Research Statement

Kyriakos MOURATIDIS
School of Computing and Information Systems
Singapore Management University
kyriakos@smu.edu.sg

November 2024

My main research area is Spatial and Multi-dimensional Databases. I have worked on topics pertaining directly to that area (e.g., location-based services and their bridging with mobile computing and data stream processing), as well as topics stemming from the application of geometric reasoning and spatial database techniques to non-spatial problems (e.g., multi-objective and preference-based queries).

1. Research Philosophy and Aspiration

The overarching principle that guides my research efforts is to build a reputation around the following axes:

- *Identity*: I would like people who work on topics related to mine to know who/where I am and what my fields of expertise are. Conversely, I would like to be one of the first people that come to the mind of a researcher who is working on a project on spatial databases or preference queries, or of a professor who is looking for reading material on these topics for a postgraduate course.
- *Quality*: Should that researcher or professor visit my publication page, I want him/her to be sure that he/she will find papers that provide well-worked out and non-trivial ideas on practical problems, that are detailed and properly presented (so that if someone were to implement them, it would be clear exactly how), and include a thorough and academically honest evaluation.
- *Signature*: Should the researcher or professor get a copy of one of my papers, he/she should expect to see a problem that includes data of large scale, which need to be processed fast in practice (as opposed to problems/solutions of a purely theoretical nature). Typically, the proposed techniques will rely on geometric observations (even if the problem may originally not be spatial).

2. Research Overview

In this section, I outline the main aspects of my work and the evolution of my interests over time. Publication references starting with 'J' correspond to journals, with 'C' to conferences, and with 'I' to invited papers/tutorials.

Most of my contributions fall under the following topics:

- ***Continuous Queries.*** Traditional database systems are designed to answer *snapshot* queries over *persistent* data, i.e., one-off queries over relatively static data. However, the evolution of wireless communications, positioning devices (e.g., GPS) and sensor technologies has given rise to a new data processing model. In this model, multiple long-running queries require continuous update of their results as the data dynamically change. These queries are called *continuous* (or standing) and find application, among others, in location-based services (e.g., "keep me updated about who are the 10 SMU students that are closest to my location at any point as I walk along Orchard Road"), network traffic monitoring (e.g., "monitor the 100 users that cause the highest network overhead”), online decision support systems (e.g., “continuously report the 5 most interesting stocks according to my investment criteria, as new market data become available”), etc. My work in this area includes highly cited papers (e.g., [C3, J2, C6, C7]) that helped me make a mark in my research community. Another noteworthy achievement is a successful collaboration with PS Solutions (a subsidiary of the SoftBank Group) that led to the licensing of the developed spatiotemporal engine for worldwide commercial use, discussed further in Section 3.1.

- ***Spatial Optimization*.** This topic includes problems that (i) arise in resource allocation applications, (ii) have a spatial nature (i.e., their optimization criteria involve the notion of distance), and (iii) are of large scale (thus, typically requiring disk-based storage). Although such problems have been dealt with in the operations research literature, the solutions proposed there are tailored to small, memory-resident datasets. My work enables efficient processing in cases where the scale is several times (or orders) larger than considered before. My output includes papers that have received significant attention [C1, J1], studies that address some of the most central operations research problems in unprecedented scale [C9, J9, J10], as well as one of the best papers of SSTD’13 [C24], which was included as such in the “GeoInformatica: Special Issue on the Best of SSTD 2013” [J17].

- ***Location Privacy and Database Authentication*.** The trend of embedding positioning systems (e.g., GPS) in mobile devices facilitates the widespread use of location-based services. For such applications, the privacy and confidentiality issues are of paramount importance. Existing techniques, like encryption, safeguard the communication channel from eavesdroppers. Nevertheless, the queries themselves may disclose the position, identity, and habits of the user. To protect the user’s privacy, I have studied techniques for location obfuscation. The most prominent example is [J5], which defines a necessary property for any approach that follows the spatial anonymity paradigm, and has thus been cited by almost every follow-up work in that model. Another impactful paper is [J11], which considers for the first time obfuscation methods for users that move in road networks. Location privacy aside, the last few years have witnessed a growing popularity of database outsourcing. In this model, a data owner publishes his/her database through a third-party server. The server is responsible for hosting the data and answering user queries on behalf of the owner. Since the server may not be trusted, or may be compromised, users need a means to verify that answers received are both authentic and complete, i.e., that the returned data have not been tampered with, and that no qualifying results were omitted. Due to the growing popularity of database-as-a-service applications, I was enticed by this problem early on, and contributed ideas that have substantial research influence, such as [C12, J7].

