

Use of knowledge graphs applied to modeling mobile telecommunications network infrastructures in Brazil^{*}

Daniele Nazaré Tavares^{1,†}, José Maria Parente de Oliveira^{1,†}

¹Instituto Tecnológico de Aeronáutica (ITA), Praça Marechal Eduardo Gomes, 50. Vila das Acácias, 12228-900. São José dos Campos/SP - Brasil.

Abstract

The information age and digital communication have transformed the way we collect, process, store, and display data. However, when dealing with georeferenced information, the challenge remains of representing spatial data in a manner consistent with its logical and semantic context. Conventional spatial tables describe attributes in isolation, failing to capture their meaning within the context in which they are embedded. This work proposes the use of knowledge graphs to integrate spatial and non-spatial databases, aiming to represent semantic relationships between geographic entities and telecommunications infrastructure across Brazil. Open data from Anatel and IBGE were used to structure the knowledge graph, which allows for the analysis of the distribution of radio base stations in each Brazilian municipality and their geoeconomic impact. The study demonstrates the feasibility of enriching analyses of mobile telecommunications infrastructure (2G, 3G, 4G, and 5G), supporting Anatel's decision-making in public policies aimed at ensuring interoperability.

Keywords

Ontology, Database, Knowledge Graphs, Knowledge Representation, Semantic Networks, Telecommunications, Mobile Networks, Geographic Information System (GIS)

1. Introduction

Mobile telecommunications network data is characterized by heterogeneous data types and complex relationships between the entities represented. However, the way in which data is currently available is limited in relation to these aspects. For this reason, a form of storage is needed that enables more efficient and meaningful organization of this data. Knowledge graphs are emerging as a promising alternative for organizing information on telecommunications infrastructure and services in Brazil.

In this paper, we present a proposal for a knowledge graph structure that represents the topology or architecture of mobile telecommunications networks in Brazil. The proposal allows for data scalability and more accurate analyses of the organization and behavior of radio base stations (ERBs) in different regions. The proposal also offers the following advantages:

- Knowledge graphs allow visualization of the network topology, understanding the relationship between its physical and logical components.
- With graph-based reasoning, it is possible to identify the root cause of failures or service degradation, streamlining the maintenance and restoration process of the telecommunication network.
- They allow the use of performance metrics related to network configurations to recommend antenna installation optimization strategies.

The geospatial database was obtained from open data from the IBGE (Brazilian Institute of Geography and Statistics), including names of states and municipalities, territorial geometries and other geospatial information about Brazil [1][2]. The data related to mobile telecommunications network infrastructure was extracted from the MOSAICO ou System Integrated Spectrum Management and Control System (Broadcasting Modules), made available by ANATEL (National Telecommunications Agency) [3].

Proceedings of the 18th Seminar on Ontology Research in Brazil (ONTOBRAS 2025) and 9th Doctoral and Masters Consortium on Ontologies (WTDO 2025), São José dos Campos (SP), Brazil, September 29 – October 02, 2025.

✉ dany.nt.14@gmail.com (D. N. Tavares); parente@ita.com.br (J. M. P. d. Oliveira)

🌐 <https://danynazaretech.github.io/danynazare> (D. N. Tavares)

🆔 0009-0001-2276-0145 (D. N. Tavares); 0000-0002-7803-1718 (J. M. P. d. Oliveira)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The rest of the document is organized as follows: **Section 2** presents related work in telecommunications that uses knowledge graphs and ontologies; **Section 3** describes the proposal for a knowledge graph for mobile telecommunications networks in Brazil; **Section 4** presents examples of queries made to the graph and discusses its benefits; finally, **Section 5** presents conclusions and suggestions for future work.

2. State of the Art

knowledge graphs began to be used in 2012 by Google [4] for the purpose of organizing data in the search for information at the conceptual level as a result of the interconnection of various types of data associated by semantic relationships. Knowledge graphs, which are data structures in graphs that combine data, relationships, and metadata to create a comprehensive understanding of the contextual object being represented, enable knowledge extraction [5][6]. This formal model is composed of nodes, which represent entities or objects, while the edges indicate the inference relationships established between them.

A notable example of this approach can be seen in Google's search engine. When you search for the name of a city, it not only displays related hyperlinks, but also presents organized information about its location, cultural aspects, health infrastructure, and geographic data. This data comes from an ontologically structured knowledge graph, which links different sources of information based on formally established semantic relationships.

