Neural Pattern Mining in Spatiotemporal Multigraphs: A Pan-European Land-Use Study

Assaad Zeghina[1[0000-0002-2649-5609]], Aurélie Leborgne[1[0000-0001-8456-2745]], Florence Le Ber[1[0000-0002-2415-7606]], and Antoine Vacavant[2[0000-0001-9616-3282]]

[1] Université de Strasbourg, CNRS, ENGEES, ICube UMR 7357, Strasbourg, France
{assaad-oussama.zeghina,aurelie.leborgne,florence.le-ber}@icube.unistra.fr
[2] Université Clermont Auvergne, Clermont Auvergne, CNRS, Institut Pascal, France
antoine.vacavant@uca.fr

Abstract. Understanding how landscapes evolve requires going beyond aggregated land-cover statistics and looking at the structural arrangement of parcels and their joint evolution. A natural way to formalise this is to model a geographical area as a spatiotemporal multigraph whose nodes are land parcels at successive time instants and whose edges encode their spatial adjacencies and topological transitions. Within this representation frequent subgraph patterns become a principled tool for environmental analysis: a recurring configuration that appears in some areas and not in others can serve as a signature of a phenomenon and as a readable label for a geographical context. We argue that mining such patterns in land-cover data opens a complementary route to area-based monitoring that characterises regions through the dynamics they share rather than the classes they contain. We exercise this idea on Corine Land Cover with Multi-SPMiner, a graph neural network-based pattern miner that scales to multigraphs mixing spatial and spatiotemporal relations, complemented by three post-hoc filters targeting environmental semantics. We illustrate the approach on a handful of recovered patterns that recur across cities of similar profile while remaining absent elsewhere, each acting as a readable label for an environmental phenomenon and supporting cross-region comparison.

Keywords: Spatiotemporal pattern mining · Corine Land Cover · Graph neural networks · Environmental monitoring · Land-use change.

1 Introduction and Related Work

Long-term monitoring of land-use dynamics is central to assessing urban pressure, agricultural change, and ecosystem transformation. Most quantitative analyses of continental land-cover inventories rely on per-class area statistics or pairwise transition matrices [6], indicators that capture first-order change but overlook the structural arrangement of parcels and their joint evolution. Capturing this kind of relations requires a representation that makes the spatial layout of parcels and their evolution across time first-class objects of analysis. A natural such representation is the spatiotemporal multigraph, in which nodes are land parcels at successive time instants, spatial edges encode adjacencies between co-temporal parcels, and spatiotemporal edges encode the topological transitions that link parcels at consecutive timestamps. Within this representation, recurring subgraph patterns become a principled tool to characterise regions: a configuration that appears with high support in some regions and is absent from others can serve as a signature of a phenomenon and as a label attached to the geographical context in which it recurs. Mining frequent subgraph patterns is, however, computationally hard. Traditional algorithms [3,4] enumerate candidate patterns and count their occurrences, an operation whose complexity grows super-exponentially with pattern size and which becomes intractable on multigraphs that combine several edge types. The rapid progress of deep learning on graphs and spatiotemporal graphs has led to powerful representation learning techniques for capturing complex structural dependencies [10]. In particular, recent neural approaches replace explicit enumeration with a learned geometric embedding in which subgraph relations are preserved. NeuroMatch [5] introduced order embedding for subgraph matching, and SPMiner [8] extended it to frequent pattern mining. Multi-SPMiner [9] generalises SPMiner to single labelled directed multigraphs and was shown to retrieve frequent patterns in near-linear time on spatiotemporal graphs.

In this work, we apply this representation-and-mining pipeline on the *Corine Land Cover*[3] (CLC) inventory of the Copernicus programme, which provides harmonised pan-European maps at five time instants (1990, 2000, 2006, 2012, 2018) and organises classes into a three-level nomenclature, from five top-level classes down to forty-four detailed ones. CLC is well suited to our goal: its continental coverage and multi-decadal horizon let us search for patterns across geographical contexts that span very different profiles, and its multi-level nomenclature lets us probe how the granularity of land classes affects the patterns recovered. Our contributions are: (i) a pan-European study at two CLC nomenclature levels (Level 1 and Level 2) over regions of contrasting profiles, including Eastern-European, Anatolian, and Caribbean territories rarely considered jointly; (ii) three post-hoc filters (frequency, change-intensity, and redundancy) designed to keep only environmentally informative patterns; (iii) a temporality-dropping protocol re-purposed as a means to surface patterns that operate on horizons longer than the native CLC step; (iv) a qualitative cross-city analysis that contrasts urbanisation speed, agricultural transitions, and coastal dynamics across European contexts.

