Graph-based Spatial Pattern Matching: A Theoretical Comparison

Nicole R. Schneider
nsch@umd.edu
University of Maryland
USA

Kent O'Sullivan
osullik@umd.edu
University of Maryland
USA

Hanan Samet
hjs@cs.umd.edu
University of Maryland
USA

Abstract

Spatial Pattern Matching is an important search problem that involves reasoning about the relative position, distance, and orientation of objects with respect to one another. Spatial relationships between objects contain a lot of information about the world, which makes them useful in applications like Point of Interest (POI) retrieval and location-based services. However, spatial pattern matching is an NP-hard problem in the worst case. This paper presents a theoretical comparison of spatial pattern matching approaches, showing how the prominent methods compare for each type of spatial relation they support. We further highlight the common techniques used to gain performance improvements and provide suggestions towards developing approximate solutions to this form of spatial search.

CCS Concepts

• **Theory of computation** → **Graph algorithms analysis**; Sorting and searching; • **Information systems** → Spatial-temporal systems.

Keywords

Spatial pattern matching, complexity analysis, graph pattern matching

ACM Reference Format:
Nicole R. Schneider, Kent O'Sullivan, and Hanan Samet. 2024. Graph-based Spatial Pattern Matching: A Theoretical Comparison. In *The 32nd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '24), October 29-November 1, 2024, Atlanta, GA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3678717.3691227

1 Introduction

Spatial Pattern Matching is a form of spatial search that seeks all combinations of database objects that satisfy a set of spatial query constraints. The constraints can include directional, topological, and metric relations, which convey critical spatial information, but present computational challenges for search.

Spatial relations are typically encoded using a graph, where nodes represent objects and (possibly multiple) edges between nodes represent the spatial relations or constraints between those objects. The quantity of spatial relationships that exist between objects is quadratic in the number of objects, making the process of matching query and data patterns NP-hard [10]. There are no good approximate methods that reduce the complexity of spatial pattern matching without altering the problem fundamentally [17, 18].

State of the art spatial pattern matching approaches typically use initial filtering steps, including using keyword-based spatial indices, to prune the set of candidates before the more costly spatial pattern matching algorithms are performed [4, 15, 22]. Binning methods and other tricks, like localizing the search to a very limited scope, can also make the problem more tractable by substantially reducing the number of comparisons needed to identify matching patterns. However, directly comparing these approaches against each other in a rigorous way is difficult to achieve. Different spatial pattern matching approaches frame the problem differently, typically supporting different combinations of spatial entity and relation types, and empirically testing the performance of their techniques in an ad-hoc manner, which makes comparison difficult since graph algorithms are sensitive to the size and shape of the graph [16].

To address this issue, we present a comparison of the complexities of the prominent graph-based spatial pattern matching approaches, discussing for each the techniques or heuristics used to gain performance improvements. Finally, we provide suggestions for future work that could lead to meaningful speedups in spatial pattern matching, including by leveraging approximate solutions to the algorithm classes typically used by the current approaches.

2 Related Work

There are several survey papers that discuss the recent approaches to spatial pattern matching and other forms of spatial search. Many of these provide useful taxonomies to organize the types of spatial queries that are possible, covering the major relations like metric, topological, and directional [2, 3]. Other surveys focus only on one type of spatial relation, such as Dylla et al., which surveys topological relations captured by qualitative spatial calculi [7]. However, existing surveys do not provide any theoretical comparison of the methods surveyed to enable a thorough understanding of the state of the problem. To fill this gap, we provide a comparison of many of the popular spatial pattern matching approaches based on the type of spatial relation that can be supported by each, finding that most methods have exponential worst-case complexity. We further synthesize the common techniques these methods share that enable them to gain performance speedups in their empirical testing.

