

Data Descriptor

Not peer-reviewed version

WorldPT: An Ontology-Driven, Directed Multilayer Repository for Computational Narratology, Structural Evaluation and Generative Modeling

[Lucas V. Souza](#)* and Leopoldo Lusquino Filho

Posted Date: 16 April 2026

doi: 10.20944/preprints202604.1126.v1

Keywords: worldbuilding; computational narratology; graph theory; Portuguese NLP; social network analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Data Descriptor

WorldPT: An Ontology-Driven, Directed Multilayer Repository for Computational Narratology, Structural Evaluation and Generative Modeling

Lucas V. Souza^{1,*} and Leopoldo Lusquino Filho^{1,2}

¹ Institute of Science and Technology, São Paulo State University, Sorocaba, Brazil

² Recod.AI, Institute of Computing, University of Campinas, Campinas, Brazil

* Correspondence: lucas.v.souza@unesp.br

Abstract

The systematic construction of expansive fictional universes, known as worldbuilding, faces significant challenges in maintaining long-range structural consistency, particularly within generative AI architectures prone to "ontological drift." This paper introduces WorldPT, a novel framework and dataset that formalizes worldbuilding through Directed Multilayer Attributed Graphs. By implementing a Grounding Directionality Axiom and a hexapartite layering system (Structural, Causal, Temporal, Social, Ontological, and Symbolic), we transition from unstructured text-centric models to machine-verifiable narrative structures. The dataset is uniquely curated in Portuguese, aiming to democratize access to advanced computational narratology resources for the Lusophone community. To evaluate the framework, we applied Social Network Analysis (SNA) metrics to a case study of Tolkien's Middle-Earth universe. Results reveal a "Small-World" topology (average path length of 2.68) and a predominant structural layer (48.7% of connections), quantitatively fingerprinting the setting as a structural-driven worldbuilding. Furthermore, we propose the Cross-Layer Coupling (CLC) metric to identify "lore-shifters" entities whose multidimensionality transcends individual layers. Our findings demonstrate that WorldPT provides a robust foundation for building ontologically stable and interconnected narrative experiences, bridging the gap between graph-based knowledge representation and creative storytelling.

Keywords: worldbuilding; computational narratology; graph theory; Portuguese NLP; social network analysis

1. Introduction

The systematic construction of expansive fictional universes, collectively known as *worldbuilding*, represents one of the most sophisticated cognitive tasks in human narratology. It requires the simultaneous orchestration of geographical, historical, socio-political, and metaphysical subsystems into a coherent, self-sustaining whole. In the era of Large Language Models (LLMs) and autonomous narrative agents, the challenge of maintaining long-range structural consistency across these subsystems has emerged as a critical bottleneck. While current generative architectures demonstrate remarkable local linguistic fluency, they frequently suffer from "ontological drift", where the internal logic of the world is sacrificed for immediate textual probability [1,2].

To address this limitation, we posit that fictional worlds should be formalized not as sequences of natural language, but as Directed Multilayer Attributed Graphs. In this framework, world-coherence is a function of the integrity of directed dependencies between heterogeneous entities. However, the existing landscape of narrative datasets is largely bifurcated: on one hand, we find unstructured text corpora that lack explicit semantic grounding; on the other, specialized knowledge graphs (KGs) that focus on factual triples but fail to capture the hierarchical "grounding" essential for narrative depth [3].

In this paper, we introduce the WorldPT, a curated repository designed to bridge the gap between computational narratology and graph-based knowledge representation. A defining feature of the WorldPT is its construction entirely in Portuguese, a deliberate choice aimed at expanding the accessibility of advanced narrative resources. By providing a high-quality structural benchmark in Portuguese, we seek to empower the vibrant Lusophone research and creative communities. This initiative democratizes access to complex semantic reasoning tools, fostering richer, more linguistically diverse studies in narrative AI and enabling native creators to model their fictional universes without the friction of translation, aligning with broader calls for language-specific datasets to mitigate linguistic centralism in AI [4].

Our dataset formalizes world-consistency through a hexapartite layering system (Structural, Causal, Temporal, Social, Ontological, and Symbolic), providing a rigorous schema for inter-entity dependencies. We adopt a Grounding Directionality Axiom, where every edge (u, v) denotes that the identity or function of entity u is semantically situated within the context of entity v . This ontological grounding ensures that the narrative graph remains an auditable and logically consistent structure, moving beyond simple co-occurrence toward formal narrative modeling [5–7].

Furthermore, by translating narrative lore into a topological framework, the WorldPT enables the direct application of Social Network Analysis (SNA) to storytelling [8,9]. We provide built-in evaluation pipelines that compute global and node-level narrative metrics. This transforms abstract literary critiques into quantifiable data, allowing researchers to diagnose the structural health of a fictional universe through metrics that reveal narrative rhythm and connectivity patterns [10].

The primary contributions of this work are fourfold: (i) we define a standardized schema for 13 ontological classes and their permissible directional relations, based on foundational principles of formal ontology [11], (ii) we introduce a method for disambiguating multifaceted relationships, separating physical containment from symbolic influence and causal triggers, treating the world as a multiplex network [12], (iii) we provide a versioned collection of canonical JSON-encoded worlds in Portuguese, validated for domain-range integrity and acyclic dependency, and (iv) we propose a suite of quantitative proxies for world-building quality based on SNA. These include measures of narrative density (sparsity vs. claustrophobia), fragmentation (isolated plot arcs), “small-world” interconnectivity (average path length), structural bottlenecks (critical bridges), and typological assortativity (world segregation).

