Geometry, Topology and *p*-Adic Numbers in Geospatial Data Modelling, Management and Processing: Review and Future Approaches

Patrick Bradley[1], Markus Jahn[1], Martin Breunig[1], Mulhim Al-Doori[2], Emmanuel Stefanakis[3], Paul Kuper[1]

[1] Geodetic Institute, Karlsruhe Institute of Technology, Germany – (bradley, markus.jahn, martin.breunig, paul.kuper)@kit.edu
[2] University of Science & Technology of Fujairah, College of Engineering and Technology, U.A.E. – m.aldoori@ustf.ac.ae
[3] Department of Geomatics Engineering, University of Calgary, Canada – emmanuel.stefanakis@ucalgary.ca

Keywords: Geometric Data Modelling, Topological Data Modelling, Topological Data Management, Simulations with Geospatial Data, *p*-Adic Numbers, Distributed Simulation

Abstract

Geometry and topology are both central concepts in geospatial data modelling and management. While geometry is the central concept for computational geometry applications, e.g. to intersect surfaces and solids in 3D space, topology is helpful for many application classes starting from city modelling to subsurface modelling and indoor navigation. Furthermore, *p*-adic numbers are useful for describing hierarchical processes on topological models. In this paper we first shortly review geometry- and topology-based approaches used for geospatial applications. We then describe our approach on turning geometry "upside down" focusing on topology during the whole process of distributed geospatial computing, data modelling, data management, and simulation. Furthermore, the way to use *p*-adic numbers for the description of hierarchical processes on topological models is shown for the example of simulation and the idea of *p*-adic analysis in distributed simulations is presented in the context of topological city and building models. The approach opens new insights such as topological relationships, components, and efficiently studying of the approximate behaviour of processes in the built environment using distributed computational systems. Finally, exemplary applications are presented to underline the importance of this new approach.

1. Introduction

Without any doubt, geometry and topology are central concepts to describe geospatial environments. Whereas geometry is used to describe the exact location of objects represented by coordinates, topology is taken to describe properties and spatial relationships between parts of an object as well as between different objects. Multiple geospatial applications deal with geometry and topology, e.g. the intersection of maps has a consequence on the topology (network of polygons), but for computing the intersections themselves, the coordinates of the geometries have to be known as well. In this case, geometry even seems to be the more important concept, because without geometry the exact geospatial result of the intersections cannot be computed. This may be the reason why the use of topology is narrowed in the GIS community, mostly just corresponding with topological relationships between objects such as "disjoint", "meets", and "overlaps". However, topology has an enormous potential for geospatial applications and simulations as we will see throughout this article. In order to name just a few, (Kovalerchuk et al., 2012) discuss the use of topology and geometry in Spatial Data Fusion and Mining (SDFM), focusing on dynamic and uncertain topological structures. This somewhat aligns with the focus on the management of dynamic geospatial data also taken in this article. The methodology of (Baldado et al., 2018) enables a high level of detail classification from combining geometric and topological information. (Xiong et al., 2014) perform graph correction on the roof topology of 3D building data. These lines of research exemplify the already existing abundance of applications in geographic data given by topology. The following section will be more detailed and specific about an abstract point of view of topology in the context of geographic information with a perspective on increasing the scope of applications of this mathematical discipline.

Another feature to be captured in the modelling of geographic information is hierarchy. This allows to break down information into pieces interconnected in different resolutions and bears a high potential to be exploited e.g. when simulating time-dependent processes on city models. In order to do this, the use of *p*-adic numbers is explained in this article which presents ongoing work on developing the necessary mathematical tools. *P*-adic numbers are introduced in a more detailed manner in later sections, but their importance is their inherent hierarchical nature. Inspired by applications in mathematical physics, complex systems and biology, cf. (Dragovich et al., 2017) for a broad review, (Bradley, 2010) explores the utility of *p*-adic numbers for image processing and computational topology. This resonates with the envisioned application of *p*-adic numbers for modelling hierarchical processes on topological models in this article. (Hua and Hovestadt, 2021) investigate the application of *p*-adic numbers to model complex networks, providing a theoretical basis that could intersect with the ongoing research on distributed simulations.

The remainder of this article is structured as follows: In section 2 we give a short review of geometry- and topology-based approaches motivated by geospatial applications. In section 3 we turn "geometry upside down" focusing on the role of topology during distributed geospatial computing, data modelling and data management. In section 4 we use topology to simulate processes on topological models and we show how *p*-adic analysis can help in distributed simulations within topological city and building models. Finally, section 5 shows application fields of the approach and in section 6 conclusions are drawn from this article.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-51-2024

2. Geometry- and Topology in Geospatial Applications

As is well known, geometry literally translated from ancient Greek means "land survey" or "measurement of the earth". Thus, geometry deals with spatial and non-spatial objects, shapes and structures as well as their dimensions, distances and other properties. In the practice of geospatial data modelling, data management and data processing, especially two- and three-dimensional spaces are of importance as they reflect objects on maps and in 3D models, respectively.

