# Graph-Based Socio-Technical Information Model for Power Distribution Network Planning and Operations

*Author:* Tomás Pérez Gutiérrez *Academic supervisor:* Bruce Stephen *Industrial supervisor:* Ciaran Higgins
*Universidad Pontificia de Comillas* *University of Strathclyde* *Scottish Power Energy Networks (SPEN)*

***Abstract*—Electricity distribution operators must balance the technical challenge of maintaining reliable networks with the social responsibility of delivering fair outcomes for the communities they serve. Achieving this is difficult because data on network assets, geography, and social factors are often fragmented across different formats and systems. This paper presents a knowledge graph framework that unifies three layers of information: (i) asset topology, (ii) geospatial context, and (iii) socio-economic indicators. A reproducible ETL pipeline standardises and links these heterogeneous datasets to a domain ontology, producing a scalable graph that supports cross-domain analysis. The framework is deployed in two environments: a local Neo4j instance for development and testing, and a secure enterprise deployment on Scottish Power Energy Networks' (SPEN) private cloud using Amazon Web Services Neptune. Its value is demonstrated through case studies, including: (i) identifying rural households with low electricity demand in high-deprivation areas, (ii) assessing accessibility of electric vehicle chargers using road-network travel times, and (iii) calculating asset-criticality scores that combine technical reliability with social vulnerability. The results show that the framework provides a practical, transparent foundation for equity-aware network planning and operation.**

***Index Terms*—knowledge graph; electricity distribution networks; social vulnerability; power grid planning; graph databases.**

## I. Introduction

Electricity distribution networks are facing increasing strain from the energy transition. Rising adoption of electric vehicles, distributed renewable generation, and shifting demand patterns are adding complexity to systems that were not originally designed for such conditions. At the same time, communities expect electricity supply to remain both reliable and fair. This requires planners and operators to consider not only technical performance but also the social and spatial contexts in which networks operate.

Conventional reliability indices such as SAIDI, CAIDI, SAIFI and CAIFI [1] provide robust technical benchmarks, but they treat all customers equally. For example, a two-hour outage is weighted the same for a hospital as for a hotel resort, ignoring differences in vulnerability. Research shows that areas with higher social vulnerability often face more frequent and prolonged outages [2, 3, 4], while rural communities supplied by long radial feeders are structurally more exposed to reliability risks. These disparities have placed energy justice at the forefront, highlighting the need for electricity-grid planning in which no community bears an unfair burden and everyone has equitable access to reliable, affordable energy [5, 6].

Recent regulatory responses reflect this shift, including Ofgem's Priority Services Register in the UK [7, 8], the U.S. Justice40 initiative [9], and Spain's *Bono Social Eléctrico* [10], which embed equity considerations into policy and operational practice. Despite these efforts, practice remains fragmented. Technical, spatial, and socio-economic data are usually stored in separate systems and formats, which prevents integrated analysis. Traditional indices overlook vulnerability, while emerging socially responsive metrics [11, 12, 13] show promise but are rarely adopted in operations.

The remainder of this paper is structured as follows: Section II reviews graph data models and related work; Section III defines the project scope and objectives; Section IV describes the system architecture and the graph-construction workflow; Section V details the implementation (data sources, preprocessing, staging tables, ontology mapping, and OpenCypher exports with loading); Section VI showcases three exploration examples; Section VII presents a case study on equity-aware node criticality; and Section VIII concludes with achievements, limitations, practical relevance, and directions for future work.

## II. Background and Related Work

### A. Graph Models

A graph is a data model consisting of nodes (vertices) and relationships (edges). Nodes represent entities, while relationships describe how these entities are connected. Graphs may be directed, where relationships indicate one-way flow, or undirected, where connections are mutual. This makes graphs a flexible and intuitive framework for representing systems in which the structure of connections is as important as the entities themselves. In practice, two graph models are most widely used: the triple-based model and the property graph model [14].

*1) Triple-based graphs (RDF):* In the triple-based model, all information is represented as a *triple* composed of a subject, predicate, and object. For example, the statement *"Transformer feeds Feeder"* is encoded with *Transformer* as the subject, *feeds* as the predicate, and *Feeder* as the object. This structure forms the basis of the Resource Description Framework (RDF), a W3C standard designed for the semantic web. Each entity is assigned a Uniform Resource Identifier (URI), ensuring unambiguous reference and enabling datasets from different organisations to be linked together. RDF data are stored in dedicated semantic databases known as *triple stores* and queried using SPARQL, the standard query language. The key strength of RDF lies in its interoperability and reasoning support, allowing integration across diverse domains and enabling inference through ontologies.

*2) Property graphs:* Property graphs provide a richer internal structure by allowing both nodes and relationships to hold properties as key–value pairs. For instance, a Transformer node can be represented as {id: T123, voltage: 11kV, location: Glasgow}. Nodes may also carry multiple labels that define their role, while relationships are assigned explicit types that capture their meaning (e.g., *supplies*, *located in*). This flexibility makes property graphs particularly well suited for modelling real-world systems in a way that is both intuitive and extensible. They are widely supported by native graph databases such as Neo4j, with query languages like Cypher designed for efficient pattern matching and interactive exploration.

*3) Comparison:* The two graph models offer complementary strengths. RDF excels in semantic consistency and cross-domain interoperability, making it valuable for data integration and reasoning across heterogeneous sources. In contrast, property graphs prioritise usability and analytical performance, providing a more direct and efficient way to represent, query, and visualise system structure. Increasingly, hybrid approaches are being adopted that map RDF triples into property graph structures, combining semantic interoperability with the efficiency of property-graph operations [14].

### B. Knowledge Graphs

A knowledge graph (KG) connects entities, relationships, and metadata to represent domain knowledge in a flexible way that can integrate structured, semi-structured, and unstructured data. Unlike generic graphs, a KG embeds explicit semantics, ensuring that connections are meaningful rather than purely structural.

*1) Semantics: taxonomies and ontologies:* Semantics are introduced through two key mechanisms: *taxonomies* and *ontologies*.

*Taxonomies.:* A taxonomy is a hierarchical classification system that organises concepts into parent–child levels. For example, within an electricity network:

- An `LV Transformer` is a subtype of `Transformer`.
- A `Transformer` is a subtype of `Electrical Equipment`.

This structure enables queries at different levels of detail. For instance, one can retrieve all items under `Electrical Equipment`, or narrow the scope to only `LV Transformers`. In this way, taxonomies support aggregation, filtering, and comparison across categories.

*Ontologies.:* An ontology goes further by formally defining *what things exist in the system, how they are related, and what attributes describe them*. For example, an ontology may specify that:

- A `Transformer` *feeds* a `Feeder`.
- A `Feeder` *supplies* one or more `ServicePoints`.
- A `ServicePoint` *is located in* a `Postcode` and has attributes such as `annualDemand` or `meterType`.

Such definitions ensure consistent interpretation across datasets: for instance, different sources that refer to a “meter point,” “supply point,” or “service point” can be aligned to the same concept. Ontologies also enable:

- **Rule checking** — e.g., a `Transformer` cannot directly supply a `Household` without a `ServicePoint`.
- **Reasoning** — new facts can be inferred based on logical definitions.

In practice, the ontology provides the shared vocabulary and logical rules, while the property graph serves as the efficient substrate for storing and querying the data.

*2) Construction steps:* Constructing a KG typically involves three main steps: (i) extracting entities and relationships from source systems, (ii) fusing and aligning knowledge to resolve duplicates and inconsistencies, and (iii) linking entities into a coherent, queryable graph. Once deployed, KGs can support a wide range of tasks, including semantic search, graph analytics, data auditing (e.g., consistency checks and change tracking), and cross-domain information management.

### C. Applications of Knowledge Graphs in the Electricity Sector

Knowledge graphs (KGs) are increasingly applied in the electricity sector, particularly within smart grids, to integrate and analyse heterogeneous data sources. Acting as a semantic layer, they connect equipment, events, customers, and their interrelationships, thereby enhancing interoperability, search, and reasoning. For example, the BKG framework in [15] integrates utility systems with external sources such as meteorological data and GIS. Entities and relations are extracted, modelled, and stored in graph or big-data platforms (e.g., Spark/Hadoop combined with Neo4j/HBase), enabling semantic search and decision support. Similarly, the workflow reviewed in [16] identifies common stages of knowledge extraction, ontology modelling, reasoning, and continuous updates, which together reflect prevailing practice.

Applications of KGs in electricity networks generally fall into four categories:

1) **Search and information services** — improving retrieval and navigation of grid-related data and operational procedures;
2) **Dispatch and fault handling** — structuring rules, case histories, and procedures to support operator decision-making;
3) **Maintenance and fault diagnosis** — integrating defect logs, condition-monitoring data, and equipment specifications for diagnosis and planning;
4) **Customer service and decision support** — linking events and assets with service processes to improve responsiveness.

Furthermore, some pilot projects demonstrate that KGs can mitigate “information islands” by aligning equipment data and their interconnections, thereby improving visualisation, analysis, and decision support [17].

