Computer Law & Security Review 61 (2026) 106280

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

journal homepage: www.elsevier.com/locate/clsr

Factor extraction from pretrial detention decisions by Italian and Brazilian Supreme Courts: A knowledge graph perspective☆

Thiago Raulino Dal Pont [a,b,*], Marco Billi [a], Isabela Cristina Sabo [c], Francesca Lagioia [a], Jomi Fred Hübner [b], Giovanni Sartor [a], Aires José Rover [c]

[a] CIRSFID - Alma AI, Alma Mater Studiorum, University of Bologna, Bologna, Italy
[b] Department of Automation and Systems Engineering, Federal University of Santa Catarina, Florianópolis, Brazil
[c] Department of Law, Federal University of Santa Catarina, Florianópolis, Brazil

ARTICLE INFO

Keywords:
Pretrial detention
Information extraction
Legal factor extraction
Knowledge graph
Explainable AI (XAI)
Judicial decisions

ABSTRACT

Pretrial detention challenges the presumption of innocence, requiring courts in countries such as Brazil and Italy to justify the deprivation of liberty based on specific legal factors. However, the volume and complexity of judicial decisions make manual analysis of the underlying reasoning difficult. We propose a factor-guided extraction pipeline combining Large Language Models (LLMs) and Knowledge Graphs (KGs) to extract, structure, and visualize critical legal factors from appeal cases in the Supreme Court. Building on datasets from both countries, we employ LLMs to generate summaries constrained by an expert-defined factor schema. These summaries are mapped into KGs to visualize the relationships between legal factors and judgment outcomes. We evaluated this approach against a raw-text baseline using qualitative expert review and quantitative metrics. Results demonstrate that factor-guided prompting significantly improves extraction quality, achieving higher semantic similarity (LabSE scores ≈ 0.47 vs. 0.28) compared to unguided methods. Structurally, the KGs reveal convergent reasoning complexity across the two legal systems, despite distinct topological signatures. These findings suggest that integrating LLMs with structured legal knowledge enhances the interpretability of judicial reasoning, offering a robust, scalable tool for legal analytics.

1. Introduction

Pretrial detention is a contested measure in various legal systems, as it challenges the fundamental principle of the presumption of innocence by depriving defendants of their liberty at an early stage of the legal process. This deprivation can only be justified by the need to prevent the defendant from absconding, committing further offenses, or interfering with ongoing investigations. Courts must rely on specific factors, to be assessed on a case-by-case basis, to determine whether pretrial detention is warranted. In Italy and Brazil, for instance, pretrial detention has been the subject of intense public and political debate, with concerns about overuse, inconsistency, and lack of transparency in judicial reasoning.

In the legal domain, efficiently analyzing judicial decisions is critical, as these decisions are often lengthy and complex, making them difficult to navigate. Information extraction plays a crucial role in expediting the comprehension of these documents by automatically identifying key facts, legal principles, and outcomes. However, manually generating these structured representations is time-consuming, highlighting the need for automated approaches. Thus, our aim in this paper is to provide a methodology for automated extraction of relevant knowledge at scale, enabling the creation of high-quality datasets. Such datasets can support descriptive analyses of judicial behavior, predictive modeling of outcomes, and provide explanations for these predictions based on the underlying factors.

Generative AI has emerged as a promising solution for automating information extraction. This AI technology encompasses a wide range of content generation capabilities, including the creation of text, images, audio, computer code, and beyond. Its name stems from its ability to produce diverse outputs, making it applicable to various tasks such as image creation, text generation, and much more [1]. Large Language Models (LLMs) are a type of Generative AI model that is specifically trained on large corpora of text data, designed primarily

☆ This article is part of a Special issue entitled: 'JUSMOD23-SI' published in Computer Law & Security Review: The International Journal of Technology Law and Practice.

* Corresponding author at: CIRSFID - Alma AI, Alma Mater Studiorum, University of Bologna, Bologna, Italy.
E-mail addresses: thiagordalpont@gmail.com (T.R. Dal Pont), marco.billi3@unibo.it (M. Billi), isabelasabo@gmail.com (I.C. Sabo), francesca.lagioia@unibo.it (F. Lagioia), jomi.fred@ufsc.br (J.F. Hübner), giovanni.sartor@unibo.it (G. Sartor), aires.rover@ufsc.br (A.J. Rover).

https://doi.org/10.1016/j.clsr.2026.106280

Available online 14 March 2026

for generating and understanding human language. They are employed to extract meaning from human language and can be used for various tasks such as sentiment analysis, named entity recognition, information extraction, text summarization, and others [2,3].