- ***Multi-objective and Preference-based Queries*.** In today’s connected world, consumers increasingly rely on online resources to decide their purchases. For everyday needs, such as finding a restaurant, or for more significant buys, such as purchasing a laptop, a user is likely to visit an online review portal to assist in his/her choice. Typically, these portals rate the options on several aspects. For instance, local website HungryGoWhere rates Singapore restaurants on food, value, service, and ambience. The relative significance of these aspects depends on the preferences of the individual user. At the same time, due to the number of available options, generally it is impractical for the user to examine them one by one; e.g., HungryGoWhere lists over 25,000 restaurants. To assist in such multi-objective decisions, database research has considered ways to shortlist the most promising options for the user. My most recent research efforts aim to complement the usability of the traditional approaches with auxiliary features and improved operators. Due to the computational geometric principles that underlie the arising research problems, they are amenable to spatial and multi-dimensional database techniques. In the last 8 years, I have produced a stream of publications in this area, appearing in the top database venues (namely, SIGMOD and VLDB), and I have assumed the role of a field expert, giving tutorials at several conferences, including VLDB [I4] and MDM [I3, I6]. Importantly, I have recently secured a major external grant (S$ 596,687 ≈ US $440,000) to further this body of work. The figure below illustrates the evolution of my research interests, and the breakdown of my publication counts at different career stages; the vertical red lines correspond to (i) my transition from PhD student to Assistant Professor (in May 2006), and (ii) my promotion to tenured Associate Professor (in July 2013). The yellow bar is broken into two pieces; the striped part corresponds to the enrichment of that segment with database authentication.

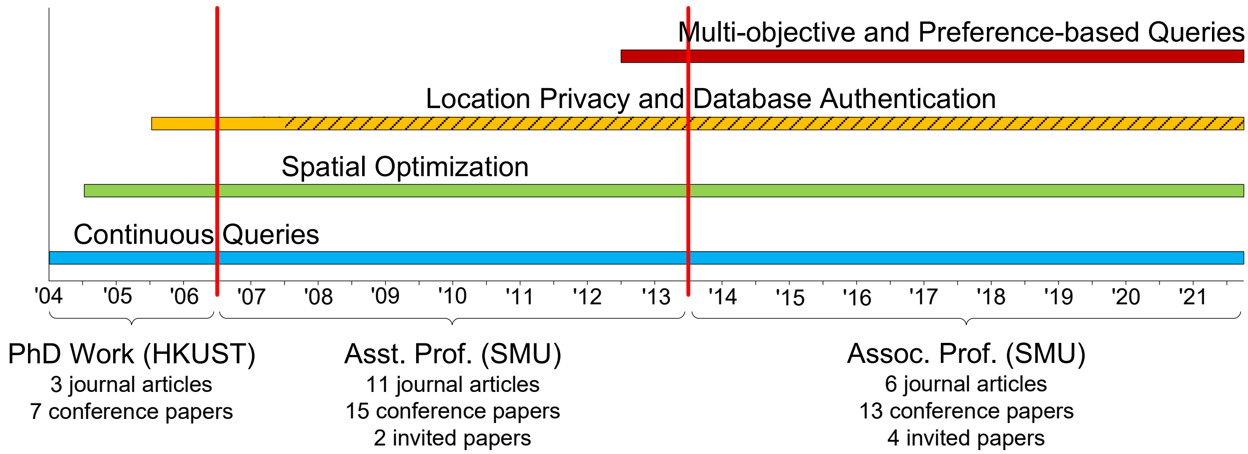

Figure 1: Evolution of research interests and publication breakdown

3. Survey of Research Work

In this section, I review representative papers from the above categories, aiming along the way to also exemplify the kind of research contributions I am looking to make, but also the way I identify and approach research problems. I conclude the section with future directions.

3.1 Continuous Queries

My work in this area has targeted mostly *spatial* continuous queries over *moving objects*. These queries consider spatial relationships between or among objects. In [C3, J2, C7, J4] we consider *k nearest neighbor* (*k*-NN) monitoring for various settings and with different performance goals. A standing *k*-NN query, defined over a set of moving objects, continuously reports the *k* objects that lie closest to a (potentially moving) query point. The query point typically corresponds to the location of a user. [C3] describes a method for Euclidean spaces that aims to minimize the *computation overhead* at the processing server for evaluating multiple such queries, while [J2] aims to reduce the *communication cost* between the processing server and the data objects.

In many real-world scenarios the data objects and the users move in a *road network*. Each road segment is associated with a cost value, typically reflecting the time required to drive from one end of the segment to the other. The shortest route (sequence of road segments) from a user to an object is called the *shortest path*, and the total cost along that path determines the *network distance* between them. In [C7] we propose efficient methods for the continuous monitoring of *k*-NNs according to network distance. In addition to object and user movements, we consider changes in the network itself, e.g., varying traffic conditions.

My work on continuous queries also includes methods for non-spatial problems. In [C6] we continuously evaluate multiple *top-k queries* over a stream of data tuples. Each query specifies a preference function *f* and requests for the *k* tuples in a sliding window that have the highest scores according to *f*. Our objective is to monitor the queries with low computation cost at the processing server. Similarly, in [J13, J19] we address the processing of *continuous text queries* over streams of documents, such as newsfeeds, email traffic, etc.