3 Spatial Pattern Matching

Spatial pattern matching is often defined as a graph matching problem, where vertices have keyword labels and edges have distance intervals and signs {*exclusion in, exclusion out, mutual exclusion* and *mutual inclusion*} [10]. Spatial pattern matching can be performed over spatial entities, including (i) *Point entities* that consist

ACM ISBN 979-8-4007-1107-7/24/10
https://doi.org/10.1145/3678717.3691227

of an (x, y) coordinate pair in Cartesian space, (ii) *Line entities* that represent the shortest path between two points, and (iii) *Region entities* that represent the area inside a polyline joining several points. Relationships between those entities can take the form of (i) *Metric relations* that describe the distances between spatial entities (ten miles, near, far, etc.), (ii) *Topological relations* that describe how regions, lines, and points interact (intersect, contain, touch, etc), and (iii) *Directional relations* that describe how entities are positioned relatively in space (north, left, behind, etc.), Most spatial pattern matching methods formulate the problem as a *qualitative spatial reasoning* (QSR) / *constraint satisfaction* problem (CSP) or a *subgraph-matching* (SGM) problem.

Constraint Satisfaction Problems. CSP for spatial pattern matching find an assignment of valid variables (database entities) given the query constraints. Most CSP are exponential in the number of entities being searched. In the worst case, where constraints are numerous and poorly ordered and specificity is low, CSP approaches perform poorly.

Subgraph Matching. SGM approaches assume the database is encoded as a graph of entities and relations, and queries are encoded as a subgraph of entities and constraints of interest. These methods seek to identify where the query pattern exists in the database graph. Spatial SGM methods typically scale poorly on large database and query graphs, or do not support complex spatial relation types, like topological and directional relations [4, 15].

Algorithm	Entities	Met	Top	Dir	Complexity
PQIS [11]	*P*	X		X	$O(n2^n)^\dagger$
M-P Join [8–10]	*P*	X			$O(m\zeta n^2 + n^Q)^\dagger$
M-S Join [8–10]	*P*	X			$O(n^4 + mn^2 + n^Q)^\dagger$
ESPM [4]	*P*	X			$O(n'^Q)^\dagger$
QQESPM [15]	*P, L, R*	X	X		$O(n'^Q)^\dagger$
QSR [6]	*P*			X	$O(2^n)$
MSJ [19]	*P, L, R*	X	X	X	$O(n^Q)$
MSJ_{WR} [19]	*P, L, R*	X	X	X	$O(n^m)$
MSJ_{JWR} [19]	*P, L, R*	X	X	X	$O(n^m)$
SketchMapia [22]	*P, L, R*		X	X	$O(n^2m^2)$

Table 1: Summary of spatial pattern matching algorithmic complexities. Entity types are represented by *P*, *L*, and *R* for Point, Line, and Regions, respectively. In our notation n is the number of spatial entities or objects in the database, m is the number of constraints or relations in the query, Q is the number of query entities, n' is a pruned subset of all possible entities n, and ζ is a sampling threshold in $[0, 1]$. The (†) denotes cases where the original paper provided the worst-case complexity.

4 Theoretical Comparison

In this section we present the major graph-based spatial pattern matching methods organized by the relation types they support. For each method, we describe the complexity of the search technique, which is summarized in Table 1.

4.1 Metric relations

Metric relations are spatial relations that describe the distances between objects (ten miles, near, far etc.). Most of the methods we outline in Table 1 support spatial pattern matching using metric relations, including the work of Folkers et al. [11] and Papadias et al. [19], and more recently Fang et al. [8–10], Chen et al. [4], and Minervino et al. [15].

Spacekey [8–10] provides the *Multi-Pair Join* and *Multi-Star Join* which use subgraph matching to enable search over metric relations that are specified as ranges of distances between graph-encoded point entities. Their Multi-Pair-Join algorithm has a worst-case complexity $O(m\zeta|D|^2+\xi)$ where ζ is a sampling threshold in range $[0, 1]$ and ξ is the maximal number of partial matches, which cannot exceed n^Q. Their Multi-Star-Join algorithm has a worst-case complexity of $O(n^4 + m|D|^2 + \xi)$ time, but in practice is faster than the Multi-Pair-Join because it uses additional pruning criteria to eliminate partial matches during the join process [10].