In addition to applications in search engines, knowledge graphs have become fundamental tools in areas such as natural language processing, recommendation systems, semantic relation extraction, knowledge engineering, and intelligent chatbots. Their structure facilitates data organization during the preprocessing and modeling stages, making them especially useful for feeding artificial intelligence algorithms with contextualized and semantically enriched data.

The paper **Ontology for IP Telephony Networks** [7] presents an ontology in the field of IP telephony networks, focusing on the development of programs for this modality of network. The proposal is based mainly on the analysis of the SIP and H.323 protocols and the dominant signaling protocols in VoIP telephony. The telephony network structure is modeled with classes representing different types of nodes in the network: end-user terminals and network infrastructure. The goal is to standardize the telephony network structure and client usage to facilitate the development of telephony applications with interoperable data.

The TOUCA Project proposes the ToCo (Telecommunication CANvas Ontology) ontology [8], developed to represent hybrid telecommunications networks that combine wired and wireless technologies, various types of devices, interfaces, links, users, services, and channel quality in a modular and reusable way, based on the Device-Interface-Link (DIL) design pattern. ToCo allows for a semantic and integrated description of the infrastructure and behavior of modern networks, supporting applications such as software-defined networks (SDN), performance monitoring, and interoperability between technologies.

Advances in the use of knowledge graphs have also been made in the direction of efficiently compressing the density of data generated by 6G mobile networks, called pervasive multilevel native AI (PML)[9]. In this work, wireless data graphs are used to *"organize and condense complex and disordered data, extracting a concise subset that represents the most effective and critical factors for network AI models that process large volumes of wireless data"*, contributing to the economy of memory in the storage of the data collected.

In the works presented above, it is worth highlighting the relationship between telecommunications networks and knowledge graphs to understand the semantic behavior between the physical and logical elements of a network for the construction of new protocols, architectural elements, topologies, and data refinement. In continental countries like Brazil, Russia, and China, where deployment costs are a barrier to its growth, this view of the network helps to build mobile telecommunications networks efficiently.

3. Knowledge graph of mobile telecommunications networks

Structuring spatial data together with non-spatial data is an inherently complex task, reflecting the wide diversity of formats, such as satellite images, cartography, vector data, matrix data, demographic data, textual data, and numerical data. Although tables help with organization, they are limited in their ability

to visualize the semantic relationships between cartographic representation, geographic attributes, and external data from multiple sources. For this reason, geographic information systems (GIS) arose from the need to capture, manipulate, analyze, and model geographic data and metadata to provide an integrated and geospatial context-oriented view [10].

An additional limitation of the conventional tabular model lies in the difficulty of performing complex data queries, thereby reducing the potential to explore semantic relationships between data elements.

To represent the structure of mobile telecommunications networks in Brazil using a knowledge graph, the first step was to define the types of data to be used. Cartographic data in Shapefile format, numerical data, and text extracted from IBGE sources and georeferenced telecommunications data from Anatel's MOSAICO system were therefore used. Geographic, cartographic, and geoeconomic data from IBGE on Brazilian territory were used to analyze the structure of radio base stations (ERBs) with ANATEL data, interpreting how MOSAICO system mobile services are impacted throughout Brazilian territory.

Based on the understanding of this data, the corresponding knowledge graph was generated, linking the georeferenced data with the geographic data from the two databases. It expresses the relationships between the data of each ERB in relation to the spectrums it controls, its influence in the regional zone, antenna coverage by region, frequencies available for use, types of mobile services, and how each region manages each mobile telecommunications service, all with the organization of the geographic entities described in the IBGE data. Thus, as a basis for constructing the graph, the defined conceptual schemes of ANATEL and IBGE data show how the ERBs are distributed and their impact on each region.

3.1. Data description of Anatel and IBGE

The geospatial data extracted from the IBGE platform are in Shapefile (.shp) format, which contains the geometric characteristics and territorial boundaries of each Brazilian municipality. From the cartographic database, 5,572 municipalities and 27 states were extracted, along with six attributes for each municipality that are implicit in the map. External geographic information on states and municipalities was obtained from the Cidades@ portal [1], which describes the socioeconomic and structural characteristics of each locality, totaling 22,558 additional data points.

The SPECTRUM MOSAICO platform from Anatel is intended to manage telecommunications services, keep track of broadcasting station records, and show whether each ERB's infrastructure conforms with Anatel regulations. The open data used on licensed telecommunications stations in Brazil is in tabular .csv format with 2,083,046 rows and 41 columns, which separates the characteristics of each station's antennas. This study considers mobile service stations (2G, 3G, 4G, and 5G), and through their geographic position (latitude and longitude), it is possible to infer the impact of each ERB in the region where it is installed and how they interact with each other.