2 Spatiotemporal Multigraphs Generation

2.1 Spatiotemporal multigraphs.

Following [2], a spatiotemporal graph (st-graph) describes a set of temporal entities and the relations that link them across a finite time domain $\mathcal{T} = \{t_1, \ldots, t_n\}$.

[3] https://land.copernicus.eu/en/products/corine-land-cover

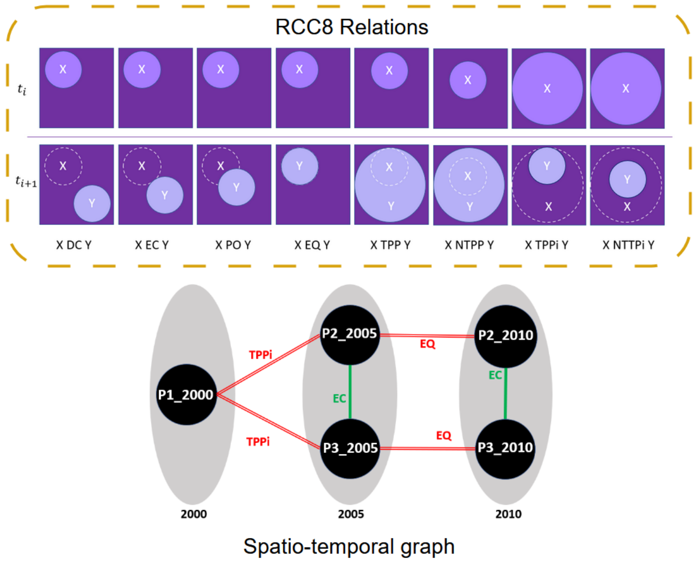

Fig. 1. Schematic spatiotemporal multigraph over three time instants.

We cast it as a labelled multigraph $\mathcal{G} = (V, E, \ell_V, \ell_E, L_V, L_E)$ where V contains nodes representing entities at specific time instants with vertex labels $L_V = \Delta \times \mathcal{T}$ for an entity set Δ, and edge labels $L_E = \Sigma$. The set Σ contains spatial and spatiotemporal topological relations, here taken from RCC8 [1] (DC, EC, PO, EQ, TPP, $NTPP$, $TPPi$, $NTPPi$). Edges only link nodes at identical or successive timestamps. Two subsets of edges can be identified: spatial edges between co-temporal entities, and spatiotemporal edges between successive timestamps. Figure 1 shows a small example.

2.2 Building the graph from CLC.

CLC is distributed as a vector inventory of polygonal parcels, each labelled with its three-level thematic class and production year. For each region of interest and each of the five CLC years we (i) select a region, (ii) create one node per polygon labelled with its class (at the chosen nomenclature level) and timestamp, (iii) insert spatial edges between adjacent co-temporal polygons with their RCC8 label, (iv) insert spatiotemporal edges between successive-year polygons that overlap geometrically, with their RCC8 label. The resulting multigraph encodes both static topological structure and land-cover transitions. This paper analyses CLC at **Level 1** (five classes: artificial, agricultural, forest/semi-natural, wetland, water) and **Level 2** (fifteen classes refining each Level 1 category, e.g. Level 2 splits *artificial* into urban fabric, industrial/commercial, mine/dump/construction, and artificial vegetated areas). Level 3 is not used here: its forty-four classes induce patterns whose support is typically too low to exceed the mining frequency threshold.

3 Pipeline and Post-Hoc Filters

3.1 Pattern mining

Multi-SPMiner extracts frequent node-anchored patterns in two phases:

- *Embedding phase.* The multigraph is decomposed into K-hop neighbourhoods around every node, and each neighbourhood is mapped into an order embedding space[7] via a multigraph convolutional network (MGCN). MGCN extends graph convolutions to graphs with multiple edge types by maintaining a separate convolution per relation type (here, spatial and spatiotemporal) and aggregating the resulting messages into a single per-node representation, so that subgraph isomorphism is preserved as elementwise vector ordering. We use the original implementation of [9] with architectural adaptation.
- *Search phase.* Starting from a seed node, the search grows the pattern greedily by adding at each step the adjacent node that increases the probability to find a frequent motif in the neighbourhood embeddings; multiple seeds yield a robust frequency estimator. For further details, the reader is referred to [9].