An achievement in this space is a successful industry project I co-led with colleague Prof. Baihua Zheng in 2015, where we developed a high-speed spatiotemporal processing engine for PS Solutions, a subsidiary of the global telecommunications and internet company SoftBank Group. The engine monitors a stream of 1 million location updates per second and continuously reports (in sub-second refresh time) the contents of 17 thousand geofences, i.e., continuous range queries. The success of the project led to the signing of a license agreement that enables the worldwide commercial use of the engine by PS Solutions (see joint press release: https://scis.smu.edu.sg/news/2016/06/14/industry-university-collaboration-iot-business-creation-smu-and-ps-solutions-signed). I was delighted to see the academic work I had done for over 10 years put to use for a real-world business problem and contribute to economic value creation.

3.2 Spatial Optimization

Regarding spatial optimization, in [C1, J1] we introduce *aggregate nearest neighbor* (ANN) queries. An ANN query retrieves the *k* data objects (from a disk-resident dataset) with the smallest *aggregate distance* from a set of query points. For example, a *sum*-ANN query identifies the facility (e.g., restaurant) where a set of users can meet by travelling the smallest total distance.

In [C9, J9] we investigate *optimal matching* (one of the oldest problems in operations research) in the context of large spatial datasets. Consider a set of *customers* (e.g., WiFi receivers) and a set of *service providers* (e.g., wireless access points), where each provider has a *capacity*, and the quality of service to its customers deteriorates with their distance from it. We compute a matching between the two sets such that (i) the greatest possible number of customers is served, and (ii) the sum of Euclidean distances across provider-customer pairs is minimized. Although max-flow algorithms exist for this problem, they are inapplicable to medium or large scale datasets, because their space and time requirements explode with the number of customers/providers. Motivated by this, we propose efficient algorithms for optimal assignment that employ novel pruning strategies, based on the geometric properties of the problem.

In [J10] we study a related problem with two important differences: (i) each service provider has a *coverage region* (i.e., it can only serve customers within a specific area), and (ii) the customers move frequently and arbitrarily. The task of the processing server in this scenario is to continuously report the optimal assignment (subject to the customers' most recent locations).

In [C13] we consider a *stable marriage* problem (yet another traditional resource allocation topic) where our task is to produce a fair assignment between a set of user preference queries and a set of available facilities. As an example, consider an internship assignment system, where at the end of each academic year, interested university students search and apply for available positions based on their preferences (e.g., nature of the job, salary, office location, etc). Although this is not a spatial problem per se, when the facility attributes are few (i.e., the dimensionality is low) geometric reasoning can be used to significantly reduce the CPU and I/O cost.

In [C19] we study a problem in the intersection of spatial databases and multi-criteria decision making. It is the first work in the area of road network databases that takes into account the coexistence of multiple distance notions in transportation decisions. For example, the different costs of a road segment could be its Euclidean length, the driving time, the walking time, possible toll fee, etc. The relative significance of these proximity notions may vary from user to user, yielding different location-based preferences over the facilities reachable via the network. In [C19] we formalize (and efficiently process) such preference queries over the facilities.

In [C24] we address a new optimization problem in road network databases. Consider a network where a subset of road segments are upgradable, i.e., the time to cross them can be reduced if a monetary cost is paid. Given a source and a destination in the network, and a monetary budget, the problem is to identify which of these edges to upgrade so that the network distance between source and destination is minimized, without exceeding the available budget. In addition to road networks, the problem arises in other domains too, such as telecommunications. [C24] was one of the best papers of SSTD'13, and was invited as such for inclusion to the "GeoInformatica: Special Issue on the Best of SSTD 2013", leading (significantly extended) to [J17].

3.3 Location Privacy and Database Authentication

On the front of location privacy, in [J5] we propose location obfuscation methods to preserve the anonymity of the users, and we design processing methods to evaluate spatial queries from obfuscated locations. A key property introduced in this work is *reciprocity*, which we show is necessary for the correctness of any spatial anonymity approach. Follow-up research in this area has widely adopted the reciprocity requirement, leading [J5] to almost 1000 citations according to Google Scholar. Our own study in [J11] belongs to this reciprocity-abiding follow-up research, proposing the first framework for anonymous query processing in road networks.

An alternative approach to location obfuscation relies on private information retrieval (PIR). PIR is a primitive that allows accessing data from a server, without the server knowing which data were accessed. Unlike location obfuscation/spatial anonymity, PIR offers cryptographic-level guarantees that no information is leaked. In [C21] (also presented as invited paper [I2]) we propose the first PIR-based method for shortest path queries.

Turning to database authentication, in [J7] we propose methods that enable the users of outsourced databases to verify the authenticity and completeness of the answers they receive from the hosting (untrusted) server. Our core idea is to decouple indexing and authentication information, and to store them into separate structures, which can then be analyzed and optimized independently, according to their individual purposes. In [C10] we consider text queries and conduct the first study on outsourced document databases. In [C18] we propose the first suite for the verification of shortest paths in outsourced graphs, while in [C12] we revisit an authentication model that had been dismissed based on obsolete hardware assumptions, and we show that it significantly outperforms alternatives for dynamic datasets on modern computers.