### D. Research Gap and Contribution of this Thesis

Despite these advances, surveys continue to characterise the use of KGs in the electricity sector as being at an early stage. Key challenges include keeping knowledge up to date, expanding reasoning capabilities, and integrating data across highly heterogeneous sources [16]. Furthermore, most existing implementations focus mainly on technical and operational datasets, with limited attention to socio-demographic indicators or links to interdependent infrastructures.

This thesis addresses these gaps by developing an ontology-aligned property KG that integrates network asset data with socio-economic indicators and spatial context (e.g., rurality and accessibility). This integration enables cross-domain queries that jointly consider technical reliability and social vulnerability, supporting analyses that extend beyond operational concerns. In doing so, the work contributes to an underexplored research area by embedding social and spatial dimensions into electricity-sector knowledge graphs, thereby broadening their value for both equity-aware planning and resilience assessment.

## III. Project Definition

### A. Motivation

The transition to a low-carbon, data-rich energy system is reshaping the role of UK Distribution Network Operators (DNOs). Organisations such as Scottish Power Energy Networks (SPEN) must plan, operate, and invest in increasingly complex infrastructures while meeting expectations of fairness, transparency, and resilience. Regulatory frameworks (e.g., RIIO-ED2) and stakeholder engagement processes further require utilities to demonstrate social impact alongside technical performance. Conventional tools and siloed databases are poorly suited to represent these multi-layered realities.

This study addresses the challenge by developing a graph-based socio-technical data model that links engineering metrics with social and spatial context. Graph databases naturally represent infrastructure components and their interdependencies, while semantic relationships connect them to the communities they serve. Aligning technical data with contextual indicators enables more equitable and evidence-based approaches to planning and operations. Specifically, the model supports: (i) identifying critical infrastructure serving vulnerable populations, (ii) simulating disruption scenarios and their social impacts, and (iii) evaluating investment strategies that are both technically robust and socially fair.

### B. Objectives

The graph model is designed around six objectives:

1) **Define scope and select data sources:** Bound the study through literature review and consultation with SPEN; select a representative pilot area and gather relevant open datasets (e.g., EV chargers, road networks, socio-economic indicators).
2) **Design an integrated graph model:** Develop a flexible property-graph schema linking physical assets, customer locations, and geo-social context in a single structure.
3) **Build reproducible pipelines:** Create workflows to clean, standardise, and prepare heterogeneous data, preserving identifiers, relationships, and spatial references to ensure repeatability as new data arrive.
4) **Enable cross-domain analytics and visualisation:** Implement queries that combine technical, spatial, and social layers; provide results in formats accessible to both technical and non-technical stakeholders.
5) **Deploy locally and in the cloud:** Validate portability by running the graph in (i) a local Neo4j instance and (ii) an Amazon Neptune deployment within SPEN's private cloud.
6) **Demonstrate value through case studies:** Apply the model to representative analyses and a real-world case study to illustrate its use for planning and operational decision-making.

### C. Project Overview

This paper presents a multi-layer knowledge-graph framework that integrates asset topology, geospatial context, and socio-economic indicators into a unified property graph. A reproducible ETL pipeline standardises and fuses heterogeneous datasets, mapping them to a domain ontology for consistent representation and cross-domain analysis.

The framework is validated in two environments: (i) a local Neo4j instance for iterative development, and (ii) an enterprise deployment on SPEN's private cloud using Amazon Neptune.

Its applicability is illustrated through case studies, including:

- identifying rural low-demand households in high-deprivation areas,
- evaluating EV-charger accessibility using road-network travel times, and
- deriving asset-criticality scores that combine technical reliability with social vulnerability.

## IV. System Architecture

The graph model is implemented using a modular system architecture that supports data ingestion, transformation, graph construction, and deployment across different environments. The design emphasises **flexibility**—to accommodate heterogeneous datasets and evolving use cases—and **scalability**—to move seamlessly from local prototypes to enterprise-grade deployments.

At its core, the architecture ingests technical and contextual datasets such as network asset records, outage logs, socio-economic indicators, and geographic metadata. These are harmonised and transformed into an ontology-aligned property-graph schema that preserves both topological and semantic integrity. The processing workflow is complemented by a deployment layer, which enables end-users to query and explore the graph through modern database tools and visual interfaces.

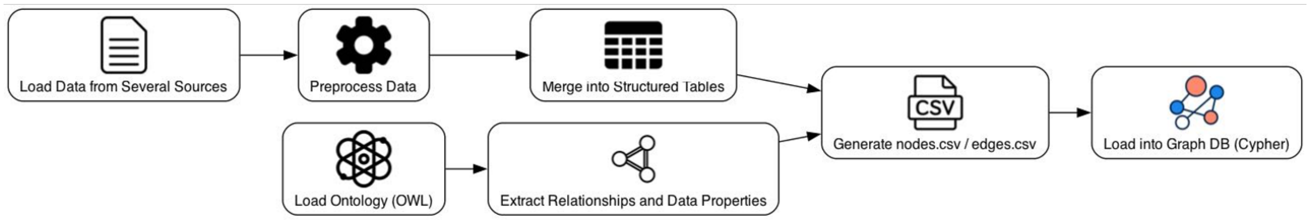


Fig. 1. End-to-end methodology for graph model construction.

Two complementary deployment configurations are supported:

- **Local Neo4j:** A Docker-based environment for development, testing, and visualisation using Cypher queries. This setup enables rapid prototyping and interactive exploration.
- **Amazon Neptune:** A secure, cloud-hosted environment within SPEN's Amazon Web Services (AWS) Virtual Private Cloud (VPC), accessed through a WireGuard VPN tunnel to meet corporate privacy and compliance requirements. This setup is optimised for scale, robustness, and multi-user collaboration.

The remainder of this section describes the graph construction workflow and the technical features of each deployment scenario.

### *A. Graph Construction Methodology*

The graph is developed through a structured pipeline that transforms raw datasets and semantic definitions into a property-graph representation of the electricity distribution network and its context. The methodology, illustrated in Figure 1, is ontology-driven to ensure that the resulting graph is not only topologically valid but also semantically meaningful. This design supports intuitive querying, scalability, and the integration of new datasets as analytical needs evolve. Section V details each pipeline stage.

### *B. Local Deployment: Neo4j*

The first deployment configuration is a local Neo4j environment (Figure 2), designed to support rapid prototyping and iterative development without relying on external infrastructure.

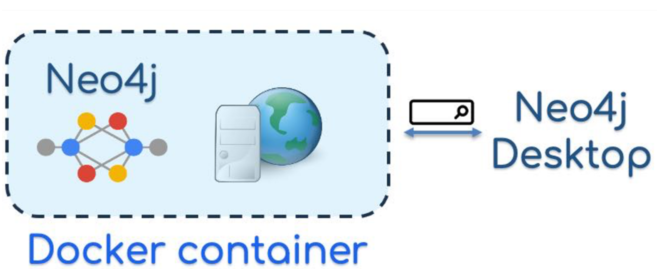


Fig. 2. Local development architecture using Neo4j and Docker.

Neo4j natively implements the property-graph model and provides the Cypher query language, which is well suited for pattern matching and exploratory analysis. Its integration with Neo4j Desktop offers built-in visualisation and schema exploration tools, making it particularly useful for designing graph structures, testing ontology mappings, and developing queries in an interactive manner. The Neo4j Community Edition is used in this work, which is free under the GPL v3 licence yet fully featured for research and prototyping purposes. This configuration prioritises accessibility and flexibility, enabling experimentation before moving to a production-grade environment.

### *C. Private Cloud Deployment: Amazon Neptune*

For enterprise deployment, the system is hosted on Amazon Neptune, a managed graph database service provided by AWS. The database is deployed within Scottish Power Energy Networks' AWS Virtual Private Cloud (VPC), ensuring isolation, scalability, and compliance with corporate security standards. Access is secured using a WireGuard VPN tunnel, which allows authenticated users to connect to the VPC while meeting SPEN's privacy and cybersecurity requirements (Figure 3).

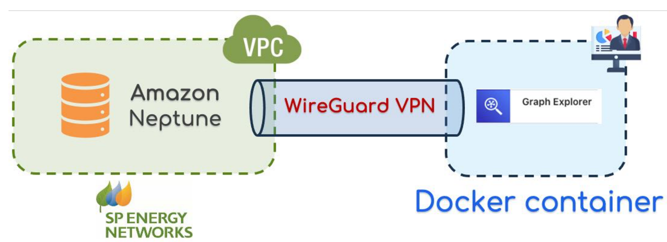


Fig. 3. Private cloud deployment architecture using Amazon Neptune and Graph Explorer.

Amazon Neptune supports multiple query languages (Gremlin, SPARQL, and OpenCypher), making it compatible with both RDF and property-graph models. OpenCypher is adopted in this work to ensure continuity with the Neo4j-based development phase. In addition, the Amazon's Graph Explorer [18]—deployed as a Docker service—provides a browser-based interface for visualising nodes, relationships, and query results, complementing programmatic access through APIs and client libraries.

Compared to the local Neo4j setup, this cloud configuration prioritises robustness and scalability. It is designed for production use, supporting larger datasets, integration with other

AWS services (e.g., storage and monitoring), and secure multi-user collaboration across distributed teams.

