REVISTA TÓPICOS
ISSN 2965-6672 | Qualis A2

CROSS-SOURCE GEOSPATIAL RECONCILIATION OF SHARED UTILITY POLE INVENTORIES FOR THE DETECTION OF UNAUTHORIZED TELECOMMUNICATION ATTACHMENTS: A CASE STUDY FROM A BRAZILIAN DISTRIBUTION NETWORK

RECONCILIAÇÃO GEOESPACIAL ENTRE FONTES DE INVENTÁRIOS DE POSTES COMPARTILHADOS PARA A DETECÇÃO DE OCUPAÇÕES DE TELECOMUNICAÇÕES NÃO AUTORIZADAS: UM ESTUDO DE CASO EM UMA REDE DE DISTRIBUIÇÃO BRASILEIRA

Engenharias • 09/08/2026

REGISTRO DOI: 10.70773/revistatopicos/786321036

Fabio Guerra[1]

Ana Bonafin[2]

Dierli Maschio[3]

João Jacobsen[4]

Kallil Zielinski[5]

Flavio Grando[6]

Vinicius Parede[7]

Vanderlei Silva[8]

Milton Ramos[9]

Fabricio Salmazo[10]

Adriane Fuhr[11]

Mariana Santini[12]

Thais Lazaro[13]

Leonardo Scabini[14]

ABSTRACT

Shared utility poles are economically significant: one pole owned by an electric utility may carry cables of several telecommunication operators under joint-use agreements. In practice, the owner's records diverge across the billing and geographic-operations systems, and many attachments exist with no contractual record — a phenomenon known locally as *revelia*, analogous to non-technical losses in electricity. We present a reproducible methodology that reconciles two heterogeneous enterprise inventories of the same pole population by coordinate normalization and nearest-neighbor spatial matching, characterizes the discrepancies under the ISO 19157 dimensions of completeness and positional accuracy, and tests whether pole congestion (cables per pole) predicts unauthorized attachments. Applied to the metropolitan network of Curitiba, Brazil (about one million attachment records per source), the pipeline matches 144,903 poles across the two systems and exposes a systematic ≈6 m positional offset between them. It surfaces more than 1.9×10^5 field-flagged unauthorized attachments across 82 operators and shows, via a Mann–Whitney U test ($p < 0.001$; median 8 versus 5 cables per pole), that congested poles are markedly more likely to host them. At the reference joint-use tariff, the billing gap corresponds to an estimated R$ 1.44 million per month. The pipeline is both a quality-assurance instrument for the geographic base and a prioritization tool for revenue-recovery inspections.

Keywords: Shared utility poles; Joint use; Unauthorized attachments; Geospatial entity resolution; Asset inventory quality.

RESUMO

Os postes compartilhados têm relevância econômica significativa: um único poste de propriedade de uma distribuidora de energia elétrica pode sustentar os cabos de diversas operadoras de

telecomunicações sob acordos de uso mútuo. Na prática, os registros da proprietária divergem entre os sistemas de faturamento e de operação geográfica, e muitas ocupações existem sem qualquer registro contratual — fenômeno conhecido localmente como *revelia*, análogo às perdas não técnicas de energia elétrica. Apresenta-se uma metodologia reprodutível que reconcilia dois inventários corporativos heterogêneos da mesma população de postes por meio da normalização de coordenadas e do pareamento espacial por vizinho mais próximo, caracteriza as discrepâncias sob as dimensões de completude e acurácia posicional da norma ISO 19157 e verifica se o congestionamento do poste (cabos por poste) prediz ocupações não autorizadas. Aplicado à rede metropolitana de Curitiba, Brasil (cerca de um milhão de registros de ocupação por fonte), o fluxo de trabalho pareia 144.903 postes entre os dois sistemas e revela um deslocamento posicional sistemático de aproximadamente 6 m entre eles. Identifica mais de $1{,}9 \times 10^5$ ocupações não autorizadas sinalizadas em campo, distribuídas entre 82 operadoras, e demonstra, por meio do teste U de Mann–Whitney ($p < 0{,}001$; mediana de 8 contra 5 cabos por poste), que postes congestionados são marcadamente mais propensos a abrigá-las. À tarifa de referência de uso mútuo, a lacuna de faturamento corresponde a uma estimativa de R$ 1,44 milhão por mês. O fluxo de trabalho constitui simultaneamente um instrumento de garantia de qualidade da base geográfica e uma ferramenta de priorização de inspeções voltadas à recuperação de receita.
Palavras-chave: Postes compartilhados; Uso mútuo; Ocupações não autorizadas; Resolução de entidades geoespaciais; Qualidade de inventário de ativos.

1. INTRODUCTION

The utility pole is a deceptively humble piece of infrastructure: a single wooden or concrete structure, owned by an electricity distribution company, that simultaneously supports the medium- and low-voltage power network and, through joint-use agreements, the access cables of multiple telecommunication operators. In a country the size of Brazil this translates into tens of millions of shared poles; the partner utility of this study alone maintains more than 3.3 million poles across its concession area. The shared use of this infrastructure is regulated: in Brazil the joint occupation of poles by power and telecommunication operators is governed by a joint resolution of the electricity, telecommunications and petroleum agencies (Aneel; Anatel; ANP, 1999), and the rental charged per attachment point follows a reference tariff established jointly by the electricity and telecommunications regulators (Aneel; Anatel, 2014). Accurate accounting of attachments therefore underpins both safety in the public space — each pole has a finite mechanical capacity, conventionally limited to a handful of cables — and a non-negligible recurring revenue stream for the pole owner.

Yet, as field crews and regulators repeatedly observe, the inventories that describe *which* attachments exist on *which* pole are seldom in agreement: the geographic information system (GIS/GEO) maintained by network operations, the customer-information system (CIS) used by the commercial area for billing, and the physical reality on the pole are three distinct views of the same population, and each carries its own positional and attribute errors. Beyond clerical drift, an explicit form of non-compliance further erodes the inventory: cables installed by operators that have not signed, or have ceased to honor, a sharing contract — referred to in Brazil as *revelia* and conceptually analogous to non-technical losses (NTL) in electricity (Coma-Puig; Carmona, 2022; Messinis;

Hatziargyriou, 2018). Because the pole owner cannot bill what it does not know exists, *revelia* couples a public-safety hazard (overloaded poles) to a direct revenue leakage, and its mitigation has historically relied on costly, exhaustive photogrammetric field surveys rather than on the data the utility already holds.