3.2 Post-hoc filters.

The raw output of Multi-SPMiner contains a non-trivial fraction of patterns that are technically frequent but environmentally uninformative. We apply three independent filters in sequence (Fig. 2).

F1 – Frequency filter. The walk-based estimator returns a noisy approximation of pattern frequency. We discard patterns whose empirical count over N_{walks}=500 independent walks is below $s_{\min}$=25, a threshold chosen so that the false-discovery rate under a uniform null is below 5%.

F2 – Change-intensity filter. For each pattern G with spatiotemporal edge set E_t, we define the *static ratio* $\rho(G) = |\{e \in E_t : \ell_E(e) = eq\}|/|E_t|$ and discard patterns for which $\rho(G) > \tau$. Setting τ=0.5 removes patterns in which a majority of temporal edges encode topological stasis, i.e. parcels that retain identical class and topology across successive timestamps. Such patterns are by construction frequent on CLC, where most parcels are unchanged, but they convey no transition information.

F3 – Redundancy filter. Greedy growth in Multi-SPMiner naturally produces nested patterns: if pattern G of size k is recovered with high frequency, several of its size-$(k-1)$ subgraphs are typically recovered too, with at least the same support. We apply a final pass that, for each pair (G, G') of recovered patterns with $G' \subset G$ and $\text{supp}(G') \leq 1.1 \cdot \text{supp}(G)$, retains only the larger one. This avoids double-counting nearly identical patterns and substantially cleans the qualitative output.

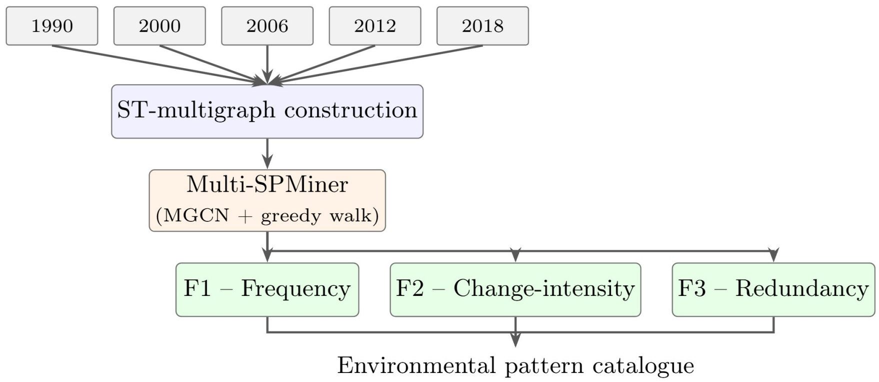

Fig. 2. End-to-end pipeline. Multi-SPMiner produces candidate patterns; the three post-hoc filters then deliver an environmental pattern catalogue.

4 Experimental Setup

We study ten geograhical areas across Europe and one overseas French territory, chosen to cover contrasting biogeographical, climatic, and socio-economic profiles. Each region is analysed at CLC **Level 1** (5 classes) and **Level 2** (15 classes). From each geographical zone we sample graphs of 1,000 nodes, patterns are searched at sizes $k \in \{4, 6, 7\}$. Beyond size 7, the estimated frequency of candidate patterns typically falls below $s_{\min}$ on the regions we considered, which makes larger patterns effectively unreachable under F1; this constrains the practical regime of pattern sizes for CLC at the considered nomenclature levels. For each region and level, we train our pattern mining method on 200 spatiotemporal graphs and validate on 10 additional graphs in order to obtain quantitative validation values. The trained model is then applied to the real CLC multigraph. Quantitative ground truth is obtained with a Mugram exhaustive enumeration [4] adapted to multigraphs. We report the *retrieval ratio*, i.e. the fraction of ground-truth patterns recovered by Multi-SPMiner, averaged over ten seeds. All experiments run on a workstation with an AMD Ryzen 7 9800X3D CPU and an NVIDIA RTX 4080 (16 GB) GPU.

5 Quantitative Validation

We briefly validate that Multi-SPMiner is reliable on CLC before turning to the qualitative analysis of the extracted patterns. Table 1 reports the retrieval ratio at the two nomenclature levels and the three pattern sizes, aggregated across the ten regions. The retrieval ratio stays in a narrow band of 0.73–0.78 across both levels and all three pattern sizes, with no systematic degradation as the pattern size grows up to seven. The results in Table 2 highlight the complementary roles of the three filters. The frequency filter F1 induces only a minor reduction (below 19%), confirming that the mining process already respects the support threshold

Table 1. Retrieval ratio under P1 (mean ± std over the ten regions and ten seeds). Performance is essentially flat across pattern sizes within the practically reachable range $k \leq 7$.

	k=4	k=6	k=7
Level 1 (5 classes)	0.77 ± 0.01	0.78 ± 0.02	0.76 ± 0.01
Level 2 (15 classes)	0.74 ± 0.01	0.75 ± 0.01	0.74 ± 0.02

Table 2. Effect of the post-hoc filters (Level 1) on found candidate patterns.