Particular attention is dedicated to Knowledge Graphs (KGs) in this paper. A KG is a structured representation of facts consisting of entities, relationships, and semantic descriptions. Entities can be real-world objects or abstract concepts, while relationships describe the connections between them, containing types and properties with well-defined meanings [4].

From that perspective, we propose a methodology that combines the generative capabilities of LLMs with the structural clarity of KGs. By leveraging raw text and a structured dataset of factors from legal decisions in Brazil and Italy on appeals, this work aims to extract and represent the critical elements that influence judicial reasoning. Unlike broad automated approaches, we leverage a "human-in-the-loop" strategy: we utilize expert-defined legal factors to guide the LLM, ensuring the extraction focuses on critical legal elements rather than procedural noise.

Specifically, we evaluate whether the use of LLMs in conjunction with structured data (i) improves the quality of knowledge graphs and (ii) enables judicial reasoning to be visualized. To this end, we built datasets of Italian and Brazilian judicial decisions, as detailed in Section 3. Section 4 describes the methodology applied, in particular for summarization and KG creation. Section 5 reports the experiments and discusses the results, evaluating the results from both a qualitative and quantitative point of view. Section 6 concludes and outlines possible future research.

This work follows recent attempts at explaining decision-making systems through relevant factors and at extracting factors from judicial decisions [5–8]. In particular, it follows previous research which applied supervised learning and eXplainable AI (XAI) techniques to explain judgments in the same legal domain [9,10].

2. Related works

Recent literature has extensively explored the integration of Large Language Models (LLMs) with Knowledge Graphs (KGs) to enhance information extraction. While interest in combining LLMs with structured representations is growing in the legal domain [11], few works explicitly address the generation of KGs from legal texts using controlled, expert-guided pipelines.

In non-legal domains, prompt engineering and zero-shot learning have shown promise for converting unstructured text into KGs. Bellan et al. [12] demonstrated that high-quality prompts significantly improve the extraction of process knowledge, while Zhao et al. [13] and Dong and Wu [14] highlighted the utility of hybrid neural-symbolic approaches for reducing reliance on labeled data in industrial contexts. Furthermore, De Bellis [15] noted that LLM latent representations align with structured KGs, suggesting potential for implicit conceptual modeling, though the lack of explicit control remains a limitation for high-stakes domains.

Within the legal field, LLMs are widely applied to summarization, reasoning, and decision support [16–20]. More recently, studies have begun integrating LLMs with legal ontologies to enhance explainability. For instance, Shi et al. [21] showed that explicit knowledge integration significantly improves legal question answering. However, most existing approaches prioritize direct QA or summarization, leaving systematic extraction for KG construction underexplored. Historically, hybrid approaches combining symbolic knowledge with statistical processing date back to early AI research [22,23], yet these methods often relied on manual engineering, limiting their scalability.

We position our work within this emerging hybrid methodology. Unlike fully automated pipelines that attempt to discover entities from scratch, we adopt an expert-guided strategy. By grounding the LLM with a predefined factors table, we ensure the output is a controlled representation that supports robust KG construction. This design choice aims to improve precision compared to generic prompting and provides explainable inputs for legal analytics, addressing a key methodological gap in cross-jurisdictional legal tech.

3. Datasets

This study utilizes four datasets developed in our previous work. Two of them are composed of unstructured judicial decisions on appeals collected from the Brazilian and Italian Supreme Courts (982 and 718 documents, respectively). The remaining two datasets (one Brazilian and one Italian) are structured counterparts, constructed using clustering techniques for extracting relevant factors (F).

Legal experts analyzed the resulting clusters to identify recurring legal factors and subsequently manually annotated each document [citation redacted for blind review].

The extracted variables were categorized into four groups: the *grounds* of appeal (defense arguments) and decision *reasons* (judicial arguments) related to the requirements of Articles 311-316 of the Brazilian Code of Criminal Procedure and Articles 272-315 of the Italian Code of Criminal Procedure (F); the category of committed *crime*, aligned between the two penal codes (C); metadata variables (*Location*, *Judge Rapporteur*, and *Date*); and the *Prisoner Status* (target variable).