## V. Implementation

### A. Input Data Sources and Formats

To examine distribution networks *in context*, the study brings together four kinds of information: (i) technical data about network assets and connections, (ii) socio-economic indicators that describe community need and vulnerability, (iii) mobility and transport layers that reflect accessibility, and (iv) geospatial classifications that characterise places. Taken together, these sources let the analysis link physical infrastructure to the people and locations it serves. Table I summarises the datasets and formats used.

### B. Pilot Region Selection

A pilot region was chosen to allow faster development while remaining representative of typical UK conditions. A suitable pilot region had to span multiple voltage levels (LV/HV/EHV), a mix of urban, peri-urban and rural settings, and reliable data coverage. Because UK distribution networks are largely radial, a mainly radial subnetwork was preferred for the initial implementation. On this basis, the subnetwork centred on *Lanark* (South Lanarkshire, South Scotland) was selected. It covers Lanark, Biggar and nearby rural communities and includes hundreds of feeders across voltage levels (Fig. 4).

### C. Preprocessing Pipeline

A reproducible Python pipeline was developed to prepare the different input datasets before graph construction. The workflow has three main stages: *cleaning*, *linking*, and *spatial resolution*.

**1. Cleaning and standardisation.** Raw files are first checked and corrected: duplicate records are removed, missing identifiers and coordinates are flagged, and measurement units are converted to common formats (volts, kilowatts, minutes). Column names are aligned across datasets. All spatial data are converted to a single coordinate system (WGS84, EPSG:4326), while postcodes are normalised to a standard format. Time fields are parsed into a single `datetime` format and used to calculate basic indicators such as event durations and averages. The cleaned outputs are exported as UTF-8 CSVs with consistent headers.

**2. Linking and enrichment.** Low-voltage metering service points (LV_MSPs) are connected to the premises they supply using property identifiers (UPRNs) and postcodes, and enriched with smart-meter readings where available. Socioeconomic indicators are also aligned: for example, the Urban–Rural Classification (2022), originally defined on Data Zone 2022 boundaries, is translated to the 2011 geography used by the Scottish Index of Multiple Deprivation (SIMD). Each postcode is further assigned a *Spatial Signature*, which classifies areas by built form and land use.

**3. Spatial resolution.** Some records have coordinates but no postcode. These are resolved using a three-step reverse-geocoding process, where each step acts as a fall-back if the previous one fails: (i) bulk lookup with the `postcodes.io` API [24]; (ii) spatial join with open post-code boundaries [25]; and (iii) nearest-centroid assignment using Code-Point Open [26]. Finally, large national datasets such as the road network and Spatial Signatures are clipped to the pilot region to avoid unnecessary pre-processing.

### D. Structured Staging Tables

Once cleaned, the datasets are organised into a set of *staging tables* that act as an intermediate layer before building the graph. Each table contains only one type of entity—for example, smart meters, addresses, postcodes, EV chargers, or road nodes—while separate link tables record how these entities are connected.

This structure has three main benefits. First, it keeps the data **clean and consistent**: separating entities avoids duplication and makes it easier to spot errors. Second, it ensures **uniqueness and speed**: every entity has a stable identifier (such as `asset_id` or, `charger_id`) with indexes that allow fast matching across datasets.

Spatial information, such as the locations of assets, chargers, and roads, is stored as geometry fields. Non-spatial entities, like socioeconomic indicators, are linked to places through

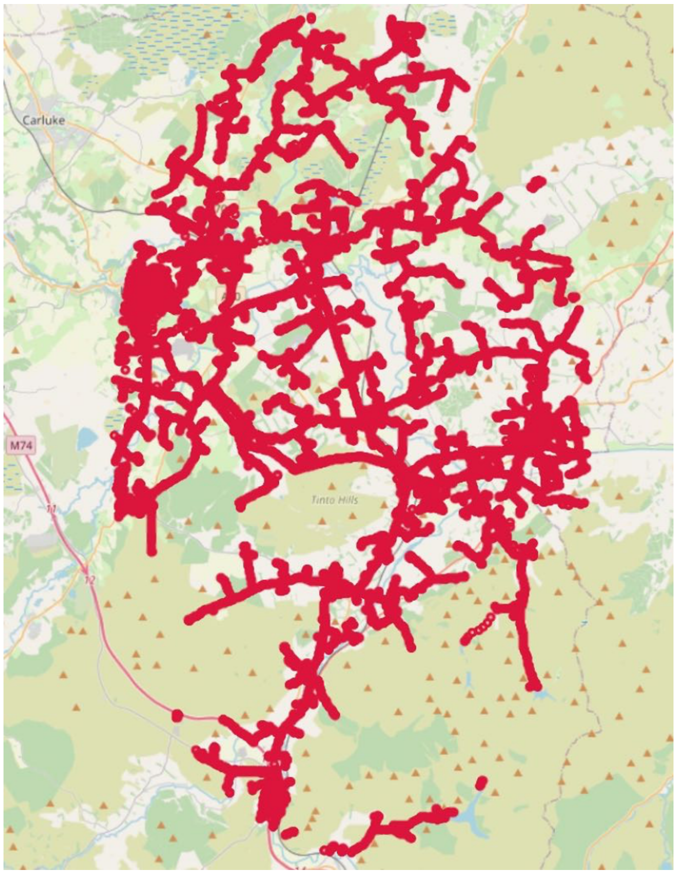


Fig. 4. Pilot area: distribution network footprint around Lanark (South Lanarkshire).

TABLE I
OVERVIEW OF DATASETS USED IN THE GRAPH MODEL

| Category | Dataset | Description | Format(s) | Source |
|---|---|---|---|---|
| **Technical Infrastructure** | SPEN Asset Registry | Inventory of distribution assets (transformers, fuses, protection devices, etc.), including IDs, types, coordinates, and metadata. | CSV/XLSX | SPEN |
| | SPEN Line and Link Data | Topology of cables and junctions connecting assets across the network. | CSV | SPEN |
| | Customer Address Mapping | Customer premises with addresses, postcodes, and smart-meter links. | CSV | SPEN |
| | Smart Meter Voltage Alerts | Half-hourly voltage readings (min/max/demand) over one week. | CSV | SPEN |
| | Smart Meter Fault Alerts | Fault event logs at the smart-meter level (e.g., supply loss). | CSV | SPEN |
| **Socioeconomic Context** | Scottish Index of Multiple Deprivation (2020) | Scores for 6,976 data zones, covering income, employment, health, education, access, housing, and crime (deciles/quintiles). | XLSX | [19] |
| **Mobility and Transport** | ChargePlace Scotland EV Infrastructure | Public EV chargers: locations, connector types, power ratings, operators, and status. | CSV/XLSX | [20] |
| | OS Open Roads | GB road network with geometries, classes, and linkages; used for accessibility analysis. | GPKG | [21] |
| **Geospatial Classification** | Urban–Rural Classification (2022) | Classification by settlement size and drive-time from settlements $\geq$10,000 people. | CSV/XLSX | [22] |
| | Spatial Signatures Framework | Built-form and land-use typology for small areas across Great Britain. | CSV/GPKG | [23] |

postcodes or data zones. The result is a normalised relational schema (Figure 5) that acts as a stable hand-off point to the graph loader. From here, data can be exported consistently as `nodes.csv` and `edges.csv` for Neo4j or Amazon Neptune. The staging schema also remains directly usable for quality checks and relational queries, making it a reliable backbone for both graph construction and validation.

### *E. Ontology and Semantic Mapping*

The staging tables provide a clean and linkable structure, but they do not explain the *meaning* of the entities or why they are related. To make this knowledge explicit and machine-readable, a domain ontology was developed in Protégé [27]. In simple terms, an ontology acts like a shared vocabulary: it defines the types of things that exist in the system, how they relate to one another, and what attributes describe them.

In this project, the ontology specifies three aspects:

- **Entities (classes)** – the main building blocks of the system, organised into hierarchies.
- **Relationships (object properties)** – the links between entities that reflect real-world connections (e.g., a smart meter `HAS_VOLTAGE_ALERT`, a charger `LOCATED_AT` an address; see Table II).
- **Attributes (data properties)** – the descriptive features of each entity (e.g., voltage level, capacity, coordinates, deprivation index).