Region	No filter	F1	F2 (τ=0.5)	F3
Ankara	1 480	1 202	280	36
Paris	1 821	1 532	195	14
Fort-de-France	1 315	1 219	103	42

and produces few spurious candidates. In contrast, the change-intensity filter F2 has a strong and region-dependent impact: it removes a large proportion of patterns in the Parisian region (around 90%), where most configurations correspond to stable land use, while having a more moderate effect in Ankara and Martinique, which exhibit more dynamic behaviors. Finally, the redundancy filter F3 prunes nested sub-patterns, retaining only representative larger structures. This leads to a strong reduction in Paris, where patterns are highly redundant, and a more limited effect in Ankara and Martinique, where the pattern sets are more diverse. End-to-end inference (graph construction excluded) takes about 23 seconds per region at Level 2, k=7, with peak GPU memory under 2 GB per graph. The exhaustive Mugram ground-truth on the same region requires approximately 30 minutes on CPU. The pipeline is therefore compatible with continental-scale exploration on a single consumer GPU.

6 Qualitative Cross-City Analysis

The remainder of the paper centres on the patterns themselves and on what they reveal about European land-use dynamics. We organise the discussion in three parts: patterns at Level 1 (broad trajectories), patterns at Level 2 (refined transitions), and patterns surfaced by temporality dropping. Throughout this section, nodes are coloured by their CLC class, green edges denote spatial adjacency at a given timestamp, and red edges denote spatiotemporal transitions across timestamps, labelled with the RCC8 relation that encodes the type of geometric change between successive states.

6.1 Level 1: broad land-use trajectories

At Level 1, patterns aggregate land cover into five broad categories and expose coarse trajectories shared by cities of similar profile (Fig. 3). The patterns reported below all occur with a frequency above 25 in each cited region. Pattern A

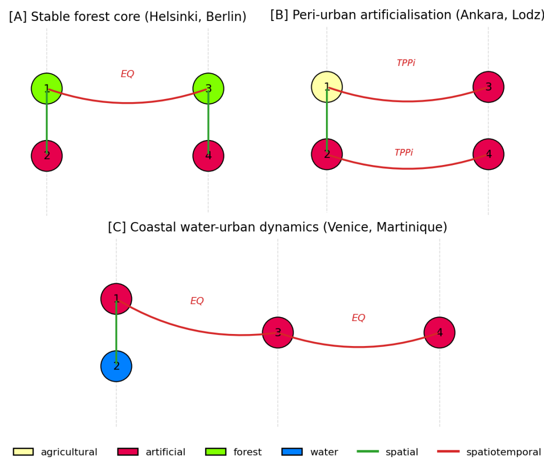

Fig. 3. Three representative Level 1 patterns. Numbers inside nodes are local node identifiers within each pattern. Green edges: spatial adjacency. Red edges: spatiotemporal transitions, labelled with the RCC8 relation.

pairs a forest parcel with an adjacent artificial neighbour at the early timestamp, with the forest persisting unchanged into the next timestamp where it remains adjacent to another artificial parcel. The pattern recurs in Helsinki and Berlin and is essentially absent from Ankara and Lódź, capturing the stability of forest cores in well-protected metropolitan green belts where the immediate urban neighbourhood reorganises around an unchanged core. Pattern B couples an agricultural parcel adjacent to an artificial one at the early timestamp, with both parcels becoming artificial at the next timestamp: the former agricultural parcel is absorbed into a larger artificial parcel that contains its earlier extent, and the adjacent artificial parcel itself expands into a larger artificial parcel that contains its own earlier extent. It appears with high support in Ankara and Lódź, and its support acts as a speed indicator for cities undergoing rapid peri-urban conversion of their agricultural rings: the simultaneous expansion of two adjacent parcels points to a coordinated wave of artificialisation rather than to isolated parcel conversions. Pattern C centres on an artificial parcel adjacent to a water body, with the artificial parcel persisting unchanged across three consecutive timestamps. It is one of the few three-timestamp patterns recovered at Level 1, and it emerges in Venice and Fort-de-France, where coastal urban parcels stay topologically stable along the shoreline.