- **Grounds and Reasons (F) (binary variables). Whether the judgment addresses:**
  - $F1$: Excess of time in prison.
  - $F2$: Suspension of time in prison or suspension of proceedings.
  - $F3$: Nullity of the interrogation or hearing of the accused.
  - $F4$: Connection between different crimes or proceedings.
  - $F5$: Remittance of proceedings to a previous stage or another jurisdiction.
  - $F6$: Presence of wiretaps.
  - $F7$: Complexity of the proceedings or flight risk.
  - $F8$: Inferred but unproven facts (e.g., behavioral aggression inferred from evidence).
  - $F9$: The victim's statement.
  - $F10$: *Flagrante delicto* (caught in the act).
  - $F11$: Defense restriction (e.g., lack of access to prosecution documents).
- **Crime categories (C) (binary variables). Whether the committed crime was:**
  - $C1$: Crimes against a person (including sexual crimes).
  - $C2$: Crimes against property.
  - $C3$: Public safety crimes (including firearms law violations).
  - $C4$: Crimes against the government, justice administration, or public economy.
  - $C5$: Crimes provided in special laws.
  - $C6$: Related to criminal organizations.
  - $C7$: Crimes provided in drug laws.
- **Others (categorical variables):**
  - Location: State or regional capital where the crime occurred.
  - Judge Rapporteur: Name of the judge responsible for reporting the case.
  - Date: Year the decision was issued.
- **Prisoner status (Target - binary variable):** Indicates whether the court ordered the release of the accused. We split each corpus based on this variable: judgments favoring the defendant (release) versus those favoring the prosecution (detention). The Italian corpus comprises 104 release orders and 614 detentions; the Brazilian corpus contains 282 and 700, respectively.

file	year	factor_1	factor_2	factor_3	crime_1	crime_2	proposed by	prisoner status
005	2021	VERO	VERO	VERO	VERO	VERO	D	not released

Fig. 1. Template of the Factor Table Structure. (Note: This figure illustrates the boolean schema used for annotation. In the running example analyzed in Section 5, specific factors such as $F1$ and $F7$ take 'TRUE' values based on the case facts.).

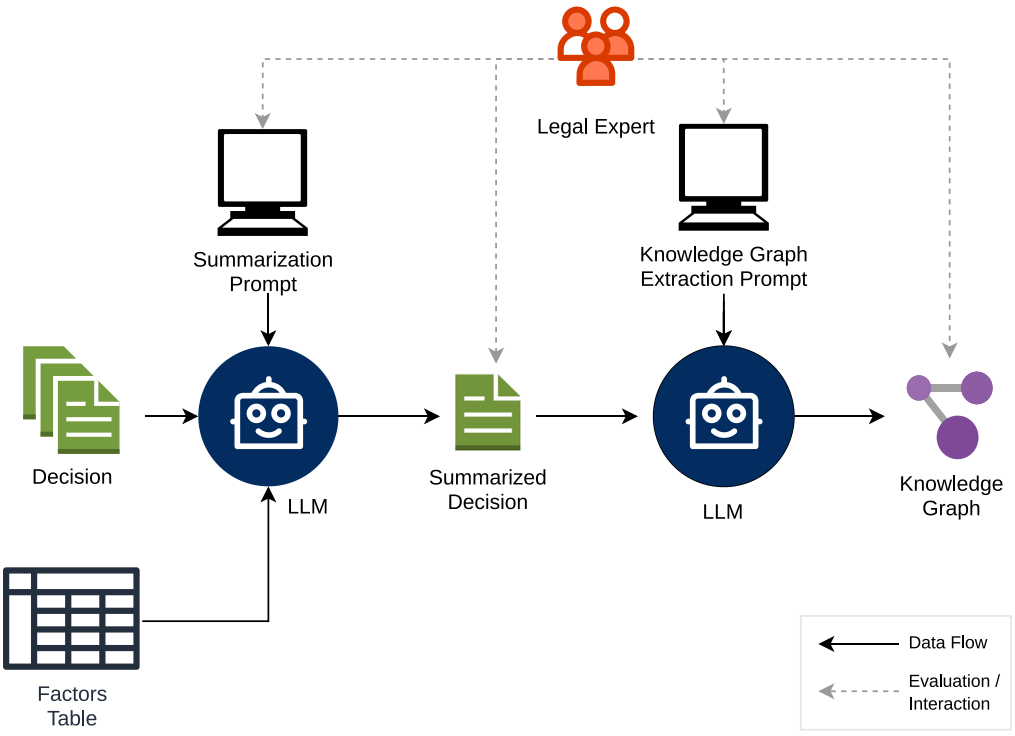

Fig. 2. Methodology pipeline.