2.1 Geometry-based Approaches

As is common in the geospatial community, (Worboys and Duckham, 2004) distinguish between field-based and object-based approaches for geospatial data modelling. A field-based model consists of a finite collection of spatial fields, whereas an object-based model decomposes space into objects. Each object in space has a unique ID. The implementations of these two approaches are raster and vector, respectively. A raster builds a regular or irregular partitioning of space. Examples are a grid of squares for a regular partitioning and Triangulated Irregular Networks (TINs) for an irregular partitioning. I.e. raster data are structured as an array of cells, referred to as pixels. A vector is a finite straight-line segment represented by its end points. Each end point is represented by its coordinates which define the location of the point. Vectors are used to approximate the geometry of real world objects. Obviously, the implementation of the object-based model needs much less storage place than the implementation of the field-based model does.

In 3D space, the two-dimensional regular partitioning of space, is extended to the voxel, i.e. the implementation of a three-dimensional partitioning of space is represented by a finite collection of cubes. Also, space-optimized solutions such as the quadtree in 2D space have their counterpart in 3D space represented by the octree. The irregular partitioning in 3D space is a finite collection of polyhedra such as the Tetrahedral Networks (TENs). In solid modelling and computer graphics, TENs are used to model the interiors of solids. Principally there are two different ways to model solids, which can be designated as direct and indirect modelling: Direct modelling of a solid means that not only the boundary of the solid is modelled, but additionally the interior of the solid is represented by a 3-cell decomposition such as a TEN. Generally speaking, the solid is then modelled by touching 3-simplexes (tetrahedra), which again consist of four 2-simplexes (triangles) each, which again consist of three 1-simplexes (straight lines), finally consisting of two 0-simplexes (points). Early implementations of 3D geodatabase management prototype systems for the geosciences considered such simplex-based 3D geometric data types (Breunig et al., 2010). Indirect modelling of solids means that only the boundary of the solid is modelled without knowledge about the interior of the solid. The boundary model is e.g. used in the CityGML standard (Gröger and Plümer, 2012), although the corresponding GML-class on which CityGML has access to, is designated as "solid". In the 3D/4D DB4GeO project, an Octree and a 3D R-Tree have been implemented (Breunig et al., 2010) to index large sets of simplex-based objects. Resilient distributed spatial indexes have been implemented in the GeoSparc distributed system architecture to support the access to big geospatial vector and raster data (Yu and Sarwat, 2021).

2.2 Topology-Based Approaches

(Egenhofer and Herring, 1990) and other authors transferred the formalization of binary topological relationships into the GIS world and shaped the narrow note of topology used in the GIS community. Their 4-Intersection- and 9-intersection model formalize the minimum amount of the binary topological relationships between two objects designated as "disjoint", "meets", "overlap", "covers/coveredBy", "contains/ containedBy" and "equals". T_0-topologies described in more detail in Section 3.1 totally contain this approach. In a T_0-space for every pair of distinct points of the topological space, at least one of them has a neighbourhood not containing the other. In the GIS community, (van Oosterom et al., 2002) reflected on the relationship between topology and geometry and on how to translate topological structures into geometric primitives. In (Breunig et al., 2020) topology in a broader sense has been identified as one of the future directions of geospatial data management research.

In geospatial application fields such as Geo-Information Science (GIS) and Building Information Modelling (BIM) and the corresponding GIS/BIM standards, topology is treated in different ways. For example, the topology in IndoorGML (OGC 2024) is based on the Poincaré duality. This duality states that for every triangulation of an n-dimensional manifold there is a dual decomposition into polyhedrals so that the k-cells of the polyhedron decomposition are unique to the (n-k)-cells of the triangulation (Seifert and Threlfall, 1934). From this in the BIM/GIS context the dual half-edge structure has been developed, which is well suited to model the neighbourhood relationships between solids along surfaces in 3D cell complexes (Boguslawski and Gold, 2016). A disadvantage of this approach is that not all topologies are covered, since the Poincaré duality only holds to closed manifolds. In dimension two, this means that exactly two surfaces border each edge. However, in multi-storey building models, for example, there are at least three wall surfaces bordering to an edge, if this edge lies between two floors. In the OGC standard CityGML (Gröger and Plümer, 2012) it is possible to explicitly define topology with so-called XLinks. However, it usually has to be extracted from the geometry (Salleh and Ujang, 2018). In the Industry Foundation Classes (IFC) of buildingSmart, the boundary relationship is implicitly modelled, and the boundary representation must first be created from IFC (buildingSMART, 2024) and (Lilis et al., 2017).

Furthermore, topological models are particularly well-suited to model the life cycle of buildings and urban structures (Jeong and Son, 2016), (Jin et al. 2019) and (Willenbacher et al., 2006). The topology serves as a "higher-level perspective" in addition to geometry and thematic classes/product model/semantics (Borrmann et al., 2009) and (Bradley and Paul, 2010). (Zlatanova et al., 2004) presented a review of models to define topological relationships of objects used in 3D GIS and 3D geo-databases. In geo-scientific modelling, so-called G-Maps (Lienhardt 1994) are used to navigate within the topology of 3D objects. However, it could be shown that the G-Maps not only require a restriction to a certain class of spaces, but they are also not memory efficient (Bradley and Paul, 2014). Finally, topology may also be helpful to advance data analysis (Bradley, 2015) beyond the shortest path search. This aspect of strengthening data analysis by topological concepts will be taken up later in this paper.

In the following, we consider the full spectrum of topology as defined in mathematics rather than the narrowed view as used in the GIS community. Thus, topology is in the centre of our interest.

3. Geometry “Turned Upside Down”: Topology put into the Centre of Interest

In the following, geometry is “turned upside down”, meaning that topology is explicitly used to find new insights for distributed geospatial computing, data modelling, data management, and even simulation. Topology then is an underlying structure of geometry.