Using data from IBGE and Anatel, it is possible to establish spatial relationships with mobile architecture and its relationship with stations, making it possible to infer which antennas are in a specific area, perform network coverage or saturated area analysis, plan the network, and detect anomalies. In addition, it is possible to integrate data from other sources to infer the operational, administrative, or regulatory characteristics of public services. The principle behind this vision is the importance of semantic enrichment of data from multiple domains [10], so that end users can understand the impact of the mobile network across Brazil and consume it abruptly. .

3.2. Conceptual data diagram

A radio base station (ERB) is a fixed installation that houses the equipment responsible for enabling communication between mobile devices and the telephone company, providing signal coverage in a geographical area.

Based on current legislation [11] [12] [13], each ERB is defined in the graph as an entity according to the technical and regulatory criteria established by the MOSAICO system data. In the graph, each ERB is configured as a component responsible for the operational management of Antennas, presenting a 1:N cardinality relationship. Each Antenna is composed of physical attributes related to its operation, using wireless communication technologies such as 2G, 3G, 4G, and 5G. In addition, each one has concrete physical

infrastructure and uses transmitting equipment to emit signals within the coverage area. Antennas installed in critical locations must be identified in order to follow an installation and operation protocol that is different from the others. Figure 1 shows the conceptual and topological model of the ERB structure with the antennas and technologies that operate, the technical characteristics under their operation, associated with the equipment used in the transmission infrastructure.

So each antenna has a concrete physical infrastructure and uses transmitter equipment to broadcast signals within the coverage area. Antennas installed in critical locations must be identified in order to follow an installation and operation protocol that is different from the others. The Figure 1 presents the conceptual model of the ERB structure with the antennas and the technical characteristics of their operation associated with the equipment used in the transmission infrastructure.

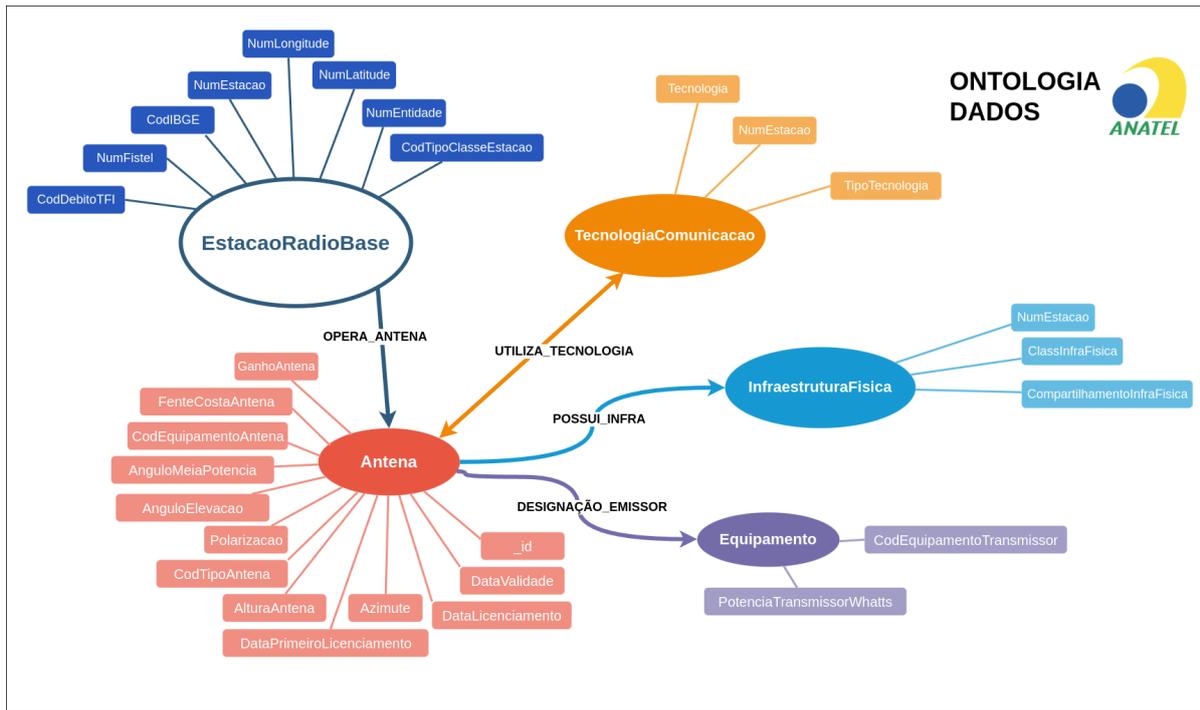


Figure 1: Conceptual model of the ERB structure.