Fig. 1 displays a sample from the created dataset, highlighting relevant legal factors. For each entry, we indicate the presence of factors and crimes (True/False) alongside the final decision outcome.

4. Methodology

We developed a multi-stage pipeline to extract, represent, and evaluate legal factors in judicial decisions on appeal. This approach integrates Large Language Models (LLMs) with expert-defined knowledge schemas to generate Knowledge Graphs (KGs) that are both structured and legally grounded. The complete pipeline is visualized in Fig. 2.[1]

To balance interpretability with robust findings, we adopted a qualitative approach, emphasizing the most relevant factors identified in previous research. These factors were extracted by using SHAP (SHapley Additive exPlanations) and decision tree techniques. From a predefined list of factors, we extracted 10 judgments containing these factors from the dataset and assessed our methodology using these cases. Additionally, to obtain a quantitative overview of the methodology, we randomly selected a sample of 200 judgments, which, as discussed later, enables us to evaluate the methodology's effectiveness for both languages in terms of graph features such as the number of nodes, edges, and density.

[1] Code, Datasets, Prompts, and Results are available at: https://osf.io/ufmqv/overview?view_only=450c69af5b0541a7b81f218df8451108

4.1. Summary

Summarizing judgments and legal decisions during the processing stage is needed to avoid redundancy and reduce the risk of errors when working with large volumes of information. Firstly, we divide the judgment into two parts: the first contains the description of the factual elements of the case and the claims of the parties, and the second part describes the legal considerations of the judge. Legal texts often contain factual and legal descriptions of the case, some of which may be irrelevant to the core issues.

Moreover, when a case reaches the highest court, it has typically undergone review by multiple lower courts. This procedural history is repeated by the judge in the text introduction, hindering the model's ability to extract the key factors that influenced the final judgment. To mitigate this, we streamline the analysis by focusing on the most critical elements, supported by a predefined table of factors that guides the summarization process. This systematic approach enhances the clarity and accuracy of the legal analysis, ensuring that the most pertinent information is highlighted in the final output.

The prompt can be seen in Listing 1, translated into English for better access and readability. The model is instructed to generate a summary of the judgment, with specific constraints on the output format. We identify and select the relevant legal factors and explicitly direct the model to focus on those in the summary.

Listing 1 Judgment Summarization Prompt

```
Summarize the following text from a judgment:
[JUDGMENT TEXT].

Focus on the following legal elements:
C6: crime related do criminal organizations
F2: suspension of time in prison or
    ↪ suspension of the proceedings
F7: the complexity of the proceedings or
    ↪ existing risks such as of the
    ↪ prisoner's flight
F8: acts inferred and not proven

The summary should be between 400 and 500
    ↪ words.
Use the following format:
- Case Name:
- Summary:
- Key Factors:
```

We observed that the LLM's ability to identify key elements was slightly higher when processing the original language of the judgment compared to an English translation. Consequently, the summary generated in the original language is utilized as input for the Knowledge Graph creation prompt.

4.2. Knowledge graph

During the KG extraction process, we concentrate on the relationships between key factors and their evaluation by judges. We utilize the generated summary as input and require the model to produce a graph, extracting relevant entities (factors, parties, crimes) and their relationships. Our objective is to extract relevant information, which enhances our understanding of the key elements the judge considered in the decision. The main prompt instructions are presented in Listing 2.