3.1 Topology and Distributed Geospatial Computing

Collaboration of specialised scientists, like any other collaboration, directly depends on suitable communication methods. This does not only include the ability of traditional communication by phone, email services or messengers which have been invented as a sort of digital twin for analogue communication based on paper and spoken words. An efficient collaboration also includes the abilities of processing, sharing and presenting scientific data as needed. A proper meta-data management also helps to support the historically grown graphs of collaboration. The intra-communication of the scientific world has different needs and purposes than the intra-communication of the political world, financial world or any other social sub-system (Luhmann, 1995). This is also true for the inter-communication of each social sub-system.

In the following paragraphs, we focus on the intra- and inter-communication of the geo-scientific social sub-system to support the geo-scientific collaboration intra- and inter-specific by the research on geo-data-management, the geo-information technology behind the curtains. The major issues which we are facing have been compromised by the FAIR principles (findability, accessibility, interoperability, and reusability). Looking at the IPO model (input-process-out) which every standard GIS (Geo-Information System) implements by its four basic components (data acquisition, data management, data analysis and data presentation) it is possible to form graphs of geo-scientific collaboration by connecting the output of one IPO-unit with the input of another distant IPO-unit by some ETL (extract transform load) approach. The bottle neck of a distributed management system for IPO-units is data transfer. We are used to think in pushing data around like letters or hand drawn maps. This is not efficient when dealing with large datasets. State offices try to digitize historical datasets to fulfil the digital revolution of modern information technology. But digitizing is not digitalizing. Digitalization also involves the revision and implementation of information technologies to manage the involved data efficiently. The communication network may not be based on a direct communication between each personal computer of each involved specialised scientist. This decentralized solution still is common practice in geo-sciences, often not even using databases and thinking in files. The reason seems to be the historical way of sharing information, by spoken words and paper.

Modern information technologies provide ways of joint work on the same datasets by utilizing OLTP (Online Transaction Processing) approaches ("working in the cloud") and ways of processing large dataset by utilizing OLAP (Online Analytic Processing) approaches (HPC - high performance computing) for a handful highly educated professionals as centralized solutions. Both solutions are needed when dealing with geo-data and the bottleneck between OLAP and OLTP is slow data transfer combined with a lot of pre-processing to transform the heterogenous geo-data from the OLTP to extremely special data structures used in OLAP technologies. Special data-structures in OLAP are needed to speed up the processing by parallelizing the processing. Contemporary communication technologies deal with that problem by implementing FOG and Edge computing which outsources pre-processing steps to the input-nodes of some specialist IPO-node. So, the data-provider can produce certain data-structures to reduce the pre-processing workload of some specialised IPO-node within the communication network. This can reduce the data transfer by implementing chains of pre-processing IPO-nodes, also. Each chain supports the final input of the special IPO-node where otherwise traditionally each IPO-node would be asked to send its output to the special IPO-node and this special IPO-node restructures the input-data in a large pre-processing step. It is always in question if a pre-processing step needs to be outsourced, since the involved nodes have different processing speeds and are connected by different band widths. As a cause of that, revising communication network topologies and automatizations are not trivial tasks.

Putting it together:

1. we deal with interconnected IPO's (GIS, CAD)
2. the graph of collaboration represents the information-flow
3. the graph of collaboration is dynamic
4. the data needs to be handled FAIR (and secure)
5. the graph of collaboration must be visible/recognizable explicitly to the end-user, the geo-scientist
6. the geo-scientist does not need to know the topology of the underlying communication network
7. the information technology needs to utilize Edge, Fog and Cloud/HPC computing supporting OLTP and OLAP

The Karlsruhe Data Infrastructure (KADI) can be used as data management system for IPO’s which focuses on the FAIR principles (Griem et al., 2022). Historically, KADI was developed for the material sciences at the Karlsruhe Institute of Technology known as KADI4Mat. In present, efforts are taken to integrate the management of GIS data. One of the major advantages is the management of workflows following a bottom-up approach where each scientist can integrate his or her work as an IPO and connect to other scientists' work. In that way the dynamically grown graphs of collaboration can be managed in a user defined environment. Services like WMS WFS or languages like CityGML are top-down approaches which force the scientist to rely on the predefined features, only. The bottom-up approach, on the other hand, can also be seen in Mediator Wrapper architecture where step by step databases can be added to form a large, distributed database management system which can be arranged by the needs of the scientific workflow/dynamic graph of collaboration. In that sense it is also possible to utilize Edge-, Fog- and Cloud/HPC-computing in a bottom-up approach.

3.2 Topology and Data Modelling

In mathematics, topology is an underlying structure of geometry. In geoinformatics, the idea emerged that the modelling of topological data ought to follow this mathematical principle that the core of the data model be topology, and geometry a way of making the underlying topology more precise in the sense that many different geometries have the same underlying topology. The second observation was that since computers can store only finite data, the topology stored on a computer must reflect this fact. These two principles would help to set the handling of data

https://doi.org/10.5194/isprs-annals-X-4-2024-51-2024 | © Author(s) 2024. CC BY 4.0 License.

in the context of geoinformatics on a sound mathematical basis. Already in (Alexandrov, 1937), it was proven that any binary relation on a finite set is topological. Hence, the customary practice in geoinformatics to call only certain binary relations “topological”, and others not, is devoid of sense from a topological perspective. This work became the leading guideline towards defining “topological databases” in (Bradley and Paul, 2010). These are a relational database implementation of a minimal representation of the partial ordering coming out of a large class of binary relations via taking the reflexive and transitive closure. And this captures in principle all possible topologies on a finite set known as T_0-topologies, according to Alexandrov. General finite topologies can be reduced to T_0-topologies by forcing relations to not have directed cycles. Besides the partially ordered sets, also cell complexes were turned into a relational database version. A comparison with *G*-maps in (Bradley and Paul, 2014) showed that those tend to be rather verbose compared with topological databases. The strength of this approach is revealed in (Paul and Bradley, 2015), where an integrated topological model for space, time and version is developed.