The present work is motivated by the operational question that follows from this gap: given two imperfect digital views of the same pole population, can a principled reconciliation procedure not only quantify their discrepancies but also surface, in a statistically defensible way, the poles where unauthorized attachments are most likely to be found? Our central hypothesis is that pole congestion — the number of cables already attached to a structure — is a strong signal for the presence of *revelia*, and that this signal can be exposed by combining geospatial entity resolution (Sehgal; Getoor; Viechnicki, 2006; Balsebre et al., 2022) with a non-parametric test on the resulting reconciled dataset; whether the signal generalizes beyond the single network studied here is a question our case study begins, but cannot fully settle, to answer. We validate this hypothesis on the metropolitan distribution network of Curitiba, Brazil, and show that the resulting pipeline can be deployed both as a quality-assurance instrument for the GEO base itself and as a prioritization tool for revenue-recovery field campaigns. Concretely, our contributions are: (i) the formulation of unauthorized-attachment detection on shared poles as a geospatial entity-resolution and data-quality problem; (ii) a reproducible two-stage pipeline that reconciles the GEO and CIS inventories and characterizes their discrepancies under the ISO 19157 quality dimensions (ISO, 2013); (iii) a statistically validated congestion signal for prioritizing inspections; and (iv) a quantification of the associated revenue leakage on a real, city-scale dataset.

2. RELATED WORK

Our contribution sits at the intersection of three traditionally separate communities, reviewed in turn below.

2.1. Geospatial Entity Resolution And Conflation

The matching of objects across heterogeneous spatial datasets has been studied for decades under the labels of *conflation*, *entity resolution* and *spatial entity linkage*. The problem inherits the probabilistic foundations of classical record linkage (Fellegi; Sunter, 1969; Christen, 2012) and was adapted to geographic data through automated map conflation (Saalfeld, 1988), in which positional, attribute and topological evidence are combined to decide whether two features denote the same real-world object. Foundational techniques explicitly combining positional, attribute and contextual similarity for geospatial integration were introduced by Sehgal, Getoor and Viechnicki (2006) and the space of measures and methods for matching geospatial vector datasets was surveyed comprehensively by Xavier, Ariza-López and Ureña-Cámara (2017). More recent work has moved toward learned models, particularly for points of interest (Deng et al., 2019) and for the joint embedding of spatial and textual attributes (Balsebre et al., 2022), while the broader entity-resolution literature has matured into end-to-end pipelines with dedicated blocking and filtering stages to make matching tractable at scale (Papadakis et al., 2020; Christophides et al., 2021). Efficient nearest-neighbor search over large point sets — the computational primitive at the core of our matching stage — rests on classical spatial-index structures such as *k*-d trees (Bentley, 1975) and R-trees (Guttman, 1984).

2.2. Spatial Data Quality And The ISO 19157 Framework

The systematic assessment of spatial data quality is a mature field (Devillers; Jeansoulin, 2006; Veregin, 1999), formalized internationally by the ISO 19157 standard (ISO, 2013) around the dimensions of completeness, logical consistency, positional accuracy, temporal accuracy and thematic accuracy. These dimensions provide the vocabulary we adopt for characterizing inventory discrepancies, and they connect naturally to the literature on assessing the quality of large, heterogeneously sourced geographic databases (Goodchild; Li, 2012). Positional comparison further presupposes a common geodetic reference; in Brazil this is the SIRGAS 2000 system, the official datum since 2015 (Sánchez; Brunini, 2009; IBGE, 2015), which we use throughout.

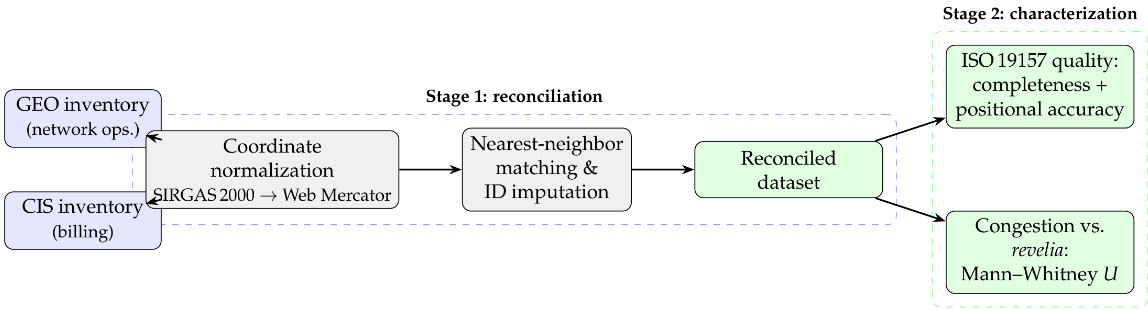
2.3. Non-Technical Losses And Infrastructure Auditing

The detection of non-technical losses (NTL) in electricity distribution is a mature line of work that frames the unauthorized consumption of a regulated resource as a class-imbalanced anomaly-detection problem and has produced a wide range of statistical and machine-learning approaches (Coma-Puig; Carmona, 2022; Nagi et al., 2010), comprehensively reviewed in Viegas et al. (2017), Glauner et al. (2017) and Messinis and Hatziargyriou (2018). The role of geographic information systems specifically in the management of jointly used utility infrastructure has long been recognized in the practitioner literature (McCoy, 2005), and recent work in the same utility setting has combined robotic process automation with artificial intelligence to support distribution-network operations (Pedretti et al., 2021). To the best of our knowledge, however, the problem of detecting unauthorized *telecommunication* attachments on shared utility poles has not been formulated as a geospatial entity-resolution problem in the peer-reviewed literature, and the existing trade press on pole audits relies almost exclusively on photogrammetric field surveys rather than on the reconciliation of data the utility already possesses. The present work bridges these three threads: it applies geospatial entity resolution to two enterprise inventories, characterizes the result under ISO 19157, and reframes *revelia* as a statistically testable, NTL-like property of the reconciled data.

3. MATERIALS AND METHODS

We frame the problem as a two-stage pipeline (Figure 1) operating on the two enterprise inventories that the pole-owning utility maintains for the same physical population. The first stage performs cross-source reconciliation; the second operates on the reconciled dataset to test our central hypothesis. The full implementation — data preprocessing, spatial joins, statistical tests and the production of operator-level and spatial summaries — is built on the open-source scientific Python stack (pandas (McKinney, 2010), NumPy (Harris et al., 2020), GeoPandas (Jordahl et al., 2020) with Shapely geometries, and SciPy (Virtanen et al., 2020) for the statistical tests); the implementation is available from the authors upon reasonable request (see the Data Availability Statement).