The taxonomy is designed to be both *clear*—each class represents a distinct and understandable concept—and *scal-*

TABLE II
ONTOLOGY OBJECT PROPERTIES LINKING SUBJECT AND OBJECT ENTITIES.

| Subject Entity | Object Entity | Relationship |
|---|---|---|
| Postcode | Datazone | `BELONGS_TO` |
| RoadNode | RoadNode | `CONNECTED_TO` |
| ChargingStation | ChargingSession | `HAS_CHARGING_SESSION` |
| SmartMeter | FaultAlert | `HAS_FAULT_ALERT` |
| SmartMeter | VoltageAlert | `HAS_VOLTAGE_ALERT` |
| LV_MSP | SmartMeter | `HAS_SMART_METER` |
| Datazone | SocioEconomicProfile | `HAS_SOCIOECONOMIC_PROFILE` |
| Address | Postcode | `IN_POSTCODE` |
| SmartMeter, ChargingStation | Address | `LOCATED_AT` |
| LV_MSP | Address | `SERVES_ADDRESS` |
| Dist_Transformer | Intersection_Point_HV, Added_Endpoint_HV | `SUPPLIES` |
| Intersection_Point_HV, Added_Endpoint_HV | Dist_Transformer | `SUPPLIED_BY` |
| HV_MSP, LV_MSP, Unmetered_SP | ElectricAsset | `SERVED_BY` |
| ElectricAsset | HV_MSP, LV_MSP, Unmetered_SP | `SERVES` |
| LV_Fuse, Protective_Device, Dynamic_Protective_Device | ElectricAsset | `PROTECTS` |
| ElectricAsset | LV_Fuse, Protective_Device, Dynamic_Protective_Device | `PROTECTED_BY` |
| Electric switches, joints, links, intersections | Electric switches, joints, links, intersections | `INTERCONNECTS` |
| Added_Endpoint_* | Intersections and Joints | `CONNECTS_TO` |

*able*—new assets, spatial units, or event categories can be added without changing the overall structure. The ontology is exported in Turtle (.ttl) format and used directly by the pipeline to generate graph nodes and relationships that are both structurally consistent and semantically meaningful.

During the mapping step, the ontology is aligned with the staging tables. Subclasses inherit the attributes of their parent classes, relationships are linked to foreign keys in the relational schema, and attributes are bound to table columns. Bidirectional relationships—such as cables connecting two joints or roads linking two nodes—are explicitly modelled as symmetric so they can be traversed in either direction.

This approach ensures that queries reflect real-world questions: for example, tracing how a specific transformer supplies households in rural areas with high deprivation scores. By combining expert knowledge with a formal semantic model, the resulting property graph is not only technically consistent but also interpretable, extensible, and directly usable for network analysis.

### *F. Entity/Relationship Extraction and CSV Generation*

The final step before loading into a graph database is to export the cleaned and semantically aligned data into two OpenCypher-compatible files: `nodes.csv` and `edges.csv`. These files can be ingested directly by both Neo4j and Amazon Neptune's OpenCypher bulk loader [28].

The exporter first reads the ontology (in Turtle format) to determine which attributes and relationships are valid for each class. Inheritance is applied automatically, so subclasses carry the properties of their parent classes, ensuring consistency across entity types.

*Nodes.:* Each staging table is transformed into rows in `nodes.csv`, with one row per entity. Every entity receives a globally unique identifier, ontology attributes (e.g., voltage level, postcode) are stored as columns, and a semicolon-separated list of labels records its class hierarchy. This design allows queries to be run at different levels of abstraction, from broad categories to specific subclasses.

```
nodes.csv (header and example row)
ID,:LABEL,asset_type,voltage_level,postcode,...
12088991,"Asset;Electric;LV;LV_MSP",LV_MSP,LV,ML11
8NT,...
```

*Edges.:* Relationships between entities are written to `edges.csv`. These come either from inline foreign keys (e.g., a voltage alert linked to a smart meter produces a `HAS_VOLTAGE_ALERT` edge) or from dedicated link tables encoding connectivity (e.g., road links and electric lines). Relationships that are naturally undirected—such as two road nodes connected by a road segment—are written in both directions so traversals work consistently across engines.

```
edges.csv (header and example row)
:ID,:START_ID,:END_ID,:TYPE,...
e68047,DG14BQ,S01007616,BELONGS_TO,...
```

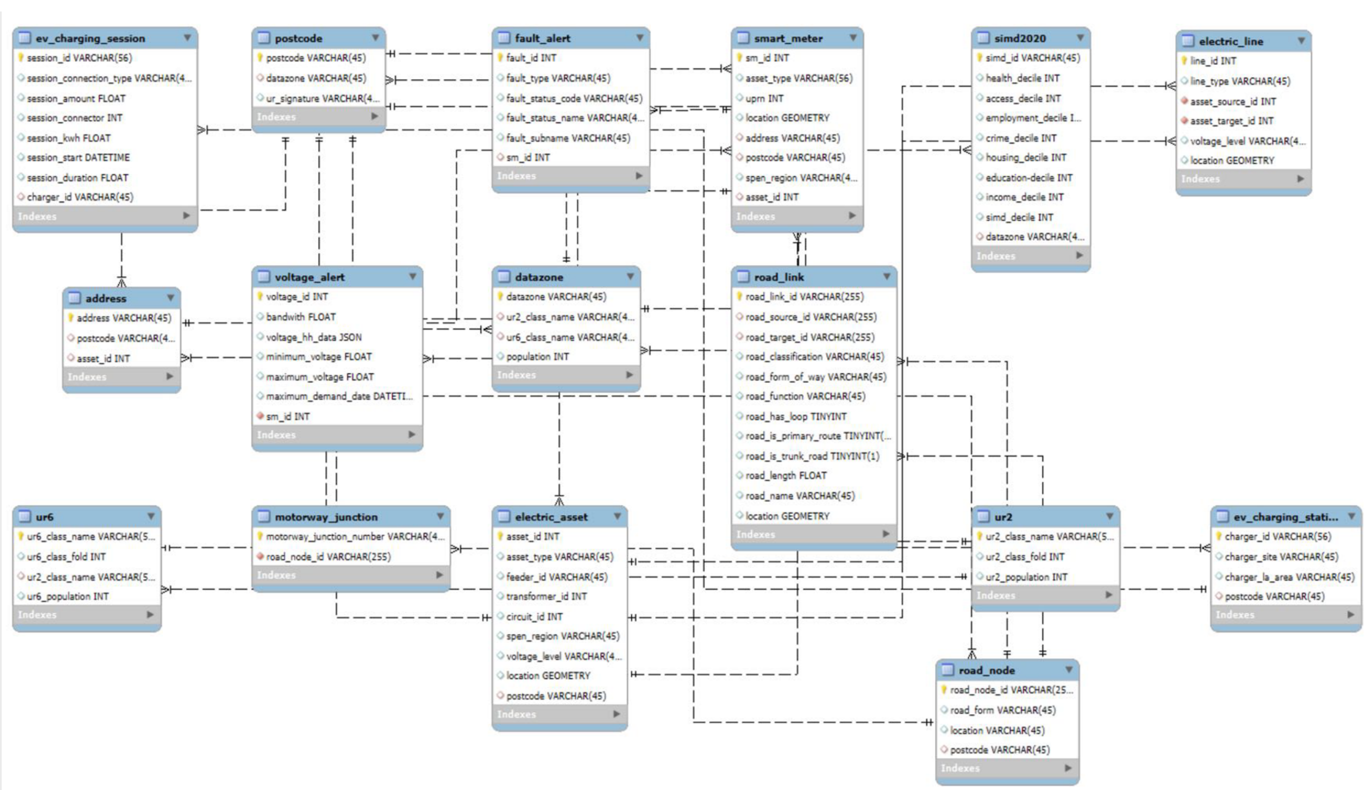


Fig. 5. Relational staging schema used to organise cleaned datasets into entity and link tables prior to graph export.

*Integrity checks.:* Before export, several quality checks are performed: duplicate nodes and edges are removed, edges without valid endpoints are discarded, and required attributes are verified for each class. The final files are UTF-8 encoded, use ISO-8601 timestamps, and store numeric/boolean values in plain types. Property names follow `lower_snake_case` conventions, while labels and relationship types follow the ontology. This guarantees that the exports are portable and immediately usable in both Neo4j and Amazon Neptune.

### G. Graph Loading and Deployment

After preprocessing, the data are exported into two files—`nodes.csv` and `edges.csv`—which capture the entities in the system and the links between them. Both follow the OpenCypher format, allowing them to be loaded directly into Neo4j and Amazon Neptune. Using a single, consistent export ensures that the same graph structure is reproduced across environments.

*Neo4j (local).:* In the local setup, a Python script loads the graph in two stages: first creating nodes (entities), then linking them with relationships. Integrity checks are applied during loading to avoid duplicates and ensure efficient queries. If a relationship cannot be created because one of its nodes is missing, it is skipped and logged. Once loading is complete, the system reports node and edge counts and runs sample queries to confirm correctness.

*Amazon Neptune (cloud).:* For enterprise deployment, the same export files are ingested into Amazon Neptune. During development, Amazon's *Simple Storage Service (S3)* was not available, so loading followed the same step-by-step process as Neo4j: adding nodes and edges in batches and validating their identifiers and labels. While this approach is slower than bulk loading, it was sufficient for testing and schema validation.

For production use, however, Neptune's high-speed bulk loader should be used. This loader operates through S3 *buckets*, which are essentially secure cloud-based folders where data files are stored. The loader reads directly from these buckets, allowing parallel ingestion, automatic retries, and built-in monitoring. This makes S3-based loading far more efficient and reliable for large datasets.

*Summary.:* Because both environments rely on the same ontology-aligned exports, they yield equivalent graphs. Neo4j is best suited for interactive exploration and rapid prototyping, while Amazon Neptune provides a secure, scalable option for production deployment.