ESPM: Efficient Spatial Pattern Matching [4] also performs SGM-based spatial pattern matching over point entities with metric relations, but extends the work of Fang et al. [10] by adding a step that uses a set of Inverted Linear Quadtrees, one per entity keyword, to prune unpromising edges before running Multi-Star-Join. By eliminating unpromising nodes early and carefully constructing the join order, ESPM scales better than the original Multi-Star-Join in their empirical testing, despite having a similar theoretical worst-case complexity which is exponential in the number of query entities.

Other methods, including *Pictorial Query Trees* [23], Multi-relation Spatial Joins (MSJ) and their window reduction variants [19], and QQESPM [15] also handle metric relations, with MSJ specifically dealing with qualitative distance relations like 'near' and 'far.' However, the focus of these methods are topological and/or directional relations, which are more challenging to search efficiently, so we discuss their complexities in those sections.

4.2 Topological relations

Topological relations are spatial relations that describe how points, lines, and regions interact with one another, using topological properties like intersection, containment, touching, and covering. Most of the methods in Table 1 that support topological relations use a CSP-based formulation of spatial pattern matching, and can support matching over point, line, and region entities.

Multi-relation Spatial Joins (MSJ) and their window reduction variants [19] support the eight major topological relations, but require exponential time complexity in the number of relations or query objects. Their Window Reduction (WR) approach is a CSP forward checking algorithm that prunes the search space over time, and their Joint Window Reduction (JWR) is a slightly more efficient version of WR that does not need to exhaustively search for a starting point. More recent work [5] builds on Multi-relation spatial joins, improving their efficiency by parallelizing them, which improves the runtime by one order of magnitude in their experiments. However, even parallelized, these methods are slower than many recent approaches that rely on pruning.

SketchMapia [13, 22] supports only four topological relations for region and point data, since their system is based on street segments and disjoint city blocks that are formed as the area between multiple connected street segments. City blocks can be *touching* or *disconnected*, while landmarks (entities) can be *inside* or *outside* of city blocks. Because SketchMapia solves a simpler version of the spatial pattern matching problem, its complexity is $O(n^2m^2)$, which is better than other methods supporting topological constraints.

QQESPM: Quantitative and Qualitative Spatial Pattern Matching [15] extends the work of Chen et al. [4] to support topological relations in addition to metric relations, checking both the quantitative and qualitative constraints at each step. The theoretical complexity is the same as ESPM with an extra constant factor for the additional qualitative checks at each step.

4.3 Directional relations

Directional relations are spatial relations that describe how objects are positioned relative to one another in space (north, left, behind etc.). These relations are typically measured using cardinal direction or relative angle. Most recent work in directional spatial pattern matching uses CSP approaches, operating either on point entities or on all entity types.

Pictorial Queries Using Isomorphic Subgraphs [11] uses a bottom-up subgraph matching approach that supports directional pattern matching over point data. It simplifies the problem of directional pattern matching by treating directional constraints as angles between vertices, and checking if they fall within some δ of the query constraint, which requires accurate relative distances between query objects. The exponential complexity in the number of database objects makes this approach prohibitively slow, and the accurate distance assumption on the query objects is unrealistic since humans are notoriously poor at estimating distance [22].

SketchMapia [22] only considers directional constraints for very localized objects and bins them into eight segments (left, half left, etc.), which limits the size of their qualitative constraint network. By simplifying the problem formulation, SketchMapia achieves a complexity of $O(n^2m^2)$, which is better than recent methods that support directional constraints.

Most recently, Duckham et al. developed a method for *Qualitative Spatial Reasoning with Uncertain Evidence* [6], which addresses probabilistic directional spatial reasoning by formulating the problem as a Markov Logic Network. Despite the noise-tolerance of this approach, it remains exponential in complexity.