Listing 2 Knowledge Graph Prompt Overview

```
Instructs an expert in knowledge extraction
    ↪ to create a detailed knowledge graph
    ↪ based on a provided text fragment.

The task focuses on identifying and
    ↪ extracting relationships between
    ↪ relevant entities in legal texts,
    ↪ particularly regarding magistrate
    ↪ decisions related to the release of
    ↪ detainees under preventive detention
    ↪ measures.

The entities include people, organizations,
    ↪ laws, documents, actions, and relevant
    ↪ factors, while relationships between
    ↪ them are described using verbs and
    ↪ other relational terms.

The output must be presented in JSON format,
    ↪ with each entry representing two
    ↪ entities (E1 and E2) and their relation
    ↪ (R), described in a maximum of four
    ↪ words.

A few examples are provided in order to make
    ↪ sure the format is respected.

Finally, the goal is stated again clearly,
    ↪ which is to ensure clarity and
    ↪ conciseness in identifying the
    ↪ interactions within the legal context.
```

One of the challenges encountered in this task is guiding the model to create relationships that focus on the key legal elements, rather than generating a Knowledge Graph that merely analyzes the nature of the judgment, such as the definition of pretrial detention or the legal validity of the appeal.

Furthermore, we enforced the prompt to be asked *iteratively*, rather than specifying a fixed number of relationships to be established. Specifically, after each Knowledge Graph extraction, we instruct the model to search for additional potentially relevant entities and relationships. However, we found that excessively repeated iterations (more than five) often lead the model to provide increasingly general or duplicated relationships.

We observed that the number or types of examples provided do not significantly impact the outcome. Furthermore, the results are qualitatively consistent across retries on the same judgment, although minor deviations are expected due to some degree of randomness from LLMs. Deviations between retries on the same judgment primarily differ in the wording used for one or two entities, which does not diminish the overall quality of the KG, nor the focus of the KG, which remains on the legal actors.

Legal experts play a vital role in the pipeline. Their primary responsibility is to iteratively refine the summarization prompt and its output, as previously discussed. Additionally, they are instrumental in crafting the KG extraction prompt and assessing the effectiveness of the resulting Knowledge Graph representation for a given judgment.

Building on the iterative KG extraction process guided by legal experts, the LLM generates structured output in the form of JavaScript Object Notation (JSON) objects. Each JSON object represents two entities (denoted as $E1$ and $E2$) and their relationship (R), which captures the relevant legal factors, parties, and crimes as outlined above. These JSON objects are parsed into a graph structure using the NetworkX library,[2] which allows us to represent entities as nodes and relationships as edges, thus reflecting the underlying legal reasoning in a structured format. The graphs are then visualized using GraphViz[3] to provide clear graphical representations of the key relationships considered in each judgment.

4.3. Evaluation criteria

A combination of qualitative and quantitative methods was used to evaluate the project. This included an expert review as well as structural and semantic analysis.

The qualitative assessment, conducted by legal experts, verified the legal correctness and logical coherence of the generated outputs. The experts evaluated whether the summaries and knowledge graphs (KGs) accurately captured the *ratio decidendi* and the correct legal factors. This ensured that the model's extraction aligned with the procedural and substantive laws of the respective jurisdictions.

The *quantitative evaluation* serves two distinct purposes: to validate the quality of the extraction and to characterize the graph topology.

First, to validate the proposed "Factor-Guided" methodology against a baseline, we measured the semantic similarity between the nodes of the generated graph and the legal concepts defined by experts. We utilized three metrics:

- *Levenshtein Distance:* measures the minimum edit distance between strings to assess morphological similarity [24].
- *Jaccard Similarity:* measures the intersection over union of token sets to assess lexical overlap [25].

[2] https://networkx.org/
[3] https://graphviz.org/

- *Semantic Similarity:* uses the LabSE[4] model [26] to measure neural semantic similarity, with a cosine measure, beyond surface-level keywords.

Secondly, we conducted an exploratory structural analysis to characterize the complexity of the resulting legal knowledge graphs. These metrics serve as proxies for *argumentative volume* and are essential for estimating the computational load of future large-scale deployments.

- *Node & Edge counts*: Represents the volume of distinct entities (e.g., factors, parties) and relationships. This proxies the density of information exchange.
- *Average node degree*: Indicates the average interconnectivity of entities. A higher degree suggests a decision where legal factors are highly interdependent rather than isolated.
- *Diameter*: defined as the *longest shortest* path between any two nodes. It measures the compactness of the legal reasoning: the larger the diameter, the more linear or fragmented the chain of argumentation.
- *Clustering Coefficient*: measures the tendency of nodes to form tightly knit groups. A high clustering score indicates that legal factors form cohesive, thematic sub-structures within the decision.
- *Degree Centrality*: identifies the most influential entities (hub nodes). In a legal KG, nodes with high centrality may represent core legal concepts or statutes that underpin a judge's reasoning.