## VI. Graph Exploration

This section presents three examples that demonstrate how the knowledge graph integrates technical, geographic, and socio-economic data within a single structure. The goal is to show how different layers of information can be explored together, enabling richer analysis and more informed decision-making.

### A. Example 1: Exploring a Distribution Transformer and Its Subgraph

To illustrate the model in practice, a specific distribution transformer (ID: 9078228) was selected. Its surrounding *subgraph* contains not only the connected electrical infrastructure (high- and low-voltage assets) but also contextual information such as smart meters, sensor alerts, addresses, postcodes, and socio-economic indicators.

Figure 6 compares two views of the same system: the traditional geospatial network model (left) and the graph-based view (right). In the graph representation, the transformer (shown in red) sits at the centre, with connections radiating outward across both engineering and social layers. This highlights how a graph structure naturally links together data that would otherwise be stored in separate silos.

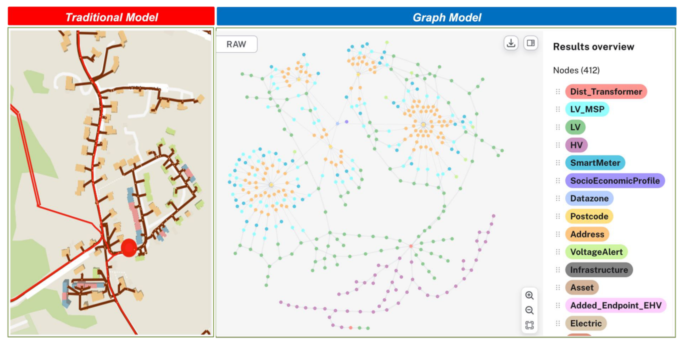


Fig. 6. Distribution transformer subgraph: traditional geospatial view (left) and graph-based view (right). The transformer (red) anchors links across engineering and social layers.

By zooming in step by step, these relationships become clearer. At the first level (Figure 7), the transformer is linked upstream to high-voltage nodes (purple) and downstream to low-voltage nodes (green). This shows how the transformer fits into the overall electrical hierarchy.

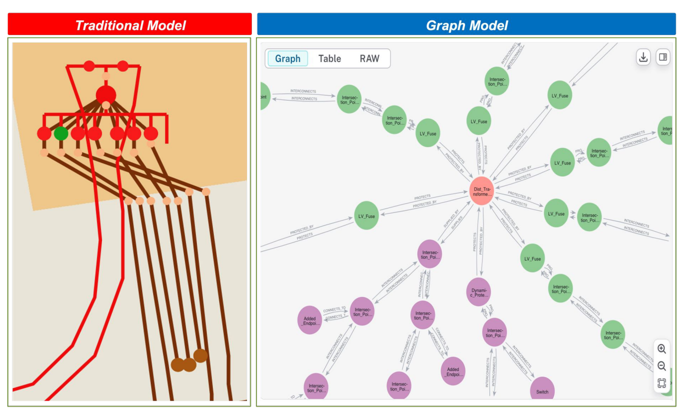


Fig. 7. Zoom 1 — Transformer context. Upstream HV nodes (purple) and downstream LV nodes (green).

The next view (Figure 8) moves closer, showing low-voltage joints (green) branching to multiple metering service points (cyan). Each of these service points is linked to specific customer addresses and postcodes, making the bridge from network infrastructure to households explicit.

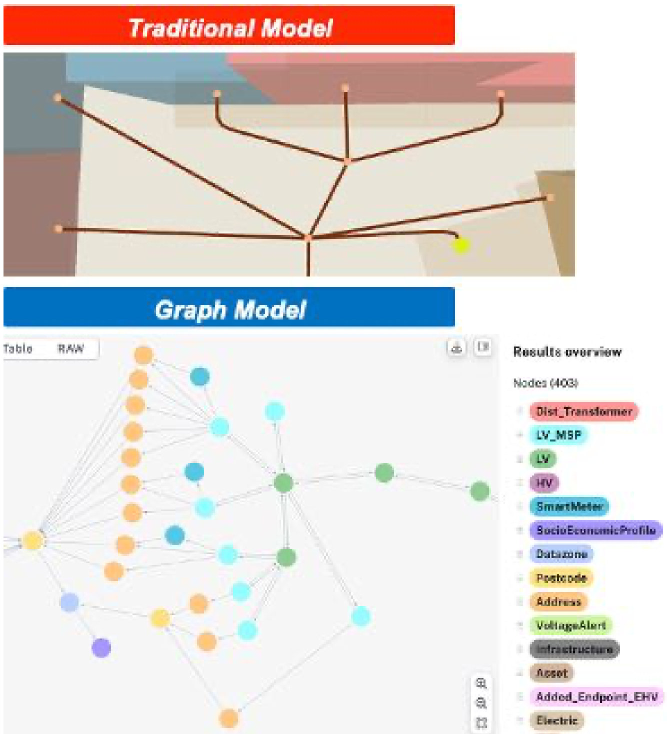


Fig. 8. Zoom 2 — LV joints (green) branching to LV_MSPs (cyan), linked to addresses and postcodes.

A further zoom (Figure 9) brings smart meters into focus. Each meter is connected to sensor-originated alerts such as voltage fluctuations or supply faults, with associated time-series readings and summary statistics. This makes it possible to move seamlessly from network topology into real-time performance data.

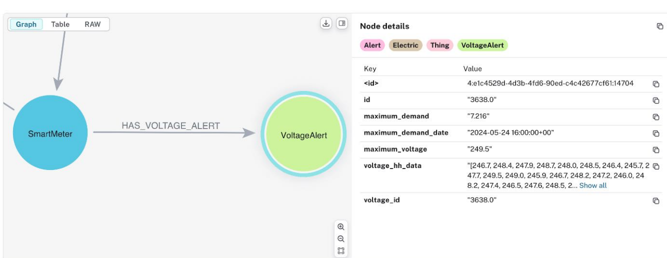


Fig. 9. Zoom 3 — Smart meters connected to sensor-originated alerts (e.g., voltage events), enriched with time-series and summary statistics.

Finally (Figure 10), the graph connects customer addresses to postcodes, postcodes to datazones, and datazones to socio-economic profiles such as Scottish Index of Multiple Deprivation (SIMD) scores.

This layered exploration shows how the graph enables end-to-end navigation: from a transformer, through its connected network, to the individual customers it serves, and finally to the socio-economic characteristics of those communities.

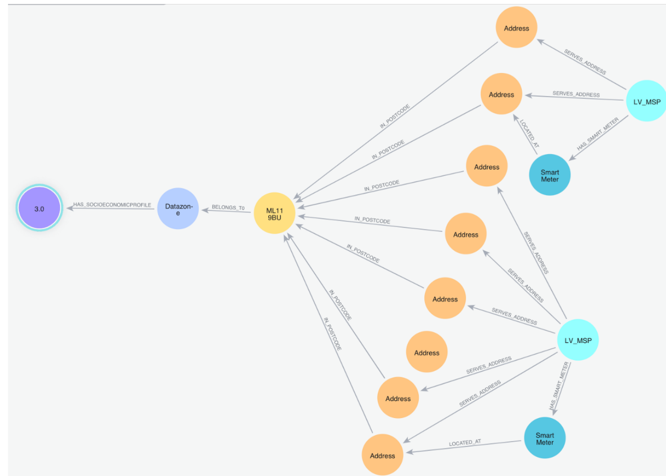


Fig. 10. Zoom 4 — Linking addresses and postcodes to datazones and socio-economic profiles (e.g., SIMD).

### B. Example 2: Identifying Rural Customers Near EV Chargers

This example examines **how accessible electric vehicle (EV) chargers are for rural customers**, using realistic *road-network distance* rather than simple straight-line distance. A customer is considered "served" if at least one charging station can be reached within **5 km by road** of their address.

The knowledge graph enables this calculation by linking customer addresses to their nearest road nodes, connecting chargers to road nodes in the same postcode, and then computing the *shortest path distance* through the road network. This produces a more practical measure of accessibility and highlights where charger provision is weaker.

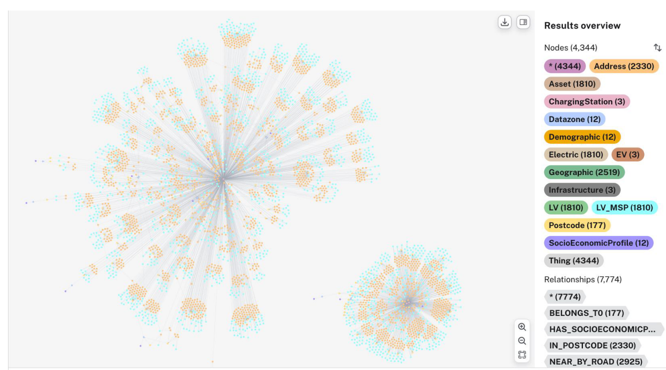


Fig. 11. Rural customers within 5 km by road of an EV charger. Orange = Address, cyan = LV MSP, pink = ChargingStation. The large rural cluster (top left) is served by only one charger, while the smaller cluster (bottom right) has two chargers.