5 Discussion

Many spatial pattern matching methods leverage techniques and tricks like binning and pruning to simplify the problem or quickly reduce the search space to a smaller set of candidates that could match the query pattern. We summarize these common techniques into a few categories, described in greater detail below: discretizing the problem to make it easier, pruning non-matches using pre-computed data structures, using heuristic methods, and localizing the scope to solve a small version of the same problem.

Discretizing and Binning Relations. For directional and metric relations, binning can be used to make the relation values discrete, enabling more efficient search methods. The MSJ [19] approach bins metric distances between entities into qualitative distance relations like 'near' and 'far,' making the relations simpler to encode and search over. Schwering et al. [22] bin directional relations into eight segments, which limits the size of their qualitative constraint network and reduces the scope of the search. Topological relations are already typically formulated as a discrete set of relationships that can exist between a pair of entities, so binning them is not useful.

Custom Data Structures and Pruning. Spatial keyword indices constructed off entity names are commonly used to prune candidates for search across all types of spatial relations. In practice this limits the number of entities evaluated which reduces the search space significantly, leading to large speedups [4, 8–11, 15]. Some methods leverage Inverted Linear Quadtrees to prune unpromising edges and carefully construct the join order for metric and topological search, gaining efficiency in practice despite a similar worst-case theoretical complexity to many other spatial pattern matching algorithms [4, 15].

Heuristic methods. Depending on the framing of the problem (subgraph matching or constraint satisfaction), heuristic methods can sometimes be used to more efficiently find entities and relations that match the query constraints. For example, Schwering et al. use a heuristic method that relies on an evaluation function to estimate the quality of matchings and to find candidates with high overlap to the query qualitative constraint network [22, 24].

Localizing the Scope. Reducing the scope of the search to only consider very localized entities can also make the search time faster. An example of this approach is shown by Schwering et al., who only consider directional constraints for very localized entities, resulting in a worst-case complexity that is only quadratic in the number of database entities [22]. However, their framing of the problem is specific to sketch map alignment to a road network.

6 Future Opportunities and Challenges

To advance the current spatial pattern matching methods and develop new approaches that scale to handle large spatial datasets, there are a few viable approaches. Reducing the scope of the search to limit the number of neighbors on each candidate node in the database is one approach that has been used in specific framings of the spatial pattern matching problem [22]. By only considering the directional spatial relationships between very localized entities, Schwering et al. achieves quadratic complexity in the number of database entities on the task of aligning a sketch map to a road network [22]. A similar technique could be applied to broader spatial pattern matching to reduce the number of edges in the database graph by dynamically choosing the number of neighbors for each node, depending on the density of entities in the spatial region (i.e. a dense urban area would require more densely connected nodes than a sparse rural area).

Alternatively, approximate methods have been discussed as a potential avenue to address the complexity issues of spatial pattern matching [21]. To match queries against the rich spatial information encoded by metric, topological, and directional relations between each database entity, subgraph matching or constraint satisfaction approaches are typically needed. One avenue to develop more efficient spatial pattern matching will be to leverage work towards approximations to those underlying algorithms [28], including neural approximations [14, 20].

Graph Neural Networks (GNNs) have been used to approximately solve generic subgraph matching tasks [14, 20, 29], but most of these methods are still too limited to apply to spatial pattern matching. GNNs typically embed graphs using sampling via random walks to aggregate neighborhood information and learn the graph structure [12]. While this embedding mechanism is efficient and enables GNNs to learn the overall structure of an input graph, it is not sufficient for subgraph matching, a substructure extraction task. Even advanced sampling strategies that guarantee coverage of certain properties are limited to target-oriented tasks like classification of a particular node in the graph [1]. To bridge this gap in embedding methods, summarization (rather than sampling) has emerged as an alternative to the typical embedding mechanisms [29].