By transitioning from text-centric to graph-centric world-modeling, we provide the computational community with a robust foundation for building more coherent, re-playable, and ontologically stable narrative experiences.

2. Formal Model and Ontology

To provide a rigorous framework for computational worldbuilding, we define the WorldPT using the formalism of directed multilayer attributed graphs. This approach allows for the representation of complex, non-linear dependencies while maintaining a clear separation between different types of narrative existence.

2.1. Mathematical Framework

A fictional world is modeled as a directed multiplex attributed graph $\mathcal{G} = (V, E, \mathcal{L}, \mathcal{T}_V, \mathcal{R})$, where:

- V is a finite set of entities (nodes).
- \mathcal{T}_V is the set of 13 ontological classes: Artifact, Figure, Race, Culture, Religion, Organization, Place, Event, Chronology, Politics, Economy, MagicSystem, Technology
- \mathcal{L} is the set of six semantic layers: $\{L_{str}, L_{cau}, L_{tem}, L_{soc}, L_{ont}, L_{sym}\}$.
- $E \subseteq V \times V \times \mathcal{L} \times \mathcal{R}$ is the set of directed edges, where \mathcal{R} represents the specific relation type (e.g., *belongs_to*, *influences*).

Every node $v \in V$ is assigned a type via the mapping function $\tau : V \rightarrow \mathcal{T}_V$. An edge $e = (u, v, \ell, r)$ connects a source node u to a target node v within a specific layer $\ell \in \mathcal{L}$ using a relation $r \in \mathcal{R}$.

2.2. The Grounding Directionality Axiom

The core innovation of the WorldPT is the Grounding Directionality Axiom. Unlike undirected graphs that merely denote association, our model enforces a strict semantic orientation:

For any directed edge $u \xrightarrow{\ell, r} v$, the entity u is ontologically, structurally, or causally grounded in entity v . Thus, v provides the necessary context, constraint, or origin for the existence or role of u .

Formally, we define the dependence as a non-symmetric relation where the target node v acts as the *independent variable* (the context) and the source node u acts as the *dependent variable* (the entity situated within that context). This ensures that traversing the graph in the direction of the edges leads toward broader systemic structures (e.g., a Character \rightarrow Organization \rightarrow Politics).

2.3. The Hexapartite Layering System

To disambiguate the multifaceted nature of narrative relations, E is partitioned into six functional layers. Each layer ℓ governs specific types of $u \rightarrow v$ dependencies:

1. **Structural Layer (L_{str}):** Defines spatial containment and institutional hierarchies.

Examples: Place_A \rightarrow Place_B (inclusion); Race \rightarrow Local (nativity).

2. **Causal Layer (L_{cau}):** Encodes mechanical triggers, technical materialization, and productive shifts.

Example: Technology \rightarrow Economy (transformation).

3. **Temporal Layer (L_{tem}):** Manages the flow of time and chronological anchoring.

Example: Event \rightarrow Chronology (temporal marking).

4. **Social Layer (L_{soc}):** Captures interpersonal bonds, group memberships, and legitimacy.

Example: Figure \rightarrow Organization (membership).

5. **Ontological Layer (L_{ont}):** Represents metaphysical laws, innate affinities, and system usage.

Example: Figure \rightarrow MagicSystem (usage).

6. **Symbolic Layer (L_{sym}):** Maps the flow of meaning, cultural values, and ideological influence.

Example: Religion \rightarrow Culture (molding).

2.4. Consistency and Domain-Range Constraints

To maintain the integrity of the WorldPT, we enforce a strict schema where each relation $r \in \mathcal{R}$ has a defined domain $\text{Dom}(r) \subseteq \mathcal{T}_V$ and range $\text{Ran}(r) \subseteq \mathcal{T}_V$. An edge (u, v, ℓ, r) is valid if and only if $\tau(u) \in \text{Dom}(r)$ and $\tau(v) \in \text{Ran}(r)$. This prevents "semantic noise", such as a *Place* being a member of a *Figure*, ensuring that the dataset is machine-verifiable.

3. Dataset Architecture

The WorldPT dataset is engineered to bridge the gap between creative worldbuilding and formal knowledge representation. The complete dataset used in this study is publicly available through the Zenodo repository (DOI: [10.5281/zenodo.18989735](https://doi.org/10.5281/zenodo.18989735)) [13]. Unlike traditional NLP datasets that rely on unstructured text, WorldPT implements a Graph-Object-Oriented (GOO) architecture, where narrative concepts are decoupled into atomic, machine-readable JSON entities. In its current version, the WorldPT repository contains a total of six fully catalogued fictional universes. Each world is structured according to the same ontological schema and directory hierarchy, allowing comparative analyses

across different narrative settings. This multi-world design enables the dataset to function not only as a single-case structural model, but also as a scalable framework for cross-universe narratological studies. Future releases aim to expand this number to include additional literary and ludic universes.

3.1. Data Sources and Curation

The WorldPT dataset was constructed through a manual data curation process. Primary information was collected from publicly available franchise wikis and community-maintained encyclopedic repositories dedicated to each fictional universe. Entities were systematically identified and categorized according to the dataset ontology, including characters, locations, organizations, races, artifacts, technologies, cultural systems, and major narrative events. Each entity was manually reviewed and structured into the dataset schema to ensure semantic consistency and ontological alignment. Manual curation was intentionally adopted in order to preserve narrative fidelity. Fictional worlds frequently contain complex contextual relationships that are difficult to extract through automated scraping pipelines alone. Human-guided structuring therefore ensured that entity descriptions and classifications accurately reflect the canonical lore of each universe.