3.4 Multi-objective and Preference-based Queries

With the advent of e-commerce, users are presented with numerous alternatives to satisfy their needs. Choosing from the available options entails the consideration of multiple, often conflicting aspects, the tradeoff among which is assessed differently by different users. Database research has considered the top-*k* query as an effective means to shortlist the most promising options for a user. In the top-*k* model, the user's preferences are typically represented by a numeric weight per data attribute, and the score of an option is defined as the weighted sum of its individual attributes. Together, the weights form a *query vector* that is specific to the user; the domain of this vector is called the *preference space*. Given a set of options, the query vector imposes an implicit ranking by score across the set and, thus, enables the shortlisting of the k top-scoring options for the user. My most recent research efforts complement the traditional top-k model.

In [C22] we develop a methodology that helps users fine-tune their preference parameters (query weights) in a result-driven fashion. In particular, along with the top-k result, we report to the user the maximal deviation to each weight in his/her query vector for which the top-k options remain the same. The derived weight ranges, called *immutable regions*, are useful for sensitivity analysis and for weight tuning. The immutable regions in [C22] are local and one-dimensional, in the sense that they assume modification in a single query weight at a time. In [C25] we drop this assumption and compute the global immutable region, i.e., the multi-dimensional region in preference space where if the query vector falls, the top-k result remains the same. In addition to supporting simultaneous changes in any of the query weights, the global immutable region (and, in particular, its volume) provides a formal sensitivity measure for the top-k result.

In [C27] we define the *maximum rank* (MaxRank) query. Given a focal option in a set of alternatives, MaxRank computes the highest rank this option may achieve under any possible user preference, i.e., for any possible query vector. Furthermore, the query reports all the regions in preference space where that rank is achieved. MaxRank finds application in market impact analysis, customer profiling, targeted advertising, etc.

In [C30] we determine in which regions of the preference space the query vector should lie so that a given option in a dataset is among the top-k. In effect, these regions capture all possible user profiles for which our focal option is highly preferable, and are therefore essential in potential customer identification, profile-based marketing, etc. The problem (also known as monochromatic reverse top-k) was previously solved only for the degenerate case of d = 2 data attributes, and remained unsolved for more dimensions since 2010. Our work is the first to solve it for $d \geq 2$.

In [C33] we consider the effective introduction of a new option into the market, with a certain type of clientele in mind. Given a target region in the consumer spectrum (preference space), we determine what attribute values the new option should have, so that it ranks among the top-k for any user in the target region. Our methodology can also be used to improve an existing option, at the minimum modification cost, so that it consistently ranks high for an intended type of customers. This is the first work on competitive option placement where no distinct user(s) are targeted, but a general clientele type, i.e., a continuum of possible preferences. Here also lies our main challenge (and contribution), i.e., dealing with the interplay between two continuous spaces: the targeted region in the preference space, and the option domain (where the new option will be placed). At the core of our methodology lies a novel interlinking between the two spaces.

In [C34] we study the *m-impact region* (mIR) problem. In a context where users look for available products via top-k queries, mIR identifies the part of the option domain that attracts the most user attention. Specifically, mIR determines the kind of attribute values that lead a (new or existing) option to the top-k result for at least a fraction of a given user population. mIR has several applications, ranging from effective marketing to product improvement. Importantly, it also leads to exact and efficient solutions for standing top-k impact problems, which were previously solved heuristically only, or whose past solutions face serious scalability limitations (namely, influence-based cost optimization, and optimal improvement strategies).

In top-k processing, the query vector represents the user's preferences and, therefore, it is key in producing meaningful recommendations. Typically, the query vector is input directly by the user or, alternatively, mined from her behavior and past choices. In either case, it is impractical to expect absolute accuracy in weight values, i.e., the query vector should be taken as a mere indication of the user's preferences. Motivated by this, in [C32] we introduce (and solve) the *uncertain top-k* (UTK) problem. Given uncertain preferences (expressed as a region in preference space), UTK reports all options that may belong to the top-*k* result. A variant of the problem additionally outputs the exact top-*k* result for each of the possible weight settings.

Although [C32] makes an important step in offering some flexibility in preference specification, it effectively replaces the query vector with a preference region, which still needs to be input by the user or the application. Deciding that region is not a trivial feat, given that the dynamics in the preference space are hard to gauge. In usability terms, specifying the output size (i.e., the number of options to be shortlisted) would arguably be more tangible and more relatable to the user/application than specifying a region in the preference space. Motivated by that, in [C35] we receive a query vector in the input as an estimate of the user's preferences, and allow that vector to relax into a gradually expanding preference region. As the expansion proceeds (thus encompassing additional weight vectors) new options make it to the output. The expansion stops when the output includes the desired number of options. We propose two operators for this general methodology, ORD and ORU. The former (ORD) employs an adaptive notion of dominance, while the latter (ORU) sticks closer to ranking by utility.