Most GNN approaches, including the ones using summarization-based embeddings, are limited to homogeneous graphs. Spatial pattern matching inherently involves graphs with node and edge attributes, which introduces additional complexity that only Heterogeneous GNNs (HGNNs) can handle. While HGNNs have been successful in a variety of graph representation learning tasks [25–27], they have yet to be applied to spatial pattern matching. While applying GNNs to SGM-based spatial pattern matching would require finding the right embedding method and model architecture, it is a promising path towards developing a flexible approximate approach to spatial pattern matching.

7 Conclusion

Spatial relations are highly descriptive of the world, which makes them challenging to represent and search over efficiently. This paper presents a theoretical comparison of existing graph-based spatial pattern matching approaches and a synthesis of the techniques and heuristics they use to gain performance improvements. We further suggest future work that could lead to meaningful speedups in spatial pattern matching, including by leveraging approximate solutions to the algorithm classes used by the current approaches.

Acknowledgments

This work was sponsored in part by the NSF under Grants IIS-18-16889, IIS-20-41415, and IIS-21-14451, and the Australian-American Fulbright Commission. Its contents do not necessarily represent the official views of the Fulbright Program.

References

[1] Hussein Abdallah, Waleed Afandi, Panos Kalnis, and Essam Mansour. 2024. Task-Oriented GNNs Training on Large Knowledge Graphs for Accurate and Efficient Modeling. (May 2024).

[2] Anderson Chaves Carniel. 2020. Spatial information retrieval in digital ecosystems: A comprehensive survey. In *Proceedings of the 12th International Conference on Management of Digital EcoSystems*. 10–17.

[3] Anderson Chaves Carniel. 2023. Defining and designing spatial queries: the role of spatial relationships. *Geo-spatial Information Science* (2023), 1–25.

[4] Hongmei Chen, Yixiang Fang, Ying Zhang, Wenjie Zhang, and Lizhen Wang. 2019. ESPM: Efficient spatial pattern matching. *IEEE Transactions on Knowledge and Data Engineering* 32, 6 (2019), 1227–1233.

[5] Zhenhong Du, Xianwei Zhao, Xinyue Ye, Jingwei Zhou, Feng Zhang, and Renyi Liu. 2017. An effective high-performance multiway spatial join algorithm with spark. *ISPRS International Journal of Geo-Information* 6, 4 (2017), 96.

[6] Matt Duckham, Jelena Gabela, Allison Kealy, Ross Kyprianou, Jonathan Legg, Bill Moran, Shakila Khan Rumi, Flora D Salim, Yaguang Tao, and Maria Vasardani. 2023. Qualitative spatial reasoning with uncertain evidence using Markov logic networks. *International Journal of Geographical Information Science* 37, 9 (2023), 2067–2100.

[7] Frank Dylla, Jae Hee Lee, Till Mossakowski, Thomas Schneider, André Van Delden, Jasper Van De Ven, and Diedrich Wolter. 2017. A survey of qualitative spatial and temporal calculi: algebraic and computational properties. *ACM Computing Surveys (CSUR)* 50, 1 (2017), 1–39.

[8] Yixiang Fang, Reynold Cheng, Gao Cong, Nikos Mamoulis, and Yun Li. 2018. On spatial pattern matching. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 293–304.

[9] Yixiang Fang, Reynold Cheng, Jikun Wang, Lukito Budiman, Gao Cong, and Nikos Mamoulis. 2018. SpaceKey: exploring patterns in spatial databases. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 1577–1580.

[10] Yixiang Fang, Yun Li, Reynold Cheng, Nikos Mamoulis, and Gao Cong. 2019. Evaluating pattern matching queries for spatial databases. *The VLDB Journal* 28 (2019), 649–673.

[11] Andre Folkers, Hanan Samet, and Aya Soffer. 2000. Processing pictorial queries with multiple instances using isomorphic subgraphs. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Vol. 4. IEEE, 51–54.

[12] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[13] Sahib Jan and Angela Schwering. 2015. SketchMapia: a framework for qualitative alignment of sketch maps and metric maps. In *Proceedings of the AGILE conference on Geographic Information Science*, Vol. 18. Association of Geographic Information Laboratories in Europe.