The notion of “topological consistency” was introduced in (Giovanella et al., 2019) in order to check whether the topology inherent in geometric models coincides with the explicit topology of such models. It simply means that space is partitioned into the “atomic” constituents of the model, i.e. without overlaps between such, as stated in the ISO 191907 standard. However, the violation of this standard seems to be ubiquitous in city models, and a partial classification of such inconsistencies is performed. The definition used here concerns not the usual correct modelling of objects, but the correct modelling of the incidence relationships. This then allows to produce volumetric models suitable for simulations of processes, which are robust (Jahn and Bradley, 2022b). On this basis, a topological access method (TOAM) was developed in (Jahn and Bradley, 2022a).

3.3 Topology and Data Management

The last two paragraphs point to the importance of topology within the geo-sciences from completely different points of view. Our practical research focuses on the integration of graph-databases into systems like KADI. We have been implementing a Framework which manages vector data for moving and morphing geo-objects in a graph called DB4GeOGraphS (Jahn, 2022). All intra- and inter-relations of a geo-object are manged by a graph. Lessons learnt from the implementation of that framework are now added to investigate how the algorithms can be expressed by a Graph Query Language (GQL) called TINKERPOP. This will end up in a set of GQL statements which can be used to manage and analyse the graphs of moving and morphing geo-objects. Finally, we hope to present a Geo-Graph-Query-Language as a Domain-Specific-Language (DSL) for moving and morphing vector based geo-objects. The graphs are semi-structured with a certain graph schema. As a cause of that, we do not focus on the use of graph databases as backends only since structured data may be managed more efficiently through other database technologies like relational databases. But the establishment of a Geo-Graph-Query-Language may ease the work of the end-users, the geo-scientists. Therefore, our research on the Geo-Graph-Query-Language as a Domain-Specific-Language for the Geoscience in the context of distributed data management systems is focused on the usability.

We took care to keep the graph schema as simple and abstract as possible to not handicap the end-user. The graph schema is derived from the mathematical point of view of topology by defining three relation types, only. These are (A) abstraction-relations, (B) aggregation-relations and (C) incidence-relations. Abstraction relations are functionally independent which means that the spatio-temporal geo-objects are not dependent from each other. This type can be used to manage Level-of-Details of geo-objects by a path from less detailed to more detailed geo-objects, as an example. On the other hand, aggregation- and incidence relations are functionally dependent. A geo-object may be part-of other geo-objects, may be made-of other geo-objects or is made-of the atomic simplices themselves, maybe triangles. Furthermore, this geo-object can intersect other geo-objects, like a triangle net intersects another by a curve. So, both nets share their inner points defined by that line. They are made-of that line, also. This may seem odd, but the idea is to strictly differentiate between geo-objects' inner points and their boundary. This is done by explicitly modelling the boundary of a geo-object through the incidence relation type. As an example, a polygon is widely defined as a set of interconnected straight-line segments forming a hopefully planar and non-intersecting ring/curve in 3D space. This is called a boundary representation of a two-dimensional geo-object. In our approach this only defines the boundary of the geo-object which is a one-dimensional geo-object in first place. A triangulation needs to be done to define its inner points by defining a suitable set of triangles filling that boundary in its planar plane. So, the whole geo-object, in common sense, is always a set of two different geo-objects, in our sense, (a) the boundary object and (b) its triangulation, in case of a polygon. This approach is strictly taken for every n-dimensional geo-object based on simplices of any dimension in 4D space. So at the end each node of our geo-object-graph represents always the inner points of the geo-object and the boundary of each geo-object is modelled explicitly with a boundary-of relation forming a completely new geo-object of one less dimension without a boundary, since there is no boundary of boundaries, or in other words, a boundary representation is always some closed hyper-hull, the ring in the polygon example. In that sense, we are not able to store Boundary Representations anymore. They contradict our approach. But we can store the boundary as an aggregation-node of line segments, take some triangulation algorithm to find the set of triangles which fills the boundary, relate that set as an aggregation node of those triangles and finally, relate both aggregation nodes (boundary-node and triangulated inner-node) with an incidence relation. This approach simplifies/splits the commonly known 9intersection model, since every object is represented as an open (without its boundary) geo-object, only.

4. Towards Simulation of Processes on Topological Models Using *p*-Adic Numbers

4.1 Topology and Simulation

The strict differentiation between boundary and inner points of geo-objects descried in Sec. 3.3 can lead to topologically consistent graphs of interconnected geo-objects via their shared boundary parts, if the computational geometry is robust. Topologically consistent graphs can be used in simulations where the boundary parts represent glued objects, like the 2D-membranes (triangle nets) between 3D-volumes (tetrahedron

https://doi.org/10.5194/isprs-annals-X-4-2024-51-2024

nets). Topologically consistent graphs are manageable by ordinary graph databases for OLTP and efficiently analysed on information technologies for OLAP (e.g. SPARK) using the Geo-Graph Query language we are working on.