In [J22] we extend [C35]. Our preliminary study in [C35] showed that the ORU problem is considerably more challenging than ORD and that, although our algorithm for ORD delivers sub-second response times even for very large problem instances, the method for ORU is not quite there. Motivated by this, in [J22] we tackle ORU further, by pursuing two orthogonal directions. Specifically, we present a new, fundamentally different approach for (exact) ORU processing, which is several times (and up to 3 orders of magnitude) faster than the one in [C35]. Furthermore, we propose an approximate ORU algorithm, which offers proven accuracy guarantees and allows control of the tradeoff between accuracy and efficiency.

In [J21] we focus on the dataset itself, and assess its competitiveness with regard to different possible preferences, i.e., different types of users. Assuming that the set of available options is represented as a multi-attribute dataset, we define measures of competitiveness, and represent them in the form of a heat-map in the preference domain. In particular, we define two categories of measures: (i) utility-based, which capture how satisfied the different user types are expected to be with the available options, and (ii) competition-based, which quantify how steep the competition among alternative products is with regard to different preferences. Applications of our heat-map include market analysis and business development. On the former front, the highest competitiveness cells indicate where the market's strength lies and what types of customers have attracted most of its efforts. On the front of business development, high competitiveness cells may indicate a saturation of the market for the respective parts of the user spectrum. When the distribution (or a representative sample) of the user preferences is known, the heat-map enables even stronger support in both aforementioned applications. We present a case study based on actual hotel data and user preferences mined from real TripAdvisor reviews, revealing actionable market insights.

Besides actively publishing in this area, I have also given a tutorial at VLDB'17 and an advanced seminar at MDM'16 on geometric approaches for rank-aware problems [I3, I4]. In 2019, I was invited to give another advanced seminar at MDM with updates since 2016 [I6]. In addition to being an honor and a recognition of my expertise, the tutorials also demonstrate the interest of the database community in this topic.

3.5 Others

In addition to the aforementioned topics, my work also includes well-received papers in the general spatial database area, such as [C2, C5, C11, C29, J20]. A representative example is [C17], which is motivated by the exploding number of applications that use road maps on mobile devices, and the limitations imposed by conventional on-demand querying (the most important pertaining to data traffic volume, which currently grows faster than mobile network capacity). In [C17] we propose the first method for shortest path computation in the wireless broadcasting model, i.e., where the server broadcasts the road network information and the users' devices tune into the channel to process their queries locally. Another example is [J16] where we consider a type of joint social and spatial search, with applications in company and friend recommendation.

Being interested in practical algorithms generally, I have also published beyond the area of databases. Such an instance is [C20] where we address a traditional bucketization problem in the general bin-packing family. We develop heuristic algorithms that produce significantly (and for certain types of data, provably) better bucket balancing than the previous state-of-the-art.

In [C37] we develop a methodology to inspect the robustness of outstanding facts (OFs) mined from knowledge graphs (KGs). OFs are truthful, striking claims which intentionally or inadvertently ignore lateral contexts and data that make them less striking. We propose measures that uncover possible cherry-picking and flag OFs for further examination before they are publicized. Such analysis is essential to avoid jumping to conclusions and disorienting the public. Our experiments, including user and case studies, demonstrate that our methodology accurately and efficiently detects frail OFs generated by state-of-the-art mining approaches.

In [C36] I participate in a very different but also very interesting project. Specifically, I join an NLP group at SMU to investigate the ability of vision-language models to reason *beyond* common sense. Given an unconventional image of a goldfish laying on the table next to an empty fishbowl, a human would effortlessly determine that the fish is not inside the fishbowl. The case, however, may be different for a vision-language model, whose reasoning may gravitate towards the common scenario that the fish is inside the bowl, despite the visual input. In [C36] we introduce a novel probing dataset to evaluate the state-of-the-art models in that regard, and reveal that they are still largely incapable of interpreting counter-intuitive scenarios.

3.6 Future Directions

The four main topics described so far lie close to my heart, and I intend to keep looking into them. However, in the last 8 years my work has largely concentrated on multi-objective queries. Framed around that topic, I have recently secured a competitive external grant for a 3-year project titled "PERFLEXO: a PERsonalized, FLExible, and controlled Output-size framework for multi-objective preference queries in large databases" (S$ 596,687 ≈ US $440,000). The grant provides for post-doc, research engineer and PhD student support and, importantly, it includes an allocation specifically for real user studies. These resources will enable me to further my work on the topic, but to also integrate my contributions into a consolidated and comprehensive framework. In the immediate future, I believe this project will guide most of my upcoming research.

References/Publications

Journals

J1. Dimitris Papadias, Yufei Tao, Kyriakos Mouratidis, Chun Kit Hui: Aggregate Nearest Neighbor Queries in Spatial Databases. *ACM Transactions on Database Systems (TODS)*, 30(2), 529-576, 2005.

J2. Kyriakos Mouratidis, Dimitris Papadias, Spiridon Bakiras, Yufei Tao: A Threshold-based Algorithm for Continuous Monitoring of *k* Nearest Neighbors. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 17(11), 1451-1464, 2005.