Figure 1 – The two-stage reconciliation-and-detection pipeline. Stage 1 resolves the GEO and CIS inventories into a common, matched pole population; Stage 2 characterizes the residual discrepancies under ISO 19157 and tests whether pole congestion predicts unauthorized attachments.



Source: prepared by the authors.

3.1. Study Area And Data Sources

We study the metropolitan distribution network of Curitiba, the capital of the state of Paraná in southern Brazil, served by the partner utility. Curitiba is the natural reference site for this work because, unlike most municipalities in the concession, essentially its entire pole population has already been visited under the utility's three-year cyclic field-inspection program — over 99.9% of records carry an “inspected” status — so the GEO base is as complete as operational data allows and the residual discrepancies are attributable to genuine inventory defects rather than to unvisited poles.

Two enterprise inventories describe the same physical pole population. The *geographic system* (GEO), maintained by network operations, stores one record per attachment, indexed by a numeric pole identifier (num_geo_ps, the primary key), georeferenced coordinates, the operator acronym occupying the point, and the cable type. The *customer-information system* (CIS), maintained by the commercial area, stores the contractual attachments billed to each operator together with the per-point rental value. Table 1 summarizes the key fields used by the pipeline. Crucially, unauthorized attachments are flagged *only* in the GEO base: when a field revision detects an operator occupying a point with no contract, the operator acronym is recorded with a REV_ prefix followed by an internal numeric operator code, a marker that by construction never appears in the billing system. We adopt this REV_ flag as the ground-truth label for *revelia*.

Table 1. Key fields of the two enterprise inventories used by the reconciliation pipeline.

System	**Field**	**Description**
GEO	*num_geo_ps*	Pole identifier (primary key), shared with CIS
GEO	*coordenadas*	Pole coordinates (latitude, longitude), SIRGAS 2000
GEO	*sigla_empresa*	Operator acronym; REV_ prefix flags an unauthorized (*revelia*) attachment
GEO	*tipo_uso*	Cable/use type (optical fiber; cables and strand; energized/de-energized equipment)
GEO	*data_confe*	Date of field confirmation/inspection
CIS	*num_geo_ps*	Pole identifier, used to join to GEO
CIS	*coordenadas*	Pole coordinates, used to impute missing identifiers by spatial matching
CIS	*CNPJ*	Operator tax identifier, used to attribute billed points
CIS	*pontos_fix.*	Number of contracted (billed) attachment points
CIS	*valor_unit.*	Reference rental value per attachment point

Source: prepared by the authors.

After cleaning (removal of records lacking a pole identifier or coordinates), the Curitiba extract comprises roughly one million attachment records per source; the corresponding distinct-pole and operator counts are reported in Table 2. Coordinates are stored as strings in the SIRGAS 2000 geographic reference (EPSG:4674); for any computation expressed in meters — nearest-neighbor distances, positional residuals — geometries are reprojected to Web Mercator (EPSG:3857).

Table 2. Dimensions of the Curitiba dataset after cleaning, by source inventory.

Quantity	**GEO**	**CIS**
Attachment records (rows)	1,048,001	1,039,754
Distinct poles (num_geo_ps)	145,593	151,029
Distinct operators (combined)	234	
Records with inspected status	>99.9%	
Optical-fiber share of cables	48.1%	

Source: prepared by the authors.

3.2. Stage 1: Cross-Source Reconciliation

Reconciliation proceeds in three steps. First, the latitude/longitude strings of both inventories are parsed into point geometries and the resulting layers are placed in a common coordinate reference system. Second, the few CIS records whose pole identifier is missing are repaired by nearest-neighbor imputation: each orphan point is matched to the closest pole in the GEO base via a spatial nearest join (an R-tree-indexed search (Guttman, 1984)), inheriting its identifier when the match lies within the local positional tolerance. Third, with identifiers harmonized, the two pole populations are compared by set operations on num_geo_ps, and the result is classified under the ISO 19157 (ISO, 2013) quality dimensions: poles present in one inventory but absent from the other are *completeness* defects (omission/commission), while paired poles whose coordinates disagree are *positional-accuracy* defects. For the latter we quantify the displacement between the two recorded positions of each matched pole; nearest-neighbor distances also bound the positional regime within which the identifier imputation above is admissible.

3.3. Stage 2: Detection Of Unauthorized Attachments

The second stage operates on the reconciled dataset to test our central hypothesis. We define the *congestion* of a pole as its number of attached cables, obtained by aggregating attachment records per num_geo_ps, and we treat the GEO REV_ flag as the ground-truth label for unauthorized attachment. The test is run over the reconciled attachment table: every pole contributes its congestion value, labelled *with revelia* once for each unauthorized attachment it carries and *without revelia* if it carries none. This construction yields 191,201 flagged and 49,807 unflagged records (241,008 in total) and, by design, weights each pole in the flagged group by its number of unauthorized cables; Section 5 revisits this choice and a pole-level re-estimation as a robustness check. Because a Shapiro–Wilk test (Shapiro; Wilk, 1965) decisively rejects normality of both congestion distributions (Section 4.3), we use the non-parametric Mann–Whitney *U* test (Mann; Whitney, 1947; Wilcoxon, 1945; Hollander; Wolfe; Chicken, 2013) to assess whether congestion is stochastically larger in the flagged group, reporting the rank-based common-language effect size alongside the test statistic. We complement the test with operator-level aggregation — the number of poles affected by each REV_ operator — and with an analysis of the spatial distribution of the flagged poles relevant to inspection-route planning.

4. RESULTS

4.1. Inventory Reconciliation And Data Quality

Once identifiers are harmonized, 144,903 poles are matched across the two inventories. The completeness analysis isolates 6,126 poles present in the CIS but absent from the GEO base and 690 poles present in the GEO base but absent from the CIS, against a net difference of 5,436 between the two distinct-pole counts (Table 3). The asymmetry is informative: omissions in the GEO base (the larger group) are the more consequential, since the GEO base is the operational source of truth used to plan inspections. Figure 2 projects the three populations geographically; the matched poles (green) blanket the built-up area, while the unmatched poles (red and blue) form thin, edge-of-network and corridor-following patterns rather than uniform noise, consistent with pole churn from street works and with projects historically mis-attributed across municipal boundaries.