3.2. Ontological Taxonomy: The 13 Pillars

The selection of our 13 ontological categories (\mathcal{T}_V) is not arbitrary; it follows a "Full-Spectrum Narrative Coverage" principle. These categories were chosen to minimize semantic overlap while ensuring that any element of a fictional universe has a dedicated class.

- Agents and Groups (Figure, Race, Organization): Define the *Who*. These categories handle agency and social stratification.
- Spatial and Temporal Anchors (Place, Event, Chronology): Define the *Where* and *When*. They provide the structural and linear backbone of the world.
- Societal Systems (Culture, Religion, Politics, Economy): Define the *How* of human (or non-human) interaction, governing the flow of symbols, power, and resources.
- Logical Constraints (MagicSystem, Technology, Artifact): Define the *Laws* of the world. These categories provide the "hard" rules that prevent narrative inconsistency.

3.3. The Multi-Parametric Relationship System

The core strength of WorldPT lies in its directed relationship model. In standard Knowledge Graphs (KGs), a triple (u, r, v) is usually binary. In WorldPT, every edge $e \in E$ is a multi-parametric object defined as:

$$e = (u, v, \ell, \omega, \kappa, r) \quad (1)$$

Where:

- Layer (ℓ): The hexapartite semantic layer ($L_{str}, L_{cau}, L_{tem}, L_{soc}, L_{ont}, L_{sym}$) that disambiguates the context.
- **Weight** ($\omega \in [0, 1]$): Represents the **intensity** or importance of the connection. For instance, a character's relationship with a "King" might have a higher weight in the L_{soc} layer than with a "Guard."
- **Reliability** ($\kappa \in [0, 1]$): A measure of **epistemic certainty**. This allows for the modeling of "Unreliable Narrators" or legends—where a connection is documented in the lore but may not be factually true within the world's internal logic.
- **Relation Type** (r): A specific string descriptor (e.g., "subordinate_to", "fueled_by").

This allows for complex queries, such as: "Find all figures connected to the Religion layer with high reliability but low social weight."

3.4. Physical Data Structure and Serialization

The physical organization of WorldPT follows a "Filesystem-as-Schema" approach. This ensures that the ontological class of an entity is intrinsically tied to its storage location, facilitating rapid filtering and batch processing without the overhead of a database engine. The dataset is rooted in a main directory, branching into specific universes, each containing 13 specialized sub-folders.

The directory structure is formalized in Table 1.

Table 1. Directory Hierarchy and Ontological Mapping of the WorldPT.

Directory (Sub-folder)	Ontological Class	Content Description
artefatos/	Artifact	Physical objects with specific properties or lore.
figuras/	Figure	Characters, NPCs, and historical agents.
linhas_cronologicas/	Chronology	Temporal markers and sequential history.
culturas/	Culture	Customs, traditions, and social norms.
economias/	Economy	Trade systems, currencies, and resource flows.
eventos/	Event	Singular occurrences that alter the world-state.
locais/	Place	Geographical, spatial, or architectural nodes.
sistemas_magia/	MagicSystem	Metaphysical laws and supernatural rules.
organizacoes/	Organization	Guilds, factions, and institutional groups.
politicas/	Politics	Power structures and governance models.
racas/	Race	Biological or species-based classifications.
religioes/	Religion	Belief systems and theological structures.
tecnologias/	Technology	Scientific advancements and technical tools.

Each entity is serialized as a JSON object, ensuring interoperability with modern LLM pipelines and graph databases. The directory structure mirrors the ontology, acting as a "physical index" of the graph:

Listing 1: Entity Example

```

1 {
2   "id": "frodo_bolseiro",
3   "mundo": "tolkien",
4   "genero": [
5     "fantasia",
6     "aventura"
7   ],
8   "titulo": "Frodo Bolseiro",
9   "tipo_conteudo": "figura",
10  "texto": "Frodo, filho do respeitavel Drogo Bolseiro e Primula Brandebuque
11    , nasceu em 22 de setembro de 2968, ano da morte de T.A. Apos a morte
12    de seus pais em um acidente de barco em 2980...",
13  "fonte": "livro",
14  "e_gerado_por_IA": false,
15  "modelo_IA": null,
16  "relacionamentos": [
17    {
18      "alvo_id": "o_um_anel",
19      "alvo_tipo": "artefato",
20      "tipo": "possui",
21      "layer": "structural",
22      "peso": 0.9,
23      "confianca": 0.95,
24      "metodo": "manual"
25    },
26    {
27      "alvo_id": "guerra_do_anel",

```

```

26         "alvo_tipo": "evento",
27         "tipo": "participa_de",
28         "layer": "social",
29         "peso": 0.8,
30         "confianca": 0.9,
31         "metodo": "manual"
32     },
33     {
34         "alvo_id": "gandalf",
35         "alvo_tipo": "figura",
36         "tipo": "aliado_de",
37         "layer": "social",
38         "peso": 0.8,
39         "confianca": 0.9,
40         "metodo": "manual"
41     }
42     ...
43 ]
44 }

```

3.5. Architectural Advantages

The WorldPT structure offers three critical advantages over traditional narrative datasets:

1. **Modular Decoupling:** Since each entity is a standalone file, worlds can be "mixed and matched". A user can import a *MagicSystem* from World A into World B without breaking the graph's integrity, provided the IDs are mapped.
2. **Algorithmic Verifiability:** The strict schema allows for automated "Inconsistency Detection". If a *Figure* is connected to a *Place* via an L_{cau} (Causal) layer instead of L_{str} (Structural), the system can flag a potential category error.
3. **LLM-Native Grounding:** By providing a "Description" field, we give LLMs a natural language context to understand the node's semantics, while the "Relationships" field provides the symbolic structure for RAG (Retrieval-Augmented Generation) applications.