5. Experiments

We experimentally validated our pipeline to assess its efficacy in generating legally relevant summaries and Knowledge Graphs (KGs). The evaluation is threefold: (1) a qualitative assessment of the summaries (Section 5.1); (2) a legal validation of the KG structures (Section 5.2); and (3) a quantitative analysis characterizing the topological complexity and semantic fidelity of the outputs (Section 5.3).

5.1. Qualitative analysis of the summary

We examined two representative judgments: *Cassazione Penale n. 16939/2009* (Italy) and *Habeas Corpus n. 156667* (Brazil).[5]

The generated summaries followed the strict structure imposed by the prompt (Listing 3). Structurally, both outputs successfully distinguished between the *factual report* (parties' claims) and the *ratio decidendi* (judge's reasoning). This separation is critical, as it prevents the model from conflating defense arguments with the final court ruling—a common failure mode in unguided legal summarization.

Content-wise, the summaries correctly identified complex legal nuances. For instance, both cases addressed Factor $F7$ (flight risk). The Italian summary correctly noted the *absence* of flight risk due to the defendant's mental incapacitation, whereas the Brazilian summary highlighted its *presence* due to the defendant's leadership role in a criminal organization. Similarly, both summaries accurately isolated Factor $F11$ (defense restrictions) within their respective procedural codes.

Listing 3 Output Summary Structure

```
- NAME
[...]

- CONCISE SUMMARY OF THE RULING
[...]

- KEY LEGAL FACTORS
[...]

- OUTCOME OF THE CASE
[...]
```

It is important to note that both the case introduction summary and the judge's reasoning summary follow a consistent structure and set of criteria. The introduction outlines the parties' claims and arguments, while the judge's reasoning focuses on the evaluation of the key legal factors that influence the judicial decision. This distinction enables the model to effectively identify and differentiate between the parties' claims and the legally significant elements that ultimately determine the outcome. It also plays a pivotal role in the next step of our methodology, where recognizing these elements is essential for providing a meaningful explanation of the decision.

In our experiments, we also investigated the impact of leaving the LLM unrestricted, i.e., requiring it to generate a summary without providing additional contextual guidance such as the set of legally relevant factors. Under these conditions, the model consistently produced summaries that focused on a limited subset of elements, often neglecting several aspects identified in our prior analysis. This behavior suggests that, in the absence of structured guidance, the LLM tends to converge on incomplete patterns, or patterns that deal with purely procedural pipelines of the judicial process, resulting in summaries that are less comprehensive and potentially biased toward irrelevant features.

We evaluated a small sample manually, as can be seen in Appendices C and B. For the Italian judgments, when summarization is guided by the Factor Table, we obtain 20 legally relevant nodes out of 30 total nodes, corresponding to 66.6%. By contrast, without access to the Factor Table, the number of legally relevant nodes drops to 16 out of 62, yielding a substantially lower ratio of 25%.

These results highlight two important effects. First, the substantial increase in the proportion of legally relevant nodes demonstrates that the Factor Table effectively steers the model toward information that is more closely aligned with expert-identified legal reasoning. Second, the sharp rise in the total number of nodes in the unguided setting indicates that, without explicit constraints, the model tends to generate more verbose and diffuse summaries, incorporating a large amount of marginal or irrelevant information.

5.2. Qualitative analysis of the knowledge graph

We evaluated the KGs to determine if the extracted entities and relationships accurately reflected the judicial logic. Figs. 3 and 4 illustrate selected subgraphs from the analysis.[6]

In Fig. 3, we can observe the following elements and relations from the Brazilian Case:

- *(F1) excesso de tempo no cárcere* (excess time in prison) *é um fator contra* (is a factor against) *prisão preventiva* (pretrial detention).

[4] We selected LabSE (Language-agnostic BERT Sentence Embedding) after a preliminary empirical evaluation against other multilingual models. In our specific context of Italian and Portuguese legal terminology, LabSE demonstrated superior *discriminative power*. While other models frequently assigned inflated similarity scores (>0.7) to semantically distinct concepts, LabSE correctly maintained lower similarity values (<0.2) for non-aligned terms, reducing false positives. The model is available at: https://huggingface.co/sentence-transformers/LaBSE

[5] Full texts and summaries are available in Appendix A.