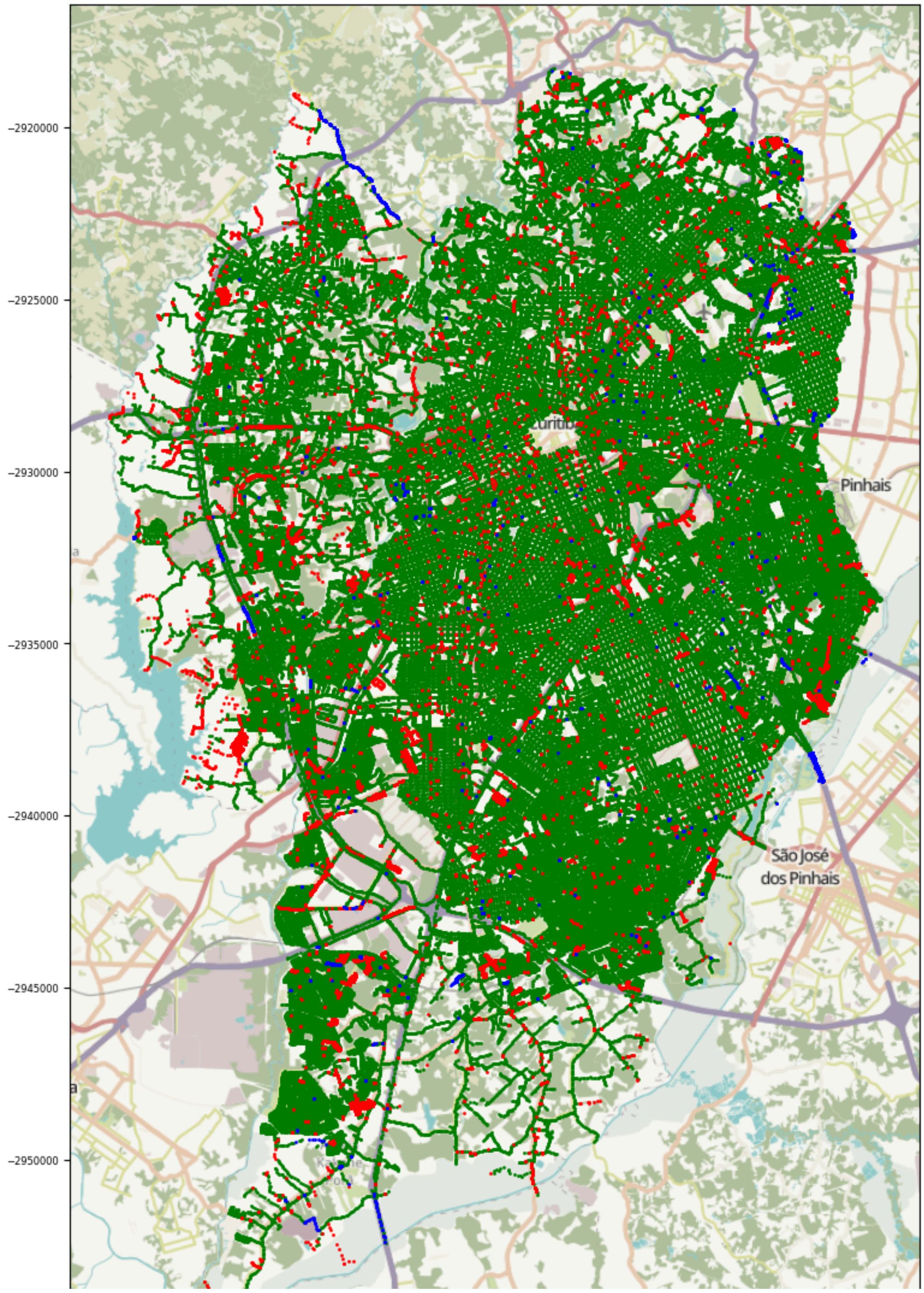
Positional accuracy is dominated by a *systematic* effect: matched poles exhibit a near-uniform displacement of approximately 6 m between the two recorded positions, with nearest-neighbor residuals for the harder, identifier-imputed matches spanning roughly 6–33 m. A single fixed matching tolerance is therefore inadequate; the tolerance must track the local positional-accuracy regime, a point we return to in Section 5.

Table 3. Cross-source reconciliation summary for the Curitiba pole population, organized by ISO 19157 quality dimension.

ISO 19157 dimension	Indicator	Interpretation	Value
Completeness	Matched poles	present in both GEO and CIS	144,903
Completeness	Omission (GEO)	in CIS, absent from GEO	6,126
Completeness	Omission (CIS)	in GEO, absent from CIS	690
Completeness	Net pole-count gap	|CIS| − |GEO|	5,436
Positional accuracy	Systematic offset	median displacement, matched poles	≈6 m
Positional accuracy	Residual range	nearest-neighbor distances, imputed matches	6–33 m

Source: prepared by the authors.

Figure 2. Geographic projection of the cross-source reconciliation over the Curitiba metropolitan area. **Green:** poles present in both inventories (144,903); **red:** poles in the CIS but absent from the GEO base (6,126); **blue:** poles in the GEO base but absent from the CIS (690). Coordinates are in Web Mercator (EPSG:3857) meters; the vertical-axis tick labels are shown.