6.2 Level 2: refined transition patterns

Level 2 brings finer thematic resolution and lets us discriminate trajectories that are merged at Level 1 (Fig. 4).

Pattern D combines an arable parcel adjacent to an urban-fabric parcel, with the arable parcel transitioning into an industrial/commercial parcel that contains

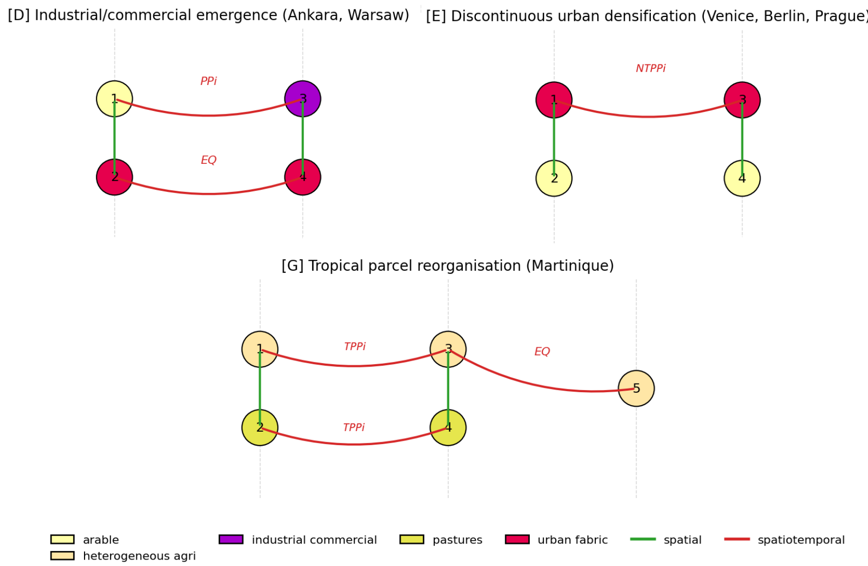

Fig. 4. Three representative Level 2 patterns.

its earlier extent, while the urban fabric persists unchanged and remains adjacent to the new industrial parcel. The pattern is found with high support in Ankara and Warsaw and refines pattern B by showing that the artificialisation observed at Level 1 is dominated by the emergence of industrial and commercial zones from arable land rather than by residential expansion of urban fabric. Pattern E pairs an urban-fabric parcel with an arable neighbour at the early timestamp, with the urban-fabric parcel growing in the next timestamp to non-tangentially contain its earlier extent, indicating that the later urban parcel envelops the earlier one without touching its boundary, encoding densification through extension on all sides, while a fresh adjacency to an arable neighbour is observed at the later timestamp. Overall, pattern E signals an expansion of the built-up area. The pattern emerges in the mainland fringe of Venice and in the peripheries of Berlin and Prague. Pattern F is specific to Fort-de-France and involves two adjacent agricultural parcels (one heterogeneous agricultural, one pasture) that both fragment between two timestamps, with the new heterogeneous and pasture parcels remaining adjacent and the heterogeneous one then persisting unchanged into a third timestamp. Beyond the fragmentation itself, the pattern points to an expansion of agricultural and pasture areas, likely at the expense of forested land or, in a context of agricultural abandonment, of arable plots converted to pasture. The combined fragmentation of two adjacent agricultural parcels at the same step does not appear in any continental region of our sample, illustrating how strongly local management practices imprint themselves on pattern catalogues.

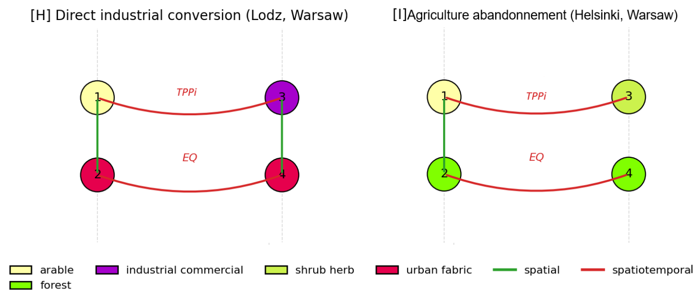

Fig. 5. Two patterns recovered after dropping the 2006 timestamp. Both encode trajectories that are masked at the native CLC step.