[6] Full visualizations are available in Appendix B (generated with summarization) and Appendix C (generated without summarization).

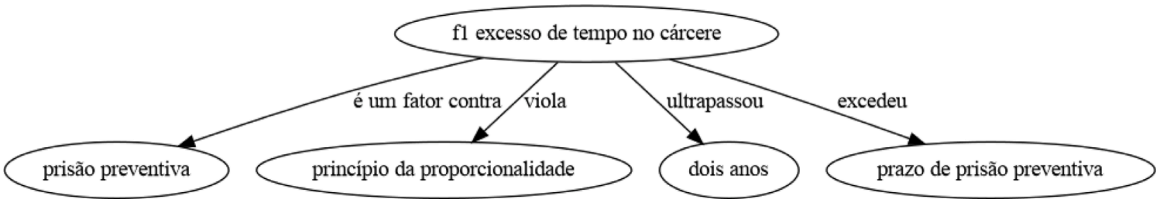

Fig. 3. Subgraph from Brazilian Case (HC n. 156667) illustrating Factor $F1$ (Excess Time in Prison).

- *(F1) excesso de tempo no cárcere* (excess time in prison) *viola* (violates) *princípio da proporcionalidade* (the principle of proportionality).
- *(F1) excesso de tempo no cárcere* (excess time in prison) *ultrapassou* (has exceeded) *dois anos* (two years).
- *(F1) excesso de tempo no cárcere* (excess time in prison) *excedeu* (has exceeded) *prazo de prisão preventiva* (the time limit for pretrial detention).

Listing 4 shows the JSON output generated by the LLM and used to construct the subgraph in Fig. 3, illustrating how entities ($E1, E2$) and relationships (R) are structured.

Listing 4 Raw JSON Output for Factor F1 (Brazilian Case)

```
[
    {
        "E1": "f1: excesso de tempo no
            ↪ cárcere",
        "E2": "dois anos",
        "R": "ultrapassou"
    },
    {
        "E1": "f1: excesso de tempo no
            ↪ cárcere",
        "E2": "prazo de prisão preventiva",
        "R": "excedeu"
    },
    {
        "E1": "f1: excesso de tempo no
            ↪ cárcere",
        "E2": "princípio da
            ↪ proporcionalidade",
        "R": "viola"
    },
    {
        "E1": "paciente",
        "E2": "f1: excesso de tempo no
            ↪ cárcere",
        "R": "alega"
    },
    {
        "E1": "f1: excesso de tempo no
            ↪ cárcere",
        "E2": "prisão preventiva",
        "R": "é um fator contra"
    }
]
```

Note that while the graph explicitly models the legal arguments (factors and relationships), the final 'Prisoner Status' (Release/Detention) is treated as a metadata attribute of the case node rather than a separate entity in this subgraph, though it is used to label the target variable in our dataset.

This partial KG effectively captures the key factor $F1$ (excess time in prison) and establishes a clear connection to the case by illustrating how prolonged detention violates the principle of proportionality. It also outlines how the facts — such as exceeding the two-year preventive detention limit — are directly linked to the legal framework, highlighting the breach of established legal thresholds.

In Fig. 4 we can observe the following elements and relations from the Italian Case:

- *(F2) è* (is associated with) *sospensione dei termini della custodia cautelare* (the suspension of time in pretrial detention).
- *(F2) è la base di* (is the basis of) *impedimento dell'imputato di partecipare al processo* (defendant's inability to attend the proceedings).
- *art. 303 c.p.p.* (Article 303 of the Italian Code of Criminal Procedure) *regola la* (regulates the) *sospensione dei termini della custodia cautelare* (suspension of time in pretrial detention).
- *sospensione dei termini della custodia cautelare* (suspension of time in pretrial detention) *è dovuta a* (is due to) *incapacità mentale dell'imputato* (mental incapacity of the defendant).
- *incapacità mentale dell'imputato* (mental incapacity of the defendant) *è equivalente a* (is equal to) *impedimento dell'imputato di partecipare al processo* (defendant's inability to attend the proceedings).
- *v.g.* (defendant's name abbreviated) *è affetto da* (is affected by) *incapacità mentale* (mental incapacity) *è* (is) *impedimento dell'imputato di partecipare al processo* (defendant's inability to attend the proceedings).