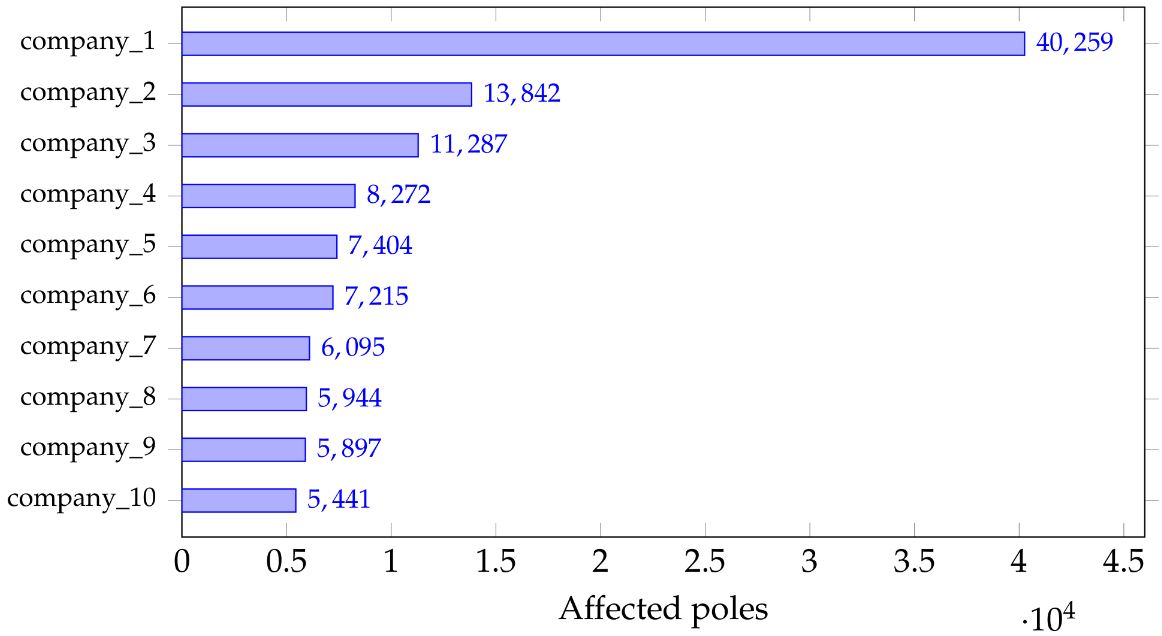


Source: prepared by the authors. Basemap © OpenStreetMap contributors.

4.2. Unauthorized Attachments Across Operators

The reconciled dataset contains more than 1.9×10^5 attachment records carrying the REV_ flag, attributable to 82 distinct telecommunication operators. The distribution across operators is strongly heavy-tailed (Figure 3): a single operator (company_1, the *revelia* record of a large national carrier) accounts for 40,259 affected poles, nearly three times the second-ranked operator, while the long tail comprises dozens of small regional providers, the smallest of which appear on only a handful of poles. The ten largest operators alone account for more than 1.1×10^5 operator–pole incidences, confirming that a small number of actors concentrate a disproportionate share of the unauthorized attachments and that an enforcement effort targeting them would capture most of the leakage.

Figure 3. The ten telecommunication operators with the largest number of poles carrying an unauthorized (*revelia*) attachment in the Curitiba dataset. Because operator identities are commercially and legally sensitive, they are replaced by neutral placeholders (company_1, company_2, ...), ordered by decreasing incidence. The distribution is heavy-tailed: the single largest operator accounts for more affected poles than the next three combined.



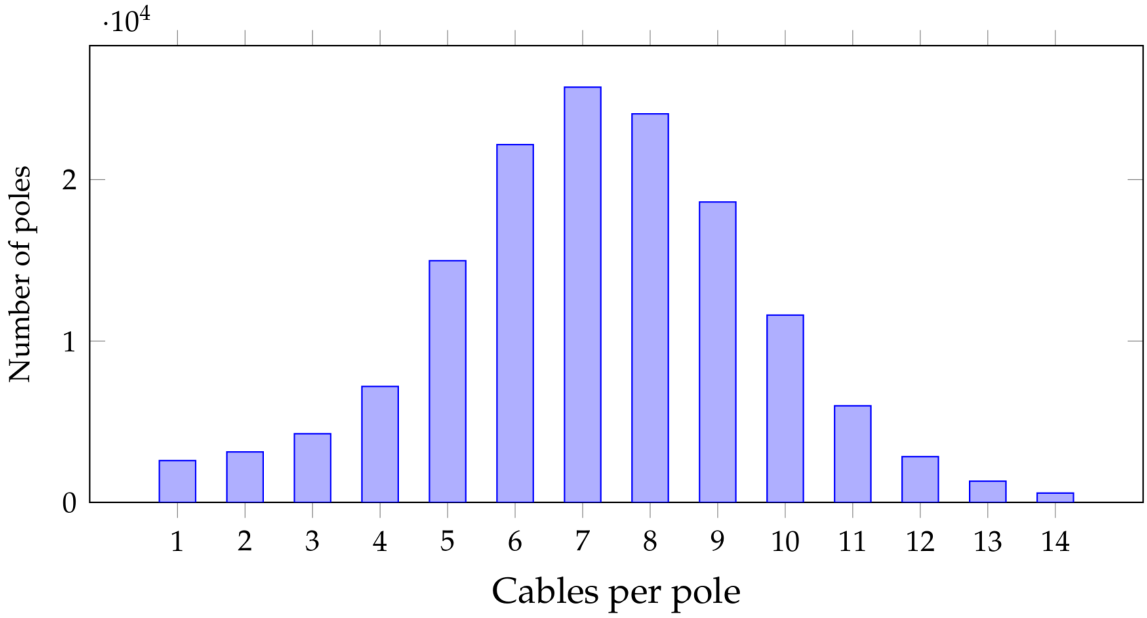
Source: prepared by the authors.

4.3. Congestion as a Signal For Unauthorized Attachments

The number of cables per pole in the reference inventory is itself concentrated in the congested regime (Figure 4): the modal pole carries 7 cables and the bulk of the population lies between 5 and 10, well above the conventional structural rule-of-thumb of a few cables per pole, underscoring the safety relevance of congestion independently of billing. Partitioning poles by the presence of *revelia* reveals a clear shift (Figure 5): poles bearing unauthorized attachments have a median of 8 cables, against a median of 5 for poles without, with the entire interquartile range displaced upward.