4. Network-Theoretic Narrative Metrics

To evaluate the structural health and complexity of fictional universes within the WorldPT, we propose a suite of quantitative proxies based on Social Network Analysis (SNA) and Multilayer Network Theory. These metrics allow researchers to move beyond qualitative literary critique toward a verifiable "narrative diagnostic".

4.1. Ontological Authority: Weighted PageRank

We adapt the PageRank algorithm [14] to identify entities that serve as the "logical anchors" of a world. In our directed graph, a high PageRank does not merely denote popularity, but Ontological Authority: an entity whose existence is fundamental to the grounding of many others.

Given the weighted nature of WorldPT edges, the score for a node u is defined as:

$$PR(u) = \frac{1-d}{N} + d \sum_{v \in \mathcal{N}_{in}(u)} \frac{w_{vu} \cdot PR(v)}{L_{out}(v)} \quad (2)$$

where d is the damping factor (typically 0.85), w_{vu} is the product of base relation strength and information reliability, and $L_{out}(v)$ is the weighted out-degree of node v . Entities with high PR are critical; their removal would cause the most significant "ontological collapse" in the narrative.

4.2. Narrative Density and Layer Specificity

The connectivity of a world is measured through its density ρ . While global density ρ_G provides an overview of how "claustrophobic" or "expansive" a world is [10], we further decompose this by the hexapartite layers:

$$\rho_\ell = \frac{|E_\ell|}{|V|(|V| - 1)} \quad (3)$$

By comparing ρ_{soc} (Social) with ρ_{cau} (Causal), we can mathematically distinguish between a "character-driven drama" (high social density) and a "hard magic/technical setting" (high causal density).

4.3. Structural Robustness: Bridges and Bottlenecks

A narrative is robust if it possesses multiple redundant paths of meaning. We identify Critical Bridges—edges $e \in E$ such that the removal of e increases the number of connected components in the graph. In worldbuilding terms, these bridges represent "single points of failure" in the lore: connections that, if severed, would leave entire plot arcs or regions isolated and disconnected from the main world-logic.

4.4. Multidimensionality: Cross-Layer Coupling (CLC)

A unique contribution of the WorldPT is the Cross-Layer Coupling (CLC) metric. This measures the versatility of an entity across the hexapartite system, a concept derived from multiplex network coupling [15]. The CLC for a node u is defined as the cardinality of the set of layers in which u participates:

$$CLC(u) = |\{\ell \in \mathcal{L} \mid \exists v \in V, (u, v, \ell) \in E \vee (v, u, \ell) \in E\}| \quad (4)$$

A high CLC score (max 6) identifies "multidimensional" entities such as a King who is simultaneously a political leader (L_{str}), a social figurehead (L_{soc}), and a religious symbol (L_{sym}). This metric allows for the automated identification of protagonists or central world-shifting elements. Figure 1 illustrates the theoretical architecture of the proposed hexapartite system. By calculating the vertical degree of a node (the dashed lines), we can quantitatively separate foundational world-building elements from superficial narrative actors.

The diagram visually confirms the quantitative findings. The horizontal planes show the intra-layer connections; the high density of the Social Layer is immediately apparent compared to the sparser Causal or Temporal layers, visually validating the "Social Drama" genre classification.

Crucially, the vertical dashed lines represent the ontological grounding in action. These lines act as the structural "glue" of the fictional universe. Entities with high CLC scores are not merely as nodes on a plane, but as vertical pillars transpiercing the entire system. Without these multi-layered entities anchoring the structure, the social narrative would risk disconnecting from the metaphysical laws (L_{ont}) and structural constraints (L_{str}), leading to the aforementioned "ontological drift".