The results (Figure 11) highlight two contrasting situations. In one case, a *large rural cluster* is covered by just **one** charger within 5 km, creating a potential vulnerability: if the charger is unavailable, all nearby customers are left without accessible charging. In contrast, a *smaller rural cluster* (Figure 12) has **two** chargers within the same distance, providing greater resilience and choice.

This analysis gives a useful first approximation of coverage, but it also has limitations. Charger positions are estimated at postcode level, and accessibility alone does not reveal the full picture: understanding the actual *EV penetration* in each zone is crucial for prioritising investment. Nevertheless, this simple view provides a clear starting point for identifying underserved areas and directing more detailed planning efforts.

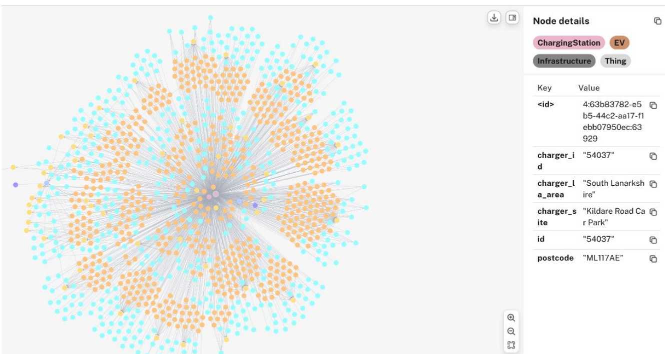


Fig. 12. Zoom into the smaller rural cluster. Customers here have access to two chargers within 5 km, offering higher resilience compared to the larger cluster.

### *C. Example 3: Identifying Vulnerable Low-Demand Rural Households*

This example shows how the graph can be used to **spot potentially vulnerable rural households** by combining electricity demand data from smart meters with socio-economic indicators from SIMD. A household is flagged as at risk if it meets all three criteria:

- **Low demand** — daily electricity use of $\leq$5 kWh.
- **High deprivation** — located in a datazone with SIMD decile $\leq$3.
- **Rural location** — classified as UR2 = RURAL.

Figure 13 shows all households that meet these conditions, with connections traced from electrical assets to addresses, postcodes, and socio-economic profiles.

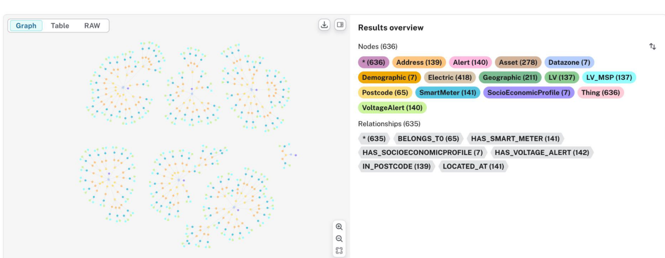


Fig. 13. Rural households meeting all three vulnerability filters: low demand ($\leq$5 kWh/day), high deprivation (SIMD $\leq$3), and rural classification.

A closer view is given in Figure 14, showing one example: a household in SIMD decile 1 (most deprived) with smart-meter readings of only 3–4 kWh/day—well below the UK average of approximately 7–10 kWh/day [29].

Low demand alone does not necessarily signal hardship—some homes may simply be small or under-occupied.

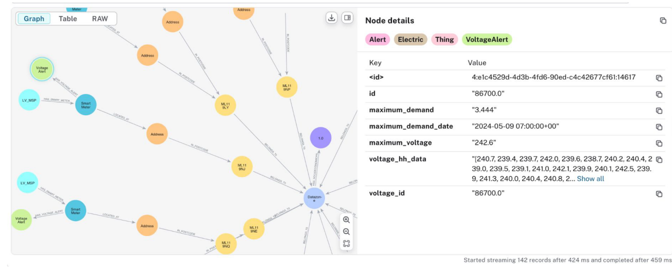


Fig. 14. Zoom into a single household in SIMD decile 1. The smart meter records only 3–4 kWh/day, far below the UK average ($\approx$7–10 kWh/day).

But when combined with **high deprivation and rural isolation**, it can point to more concerning issues such as self-rationing of energy, pre-payment constraints, or inefficient housing. By linking engineering data with social context, the graph provides a simple but powerful way to flag cases where **technical risk and social vulnerability overlap**, offering a foundation for more targeted customer support and policy interventions.

## VII. Case Study: Topological Vulnerability and Socioeconomic Impact of Node Failures

### *A. Overview and Related Work*

Power interruptions do not affect all customers equally: the same outage can impose very different burdens depending on who experiences it and how resilient the surrounding network is. Classical graph-theoretic studies of power grids typically focus on structural properties—such as betweenness centrality or degree distributions—and consistently show that removing high-centrality components fragments the network more severely than random failures [30, 31, 32, 33]. Yet such purely topological measures neglect three critical aspects: (i) the number of customers actually disconnected, (ii) the time required for restoration, and (iii) the social vulnerability of the affected population. To overcome these limitations, recent research has begun to augment structural analysis with impact-based metrics (e.g., Customer Hours of Outage) and socioeconomic indicators, producing assessments that more closely reflect real-world consequences [3, 13, 34]. Building on this trajectory, the present case study employs the knowledge graph to integrate failure likelihood, technical impact, and social vulnerability within a single, queryable model—enabling analyses that account for both engineering risk and equity considerations.

### *B. Objectives*

The case study focuses on a section of SPEN's distribution network (the pilot region) with four aims:

1) **Quantify reliability risks** by estimating failure probabilities and expected repair durations for different asset types.
2) **Assess customer impacts** by simulating single-node failures and calculating Customer Minutes of Interruption (CMI).

3) **Measure social context** through a Social Vulnerability Index (SVI) derived from SIMD 2020 scores and the Urban/Rural Classification.
4) **Integrate technical and social factors** by combining CMI and SVI into a composite criticality score that supports planning decisions balancing reliability and equity.

### *C. Methodology*

*1) Failure Likelihood from External Outage Data:* Unplanned outage records from Electricity North West (ENWL) [35] are used as a proxy to estimate annual failure rates for SPEN. This approach relies on the assumption that the underlying failure processes—such as asset ageing, weather-related events, and operational practices—are sufficiently similar across UK DNOs to allow transferability of statistics. To ensure consistency with RIIO-ED1 reporting standards, only events after 2015 are included. Each event is classified by asset category $a$, cause $c$, and voltage level $v \in \mathrm{LV}, \mathrm{HV}$.

The mean annual outage frequency for each class $(a, c, v)$ is:

$$\lambda_{a,c,v}^{\mathrm{ENWL}} \;=\; \frac{N_{a,c,v}^{\mathrm{ENWL}}}{T}, \tag{1}$$

where $N_{a,c,v}^{\mathrm{ENWL}}$ is the number of ENWL outages observed in that class over the period, and $T$ is the number of years of data.

Because ENWL and SPEN serve different customer bases, the rate is scaled by the ratio of total customers:

$$\lambda_{a,c,v}^{\mathrm{SPD}} \;=\; \frac{C^{\mathrm{SPD}}}{C^{\mathrm{ENWL}}}\, \lambda_{a,c,v}^{\mathrm{ENWL}}, \tag{2}$$

where $C^{\mathrm{SPD}}$ is the number of customers in the pilot area and $C^{\mathrm{ENWL}}$ is ENWL's total customer base. This scaling assumes that outage exposure is proportional to customer numbers, meaning that larger customer bases imply proportionally more assets and therefore more outage events.

Finally, to obtain a per-asset failure rate (failures per asset per year), the scaled frequency is divided by the number of relevant SPD assets:

$$r_{a,c,v} \;=\; \frac{\lambda_{a,c,v}^{\mathrm{SPD}}}{A_{a,c,v}^{\mathrm{SPD}}}, \tag{3}$$

where $A_{a,c,v}^{\mathrm{SPD}}$ is the count of SPD assets of type $a$ at voltage $v$ subject to cause $c$. This formulation assumes that all assets within a given class $(a, v)$ are statistically homogeneous with respect to the failure causes considered.

Thus, $r_{a,c,v}$ gives a normalised failure likelihood that can be applied across the pilot network in subsequent simulations.

*2) Repair Duration and Downtime:* Repair durations are taken from the same ENWL outage dataset. For each asset class $(a, c, v)$, the 75th percentile duration $D_{a,c,v}^{75}$ is used rather than the mean, providing a conservative estimate that better reflects tail risks of long repairs. The transfer of repair times from ENWL to SPD assumes that repair time distributions in the two regions are comparable, and that the 75th percentile provides a suitable balance between typical and worst-case outcomes.

The expected annual downtime for an asset of type $a$ at voltage $v$ is then:

$$\mathrm{Downtime}_{a,v} \;=\; \sum_{c} r_{a,c,v}\, D_{a,c,v}^{75}. \tag{4}$$

Here, the sum covers all relevant failure causes, producing the expected number of outage-minutes per year for a single asset.

*3) Customer-Impact Simulation:* In the distribution network, supply must always follow a valid path: current flows downstream from a *Primary Transformer* (EHV), through one or more *Distribution Transformers* (HV/LV), and finally to the individual *Low-Voltage Metering Service Points* (MSPs). A path is therefore considered valid only if it connects an MSP back to a primary transformer through this sequence of assets. This assumes radial operation of feeders and does not account for normally open points or real-time reconfiguration beyond the static connectivity recorded in the dataset.