Simulating a process, e.g. a heat flow, needs correctly modelled incidences at all resolutions. Hence, the topological model must be topologically consistent, cf. Sec. 3., since incidences not explicitly modelled prevent the passage of a physical quantity through the simulation model. The same holds true for fake incidences. So, an incorrect incidence model leads to incorrect simulation results. The Alexandrov property (cf. Sec. 3.1), allows graphs to model processes in a topologically consistent model. Integrating over a group of “atoms” allows the passage to a lower resolution. Using aggregation hierarchies leads to a tree structure usable to locally approximate simulations at the microscopic level, if the flows on a macroscopic level are of interest. In theory, a speed-up of calculations on large Laplacian matrices which are “hierarchical matrices” can be obtained. This is ongoing research in the project in more detail in Sec. 3.5. Hierarchical matrices by (Hackbusch, 2009) have been developed precisely with the aim of speeding up simulations by approximating differential operators. The ongoing project uses a special case of hierarchical matrices.

For understanding the main idea, assume that a heat source increases the temperature of a city (e.g. the sun is shining). Then the temperature increases on the surface of the buildings, and this increase spreads throughout the whole construction of all buildings, even the smallest constructive parts of the buildings, if connected with the area heated by the heat source. The flow depends on the local diffusion parameters between all parts of the model at the various resolution levels. In principle, one could model the temperature flow at the microscopic level of individual bolts, screws, fillings materials etc. But if the city has any reasonable size, the calculations would be too cumbersome on even large and fast computing systems. And if the resolution of interest is significantly lower (say, building complexes as units), then this huge calculation would be an overkill, and would not lead to any added value compared to the faster calculation at the lower resolution. However, assuming that the diffusion parameters are known at the higher resolution, it does make sense to aggregate the data, and then calculate an aggregated flow of temperature at the desired resolution. The idea now is to use a diffusion process on the leaves of a suitable tree representing the finer constituents of the low-resolution components of the system. The reason behind this is that calculations on a tree are much faster than on a more complex topological structure. And the hypothesis is that the diffusion at the microscopic level does not really matter for the diffusion at the lower resolution, as long as it is approximatively close to it. That is why the microscopic diffusion is replaced by a diffusion on a tree. When (in the particle model of diffusion) a random particle leaves a tree and jumps to another part of the system, then it enters through the leaf of another tree which also approximates that other local part of the system. And if a greater resolution is needed, then the parts represented by a tree are subdivided into smaller parts, again each of which is modelled by a tree on which a diffusion takes place. In Sec. 4.2, it will be explained how p-adic numbers are helpful for realising this idea. And since the analysis on p-adic numbers lags behind classical mathematical analysis, there is a need to develop the mathematical tools for such an approach. The aim of this ongoing work is to develop the necessary mathematical theory, the numerical tools, and to investigate their suitability in the context of simulating processes in city and building models, following the idea described here.

4.2 *p*-Adic Analysis in Distributed Simulations on Topological City and Building Models

In an ongoing research project *Distributed Simulation of Processes in Buildings and City Models,* funded by the German Research Foundation, one important focus is on developing new mathematical methods for simulating processes like heat flows on city models and buildings using distributed computational systems. The key idea is that using hierarchical structures, i.e. trees, should not simply be restricted to the data itself, but also used in the distributed computation process. This means that the used quantities (e.g. a Laplacian matrix) should be partitioned into pieces which are each approximated by a hierarchical version of this piece. This idea relies on the fact that processing a rooted tree is much faster than more complex structures. A coarse version of a graph structure at the macroscopic level is obtained by aggregation, whereas at the internal level, the microscopic processes each are replaced by a process on the leaves of a tree. The jumps between the local tree-like patches do occur between the corresponding leaf nodes, but governed by the coarse graph which is meant to be relatively small in comparison with the total number of data points.

For simulating the dynamics within and between hierarchically organised point-sets, suitable mathematics needs to be developed. Diffusion processes, describing heat flows on various media, have been studied also for hierarchically organised structures, in particular if the corresponding tree is regular. If the regularity is a prime number p, then the tools of *p-adic analysis* come into play (Vladimirov et al., 1994). This mathematical toolbox is almost 50 years old (Taibleson, 1975), and has found various applications in mathematical physics, biology, complex systems, and also deep learning (Dragovich et al., 2017). It relies on the hierarchical structure of p-adic numbers, introduced in (Hensel, 1897), which are expressions of the form

$$a_{-m}\, p^{-m} + \ldots + a_{-1}\, p^{-1} + a_0 + a_1\, p + a_2\, p^2 + \ldots \qquad (1)$$

for some natural number m, and where the coefficients a_i are integers between 0 and p-1. The importance of the primality of p lies in the fact that then, and only then, the usual properties of arithmetic hold true, including convergence with respect to a special metric coming from the prime number p. In this p-adic metric, any two such series (called *p-adic numbers*) are nearby, if they have an equal common initial term up to a high power of p, i.e. if their difference is divisible by a high power of p. The norm of high powers of p is deemed small for this metric. Another peculiarity is that no two p-adic discs overlap unless one is contained in the other. This amounts to a hierarchical organisation of p-adic discs which can be depicted as a rooted tree: the root is a large p-adic disc, and the children are the (disjoint!) maximal sub-discs contained in the root disc, and so on. This works because the only possible radii of discs are integer powers of p. This tree is regular, each vertex having precisely p children, because there are precisely p possible values for each a_i in the expansion above. This is an infinite regular rooted tree whose ends correspond to the p-adic numbers contained in the disc D associated with the root. Pruning at a certain finite level yields a finite regular tree whose leaf nodes correspond to small p-adic discs inside D, and these form a partition of D into p^n discs of equal radius for some natural n.