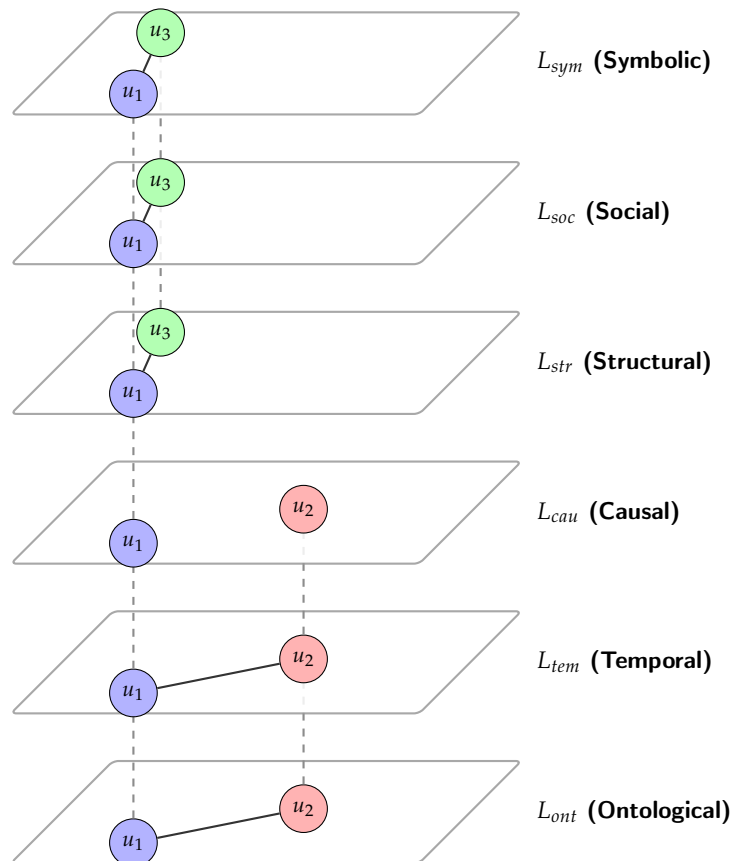


Figure 1. Generic visualization of the Hexapartite Multiplex framework. The solid lines represent intra-layer semantic relationships, while the dashed vertical lines denote the Cross-Layer Coupling (CLC). Entity u_1 represents a highly grounded, multidimensional node ($CLC = 6$), anchoring the narrative from its physical/ontological rules up to its symbolic meaning. Entity u_2 is restricted to base mechanics ($CLC = 3$), and entity u_3 exists purely as a higher-level superstructural construct ($CLC = 3$).

4.5. Topological Small-Worldness and Assortativity

Finally, we calculate the average shortest path length L to determine the "degrees of separation" within the lore [16]. A "Small-World" narrative (low L) suggests that a change in one part of the world (e.g., a magic system shift) will rapidly propagate through the entire system. Furthermore, we measure **Typological Assortativity** to detect segregation [17]. A high assortativity coefficient indicates a "siloed" world where entities only interact with their own kind (e.g., Places only connecting to Places), whereas negative assortativity reveals a rich, heterogeneous ecosystem where figures, technologies, and ideologies are deeply intertwined.

5. Case Study

In this section, we present the empirical findings obtained by applying the WorldPT analysis framework to the case study universe: *Tolkien's Middle-Earth lore*.

5.1. Network Topology and Visual Distribution

The global structure of the narrative was first visualized to identify clusters and isolated components. As shown in Figure 2, the world exhibits a centralized topology, with visible narrative cores that are slightly dispersed.

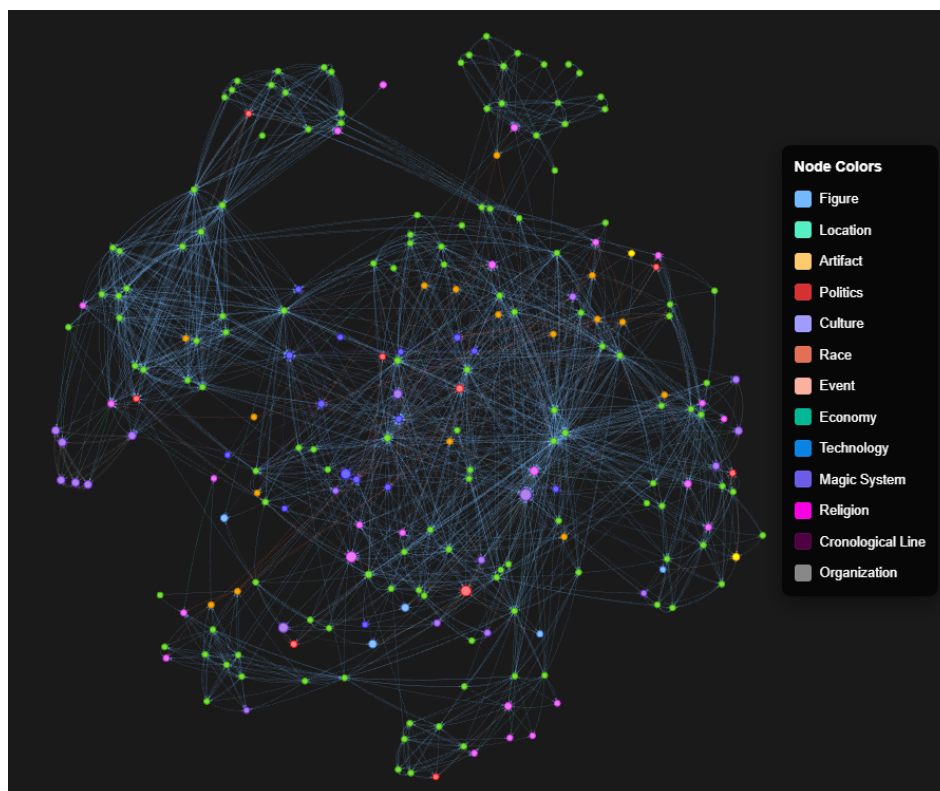


Figure 2. Full Systemic Map of Tolkien lore. Node size represents degree centrality; colors represent ontological categories.

5.2. Global Network Statistics

The graph reconstruction of *Tolkien lore* resulted in a complex network with a total of $N = 141$ nodes and $E = 520$ directed relationships.

The global topological indicators presented in Table 2 reveal intrinsic structural properties that underpin the narrative cohesion of the *Tolkien* universe. The small-world phenomenon, evidenced by an Average Path Length (L) of only 2.68 steps, demonstrates remarkable causal efficiency, indicating that the propagation of events between distinct ontological layers occurs almost immediately. This average proximity suggests that the system lacks isolated narrative peripheries; on the contrary, any entity, however obscure, is less than three connections away from the central cores of authority, ensuring a systemic reactivity that is vital for maintaining tension in investigative horror settings.