Using the knowledge graph, these supply paths are traced for every MSP. Assets that appear on *all* valid paths to an MSP are classified as *critical nodes*, since their failure necessarily disconnects the MSP (no alternative route exists).

The number of households dependent on node $n$ is then:

$$N_{\mathrm{cust},n} \;=\; \sum_{m \in N_{\mathrm{MSP},n}} H_m, \tag{5}$$

where $N_{\mathrm{MSP},n}$ is the set of MSPs whose supply always traverses $n$, and $H_m$ is the number of households served at MSP $m$. The calculation assumes that household counts remain constant over the analysis period and that each household contributes equally to interruption impact.

The annual *Customer Minutes of Interruption* (CMI) attributable to node $n$ is:

$$\mathrm{CMI}_n \;=\; N_{\mathrm{cust},n}\, \mathrm{Downtime}_{a,v}, \tag{6}$$

where $(a, v)$ denotes the asset type and voltage level of $n$. This formulation combines asset reliability with network structure and customer counts, providing a practical measure of service impact.

*4) Social Vulnerability Index (SVI):* Beyond technical outage impacts, a *social vulnerability index* was developed to reflect how different populations experience inequities during interruptions. The construction followed two steps. First, SIMD 2020 deciles (income, employment, health, education, housing, access, crime) and the Urban–Rural Classification (UR6) were rescaled to $[0, 1]$ using indicator-specific curves, capturing nonlinear effects such as the heightened risks of poor health or geographic remoteness. It is assumed that these indicators remain representative over the study horizon, and that the chosen rescaling functions adequately capture the marginal risks of deprivation. Second, LV MSPs were ranked with multi-objective Pareto dominance (`pymoo.NonDominatedSorting()` [36]), ensuring that high deprivation in one dimension cannot be masked by better performance in another.

Since a single node can affect many MSPs with different profiles, results were aggregated by computing a household-weighted Pareto score. This assumes that vulnerability contributions are additive across households and that weighting by household count provides a fair proxy for aggregate social exposure. The result is a node-level SVI in $[0, 1]$ that reflects both the vulnerability of the affected areas and the size of the population impacted.

*5) Unified Criticality Score:* To balance technical and social aspects, both $\mathrm{CMI}_n$ and $\mathrm{SVI}_n$ are min–max scaled to the interval $[0, 1]$:

$$\widehat{\mathrm{CMI}}_n,\ \widehat{\mathrm{SVI}}_n \in [0, 1].$$

The unified *criticality score* is then defined as:

$$\mathrm{Crit}_n \;=\; \alpha\,\widehat{\mathrm{CMI}}_n \;+\; \beta\,\widehat{\mathrm{SVI}}_n, \qquad \alpha + \beta = 1, \tag{7}$$

where $\alpha$ controls the weight on technical interruption impact and $\beta$ controls the weight on social vulnerability. Equal weighting ($\alpha = \beta = 0.5$) is used in the base case, though sensitivity analysis can explore alternative policy preferences. Nodes with high $\mathrm{Crit}_n$ represent locations where both outage risk and population vulnerability are concentrated, making them natural priorities for resilience planning and investment.

### D. Results and Discussion

In the graph, every electrical asset is linked to a `CriticalityScore` node that captures not only the number of customers affected but also their social vulnerability (see Figure 15). This means high-risk components can be identified directly through queries, without the need to rerun simulations. The stored values also make it straightforward to reproduce results and experiment with different weighting schemes.

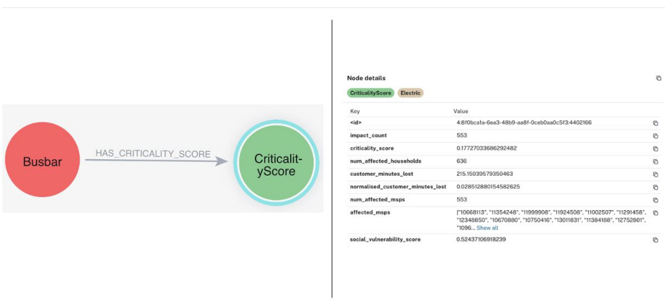


Fig. 15. Graph representation of an electric node (Busbar) and its associated `CriticalityScore` node. The criticality score shown was calculated using a weight of $\alpha = 0.7$ for technical impact and $\beta = 0.3$ for social vulnerability.

*Sensitivity to weighting.:* Changing the balance between technical and social factors produces different priorities. When technical impact is emphasised ($\alpha > 0.5$), large assets such as Primary Transformers dominate the ranking because of their wide reach and longer repair times. When social vulnerability is emphasised ($\beta > 0.5$), smaller assets in deprived or remote areas move up the list, despite affecting fewer customers. A balanced weighting produces a mixed shortlist, highlighting both types of risk.

*Archetypes and planning implications.:* Across the network, three broad patterns appear. Some assets have high technical impact but low vulnerability; these are systemic hubs best managed through reinforcement, switching options, or stronger protection. Others serve fewer customers but are located in socially vulnerable areas; here, targeted interventions such as faster restoration or backup supply may be more appropriate. A third group combines both high technical and high social risk; these are few in number but deserve the highest priority for investment and contingency planning.

*Geographic patterns.:* The two perspectives also highlight different areas on the map. Technical weighting points to hubs around substations and main feeders. Social weighting shifts attention to remote rural areas and pockets of deprivation in urban settings. This contrast underlines the need to align engineering reinforcement with community-focused measures.

*Operational use.:* Because both the technical ($\mathrm{CMI}_n$) and social ($\mathrm{SVI}_n$) components are stored alongside the combined score, operators can flexibly adjust the weights to match policy goals, generate alternative rankings under different assumptions, and link the results to planned maintenance programmes. In practice, giving more weight to $\alpha$ prioritises security of supply, while raising $\beta$ reflects commitments to equity.

*Robustness and limitations.:* The rankings remain stable under different assumptions about repair times and outage causes, which makes the approach reliable for first-order screening. Still, the method simplifies reality by assuming independent single-node failures, ignoring cascading effects, and not modelling power flows. Future extensions could include load-flow analysis, restoration strategies, or asset-condition data to refine failure probabilities.

## VIII. Conclusions

This paper has shown that electricity distribution networks can be represented as multi-layer knowledge graphs that bring together technical assets, geospatial context, and socio-economic indicators in a single, ontology-aligned model. Heterogeneous datasets were mapped to a domain ontology and processed through a reproducible ETL pipeline, creating a shared vocabulary and relationship structure across domains. The resulting graph supports flexible, cross-domain queries—such as identifying vulnerable households or assessing EV-charger accessibility—without changing query logic.

Deployment was demonstrated in both Neo4j (for local prototyping) and Amazon Neptune (for enterprise-scale use), using a common OpenCypher export to guarantee portability across engines. In the case study on node criticality, the graph combined failure likelihood, expected downtime (CMI), and a Social Vulnerability Index (SVI) into a single tunable score. Adjusting the balance between technical and social weights produced complementary insights: large hubs emerged as priorities under technical weighting, while smaller assets in deprived or rural areas were highlighted when social vulnerability was emphasised.

The study has some limitations. The network topology is simplified, power-flow constraints are not included, and the model operates on static snapshots rather than real-time data. These choices ensured tractability but mean the framework is not yet a full digital twin. Future work should integrate operational parameters such as transformer capacities, extend spatial hierarchies, test the approach on meshed networks, and support near real-time data ingestion from smart meters and EV infrastructure. Incorporating utility-governed datasets would further improve coverage and accuracy.

**Practical relevance.** The framework has value for both public and private stakeholders. Policymakers can use it as a transparent and flexible tool to plan interventions that balance infrastructure resilience with social fairness, helping to target investment where technical and social risks overlap. For distribution network operators, it acts as a decision-support system for asset management, reinforcement planning, and demand forecasting, while also strengthening trust with communities.

In summary, ontology-aligned knowledge graphs offer a portable and extensible way to bridge engineering and socio-economic perspectives. They enable decisions that are not only technically robust but also socially equitable, laying the foundation for more resilient energy systems.

## References


[1] "IEEE Guide for Electric Power Distribution Reliability Indices". In: *IEEE Std 1366-2022 (Revision of IEEE Std 1366-2012)* (2022), pp. 1–44. DOI: 10.1109/IEEESTD.2022.9955492.

[2] Bo Li et al. *Recent Decade's Power Outage Data Reveals the Increasing Vulnerability of U.S. Power Infrastructure*. 2024. arXiv: 2408.15882 [physics.soc-ph]. URL: https://arxiv.org/abs/2408.15882.

[3] Jesse Dugan, Dahlia Byles, and Salman Mohagheghi. "Social vulnerability to long-duration power outages". In: *International Journal of Disaster Risk Reduction* 85 (2023), p. 103501. ISSN: 2212-4209. DOI: https://doi.org/10.1016/j.ijdrr.2022.103501. URL: https://www.sciencedirect.com/science/article/pii/S2212420922007208.