Diffusion and the heat equation on p-adic numbers is a well-studied topic in the meantime, cf. (Dragovich et al., 2017) and the references therein, but what is much less developed is the theory of such processes on subspaces of the p-adic numbers or

https://doi.org/10.5194/isprs-annals-X-4-2024-51-2024 |

even p-adic manifolds, except for the case of a single p-adic disc, or finitely many p-adic discs with constant transition rates between them (Zúñiga-Galindo, 2020). In the manifold case in dimension 1, the presence of an algebraic structure is used for developing a corresponding Laplacian operator and study the corresponding heat equation (Bradley, 2023). In the present context of this article, the finitely many patches of a large graph on which a diffusion occurs each correspond to a p-adic disc, but in contrast to the approach in (Zúñiga-Galindo, 2020), the overall process is that of a non-constant diffusion along the many edges connecting the relatively few patches. This means that the diffusion process is only locally of the diffusion on the leaves of a regular tree.

In the known cases of p-adic diffusion, there is a corresponding stochastic process of Markov type, and the solution to the heat equation is unique for a given initial condition and can be explicitly worked out like in the classical setting. This can with good reason be expected to be the case also in the more general p-adic setting needed for the efficient distributed computing of heat flows on city models, at least approximately, if the resolution is not too large. The ongoing work is to prove the necessary mathematical theorems, to be followed by developing numerical schemes for being able to do actual calculations on distributed systems, and for error estimation. Developing algorithms on distributed graph database systems, and dealing with the computational aspects like indexing, partitioning schemes and graph database query languages, is the ongoing applied part of this current project in the intersection of mathematics, computer science and geo-information.

A benefit of an additional p-adic structure, is the facilitation of the “hearing” of shapes, at least theoretically (Bradley and Ledezma, 2023). This is the question about whether important properties, or the whole structure itself, can be recovered from the spectrum of a suitable Laplacian operator. In general, the answer is negative, as e.g. the graph Laplacian can “hear” a graph only up to iso-spectrality, but not the desired isomorphy, and many examples of non-isomorphic but iso-spectral pairs of graphs have been found in the literature. Adding a p-adic structure leads to an operator whose spectrum can identify the isomorphism class of the graph. This does not solve the graph isomorphism problem, however, since the spectra of such operators are infinite, and the computational costs are deemed high, even as one does not need all eigenvalues of the p-adic Laplacian operator. However, if the question is more humble and asks only for topological information about the graphs, then a heat flow approach is e.g. possible for extracting the Euler characteristic of a graph (Roth, 1984). A desideratum for the future is such an approach via p-adic analysis which will hopefully enable the development of efficient methods for extracting topological information from topological data on distributed systems. A step towards this aim is the p-adic version of space-filling curves in higher dimension introduced and studied in (Bradley and Jahn, 2022) which now needs extension to topological datasets.

5. Advanced Applications

5.1 Semantic 3D City Models for Smart Cities

Semantic 3D city models are an advanced form of 3D modelling that includes additional layers of information about the objects and features within the model, beyond their physical dimensions and geographic locations (Ohori et al., 2022). These models are essential for the development of smart cities as they facilitate the integration of various types of data, including GIS data, building information modelling (BIM) data, and other environmental data to create a comprehensive view of the city landscape. Semantic 3D city models provide planners and managers with detailed insights into urban spaces, allowing for better decision-making regarding land use, infrastructure development, service allocation, and asset/resource management.

In recent years, there has been an increasing interest in the development of semantic 3D city models. The relevant activities are focusing on the following four directions:

(a) collection of feature data for semantic 3D city models using Geomatics Engineering technologies and tools (such as LiDAR imaging and other active and passive sensing systems);

(b) implementation of a series of pilot semantic 3D city models for the exterior and interior (rooftops, openings, architectural, electrical, mechanical, and piping infrastructure) of municipal buildings, as well as ground and underground street infrastructure (trees, traffic lights, curbs, gutters, road surface, utility lines);

(c) advanced visualization and analytical tools for semantic 3D city models to support planners and managers in the decision making process; and

(d) encoding and dissemination of the semantic 3D city models data using international standards, including the Open Geospatial Consortium, buildingSMART Consortium, and ISO geospatial and BIM formats (e.g. CityGML/JSON, IFC) and integrated web services (e.g. WMS, WFS, CSW, WPS); to be hosted by the City of Calgary and other Open Data servers.