To model the information related to licensed ERB concessions, latitude and longitude coordinates were used as a basis for establishing an indirect relationship between stations. This approach prevents data explosion if the ERBs were directly connected to other stations. With the support of thematic cartography extracted from the IBGE map, it is possible to indirectly associate each station with its structural characteristics and infer its area of influence in the geographical region where the ERB provides signal coverage.

Thus, to represent the distribution of ERBs in cities and organize them administratively in accordance with Anatel, only neighbouring cities in the same state were connected directly, and the state module was connected to the capital city, as illustrated in Figure 2. This connection between cities is useful for solving the travelling salesman problem when Anatel’s regional superintendence needs to perform maintenance and inspections of its internal state mobile infrastructure. The states were connected directly to their neighbours to formalize the connections between states, according to IBGE data.

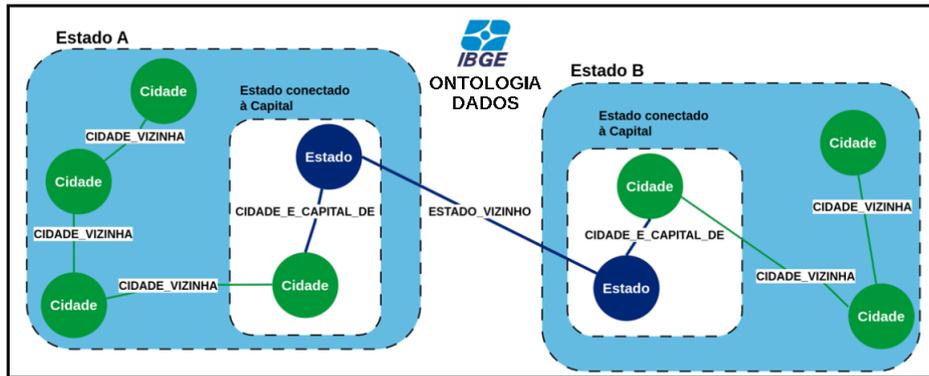


Figure 2: Conceptual model of the organization of geographic entities.

By way of illustration, Figure 3 shows the relationship between spatial data from ERBs. It helps illustrate the argument that integrating data from the nearest stations with their geographic position contributes to the study of improving the architecture of mobile telecommunications networks for the provision of services in the coverage area, with the geographic and geodetic conditions presented in the IBGE data [14]. Figure 3 (a) shows the geometric distance relationship between ERBs in Rio Branco, capital of Acre, and Figure 3 (b) shows the cartographic relationship between ERBs. Therefore, it is possible to visualize the concentration of ERBs in the east of the capital due to a large part of its area being covered by the Amazon Rainforest.

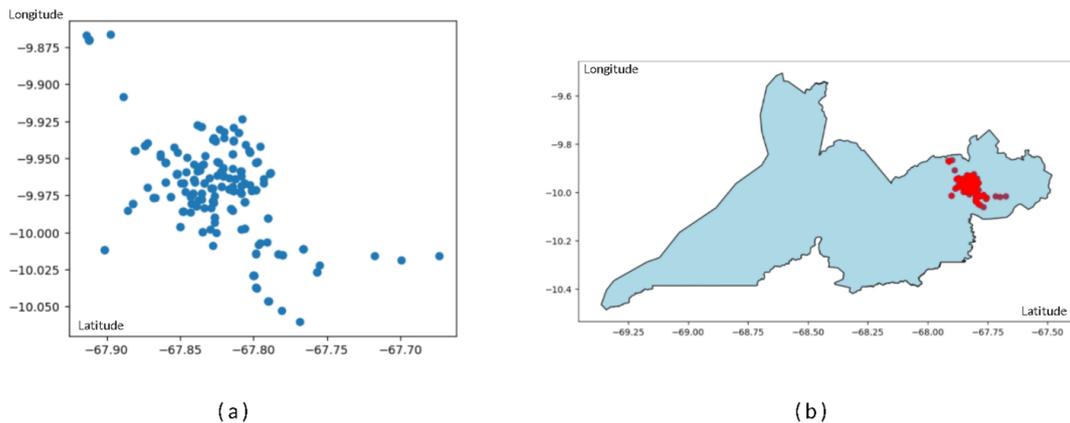


Figure 3: (a) Graphical representation of the cartographic distribution of ERBs in the metropolitan area of the East Zone of Rio Branco, and (b) overlay of these ERBs on the cartographic base of the capital.