[4] Idowu Ajibade et al. "Who bears the burden? An assessment of vulnerability and resilience to consecutive disasters in the Portland metro region". In: *Environmental Research Letters* 20.8 (2025), p. 084006. DOI: 10.1088/1748-9326/ade459.

[5] Initiative for Energy Justice. *Section1– Defining Energy Justice: Connections to Environmental Justice, Climate Justice, and the Just Transition*. https://iejusa.org/section-1-defining-energy-justice/. Accessed 28 July 2025. 2019.

[6] Benjamin K. Sovacool and Michael H. Dworkin. "Energy justice: Conceptual insights and practical applications". In: *Applied Energy* 142 (2015), pp. 435–444. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2015.01.002. URL: https://www.sciencedirect.com/science/article/pii/S0306261915000082.

[7] Ofgem (Office of Gas and Electricity Markets). *Equality and diversity*. Web page. Accessed July 28, 2025. 2025. URL: https://www.ofgem.gov.uk/about-us/working-ofgem/equality-and-diversity.

[8] Ofgem (Office of Gas and Electricity Markets). *Join your supplier's Priority Services Register*. Web page. Accessed July29,2025. 2025. URL: https://www.ofgem.gov.uk/join-your-suppliers-priority-services-register.

[9] U.S. Department of Energy. *2023 Equity Action Plan and Update to 2022 Plan*. Equity Action Plan pursuant to Executive Order 14091. Performance.gov publication, update issued January 5, 2024. U.S. Department of Energy, 2024. URL: https://assets.performance.gov/cx/equity-action-plans/2023/EO_14091_DOE_EAP_2023.pdf.

[10] Government of Spain / Ministry for the Ecological Transition and Demographic Challenge. *Integrated National Energy and Climate Plan (INECP) 2021–2030*. Public official document. English language version. 2020. URL: https://energy.ec.europa.eu/system/files/2020-06/es_final_necp_main_en_0.pdf.

[11] Barry Flanagan et al. "A Social Vulnerability Index for Disaster Management". In: *Journal of Homeland Security and Emergency Management* 8 (Jan. 2011). DOI: 10.2202/1547-7355.1792.

[12] Juan P. Montoya-Rincon et al. "A socio-technical approach for the assessment of critical infrastructure system vulnerability in extreme weather events". In: *Nature Energy* 8.9 (Sept. 2023), pp. 1002–1012. ISSN: 2058-7546. DOI: 10.1038/s41560-023-01315-7. URL: https://doi.org/10.1038/s41560-023-01315-7.

[13] Farzane Ezzati et al. "Power outage-risk integrated social vulnerability analysis highlights disparities in small residential communities". In: *Communications Earth Environment* 6 (Apr. 2025). DOI: 10.1038/s43247-025-02278-1.

[14] Carlos Ramonell, Rolando Chacón, and Hector Posada. "Knowledge graph-based data integration system for digital twins of built assets". In: *Automation in Construction* 156 (Oct. 2023). DOI: 10.1016/j.autcon.2023.105109.

[15] Jian Wang et al. "A survey on the development status and application prospects of knowledge graph in smart grids". In: *IET Generation, Transmission Distribution* 15 (3 2021), pp. 383–407. DOI: 10.1049/gtd2.12040. eprint: https://digital-library.theiet.org/doi/pdf/10.1049/gtd2.12040. URL: https://digital-library.theiet.org/doi/abs/10.1049/gtd2.12040.

[16] Han Ding et al. "A Review of the Construction and Application of Knowledge Graphs in Smart Grid". In: *2021 IEEE Sustainable Power and Energy Conference (iSPEC)*. 2021, pp. 3770–3775. DOI: 10.1109/iSPEC53008.2021.9736038.

[17] Haichao Huang et al. "Constructing Knowledge Graph from Big Data of Smart Grids". In: *2019 10th International Conference on Information Technology in Medicine and Education (ITME)*. 2019, pp. 637–641. DOI: 10.1109/ITME.2019.00147.

[18] Amazon Web Services, Inc. *Visualizing Graphs with Neptune Graph Explorer*. https://docs.aws.amazon.com/neptune/latest/userguide/visualization-graph-explorer.html. Accessed: 2025-08-08. 2025.

[19] Scottish Government. *Scottish Index of Multiple Deprivation 2020: Introduction*. Online booklet. Scottish Government, Jan. 2020. URL: https://www.gov.scot/publications/scottish-index-multiple-deprivation-2020/.

[20] Scottish Government / Transport Scotland. *ChargePlace Scotland*. https://chargeplacescotland.org. Accessed 7 August 2025. 2025.

[21] Ordnance Survey. *OS Open Roads*. https://www.ordnancesurvey.co.uk/products/os-open-roads. Data updated every six months; accessed 7 August 2025. 2025.

[22] Scottish Government. *Scottish Government Urban Rural Classification 2022*. Tech. rep. Current version as of December 2024, based on Census Day 2022 data; published via gov.scot. Scottish Government, Dec. 2024. URL: https://www.gov.scot/publications/scottish-government-urban-rural-classification-2022/.

[23] Martin Fleischmann and Daniel Arribas-Bel. "Geographical characterisation of British urban form and function using the spatial signatures framework". In: *Scientific Data* 9.1 (Sept. 2022), p. 546. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01640-8. URL: https://doi.org/10.1038/s41597-022-01640-8.

[24] Ideal Postcodes. *Postcodes.io: UK Postcode Geolocation API (Overview)*. https://postcodes.io/docs/overview/. Accessed 7 August 2025. 2025.

[25] Mark. *Open Data GB Postcode Unit Boundaries*. https://longair.net/blog/2021/08/23/open-data-gb-postcode-unit-boundaries/. Accessed 7August2025. Aug. 2021.

[26] Ordnance Survey. *Code-Point Open*. https://www.ordnancesurvey.co.uk/products/code-point-open. Updated quarterly (February, May, August, November); accessed 7 August 2025. 2025.

[27] Mark A. Musen. "The Protégé Project: A Look Back and a Look Forward". In: *AI Matters* 1.4 (June 2015), pp. 4–12. DOI: 10.1145/2757001.2757003. URL: https://doi.org/10.1145/2757001.2757003.

[28] Amazon Web Services. *Load format for openCypher data*. https://docs.aws.amazon.com/neptune/latest/userguide/bulk-load-tutorial-format-opencypher.html. Amazon Neptune User Guide – accessed 8 Aug 2025. 2025.

[29] Louise Horscroft. *How much electricity does the average home use per day in the UK?* AquaSwitch blog post; accessed 2025-08-09. 2023. URL: https://www.aquaswitch.co.uk/blog/average-electricity-usage/.

[30] Momhammad Ali Saniee Monfared, Mahdi Jalili, and Zohreh Alipour. "Topology and vulnerability of the Iranian power grid". In: *Physica A: Statistical Mechanics and its Applications* 406 (2014), pp. 24–33. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2014.03.031. URL: https://www.sciencedirect.com/science/article/pii/S037843711400226X.

[31] Réka Albert, István Albert, and Gary L. Nakarado. "Structural vulnerability of the North American power grid". In: *Physical Review E* 69.2 (Feb. 2004). ISSN: 1550-2376. DOI: 10.1103/physreve.69.025103. URL: http://dx.doi.org/10.1103/PhysRevE.69.025103.

[32] Paolo Crucitti, Vito Latora, and Massimo Marchiori. "A topological analysis of the Italian electric power grid". In: *Physica A: Statistical Mechanics and its Applications* 338.1 (2004). Proceedings of the conference A Nonlinear World: the Real World, 2nd International Conference on Frontier Science, pp. 92–97. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2004.02.029. URL: https://www.sciencedirect.com/science/article/pii/S0378437104002249.

[33] Banghua Xie et al. "The Vulnerability of the Power Grid Structure: A System Analysis Based on Complex Network Theory". In: *Sensors* 21.21 (2021). ISSN: 1424-8220. DOI: 10.3390/s21217097. URL: https://www.mdpi.com/1424-8220/21/21/7097.

[34] Alexys H Rodríguez A, Abdollah Shafieezadeh, and Alper Yilmaz. "System outage fragility for power systems: A robust data-driven framework for disparity analysis using multiple hurricane events". In: *International Journal of Disaster Risk Reduction* 118 (2025), p. 105240. ISSN: 2212-4209. DOI: https://doi.org/10.1016/j.ijdrr.2025.105240. URL: https://www.sciencedirect.com/science/article/pii/S2212420925000640.

[35] Electricity North West. *Unplanned Outages*. https://electricitynorthwest.opendatasoft.com. Accessed: 2025-07-25. 2025. URL: https://electricitynorthwest.opendatasoft.com/explore/dataset/unplanned-outages/information/?disjunctive.main_equipment_involved_1&disjunctive.direct_cause_category&disjunctive.district_name&disjunctive.primary_substation&disjunctive.voltage&disjunctive.incident_reporting_year.

[36] J. Blank and K. Deb. "pymoo: Multi-Objective Optimization in Python". In: *IEEE Access* 8 (2020), pp. 89497–89509.