J3. Jun Zhang, Dimitris Papadias, Kyriakos Mouratidis, Manli Zhu: Query Processing in Spatial Databases Containing Obstacles. *International Journal of Geographic Information Science (IJGIS)*, 19(10), 1091-1111, 2005.

J4. Kyriakos Mouratidis, Dimitris Papadias: Continuous Nearest Neighbor Queries over Sliding Windows. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 19(6), 789-803, 2007.

J5. Panos Kalnis, Gabriel Ghinita, Kyriakos Mouratidis, Dimitris Papadias: Preventing Location-based Identity Inference in Anonymous Spatial Queries. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 19(12), 1719-1733, 2007.

J6. Kyriakos Mouratidis, Dimitris Papadias, Spiros Papadimitriou: Tree-based Partition Querying: A Methodology for Computing Medoids in Large Spatial Datasets. *Very Large Data Bases Journal (VLDBJ)*, 17(4), 923-945, 2008.

J7. Kyriakos Mouratidis, Dimitris Sacharidis, HweeHwa Pang: Partially Materialized Digest Scheme: An Efficient Verification Method for Outsourced Databases. *Very Large Data Bases Journal (VLDBJ)*, 18(1), 363-381, 2009.

J8. Kyriakos Mouratidis, Spiridon Bakiras, Dimitris Papadias: Continuous Monitoring of Spatial Queries in Wireless Broadcast Environments. *IEEE Transactions on Mobile Computing (TMC)*, 8(10), 1297-1311, 2009.

J9. Leong Hou U, Kyriakos Mouratidis, Man Lung Yiu, Nikos Mamoulis: Optimal Matching between Spatial Datasets under Capacity Constraints. *ACM Transactions on Database Systems (TODS)*, 35(2), 2010.

J10. Leong Hou U, Kyriakos Mouratidis, Nikos Mamoulis: Continuous Spatial Assignment of Moving Users. *Very Large Data Bases Journal (VLDBJ)*, 19(2), 141-160, 2010.

J11. Kyriakos Mouratidis, Man Lung Yiu: Anonymous Query Processing in Road Networks. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 22(1), 2-15, 2010.

J12. Dimitris Sacharidis, Kyriakos Mouratidis, Dimitris Papadias: *k*-Anonymity in the Presence of External Databases. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 22(3), 392-403, 2010.

J13. Kyriakos Mouratidis, HweeHwa Pang: Efficient Evaluation of Continuous Text Search Queries. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 23(10), 1469-1482, 2011.

J14. HweeHwa Pang, Jilian Zhang, Kyriakos Mouratidis: Enhancing Access Privacy of Range Retrievals over B^+-Trees. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 25(7), 1533-1547, 2013.

J15. Jilian Zhang, Kyriakos Mouratidis, HweeHwa Pang: Direct Neighbor Search. *Information Systems (IS)*, 44, 73-92, 2014.

J16. Kyriakos Mouratidis, Jing Li, Yu Tang, Nikos Mamoulis: Joint Search by Social and Spatial Proximity. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 27(3), 781-793, 2015.

J17. Yimin Lin, Kyriakos Mouratidis: Best Upgrade Plans for Single and Multiple Source-Destination Pairs. *GeoInformatica: Special Issue on the Best of SSTD 2013*, 19(2), 365-404, 2015.

J18. Pui Hang Li, Man Lung Yiu, Kyriakos Mouratidis: Discovering Historic Traffic-tolerant Paths in Road Networks. *GeoInformatica*, 21(1), 1-32, 2017.

J19. Leong Hou U, Junjie Zhang, Kyriakos Mouratidis, Ye Li: Continuous Top-k Monitoring on Document Streams. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 29(5), 991-1003, 2017.

J20. Bo Tang, Man Lung Yiu, Kyriakos Mouratidis, Jiahao Zhang, Kai Wang: On Discovering Motifs and Frequent Patterns in Spatial Trajectories with Discrete Fréchet Distance. *GeoInformatica*, 26(1): 29-66, 2022.

J21. Kyriakos Mouratidis, Keming Li, Bo Tang: Quantifying the Competitiveness of a Dataset in Relation to General Preferences. *Very Large Data Bases Journal (VLDBJ)*, 33(1): 231-250, 2024.

J22. Kyriakos Mouratidis, Keming Li, Bo Tang: Marrying Top-k with Skyline Queries: Operators with Relaxed Preference Input and Controllable Output Size. *ACM Transactions on Database Systems (TODS)*, to appear.

Conferences

C1. Dimitris Papadias, Qiongmao Shen, Yufei Tao, Kyriakos Mouratidis: Group Nearest Neighbor Queries. Proceedings of the *IEEE International Conference on Data Engineering (ICDE)*, pp. 301-312, 2004.

C2. Jun Zhang, Dimitris Papadias, Kyriakos Mouratidis, Manli Zhu: Spatial Queries in the Presence of Obstacles. Proceedings of the *International Conference on Extending Database Technology (EDBT)*, pp. 366-384, 2004.