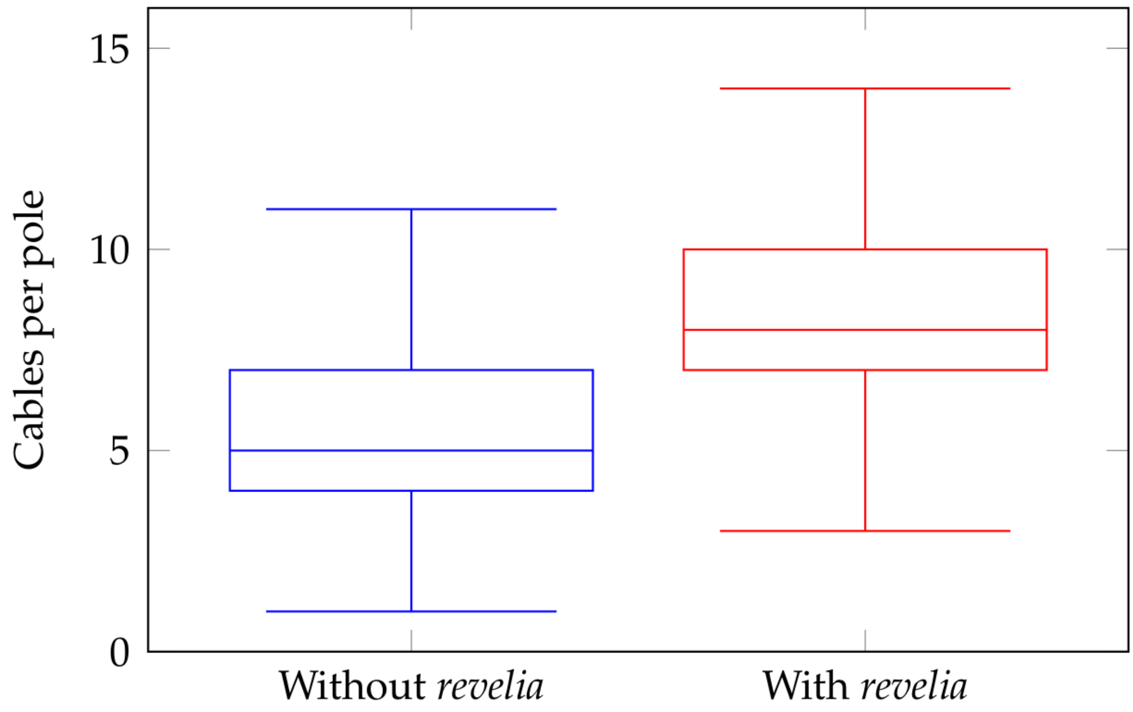
A Shapiro–Wilk test rejects normality of the congestion distribution for both groups ($W = 0.973$ for records with *revelia* and $W = 0.965$ for those without; $p < 10^{-70}$ in both cases), so we test the shift with the non-parametric Mann–Whitney U test. The test is unambiguous (Table 4): the congestion of poles with unauthorized attachments is stochastically larger than that of poles without, with $U = 7.70 \times 10^9$ and a p-value below the floating-point precision of the implementation ($p < 0.001$). Beyond significance — which is near-guaranteed at this sample size — the effect is large in magnitude: the common-language effect size $f = U/(n_1 n_2) \approx 0.81$ means that a randomly chosen flagged record sits on a more congested pole than a randomly chosen unflagged one about four times out of five. Pole congestion is therefore a strong, statistically defensible predictor of *revelia*, available at no additional sensing cost from data the utility already holds.

Figure 4. Distribution of cables per pole over the Curitiba GEO inventory. The population is concentrated in the congested regime (5–10 cables), with a mode at 7. The 523 poles carrying more than 14 cables (<0.4% of the population) are omitted from the plot for readability.



Source: prepared by the authors.

Figure 5. Distribution of pole congestion by the presence of unauthorized attachments. Poles bearing *revelia* (median 8 cables) are markedly more congested than those without (median 5); the difference is highly significant under a Mann–Whitney *U* test (Table 4). Whiskers extend to the most extreme values within 1.5× the interquartile range; individual outliers beyond the whiskers are omitted.



Source: prepared by the authors.

Table 4. Statistical comparison of pole congestion between poles with and without unauthorized attachments.

Quantity	Without *revelia*	With *revelia*
Records (n)	49,807	191,201
Median cables per pole	5	8
Interquartile range	[4, 7]	[7, 10]
Shapiro–Wilk W (p)	0.965 ($<10^{-70}$)	0.973 ($<10^{-70}$)
Mann–Whitney $U = 7.70 \times 10^9$; $p < 0.001$ (one-sided, *revelia* stochastically larger)		

Source: prepared by the authors.

4.4. Spatial Structure And Economic Impact

Projected geographically, the flagged poles tend to cluster at the neighborhood scale rather than scatter uniformly, a qualitative pattern consistent with the service footprints of individual operators (a per-operator choropleth is withheld here to preserve operator confidentiality). This structure is directly exploitable for inspection-route planning, since a crew dispatched to a congested cluster can expect a high yield of unauthorized attachments per visit. Finally, the same reconciliation supports a direct economic reading. Comparing, per operator, the number of cables recorded in the GEO base against the number actually billed in the CIS and valuing the gap at the reference joint-use tariff of R$ 7.50 per attachment point yields an estimated R$ 1.44 million per month of unbilled attachments over the study area, with a single large carrier — distinct from company_1 of Figure 3, and whose gap stems chiefly from under-reported billed attachments rather than from formally flagged *revelia* — responsible for roughly R$ 0.9 million of that total. Whether arising from formally flagged *revelia* or from mere under-reporting, this gap quantifies the revenue at stake and provides the business case for the prioritized inspection campaigns that the pipeline enables.

5. DISCUSSION

The findings carry three interlocking implications. From a methodological standpoint, they show that the geospatial entity-resolution toolbox (Sehgal; Getoor; Viechnicki, 2006; Xavier; Ariza-López; Ureña-Cámara, 2017; Balsebre et al., 2022), originally developed for points of interest and cartographic features, transfers cleanly to the reconciliation of utility-asset inventories and yields actionable quality metrics under the ISO 19157 framework (ISO, 2013) when adapted to the dense, identifier-rich setting of shared poles. From an operational standpoint, the strong association between pole congestion and unauthorized attachments ($U = 7.70 \times 10^9$, $p <$

0.001; median 8 vs. 5 cables) translates directly into a prioritization rule that field crews can apply with no additional sensing: targeting the most congested poles in a given region maximizes the expected yield of any audit campaign, an effect amplified by the heavy-tailed operator distribution of Figure 3 and the neighborhood-level clustering of Section 4.4. From a regulatory and economic standpoint, the same signal redefines the discussion of revenue leakage in shared-infrastructure ecosystems: the estimated R$ 1.44 million per month of unbilled attachments recasts *revelia*, in the spirit of the non-technical-losses literature in electricity (Coma-Puig; Carmona, 2022; Messinis; Hatziargyriou, 2018), as a quantifiable and statistically testable property of the inventory rather than a diffuse compliance concern, with a direct link to the reference-tariff mechanism of the sector regulators (Aneel; Anatel, 2014).