3.3. Graph of knowledge

Based on the conceptual models presented, a large graph was constructed representing the MOSAICO system ERB data and IBGE data using Neo4J's CYPHER language. The Figure 4 shows part of the graph referring to the structure of ERB 64300. This figure shows the main mobile technologies: GSM (2G), WCDMA (3G), LTE (4G), and NR (5G); the antennas associated with them; equipment that operates their respective antennas; and which antennas are installed in a new project without legacy infrastructure (Greenfield). By looking at the graph, it is possible to understand the structure of how the ERB operates, how its antennas are connected, and how the spectrum of one antenna affects that of another, knowing at what power it operates.

Given its size and scope of representation, viewing the graph in its entirety is not very informative. For this reason, navigation through the graph is done through specific queries written in the CYPHER language.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4, Language Tools and Grammarly in order to: Grammar, spelling and plagiarism check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] IBGE, Cidades e estados do brasil, 2022. URL: <https://cidades.ibge.gov.br>, acesso em: 5 nov. 2024.
- [2] IBGE, Malhas territoriais, 2022. URL: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?edicao=36516&t=downloads>, acesso em: 11 out. 2024.
- [3] ANATEL, Mosaico - sistema integrado de gestão e controle do espectro, 2016. URL: <https://sistemas.anatel.gov.br/se/public/view/b/licenciamento.php?view=licenciamento>, acesso em 19 de outubro de 2024.
- [4] A. Singhal, Introducing the knowledge graph: Things not strings, 2012. URL: <https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>, 16/05/2012.
- [5] M. A. Rodriguez, The rdf virtual machine, Knowledge-Based Systems 24 (2011) 890–903. URL: <https://www.sciencedirect.com/science/article/pii/S0950705111000860>. doi:10.1016/j.knosys.2011.04.004.
- [6] J. F. Sowa, Knowledge representation: Logical, philosophical, and computational foundations, Brooks Cole, Pacific Grove, CA, 2000.
- [7] I. Basicovic, M. Popovic, N. Cetic, Ontology for ip telephony networks, ICDT 2013 (2013) 9.
- [8] Q. Zhou, A. J. Gray, S. McLaughlin, Toco: an ontology for representing hybrid telecommunication networks, in: The Semantic Web: 16th International Conference, ESWC 2019, Portorož, Slovenia, June 2–6, 2019, Proceedings 16, Springer, 2019, pp. 507–522.
- [9] Y. Huang, X. You, H. Zhan, S. He, N. Fu, W. Xu, Learning wireless data knowledge graph for green intelligent communications: Methodology and experiments, arXiv preprint arXiv:2404.10365v1 (2024). URL: <https://arxiv.org/abs/2404.10365v1>. doi:10.48550/ARXIV.2404.10365, submitted April 16, 2024.
- [10] F. Aranha, Sistema de informação geográfica: Uma arma estratégica para o database marketing, Fundação Getúlio Vargas 36 (1996) 12–16. URL: <https://doi.org/10.1590/s0034-75901996000200003>. doi:10.1590/s0034-75901996000200003.
- [11] A. N. de Telecomunicações (Anatel), Glossário de termos da anatel: Estação rádio base (erb), 2022. URL: <https://informacoes.anatel.gov.br/legislacao/glossario?catid=5&faqid=2902>, acesso em 1 de julho de 2025.
- [12] Agência Nacional de Telecomunicações (Anatel), Ato nº 9727, de 06 de junho de 2022: Aprova os requisitos técnicos para avaliação da conformidade de equipamentos dos sistemas de acesso fixo sem fio para a prestação do serviço telefônico fixo comutado – stfc, <https://informacoes.anatel.gov.br/legislacao/atos-de-certificacao-de-produtos/2022/1681-ato-9727>, 2022. Publicado em 8 de julho de 2022; última atualização em 25 de outubro de 2022.
- [13] Agência Nacional de Telecomunicações (Anatel), Portaria Anatel nº 2453, de 6 de setembro de 2022: Aprova o procedimento de fiscalização para verificação do cumprimento dos compromissos de abrangência e da Área de cobertura do serviço móvel pessoal (smp), <https://informacoes.anatel.gov.br/legislacao/procedimentos-de-fiscalizacao/1724-portaria-2453>, 2022. Publicado em 9 de setembro de 2022; última atualização em 26 de novembro de 2024.
- [14] D. N. Tavares, Abordagem para análise de dados da infraestrutura de telecomunicação móvel do Brasil baseada em grafo de conhecimento, Master's thesis, Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, SP, Brasil, 2025.