[14] Zixun Lan, Limin Yu, Linglong Yuan, Zili Wu, Qiang Niu, and Fei Ma. 2023. Subgmn: The neural subgraph matching network model. In *2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 1–7.

[15] Carlos Minervino, Claudio Campelo, Maxwell Oliveira, and Salatiel Silva. 2023. QQESPM: A Quantitative and Qualitative Spatial Pattern Matching Algorithm. *arXiv preprint arXiv:2312.08992* (2023).

[16] Kent O'Sullivan. 2024. *A Framework for Benchmarking Graph-Based Artificial Intelligence*. Master's thesis. University of Maryland.

[17] Kent O'Sullivan, Nicole R. Schneider, Aleeza Rasheed, and Hanan Samet. 2023. GESTALT: Geospatially Enhanced Search with Terrain Augmented Location Targeting. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Searching and Mining Large Collections of Geospatial Data* (Hamburg, Germany) *(GeoSearch'23)*. Association for Computing Machinery.

[18] Kent O'Sullivan, Nicole R. Schneider, and Hanan Samet. 2023. COMPASS: Cardinal Orientation Manipulation and Pattern-Aware Spatial Search. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Searching and Mining Large Collections of Geospatial Data* (Hamburg, Germany) *(GeoSearch'23)*. Association for Computing Machinery.

[19] Dimitrios Papadias, Nikos Mamoulis, and B Delis. 1998. Algorithms for querying by spatial structure. In *Proceedings of Very Large Data Bases Conference (VLDB), New York*.

[20] Indradyumna Roy, Venkata Sai Baba Reddy Velugoti, Soumen Chakrabarti, and Abir De. 2022. Interpretable neural subgraph matching for graph retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 8115–8123.

[21] Nicole R. Schneider, Kent O'Sullivan, and Hanan Samet. 2024. The Future of Graph-based Spatial Pattern Matching (Vision Paper). In *40th IEEE International Conference on Data Engineering, ICDE 2024 – SEAGraph Workshop*. Utrecht, Netherlands.

[22] Angela Schwering, Jia Wang, Malumbo Chipofya, Sahib Jan, Rui Li, and Klaus Broelemann. 2014. SketchMapia: Qualitative representations for the alignment of sketch and metric maps. *Spatial cognition & computation* 14, 3 (2014), 220–254.

[23] Aya Soffer and Hanan Samet. 1999. Query processing and optimization for pictorial query trees. In *International Conference on Advances in Visual Information Systems*. Springer, 60–68.

[24] Jan Oliver Wallgrün, Diedrich Wolter, and Kai-Florian Richter. 2010. Qualitative matching of spatial information. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 300–309.

[25] Xiaocheng Yang, Mingyu Yan, Shirui Pan, Xiaochun Ye, and Dongrui Fan. 2023. Simple and efficient heterogeneous graph neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 10816–10824.

[26] Hanqing Zeng, Muhan Zhang, Yinglong Xia, Ajitesh Srivastava, Andrey Malevich, Rajgopal Kannan, Viktor Prasanna, Long Jin, and Ren Chen. 2021. Decoupling the depth and scope of graph neural networks. *Advances in Neural Information Processing Systems* 34 (2021), 19665–19679.

[27] Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor Prasanna. 2019. Graphsaint: Graph sampling based inductive learning method. *arXiv preprint arXiv:1907.04931* (2019).

[28] Shijie Zhang, Jiong Yang, and Wei Jin. 2010. SAPPER: subgraph indexing and approximate matching in large graphs. *Proc. VLDB Endow.* 3, 1–2 (sep 2010), 1185–1194. https://doi.org/10.14778/1920841.1920988

[29] Zhiqiang Zhong, Cheng-Te Li, and Jun Pang. 2024. Multi-grained Semantics-aware Graph Neural Networks (Extended abstract). In *40th IEEE International Conference on Data Engineering, ICDE 2024*. Utrecht, Netherlands.