5.2 Integrated autonomous platforms for smart cities in the United Arab Emirates

A need for efficient data management solutions has arisen as a result of the ongoing expansion of connected and electro-mobility goods and services, which has resulted in their capacity to create extremely large volumes of data in a relatively short period of time. The necessity for society to develop educated policies and judgments that can appropriately support their rising growth is another factor that contributes to the acceleration of this phenomenon. Data systems and platforms do exist; however, they are often proprietary, meaning that they are only compatible with the respective products for which they were developed. The inability of these systems to communicate with one another would make decision making more difficult, as the data from each system would need to be analyzed separately.
The design and development of a highly dynamic data platform delivering additional business intelligence capabilities that will enable the promotion of other parties to the utilization of transport data. These partners will be able to interact with the platform and define and customize data sets to be used or sold through. This approach requires interaction between different concerned departments and parties involved in the data utilization or monetization process. The installation and configuration of the analytics and business intelligence platform at this stage of evolution, will enable users to get exceptionally flexible and adaptive real-time data analysis. People should be able to work with data from any source and in any format due to its accessibility both on-premises and in the cloud, and compatibility with the widest range of data formats. Specialized actionable analytic applications with seamless app integration may be developed by development teams, opening up new revenue sources and giving them a major competitive advantage (Zhang et al, 2019).

6. Conclusions

In this paper we reviewed geometry- and topology-based approaches being especially relevant for geospatial applications. We then presented our advanced approach on turning geometry "upside down" to focus on topology during the whole process of distributed geospatial computing, data modelling, data management, and simulation. Furthermore, it was shown how to use *p*-adic numbers for the description of hierarchical processes on topological models using the example of simulation. The idea of *p*-adic analysis in distributed simulations was explained in the context of topological city and building models. The approach yielded novel insights including topological relationships, the identification of components, and the efficient examination of the approximate behaviour of processes within the built environment through the utilization of distributed computational systems. Finally, exemplary smart city applications were presented to underline the significance of this new approach.

Acknowledgements

We thank the anonymous referees for their helpful suggestions, as well as the German Research Foundation (DFG) for the support of our research (BR 2128 /11-1, BR 2128/18.1, BR 2128/21.1) under project numbers 326802322, 469999674 et al.

References

Alexandrov, P., 1937. Diskrete Räume. *Mat. Sb.*, 44, 501– 519.

Balado, J., Díaz-Vilariño, L., Arias, P., H. González-Jorge, H., 2018. Automatic classification of urban ground elements from mobile laser scanning data. *Automation in Construction*, 86, 226-239

Boguslawski, P., Gold, C., 2016. The Dual Half-Edge-A Topological Primal/Dual Data Structure and Construction Operators for Modelling and Manipulating Cell Complexes. *ISPRS Int. J. Geo-Inf.* 5(2), 19.

Borrmann A and Rank E., 2009. Topological analysis of 3D Building Models using a Spatial Query Language. *Advanced Engineering Informatics* 23(4), 370–385.

Bradley, P.E., 2010. From image processing to topological modelling with p-adic numbers. *p-Adic Numbers, Ultrametric Analysis, and Applications*, 2(4), 293, 2010.

Bradley, P.E., 2015. Supporting Data Analytics for Smart Cities: An Overview of Data Models and Topology. In: Gammerman, A., V. Vovk und H. Papadopoulos (Ed.): *Statistical Learning and Data Sciences SLDS 2015*, Lecture Notes in Computer Science 9047, Springer, 406–413.

Bradley, P.E., 2023. Heat equations and wavelets on Mumford curves and their finite quotients. *Journal of Fourier Analysis and Applications*, 29, 62.

Bradley, P.E.; Jahn, M.W., 2022. On the behaviour of p-adic scaled space filling curve indices for high-dimensional data. *The computer journal*, 65 (2), 310–330.

Bradley, P.E.; Ledezma, Á.M., 2023. Hearing shapes via p-adic Laplacians. *Journal of Mathematical Physics*, 64 (11), Art.-Nr. 113502.

Bradley, P.E.; Paul, N., 2010. Using the relational model to capture topological information of spaces. *The Computer Journal*, 53 (1), 69–89.

Bradley, P.E.; Paul, N., 2014. Comparing G-maps with other topological data structures. *Geoinformatica*, 18 (3), 595–620.

Breunig, M., Bradley, P.E., Jahn, M., Kuper, P., Mazroob, N., Rösch, N., Al-Doori, M., Stefanakis, E., Jadidi, M., 2020. Geospatial Data Management Research: Progress and Future Directions, *ISPRS International Journal of Geo-Information*, 9(2), 95, 20p., https://doi.org/10.3390/ijgi9020095, 2020.

Breunig, M., Schilberg, J., Thomsen, A., Kuper, P.V., Jahn, M., Butwilowski, E., 2010: DB4GeO, a 3D/4D Geodatabase and its Application for the Analysis of Landslides, *Geographic Information and Cartography for Risk and Crisis Management: Towards Better Solutions*, Springer Berlin Heidelberg, 83-101.

buildingSMART, 2024. Technical Resources. https://technical.buildingsmart.org/ (19 February 2024).

Dragovich, B., Khrennikov A.Y., Kozyrev, S.V., Volovich, I.V. and Zelenov, E.I., 2017. p-Adic Mathematical Physics: The First 30 Years. *p-Adic Numbers, Ultrametric Analysis and Applications*, 9(2), 87–121.

Egenhofer, M.J., Herring, J.R., 1990. A mathematical framework for the definition of topological relationships, In: *Proceedings of Fourth International Symposium on Spatial Data Handling*, Zurich, Switzerland, 803–813.

Giovanella, A.; Bradley, P.E.; Wursthorn, S., 2019. Evaluation of Topological Consistency in CityGML. *ISPRS International Journal of Geo-Information*, 8 (6), Article: 27.