A Global Density of 0.0477, while numerically low, characterizes a sparse yet highly functional network. In complex worldbuilding systems, excessive density would result in informational redundancy and the loss of mystery, whereas the density observed provides an ideal balance between systemic consistency and the cognitive gaps necessary for exploration and discovery. Complementarily, a Network Diameter of 6 steps establishes the network's maximum separation limit, confirming that the scope of the setting is finite and controlled, thus preventing narrative drift where sub-plots lose relevance to the central canon.

The Typological Assortativity ($r = 0.2087$) offers the deepest insight into the world's social and ontological organization. The positive value indicates a trend of assortative mixing, where entities of similar categories tend to establish connections with one another, forming well-defined nuclei of identity. However, the moderate magnitude of this coefficient reveals that the world is not segregated into hermetic silos, there is a structural porosity that allows for integration between human agents, cursed locations, and metaphysical laws. Finally, the existence of a Total Connected Component of 1 confirms the ontological integrity of the analyzed universe, mathematically proving that all lore elements belong to a single, unified reality structure, devoid of logical fragments or isolated systemic contradictions.

Table 2. Global Topological Metrics for the Tolkien Case Study.

Metric	Observed Value
Average Path Length (L)	2.68 steps
Network Diameter	6 steps
Total Connected Components	1 (Fully Connected)
Global Density	0.0477
Typological Assortativity (r)	0.2087

The observed assortativity of 0.2087 indicates a slight assortative mixing. This suggests that while different ontological classes (e.g., Figures and Locations) are interconnected, there is a subtle tendency for entities of the same type to cluster.

5.3. Dominance and Narrative Trade-Offs

To identify the core elements of the world's logic, we analyzed the interplay between absolute authority and topological versatility, as synthesized in Figure 3.

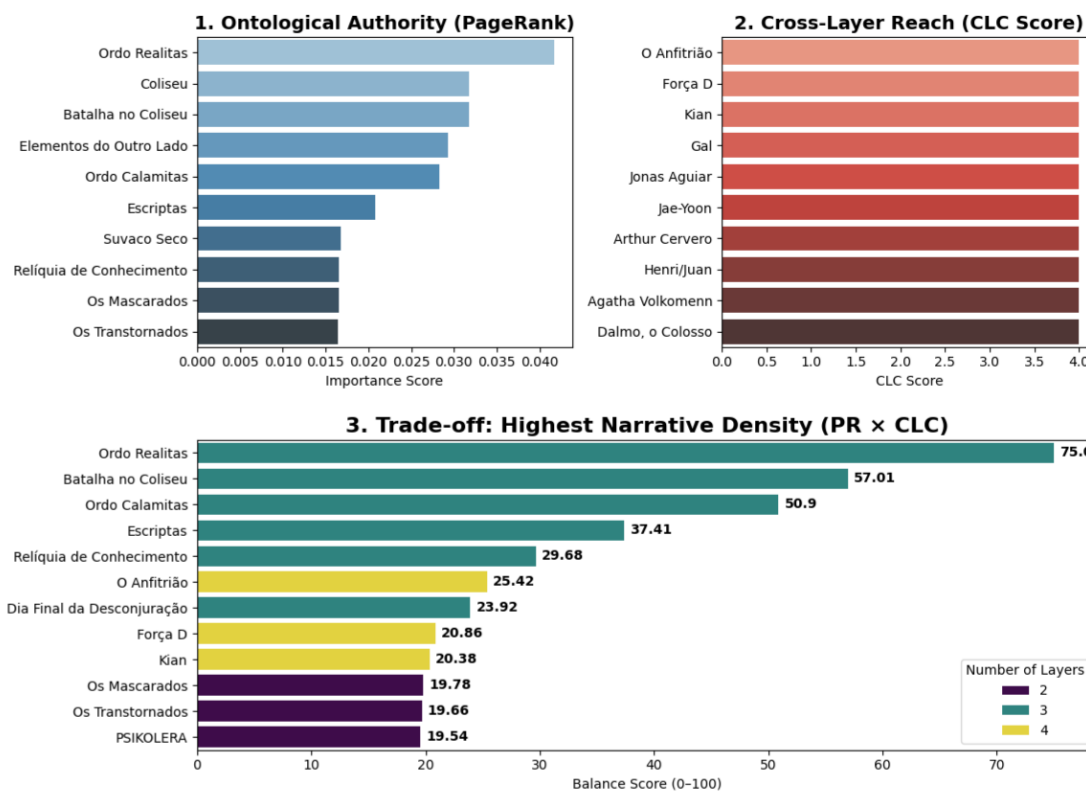


Figure 3. Analysis of Narrative Authority and Versatility: (1) Ontological Authority via PageRank; (2) Cross-Layer Reach via CLC Score; (3) Narrative Density Trade-off ($PR \times CLC$).

As shown in Panel 1, *Terra-Média* (*Middle-Earth*) ($PR=0.08$) serves as the absolute gravity center. However, the Cross-Layer Coupling (CLC) metric (Panel 2) reveals a different layer of importance. Entities like *Guerra do Anel* (*War of the Ring*), *Sauron*, and *Anéis de Poder* (*Power Rings*) achieve the maximum CLC score of 4, indicating they are "Lore-shifters" that connect Social, Ontological, Structural, and Causal layers simultaneously.

The Narrative Density Trade-off (Panel 3) demonstrates that the most "balanced" entities are those that combine institutional weight with cross-layer presence. *Terra-Média* (*Middle-Earth*) maintains the highest equilibrium (75.0), but the yellow bars in the chart highlight that elements like *Guerra do Anel* (*War of the Ring*) and *Sauron* possess superior versatility relative to their PageRank, acting as systemic pivots that bridge otherwise disparate narrative sectors.

