisms in practice that lead to effective IT governance in organizations [4]. While governance helps to motivate appropriate IT usage, it is a periodic systems audit that proves to regulatory bodies that inappropriate usage is not taking place. Thus, in our research we have also studied information systems audit issues [11] as well as the changing role of the IS audit.

6. E-RESEARCH

We are involved in a large number of data centric applications in science and humanities, and are investigating innovative approaches to the management, analysis and visualization services of large scale data collections, to accelerate scientific discovery. Members of the group are working on cross-disciplinary research projects across a broad range of topics including environmental informatics, dig-

ital humanities and biomedical sciences. The common aim across all of these projects is to expedite research outcomes through the sharing, integration and analysis of open access data, using Semantic Web approaches. Brief summaries of some of the current projects are presented below:

Online Environmental Report Cards Project provides GIS interfaces for online interactive access to integrated environmental monitoring data and environmental models, for both South East Queensland and the Great Barrier Reef, enabling decision support and adaptive resource management [19].

Twentieth Century in Paint Project is for painting conservators and links information about artists techniques, paint chemistry, characterization data, structured knowledge extracted from publications and experiments on paint degradation, to answer queries associated with paint conservation [21].

Semantic Annotation Services Project includes exploring new tools to support the annotation of 3D digital representations of cultural heritage artefacts, using the recently developed Open Annotation (OA) data model [20].

OzTrack Project develops a national repository for the storage, analysis, sharing and exchange of GPS-based animal tracking data. It combines a GIS interface (built on OGC standards) with R statistical analysis tools to calculate home ranges and understand interactions within and between species in the wild, and to assist with conservation management planning for endangered species [10].

Skeletome Project builds a community-driven knowledge base for the Skeletal Dysplasia domain. The research focus is the development of improved methods for extracting phenotype-disorder associations from medical literature, to improve the diagnosis and treatment of skeletal dysplasias [30, 14].

7. OUTLOOK

Data centric research is increasingly become a key element in addressing some of the most wicked problems of our times. We have been particularly fortunate to have the benefit of a team with diverse interests, expertise and networks. Our groups diversity has been instrumental in assisting us with the ability to tackle the whole data pipeline. As we move forward, we see that such multi-disciplinary teams have a growing importance, due to the need for database research and technical contributions to be positioned within a clear and evident business or scientific need on one hand, and a deployment blueprint on the other, in order to realize impact of our contributions on technological, economic and societal level.

We believe that such a creation of research impact requires genuine partnerships with user communities so as to bring true big (and dirty) datasets into the realm of database research community [1]. To this end, we advocate improved mechanisms for data sharing. This also requires closer relationship with hardware configurations in database research to fully capitalize on new in-memory and multi-core computing infrastructures, so as to improve experimental repeatability and generate a stronger value chain.

8. ADDITIONAL AUTHORS

Additional authors: Heng Tao Shen, Xue Li, Mohamed A. Sharaf, Zi Huang, Kai Zheng, Jane Hunter, Peter Green, and Marta Indulska.

9. REFERENCES

[1] N. S. Abdullah, S. W. Sadiq, and M. Indulska. Emerging challenges in information systems research for regulatory compliance management. In *CAiSE*, 2010.
[2] N. S. Abdullah, S. W. Sadiq, and M. Indulska. A compliance management ontology: Developing shared understanding through models. In *CAiSE*, 2012.
[3] D. J. Abel. SIRO-DBMS: a database tool-kit for geographical information systems. *International Journal of Geographical Information System*, 3(2):103–116, 1989.
[4] S. Ali and P. Green. Effective information technology (IT) governance mechanisms: An IT outsourcing perspective. *Information Systems Frontiers*, 14(2):179–193, 2012.
[5] S. Ali, P. Green, and A. Robb. Measuring top management's IT governance knowledge absorptive capacity. *Journal of Information Systems*, 27(1):137–155, 2013.
[6] A. Burton-Jones and C. Grange. From use to effective use: a representation theory perspective. *Information Systems Research*, pages 1–27, 2012.
[7] Z. Chen, H. T. Shen, X. Zhou, Y. Zheng, and X. Xie. Searching trajectories by locations: an efficiency study. In *SIGMOD Conference*, 2010.
[8] K. Deng, L. Wang, X. Zhou, S. W. Sadiq, and G. P. C. Fung. Active duplicate detection. In *DASFAA (1)*, 2010.
[9] K. Deng, X. Zhou, H. T. Shen, K. Xu, and X. Lin. Surface k-nn query processing. In *ICDE*, 2006.
[10] L. Gao, H. A. Campbell, O. R. Bidder, and J. Hunter. A web-based semantic tagging and activity recognition system for species' accelerometry data. *Ecological Informatics*, 13:47–56, 2013.
[11] P. Green, P. Best, M. Indulska, and T. Rowlands. Information systems audit and control issues for enterprise management systems: Qualitative evidence. *Australian Accounting Review*, 15(3 Supplement 37):67–77, 2005.
[12] P. F. Green and M. Rosemann. Integrated process modeling: An ontological evaluation. *Inf. Syst.*, 25(2):73–87, 2000.