Gröger, G. and Plümer, L., 2012. CityGML—Interoperable semantic 3D city models. *ISPRS J. Photogramm. Remote Sens.*, 71 (2012), 12–33.

Griem, L., Zschumme, P., Laqua, M., Brandt, N., Schoof, E., Altschuh, P., Selzer, M., 2022. KadiStudio: FAIR Modelling of Scientific Research Processes. *Data Science Journal*, 21. 10.5334/dsj-2022-016.

Hackbusch, W., 2009. *Hierarchische Matrizen. Algorithmen und Analysis*. Heidelberg: Springer.

Hensel, K., 1897. Über eine neue Begründung der Theorie der algebraischen Zahlen. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 6 (3), 83–88.

Hua, H., Hovestadt, L., 2021. p-adic numbers encode complex networks. *Sci Rep*, 11, 17.

Jahn, M.W., 2022. Distributed & Parallel Data Management to Support Geo-Scientific Simulation Implementations. PhD thesis. Karlsruhe Institute of Technology.

Jahn, M.W., Bradley, P.E., 2022a. Topological Access Methods for Spatial and Spatiotemporal Data. *ISPRS International Journal of Geo-Information*, 11 (10), 533.

Jahn, M.W., Bradley, P.E., 2022b. A Robustness Study for the Extraction of Watertight Volumetric Models from Boundary Representation Data. *ISPRS International Journal of Geo-Information*, 11 (4), Artk.Nr.: 224.

Jeong, W.S. and Son, J.W., 2016. An Algorithm to Translate Building Topology in Building Information Modeling into Object-Oriented Physical Modeling-Based Building Energy Modeling. *Energies* 9, 50.

Jin, R. et al., 2019. Integrating BIM with building performance analysis in project life-cycle. *Automation in Construction,* 106, 102861.

Lienhardt, P., 1994. N-dimensional generalized combinatorial maps and cellular quasi-manifolds. *International Journal on Computational Geometry and Applications,* 4(3), 275–324.

Lilis, G.N., Giannakis, G.I. and Rovas, D.V., 2017. Automatic generation of second-level space boundary topology from IFC geometry inputs. *Automation in Construction,* 76, 108–124.

Luhmann, N., 1995. *Social Systems*. Timothy Lenoir and Hons Ulrich Gumbreeht Writing science. Stanford University Press, 1995.

Kovalerchuk, B., Perlovsky, L., and Kovalerchuk, M., 2012. Modeling spatial uncertainties in geospatial data fusion and mining. Proceedings Volume 8396, Geospatial InfoFusion II; 83960N.

OGC. IndoorGML, 2024. OGC Standard for Indoor Spatial Information. http://indoorgml.net/ (19 February 2024).

Ohori, K.A., Ledoux, H., Peters, R., 2022. *3D modelling of the built environment*, v0.8.

Paul, N., Bradley, P. E., 2015. Integrating Space, Time, Version, and Scale using Alexandrov Topologies. *International journal of 3-D information modeling*, 4 (4), 64–85.

Roth, J.P., 1984. Le spectre du Laplacien sur un graphe. In: Mokobodzki, G., Pinchon, D. (eds) *Théorie du Potentiel*. Lecture Notes in Mathematics, vol 1096. Springer, Berlin, Heidelberg.

Salleh, S., Ujang, U., 2018. Topological information extraction from buildings in CityGML. In: *IOP Conf. Series: Earth and Environmental Science*. Vol. 169. IOP Publishing Ltd, 2018, 012088.

Seifert, H., Threlfall, W., 1934. *Lehrbuch der Topologie*. Teubner.

Taibleson, M.H., 1975. Fourier analysis on local fields, Mathematical Notes, Princeton University Press, Princeton, New Jersey.

Van Oosterom. P., Stoter, J., Quak, W., Zlatanova, S., 2002. The Balance Between Geometry and Topology, *Advances in Spatial Data Handling*, 209-224.

Vladimirov, V.S., Volovich, I.V., Zelenov, E.I., 1994. *p-Adic Analysis and Mathematical Physics*, World Scientific, Singapore.

Willenbacher, H., Hubler, R., Wender, K., 2006. GIS-Functionality for building model integration and analysis. In: *Joint Int. Conf. Comp. Dec. Mak. Civ. Build. Eng*. 2191–2199.

Worboys, M., Duckham, M., 2004. *GIS – A Computing Perspective*, Second Edition, CRC Press, 426p.

Xiong, B., Oude Elberink, S., Vosselman, G., 2014. A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93, 227-242.

Yu, J. Sarwat, M., 2021: Big Geospatial Data Processing Made Easy: A Working Guide to GeoSpark, M. Werner and Y.-Y Chiang: *Handbook of Big Geospatial Data*, 35-53.

Zhang D.-G., Zhao P.-Z., Cui Y.-Y., Chen L., Zhang T., Wu H., 2019. A New Method of Mobile Ad Hoc Network Routing Based on Greed Forwarding Improvement Strategy, *IEEE Access*, vol. 7, pp. 158514-158524, doi: 10.1109/ACCESS.2019.2950266.

Zlatanova, S., Rahman, A.A., Shi, W., 2004. Topological models and frameworks for 3D spatial objects. *Computers & Geosciences* 30(4), 419–428.

W. Zúñiga-Galindo, 2020. Reaction-diffusion equations on complex networks and Turing patterns, via p-adic analysis. *Journal of Mathematical Analysis and Applications*, 491(1), 124239.