Several limitations temper these conclusions and frame future work. First, as Section 4.1 shows, positional accuracy is governed by a local regime — a systematic ≈6 m offset overlaid with larger residuals — so a single fixed matching tolerance is inadequate and the imputation of missing identifiers should be calibrated per region; an adaptive tolerance, or a probabilistic matcher in the Fellegi–Sunter tradition (Fellegi; Sunter, 1969), would make the reconciliation more robust. Second, the *revelia* flag is itself a product of field-revision practice and of operator-side reporting, so it is best read as a high-precision but not necessarily high-recall label; poles never revised may harbor undetected unauthorized attachments, which would make our congestion signal a conservative one. Third, the hypothesis test is computed on the reconciled attachment table, in which poles bearing several unauthorized cables contribute proportionally more records to the *revelia* group; while this does not affect the direction or the overwhelming significance of the result, a pole-level re-estimation is a natural robustness check. Finally, the case study is confined to a single, fully inspected metropolitan network chosen precisely for its data maturity; the extent to which the congestion signal and the ≈6 m positional regime generalize to rural feeders and to less thoroughly inspected concessions remains to be validated.

6. CONCLUSION

We have presented a reconciliation-and-detection pipeline that turns two imperfect enterprise inventories of shared utility poles into a quality-assured dataset and a statistically defensible prioritization tool for the detection of unauthorized telecommunication attachments. The methodology is principled in the sense that each of its stages is grounded in a well-established line of work — geospatial entity resolution, ISO 19157-aligned data-quality assessment, and non-parametric hypothesis testing — and is operational in the sense that all of its outputs are directly consumable by the planning, commercial and field-inspection areas of a pole-owning utility. Applied to the Curitiba metropolitan network, it matched 144,903 poles across the two systems, characterized their completeness and positional discrepancies under ISO 19157, and provided quantitative evidence that pole congestion is a strong and significant indicator of *revelia* (median 8 vs. 5 cables; Mann–Whitney $p < 0.001$) across 82 operators and more than 1.9×10^5 flagged attachments, corresponding to an estimated R$ 1.44 million per month of unbilled use. The pipeline is expressed entirely in terms of standard, source-agnostic operations, making it reproducible on any pair of CIS-style and GEO-style inventories. Future work will extend the pipeline along three axes. First, temporal information will be incorporated to track how discrepancies evolve between successive inventory snapshots. Second, the binary congestion signal will be replaced by a learned anomaly score that aggregates additional contextual features, notably the network topology obtained by modeling the feeder as a graph (Hagberg; Schult; Swart, 2008) aligned to the surrounding street network (Boeing, 2017), allowing sequences of unauthorized attachments to be classified rather than scored pole by pole. Third, the prioritization rule will be validated prospectively through controlled field-inspection campaigns in collaboration with the partner utility.

ACKNOWLEDGEMENTS AND FUNDING

This research was carried out under the Research and Development Program of the Brazilian Electricity Regulatory Agency (ANEEL/PROPDI) (Aneel, 2022), project PD-02866-0544/2023.

REFERENCES

ANEEL – AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Procedimentos do Programa de Pesquisa, Desenvolvimento e Inovação – PROPDI.** Brasília, DF: Aneel, 2022. Aprovado pela Resolução Normativa Aneel n. 1.045, de 4 de outubro de 2022.

ANEEL; ANATEL; ANP. **Resolução Conjunta n. 001, de 24 de novembro de 1999:** aprova o Regulamento Conjunto para Compartilhamento de Infraestrutura entre os Setores de Energia Elétrica, Telecomunicações e Petróleo. Brasília, DF: Aneel; Anatel; ANP, 1999.

ANEEL; ANATEL. **Resolução Conjunta n. 4, de 16 de dezembro de 2014:** aprova o preço de referência para o compartilhamento de postes entre distribuidoras de energia elétrica e prestadoras de serviços de telecomunicações e estabelece regras de uso e ocupação dos pontos de fixação. Brasília, DF: Aneel; Anatel, 2014. Disponível em: https://informacoes.anatel.gov.br/legislacao/resolucoes/resolucoes-conjuntas/820-resolucaoconjunta-4. Acesso em: 15 jul. 2025.

BALSEBRE, Pasquale *et al.* Geospatial entity resolution. In: THE ACM WEB CONFERENCE, 2022, Lyon. **Proceedings of the ACM Web Conference 2022.** New York: Association for Computing Machinery, 2022. p. 3061-3070. DOI: 10.1145/3485447.3512026.

BENTLEY, Jon Louis. Multidimensional binary search trees used for associative searching. **Communications of the ACM,** New York, v. 18, n. 9, p. 509-517, 1975. DOI: 10.1145/361002.361007.

BOEING, Geoff. OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. **Computers, Environment and Urban Systems,** Amsterdam, v. 65, p. 126-139, 2017. DOI: 10.1016/j.compenvurbsys.2017.05.004.

CHRISTEN, Peter. **Data matching:** concepts and techniques for record linkage, entity resolution, and duplicate detection. Berlin: Springer, 2012. (Data-Centric Systems and Applications). DOI: 10.1007/978-3-642-31164-2.

CHRISTOPHIDES, Vassilis *et al.* An overview of end-to-end entity resolution for big data. **ACM Computing Surveys,** New York, v. 53, n. 6, p. 1-42, 2021. DOI: 10.1145/3418896.

COMA-PUIG, Bernat; CARMONA, Josep. Non-technical losses detection in energy consumption focusing on energy recovery and explainability. **Machine Learning,** Berlin, v. 111, n. 2, p. 487-517, 2022. DOI: 10.1007/s10994-021-06051-1.

DENG, Yue *et al.* Point of interest matching between different geospatial datasets. **ISPRS International Journal of Geo-Information,** Basel, v. 8, n. 10, p. 435, 2019. DOI: 10.3390/ijgi8100435.

DEVILLERS, Rodolphe; JEANSOULIN, Robert (ed.). **Fundamentals of spatial data quality.** London: ISTE, 2006. DOI: 10.1002/9780470612156.

FELLEGI, Ivan P.; SUNTER, Alan B. A theory for record linkage. **Journal of the American Statistical Association,** Alexandria, v. 64, n. 328, p. 1183-1210, 1969. DOI: 10.1080/01621459.1969.10501049.

GLAUNER, Patrick *et al.* The challenge of non-technical loss detection using artificial intelligence: a survey. **International Journal of Computational Intelligence Systems,** Amsterdam, v. 10, n. 1, p. 760-775, 2017. DOI: 10.2991/ijcis.2017.10.1.51.

GOODCHILD, Michael F.; LI, Linna. Assuring the quality of volunteered geographic information. **Spatial Statistics,** Amsterdam, v. 1, p. 110-120, 2012. DOI: 10.1016/j.spasta.2012.03.002.