[1] http://wp.sigmod.org/?p=786

[13] P. F. Green, M. Rosemann, M. Indulska, and C. Manning. Candidate interoperability standards: An ontological overlap analysis. *Data Knowl. Eng.*, 62(2):274–291, 2007.
[14] T. Groza, J. Hunter, and A. Zankl. Mining skeletal phenotype descriptions from scientific literature. *PloS one*, 8(2):e55656, 2013.
[15] J. Heales. Factors affecting information system volatility. In *ICIS*, 2000.
[16] Z. Huang, H. T. Shen, J. Liu, and X. Zhou. Effective data co-reduction for multimedia similarity search. In *SIGMOD Conference*, 2011.
[17] Z. Huang, H. T. Shen, J. Shao, X. Zhou, and B. Cui. Bounded coordinate system indexing for real-time video clip search. *ACM Trans. Inf. Syst.*, 27(3), 2009.
[18] Z. Huang, X. Zhou, D. Song, and P. Bruza. Dimensionality reduction in patch-signature based protein structure matching. In *ADC*, 2006.
[19] J. Hunter, P. Becker, A. Alabri, C. van Ingen, and E. G. Abal. Using ontologies to relate resource management actions to environmental monitoring data in south east queensland. *IJAEIS*, 2(1):1–19, 2011.
[20] J. Hunter and A. Gerber. Harvesting community annotations on 3d models of museum artefacts to enhance knowledge, discovery and re-use. *Journal of Cultural Heritage*, 11(1):81–90, 2010.
[21] J. Hunter and S. Odat. Building a semantic knowledge-base for painting conservators. In *eScience*, 2011.
[22] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou. A hybrid prediction model for moving objects. In *ICDE*, 2008.
[23] H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen. Discovery of convoys in trajectory databases. *PVLDB*, 1(1):1068–1080, 2008.
[24] H. A. Khan, M. Drosou, and M. A. Sharaf. Dos: An efficient scheme for the diversification of multiple search results. In *SSDBM*, 2013.
[25] H. Köhler, X. Zhou, S. W. Sadiq, Y. Shu, and K. L. Taylor. Sampling dirty data for matching attributes. In *SIGMOD Conference*, 2010.
[26] Z. Li, M. A. Sharaf, L. Sitbon, S. W. Sadiq, M. Indulska, and X. Zhou. Webput: Efficient web-based data imputation. In *WISE*, 2012.
[27] J. Liu, Z. Huang, L. Chen, H. T. Shen, and Z. Yan. Discovering areas of interest with geo-tagged images and check-ins. In *ACM Multimedia*, 2012.
[28] J. Liu, Z. Huang, H. C. H. T. Shen, C.-W. Ngo, and W. Wang. Near-duplicate video retrieval: Current research and future trends. *ACM Computing Surveys*, 45(4), 2013.
[29] G. M. Nijssen and T. A. Halpin. *Conceptual schema and relational database design - a fact oriented approach*. Prentice Hall, 1989.
[30] R. Paul, T. Groza, J. Hunter, and A. Zankl. Decision support methods for finding phenotypedisorder associations in the bone dysplasia domain. *PloS one*, 7(11):e50614, 2012.
[31] J. Recker, M. Rosemann, P. F. Green, and M. Indulska. Do ontological deficiencies in modeling grammars matter? *MIS Quarterly*, 35(1):57–79, 2011.
[32] M. Rosemann and P. F. Green. Integrating multi-perspective views into ontological analysis. In *ICIS*, 2000.
[33] S. Sadiq. *Handbook of Data Quality: Research and Practice*. Springer-Verlag New York Incorporated, 2013.