C3. Kyriakos Mouratidis, Marios Hadjieleftheriou, Dimitris Papadias: Conceptual Partitioning: An Efficient Method for Continuous Nearest Neighbor Monitoring. Proceedings of the *ACM Conference on Management of Data (SIGMOD)*, pp. 634-645, 2005.

C4. Kyriakos Mouratidis, Dimitris Papadias, Spiros Papadimitriou: Medoid Queries in Large Spatial Databases. Proceedings of the *International Symposium on Spatial and Temporal Databases (SSTD)*, pp. 55-72, 2005.

C5. Manolis Terrovitis, Spiridon Bakiras, Dimitris Papadias, Kyriakos Mouratidis: Constrained Shortest Path Computation. Proceedings of the *International Symposium on Spatial and Temporal Databases (SSTD)*, pp. 181-199, 2005.

C6. Kyriakos Mouratidis, Spiridon Bakiras, Dimitris Papadias: Continuous Monitoring of Top-k Queries over Sliding Windows. Proceedings of the *ACM Conference on Management of Data (SIGMOD)*, pp. 635-646, 2006.

C7. Kyriakos Mouratidis, Man Lung Yiu, Dimitris Papadias, Nikos Mamoulis: Continuous Nearest Neighbor Monitoring in Road Networks. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 43-54, 2006.

C8. Stavros Papadopoulos, Dimitris Sacharidis, Kyriakos Mouratidis: Continuous Medoid Queries over Moving Objects. Proceedings of the *International Symposium on Spatial and Temporal Databases (SSTD)*, pp. 38-56, 2007.

C9. Leong Hou U, Man Lung Yiu, Kyriakos Mouratidis, Nikos Mamoulis: Capacity Constrained Assignment in Spatial Databases. Proceedings of the *ACM Conference on Management of Data (SIGMOD)*, pp. 15-28, 2008.

C10. HweeHwa Pang, Kyriakos Mouratidis: Authenticating the Query Results of Text Search Engines. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 126-137, 2008.

C11. Dimitris Sacharidis, Kostas Patroumpas, Manolis Terrovitis, Verena Kantere, Michalis Potamias, Kyriakos Mouratidis, Timos K. Sellis: On-Line Discovery of Hot Motion Paths. Proceedings of the *International Conference on Extending Database Technology (EDBT)*, pp. 392-403, 2008.

C12. HweeHwa Pang, Jilian Zhang, Kyriakos Mouratidis: Scalable Verification for Outsourced Dynamic Databases. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 802-813, 2009.

C13. Leong Hou U, Nikos Mamoulis, Kyriakos Mouratidis: A Fair Assignment Algorithm for Multiple Preference Queries. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 1054-1065, 2009.

C14. Kyriakos Mouratidis, HweeHwa Pang: An Incremental Threshold Method for Continuous Text Search Queries. Proceedings of the *IEEE International Conference on Data Engineering (ICDE)*, pp. 1187-1190 (short paper), 2009.

C15. Leong Hou U, Nikos Mamoulis, Kyriakos Mouratidis: Efficient Evaluation of Multiple Preference Queries. Proceedings of the *IEEE International Conference on Data Engineering (ICDE)*, pp. 1251-1254 (short paper), 2009.

C16. Kar Way Tan, Yimin Lin, Kyriakos Mouratidis: Spatial Cloaking Revisited: Distinguishing Information Leakage from Anonymity. Proceedings of the *International Symposium on Spatial and Temporal Databases (SSTD)*, pp. 117-134, 2009.

C17. Georgios Kellaris, Kyriakos Mouratidis: Shortest Path Computation on Air Indexes. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 747-757, 2010.

C18. Man Lung Yiu, Yimin Lin, Kyriakos Mouratidis: Efficient Verification of Shortest Path Search via Authenticated Hints. Proceedings of the *IEEE International Conference on Data Engineering (ICDE)*, pp. 237-248, 2010.

C19. Kyriakos Mouratidis, Yimin Lin, Man Lung Yiu: Preference Queries in Large Multi-Cost Transportation Networks. Proceedings of the *IEEE International Conference on Data Engineering (ICDE)*, pp. 533-544, 2010.

C20. Jilian Zhang, Kyriakos Mouratidis, HweeHwa Pang: Heuristic Algorithms for Balanced Multi-way Number Partitioning. Proceeding of the *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 693-698 (full paper, full presentation), 2011.

C21. Kyriakos Mouratidis, Man Lung Yiu: Shortest Path Computation with No Information Leakage. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 692-703, 2012.

C22. Kyriakos Mouratidis, HweeHwa Pang: Computing Immutable Regions for Subspace Top-k Queries. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 73-84, 2013.

C23. Yimin Lin, Kyriakos Mouratidis: Shortlisting Top-K Assignments. Proceeding of the *International Conference on Scientific and Statistical Database Management (SSDBM)*, article no. 21, 2013.

C24. Yimin Lin, Kyriakos Mouratidis: Best Upgrade Plans for Large Road Networks. Proceedings of the *International Symposium on Spatial and Temporal Databases (SSTD)*, pp. 223-240, 2013.