5.4. Layer Density and Genre Fingerprinting

The nature of the universe is further defined by its categorical and semantic distribution, illustrated in Figure 4. The Edge Distribution by Layer (Panel 4) reveals that the Structural (L_{str}) layer accounts for 48.7% of connections. This dominance is explained by the Ontology Category Composition (Panel 5), where *Location* is the most frequent category.

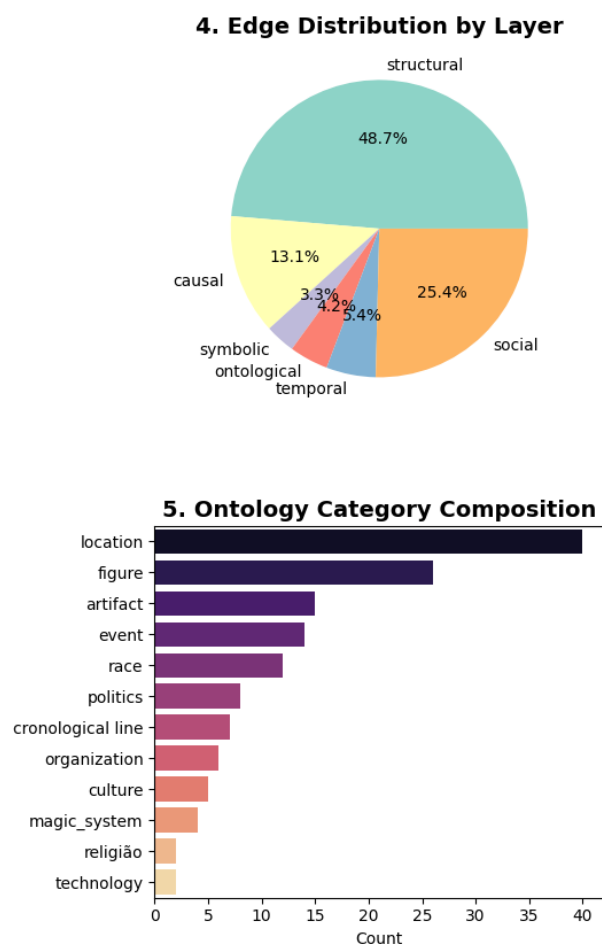


Figure 4. Ontological and Semantic DNA: (4) Edge Distribution by Hexapartite Layer; (5) Categorical Composition of the World-Graph.

This quantitative profile classifies the Tolkien universe as a Structurally-Driven Mythological Setting. The high Structural density (48.7%), combined with a significant Causal component (13.1%), indicates that the narrative world is primarily organized around ontological hierarchies, cosmological structures, and rule-based relationships between entities. Social interactions (25.4%) are relevant but do not dominate the relational landscape, suggesting that character dynamics operate within a broader mythological and systemic framework rather than serving as the primary driver of narrative tension.

5.5. Ontological Interaction Patterns

While the layer distribution identifies the "macro" genre of the world, the *Ontological Interaction Matrix* (Figure 5) reveals the "micro" dynamics of how these categories exchange information. This matrix is essential to verify if the world-building is balanced or concentrated in specific semantic silos.

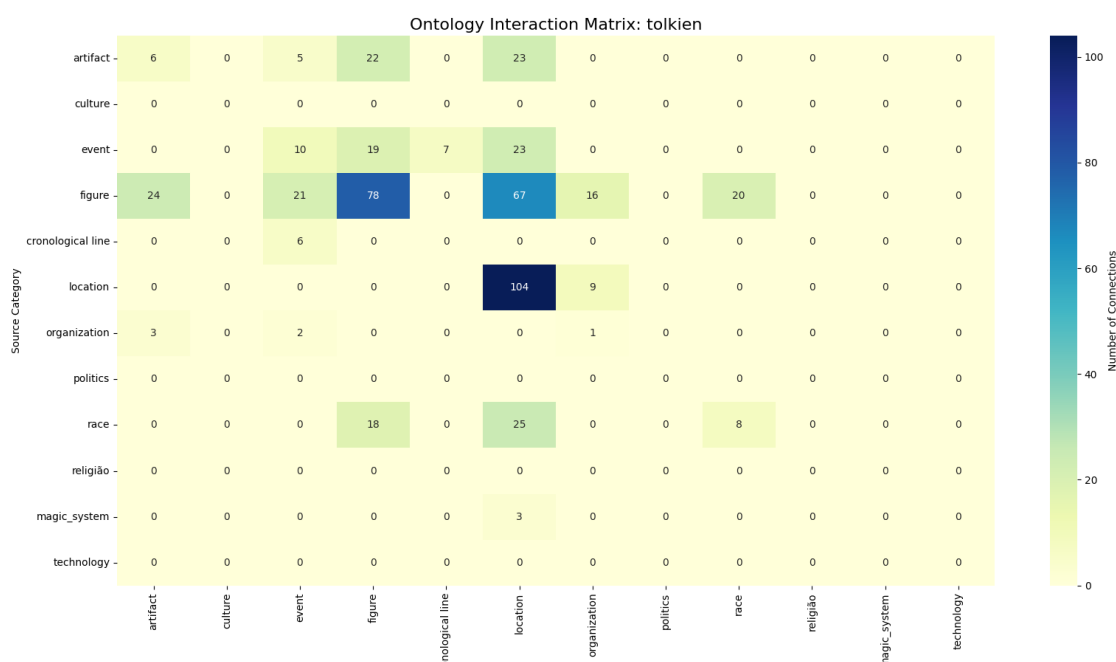


Figure 5. Ontological Interaction Heatmap. The color intensity and numerical values represent the volume of directed edges between origin (y-axis) and target (x-axis) categories.