[34] S. W. Sadiq, M. E. Orlowska, and W. Sadiq. Specification and validation of process constraints for flexible workflows. *Inf. Syst.*, 30(5):349–378, 2005.
[35] S. W. Sadiq, N. K. Yeganeh, and M. Indulska. 20 years of data quality research: Themes, trends and synergies. In *ADC*, 2011.
[36] W. Sadiq and M. E. Orlowska. Analyzing process models using graph reduction techniques. *Inf. Syst.*, 25(2):117–134, 2000.
[37] M. A. Sharaf, C. Xu, and X. Zhou. Finding the silver lining for data freshness on the cloud. In *CloudDb*, 2012.
[38] H. T. Shen, B. C. Ooi, and X. Zhou. Towards effective indexing for very large video sequence database. In *SIGMOD Conference*, 2005.
[39] H. T. Shen, X. Zhou, and A. Zhou. An adaptive and dynamic dimensionality reduction method for high-dimensional indexing. *VLDB J.*, 16(2):219–234, 2007.
[40] J. Song, Y. Yang, Y. Yang, and Z. H. H. T. Shen. Inter-media hashing for large-scale retrieval from heterogenous data sources. In *SIGMOD Conference*, 2013.
[41] H. Su, K. Zheng, H. Wang, J. Huang, and X. Zhou. Calibrating trajectory data for similarity-based analysis. In *SIGMOD Conference*, 2013.
[42] Y. Wand and R. Weber. On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal*, 3(4):217–237, 1993.
[43] K. Xie, K. Deng, S. Shang, X. Zhou, and K. Zheng. Finding alternative shortest paths in spatial networks. *ACM Trans. Database Syst.*, 37(4):29, 2012.
[44] C. Xu, M. A. Sharaf, M. Zhou, A. Zhou, and X. Zhou. Adaptive query scheduling in key-value data stores. In *DASFAA (1)*, 2013.
[45] K. Xu, X. Zhou, and X. Lin. Direct mesh: a multiresolution approach to terrain visualization. In *ICDE*, 2004.
[46] Y. Yang, F. Wu, F. Nie, H. T. Shen, Y. Zhuang, and A. G. Hauptmann. Web and personal image annotation by mining label correlation with relaxed visual graph embedding. *IEEE Transactions on Image Processing*, 21(3):1339–1351, 2012.
[47] Y. Yang, Y. Yang, Z. Huang, H. T. Shen, and F. Nie. Tag localization with spatial correlations and joint group sparsity. In *CVPR*, 2011.
[48] N. K. Yeganeh, S. W. Sadiq, K. Deng, and X. Zhou. Data quality aware queries in collaborative information systems. In *APWeb/WAIM*, 2009.
[49] X. Zhao, X. Li, C. Pang, X. Zhu, and Q. Z. Sheng. Online human gesture recognition from motion data streams. In *ACM Multimedia*, 2013.
[50] K. Zheng, P. C. Fung, and X. Zhou. K-nearest neighbor search for fuzzy objects. In *SIGMOD Conference*, 2010.
[51] K. Zheng, S. Shang, N. J. Yuan, and Y. Yang. Towards efficient search for activity trajectories. In *ICDE*, 2013.
[52] K. Zheng, Y. Zheng, X. Xie, and X. Zhou. Reducing uncertainty of low-sampling-rate trajectories. In *ICDE*, 2012.
[53] X. Zhou, D. J. Abel, and D. Truffet. Data partitioning for parallel spatial join processing. *GeoInformatica*, 2(2):175–204, 1998.
[54] X. Zhou, D. Truffet, and J. Han. Efficient polygon amalgamation methods for spatial olap and spatial data mining. In *SSD*, 1999.
[55] X. Zhu, Z. Huang, J. Cui, and H. T. Shen. Video-to-shot tag propagation by graph sparse group lasso. *IEEE Transactions on Multimedia*, 15(3):633–646, 2013.