C25. Jilian Zhang, Kyriakos Mouratidis, HweeHwa Pang: Global Immutable Region Computation. Proceedings of the *ACM Conference on Management of Data (SIGMOD)*, pp. 1151-1162, 2014.

C26. Pui Hang Li, Man Lung Yiu, Kyriakos Mouratidis: Historical Traffic-Tolerant Paths in Road Networks. Proceedings of the *ACM Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL)*, pp. 477-480 (short paper), 2014.

C27. Kyriakos Mouratidis, Jilian Zhang, HweeHwa Pang: Maximum Rank Query. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 1554-1565, 2015.

C28. Kyriakos Mouratidis, Jing Li, Yu Tang, Nikos Mamoulis: Joint Search by Social and Spatial Proximity [Extended Abstract]. Proceedings of the *IEEE International Conference on Data Engineering (ICDE)*, pp. 1578-1579, 2016 (in poster session devoted to TKDE articles published in 2015).

C29. Bo Tang, Man Lung Yiu, Kyriakos Mouratidis, Kai Wang: Efficient Motif Discovery in Spatial Trajectories Using Discrete Fréchet Distance. Proceedings of the *International Conference on Extending Database Technology (EDBT)*, pp. 378-389, 2017.

C30. Bo Tang, Kyriakos Mouratidis, Man Lung Yiu: Determining the Impact Regions of Competing Options in Preference Space. Proceedings of the *ACM Conference on Management of Data (SIGMOD)*, pp. 805-820, 2017.

C31. Leong Hou U, Junjie Zhang, Kyriakos Mouratidis, Ye Li: Continuous Top-k Monitoring on Document Streams (Extended Abstract). Proceedings of the *IEEE International Conference on Data Engineering (ICDE)*, pp. 1803-1804, 2018 (in poster session devoted to TKDE articles published in 2017).

C32. Kyriakos Mouratidis, Bo Tang: Exact Processing of Uncertain Top-k Queries in Multi-criteria Settings. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 866-879, 2018.

C33. Bo Tang, Kyriakos Mouratidis, Man Lung Yiu, Zhenyu Chen: Creating Top Ranking Options in the Continuous Option and Preference Space. Proceedings of the *Very Large Data Bases Conference (VLDB)*, pp. 1181-1194, 2019.

C34. Bo Tang, Kyriakos Mouratidis, Mingji Han: On m-Impact Regions and Standing Top-k Influence Problems. Proceedings of the *ACM Conference on Management of Data (SIGMOD)*, pp. 1784-1796, 2021.

C35. Kyriakos Mouratidis, Keming Li, Bo Tang: Marrying Top-k with Skyline Queries: Relaxing the Preference Input while Producing Output of Controllable Size. Proceedings of the *ACM Conference on Management of Data (SIGMOD)*, pp. 1317-1330, 2021.

C36. Kankan Zhou, Eason Lai, Wei Bin Au Yeong, Kyriakos Mouratidis, Jing Jiang: ROME: Evaluating Pre-trained Vision-Language Models on Reasoning beyond Visual Common Sense. *Findings of the Association for Computational Linguistics: EMNLP*, pp. 10185-10197, 2023.

C37. Hanhua Xiao, Yuchen Li, Yanhao Wang, Panagiotis Karras, Kyriakos Mouratidis, Natalia-Rozalia Avlona: How to Avoid Jumping to Conclusions: Measuring the Robustness of Outstanding Facts in Knowledge Graphs. Proceedings of the *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 3539-3550, 2024.

Keynote Speeches, Tutorials & Invited Papers

I1. Kyriakos Mouratidis: Spatial Queries in Wireless Broadcast Environments. Keynote at *International ACM Workshop on Data Engineering for Wireless and Mobile Access (MobiDE)*, pp. 39-44, 2012.

I2. Kyriakos Mouratidis: Strong Location Privacy: A Case Study on Shortest Path Queries. Invited paper at *International Workshop on Mobile Data Analytics (MoDA)*, (ICDE Workshops) pp. 136-143, 2013.

I3. Kyriakos Mouratidis: Geometric Aspects and Auxiliary Features to Top-k Processing. Advanced seminar at *International Conference on Mobile Data Management (MDM)*, pp. 1-3, 2016.

I4. Kyriakos Mouratidis: Geometric Approaches for Top-k Queries. Tutorial at *Very Large Data Bases Conference (VLDB)*, pp. 1985-1987, 2017.

I5. Kyriakos Mouratidis: Applying Spatial Database Techniques to Other Domains: a Case Study on Top-k and Computational Geometric Operators. Invited seminar at *ACM Workshop on Managing and Mining Enriched Geo-Spatial Data (GeoRich)*, pp. 25-26, 2018.

I6. Kyriakos Mouratidis: Geometric Top-k Processing: Updates since MDM'16. Advanced seminar at *International Conference on Mobile Data Management (MDM)*, pp. 1-3, 2019.