6.3 Long-horizon patterns from temporality dropping

Several environmental processes evolve on horizons longer than the six-to-ten-year CLC step and are fragmented by intermediate timestamps that capture transient states, splitting a long trajectory into several shorter patterns (Fig. 5). For instance, pattern H, obtained by removing the 2006 timestamp before mining, is structurally identical to pattern D but expresses the arable-to-industrial transition across two native CLC steps; in Łódź and Warsaw, the end-to-end conversion is then recovered with higher frequency than at the native step, indicating that widening the time window makes such patterns more visible. Pattern I pairs an arable parcel with a forest neighbour and captures the transition of the arable parcel into a shrub/herbaceous cover that contains its earlier extent, while the forest persists unchanged. The arable-to-shrub trajectory over a long period is characteristic of agricultural abandonment. It emerges in rural areas of Helsinki and Warsaw and is masked at the native step by short patterns involving pastures or heterogeneous agricultural areas as transient states. These observations indicate that temporality dropping is best treated as a deliberate multi-resolution exploration of the same data: the native step and a sparser temporal grid yield complementary patterns that should be analysed jointly.

6.4 Cross-region signatures

Beyond the individual patterns, the recovered set can itself be used as a basis for cross-region analysis: each region is summarised by the subset of patterns it exhibits, and regions can then be compared, contrasted, or grouped by the overlap of their pattern sets. The handful of patterns reported here already illustrates how this works. Helsinki and Berlin both exhibit pattern A and do not appear in B or D. Warsaw, Łódź, and Ankara share patterns B and D. Venice and Fort-de-France share pattern C, with Fort-de-France additionally exhibiting pattern F. Paris, Prague, and Budapest sit between these groups. Scaling this approach across more regions and more patterns would provide a structural complement to area-based monitoring by characterising regions through the dynamics they share rather than through the classes they contain.

7 Discussion and Conclusion

We applied Multi-SPMiner to Corine Land Cover at two nomenclature levels across ten European regions. The recovered spatiotemporal patterns reflect well-known dynamics of European land use and expose finer effects that are not visible in standard CLC analyses. The proposed post-hoc filters keep the output manageable, help analyse evolution in given areas, and discard patterns dominated by topological stasis. Treating temporality dropping as a discovery device further surfaces long-horizon patterns that the native CLC step fragments. Several directions are worth pursuing. The pattern sizes reachable in practice are currently bounded above by seven at the considered nomenclature levels, and we aim to extend the approach to larger patterns by improving the frequency estimator. Pattern interpretation is presently manual, and a natural next step is to cluster recovered patterns by transition type to support automated reading of the output.

References

1. Cohn, A., Bennett, B., Gooday, J., Gotts, N.: Qualitative spatial representation and reasoning with the region connection calculus. GeoInformatica **1**(3), 275–316 (1997)
2. Del Mondo, G., Rodríguez, M., Claramunt, C., Bravo, L., Thibaud, R.: Modeling consistency of spatio-temporal graphs. Data & Knowledge Engineering **84**, 59–80 (2013)
3. Elseidy, M., Abdelhamid, E., Skiadopoulos, S., Kalnis, P.: GrAMI: Frequent subgraph and pattern mining in a single large graph. Proceedings of the VLDB Endowment **7**(7), 517–528 (2014)
4. Ingalalli, V., Ienco, D., Poncelet, P.: Mining frequent subgraphs in multigraphs. Information Sciences **451–452**, 50–66 (2018)
5. Lou, Z., You, J., Wen, C., Canedo, A., Leskovec, J.: Neural subgraph matching. arXiv preprint arXiv:2007.03092 (2020)
6. Tomov, P.: Land use dynamics in the south-central planning region (2000–2018): A statistical analysis of corine land cover. In: Proceedings of the Bulgarian Academy of Sciences. pp. 785–794 (2025)
7. Vendrov, I., Kiros, R., Fidler, S., Urtasun, R.: Order-embeddings of images and language. In: Proc. Int. Conf. on Learning Representations (ICLR) (2016)
8. Ying, R., Wang, A., You, J., Leskovec, J.: Frequent subgraph mining by walking in order embedding space. Proc. Int. Conf. on Machine Learning (ICML) (2020)
9. Zeghina, A., Leborgne, A., Le Ber, F., Vacavant, A.: Multi-SPMiner: A deep learning framework for multi-graph frequent pattern mining with application to spatiotemporal graphs. Procedia Computer Science **225**, 1094–1103 (2023)
10. Zeghina, A., Leborgne, A., Le Ber, F., Vacavant, A.: Deep learning on spatiotemporal graphs: A systematic review, methodological landscape, and research opportunities. Neurocomputing **594**, 127861 (2024)