GUTTMAN, Antonin. R-trees: a dynamic index structure for spatial searching. **ACM SIGMOD Record,** New York, v. 14, n. 2, p. 47-57, 1984. DOI: 10.1145/971697.602266.

HAGBERG, Aric A.; SCHULT, Daniel A.; SWART, Pieter J. Exploring network structure, dynamics, and function using NetworkX. In: PYTHON IN SCIENCE CONFERENCE, 7., 2008, Pasadena. **Proceedings of the 7th Python in Science Conference.** Pasadena: SciPy, 2008. p. 11-15.

HARRIS, Charles R. *et al.* Array programming with NumPy. **Nature,** London, v. 585, p. 357-362, 2020. DOI: 10.1038/s41586-020-2649-2.

HOLLANDER, Myles; WOLFE, Douglas A.; CHICKEN, Eric. **Nonparametric statistical methods.** 3. ed. Hoboken: Wiley, 2013. DOI: 10.1002/9781119196037.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Resolução do Presidente n. 01/2015:** define a data de término do período de transição definido na RPR 01/2005 e dá outras providências sobre a transformação entre os referenciais geodésicos adotados no Brasil. Rio de Janeiro: IBGE, 2015.

ISO – INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19157:2013 geographic information – data quality.** Geneva: ISO/TC 211, 2013.

JORDAHL, Kelsey *et al.* **geopandas/geopandas: v0.8.1.** [S. l.]: Zenodo, 2020. Software. DOI: 10.5281/zenodo.3946761.

MANN, Henry B.; WHITNEY, Donald R. On a test of whether one of two random variables is stochastically larger than the other. **The Annals of Mathematical Statistics,** Beachwood, v. 18, n. 1, p. 50-60, 1947. DOI: 10.1214/aoms/1177730491.

MCCOY, J. L. GIS and joint use management: a productive combination. In: IEEE RURAL ELECTRIC POWER CONFERENCE, 2005, San Antonio. **Proceedings.** San Antonio: IEEE, 2005. p. C2-1-C2-4. DOI: 10.1109/REPCON.2005.1436318.

MCKINNEY, Wes. Data structures for statistical computing in Python. In: PYTHON IN SCIENCE CONFERENCE, 9., 2010, Austin. **Proceedings of the 9th Python in Science Conference.** Austin: SciPy, 2010. p. 56-61. DOI: 10.25080/Majora-92bf1922-00a.

MESSINIS, George M.; HATZIARGYRIOU, Nikos D. Review of non-technical loss detection methods. **Electric Power Systems Research,** Amsterdam, v. 158, p. 250-266, 2018. DOI: 10.1016/j.epsr.2018.01.005.

NAGI, Jawad *et al.* Nontechnical loss detection for metered customers in power utility using support vector machines. **IEEE Transactions on Power Delivery,** Piscataway, v. 25, n. 2, p. 1162-1171, 2010. DOI: 10.1109/TPWRD.2009.2030890.

PAPADAKIS, George *et al.* Blocking and filtering techniques for entity resolution: a survey. **ACM Computing Surveys,** New York, v. 53, n. 2, p. 1-42, 2020. DOI: 10.1145/3377455.

PEDRETTI, André *et al.* Robotic process automation extended with artificial intelligence techniques in power distribution utilities. **Brazilian Archives of Biology and Technology,** Curitiba, v. 64, n. spe, e21210217, 2021. DOI: 10.1590/1678-4324-75years-2021210217.

SAALFELD, Alan. Conflation: automated map compilation. **International Journal of Geographical Information Systems,** London, v. 2, n. 3, p. 217-228, 1988. DOI: 10.1080/02693798808927897.

SÁNCHEZ, Laura; BRUNINI, Claudio. Achievements and challenges of SIRGAS. In: DREWES, Hermann (ed.). **Geodetic reference frames.** Berlin: Springer, 2009. p. 161-166. (International Association of Geodesy Symposia, v. 134). DOI: 10.1007/978-3-642-00860-3_25.

SEHGAL, Vivek; GETOOR, Lise; VIECHNICKI, Peter D. Entity resolution in geospatial data integration. In: ACM INTERNATIONAL SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 14., 2006, Arlington. **Proceedings.** New York: Association for Computing Machinery, 2006. p. 83-90. DOI: 10.1145/1183471.1183486.

SHAPIRO, Samuel S.; WILK, Martin B. An analysis of variance test for normality (complete samples). **Biometrika,** Oxford, v. 52, n. 3-4, p. 591-611, 1965. DOI: 10.1093/biomet/52.3-4.591.

VEREGIN, Howard. Data quality parameters. In: LONGLEY, Paul A. *et al.* (ed.). **Geographical information systems:** principles, techniques, management and applications. New York: Wiley, 1999. p. 177-189.

VIEGAS, Joaquim L. *et al.* Solutions for detection of non-technical losses in the electricity grid: a review. **Renewable and Sustainable Energy Reviews,** Amsterdam, v. 80, p. 1256-1268, 2017. DOI: 10.1016/j.rser.2017.05.193.

VIRTANEN, Pauli *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods,** London, v. 17, p. 261-272, 2020. DOI: 10.1038/s41592-019-0686-2.

WILCOXON, Frank. Individual comparisons by ranking methods. **Biometrics Bulletin,** Hoboken, v. 1, n. 6, p. 80-83, 1945. DOI: 10.2307/3001968.

XAVIER, Emerson M. A.; ARIZA-LÓPEZ, Francisco J.; UREÑA-CÁMARA, Manuel A. A survey of measures and methods for matching geospatial vector datasets. **ACM Computing Surveys,** New York, v. 49, n. 2, p. 1-34, 2017. DOI: 10.1145/2963147.

---

[1] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil. E-mail: acesse o artigo

original para visualizar o e-mail

[2] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.

[3] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.

[4] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.

[5] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.

[6] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.

[7] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.

[8] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.

[9] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.

[10] COPEL, Rua José Izidoro Biazetto, 158, Curitiba, 81200-240, Paraná, Brazil.

[11] COPEL, Rua José Izidoro Biazetto, 158, Curitiba, 81200-240, Paraná, Brazil.

[12] COPEL, Rua José Izidoro Biazetto, 158, Curitiba, 81200-240, Paraná, Brazil.

[13] COPEL, Rua José Izidoro Biazetto, 158, Curitiba, 81200-240, Paraná, Brazil.

[14] Tech2Think – Inteligência Artificial Além da Nuvem, R. Gérson da Graça, 94, Curitiba, 80.740-690, Paraná, Brazil.