The heatmap reinforces the structural profile identified in the quantitative analysis. The dominant interaction in the system is **Location** → **Location** (104), indicating that spatial and geographic relationships constitute a central organizing principle of the Tolkien universe. Rather than being driven primarily by interpersonal dynamics, the network reveals a strong emphasis on the configuration of places and the spatial structure of the world.

Character-related interactions remain significant but appear embedded within this broader structural framework. The interactions **Figure** → **Figure** (78) and **Figure** → **Location** (67) form the secondary layer of the network, suggesting that characters operate mainly as agents moving within an already well-defined geographic and ontological landscape. Additional contributions from **Race** → **Location** (25) and **Race** → **Figure** (18) further highlight the importance of lineage and group identity in structuring relationships.

Finally, the relatively sparse presence of categories such as Culture, Religion, and Technology indicates that these elements play a limited role in generating relational connections in the current dataset. Overall, the interaction matrix depicts a world where spatial structure, lineage, and hierarchical relations form the backbone of the narrative environment.

6. Conclusions

In this paper, we introduced WorldPT, a graph-centric framework and dataset designed to move computational narratology beyond the limitations of unstructured text and towards formal topological grounding. By modeling fictional universes as Directed Multilayer Attributed Graphs, we provided a rigorous alternative to the "ontological drift" often observed in large language models. The introduction of the Grounding Directionality Axiom and the hexapartite layering system allows for a machine-verifiable representation of narrative consistency, ensuring that every entity is semantically situated within a broader systemic context.

Our case study on Tolkien universe demonstrated the empirical utility of our Social Network Analysis (SNA) metrics. The discovery of a "Small-World" topology ($L = 2.68$) and a dominant structural layer (48.7%) provided a quantitative "genre fingerprint," classifying the setting as a Structurally-Driven story. Furthermore, the Cross-Layer Coupling (CLC) metric successfully identified systemic

pivots, entities like *Guerra do Anel (War of the Ring)* and *Sauron*, whose high multidimensionality makes them more topologically versatile than traditional narrative anchors. This proves that WorldPT can effectively distinguish between institutional authority and narrative agency through objective graph metrics.

The choice of Portuguese as the primary language for WorldPT is a strategic contribution to the democratization of AI resources. By providing a high-quality structural benchmark for the Lusophone community, we enable researchers and creators to develop narrative agents and procedural worldbuilding tools that are linguistically native and ontologically stable.

Future work will focus on the integration of WorldPT with Retrieval-Augmented Generation (RAG) pipelines, exploring how structured graph grounding can mitigate hallucinations in long-form storytelling. Additionally, we intend to expand the repository with diverse literary and ludic universes, refining our "Lore Collapse" diagnostic tools to aid authors in identifying structural vulnerabilities before they manifest as plot holes. Ultimately, WorldPT provides the necessary scaffolding for a new generation of narrative systems where the art of worldbuilding is as mathematically sound as it is creatively expansive.

References

1. Marcus, G. The next decade in AI: four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177* **2020**.
2. Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Proceedings of the Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 2021, pp. 610–623.
3. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Philip, S.Y. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems* **2021**, *33*, 494–514.
4. Ponti, E.M.; Glavaš, G.; Majumder, O.; Bansal, M.; Vulić, I.; Korhonen, A. Towards multilingual commonsense reasoning. *arXiv preprint arXiv:1910.01185* **2019**.
5. Meghini, C.; Bartalesi, V.; Metilli, D. Narrative as a formal object. *PeerJ Computer Science* **2021**, *7*, e540.
6. Gervás, P. Computational approaches to storytelling and creativity. *Fundamenta Informaticae* **2014**, *129*, 117–138.
7. Ammanabrolu, P.; Riedl, M. Story Realization: Expanding Plot Events into Sentences. In Proceedings of the Proceedings of the 13th International Conference on Natural Language Generation, 2020.
8. Moretti, F. Network Theory, Plot Analysis. *Stanford Literary Lab Pamphlets* **2011**, *2*.
9. Labatut, V.; Bost, X. Extraction and analysis of fictional character networks: A survey. *ACM Computing Surveys (CSUR)* **2019**, *52*, 1–40.
10. Bearman, P.S.; Stovel, K. The Structure of Storytelling. *Theory and Society* **2004**.
11. Arp, R.; Smith, B.; Spear, A.D. *Building ontologies with basic formal ontology*; MIT Press, 2015.
12. Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J.P.; Moreno, Y.; Porter, M.A. Multilayer networks. *Journal of complex networks* **2014**, *2*, 203–271.
13. Souza, L.V.; Lusquino Filho, L. WorldPT: A Structured Dataset of Fictional Worlds in Portuguese, 2026. DOI: 10.5281/zenodo.18989735.
14. Brin, S.; Page, L. The anatomy of a large-scale hypertextual Web search engine. In Proceedings of the Proceedings of the seventh international conference on World Wide Web, 1998, pp. 107–117.
15. Battiston, F.; Nicosia, V.; Latora, V. Structural measures for multiplex networks. *Physical Review E* **2014**, *89*, 032804.
16. Watts, D.J.; Strogatz, S.H. Collective dynamics of 'small-world' networks. *nature* **1998**, *393*, 440–442.
17. Newman, M.E. Mixing patterns in networks. *Physical review E* **2003**, *67*, 026126.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.