This partial KG describes the relationships between the key factor $F2$ (suspension of prison time) and the conditions for its occurrence in the specific case, according to the Italian Code of Criminal Procedure. It focuses on the defendant's inability to participate in court proceedings due to mental incapacity, also identifying the defendant as one of the entities.

We can imagine a system employed in the judicial domain that enables judges to more efficiently identify and connect the relevant elements of a decision. This example shows that the chosen approach can extract and reconstruct the main elements of the decision, analyzing the immediate connection between the cited norms, the domain, goal, and factor extracted. It serves both as a validation of the extracted factors and a way to visualize the connection between abstract factors and concrete elements of the case law.

In general, both KGs accurately describe the relations between legal elements and their factual premise, and therefore, one of the goals we set out has been achieved. Looking at the entire graph,[7] instead of one single section, we encountered a few issues.

Firstly, in the Italian dataset, we observed a stronger tendency than in the Brazilian dataset to generate sections of the KG that contained isolated nodes and relationships, which were disconnected from the main graph structure. This occurred particularly when more than three factors influenced the outcome, and the model opted to treat each factor as a separate sub-graph, failing to establish connections between them. Even attempts to prompt the model to correct this behavior did not result in significant improvements.

[7] Which can be observed in the Appendix B.

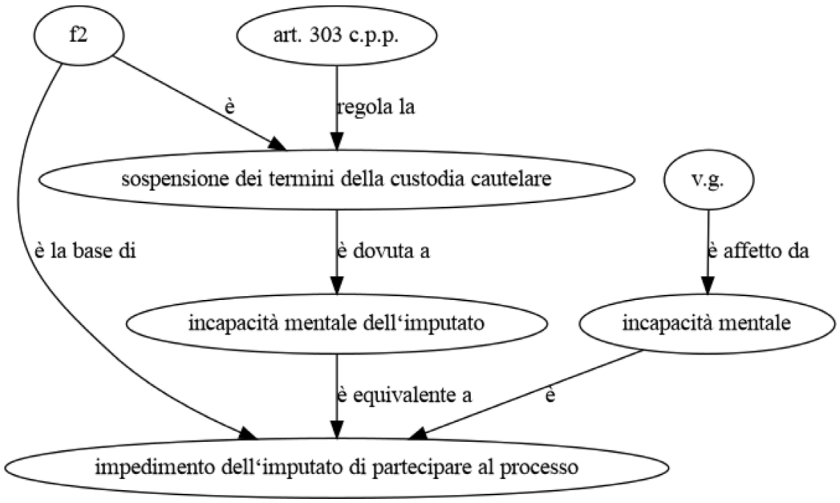

Fig. 4. Factor 2 from Italian Case (Cassazione Penale n. 16939/2009): Suspension of pretrial detention terms.

Table 1
Median graph metrics by jurisdiction.

Dataset	Nodes	Edges	Avg. Degree	Diameter	Clustering Coeff.	Degree Cent.
BR	29	33	2.167	6	0.061	0.076
IT	32	33	2.123	8	0.055	0.069

Second, the aforementioned issue – centered on the key legal elements – was notably mitigated by inputting a predefined list of factors, thus limiting the range of legal elements the model could extract. This led to marked improvements, shifting from abstract KGs, which primarily featured statutory law articles as entities, to more substantial outcomes that concentrated on legally relevant factors.

Despite differences in legal terminology, the KG structures for both the Italian and Brazilian legal systems share a similar entity-relation modeling framework. In both cases, core entities such as legal provisions (e.g., specific articles of criminal procedure) and legally relevant elements (e.g., mental incapacity) are connected through relationships that provide a comprehensive overview of the decision-making process. The factors typically serve as central nodes, around which specific details are introduced.

5.3. Quantitative analysis

We performed a quantitative analysis with two distinct objectives: (1) an *exploratory structural analysis* to characterize the complexity of the legal graphs and estimate scalability requirements, and (2) a *semantic fidelity analysis* to measure how effectively our method aligns with expert legal concepts compared to a baseline (the factors table).

Exploratory structural analysis. Understanding the topological properties of KGs is important for anticipating data load and performance constraints in real-world deployments. We analyzed a stratified sample of 200 judicial decisions (100 Italian, 100 Brazilian). We extracted the five key metrics to proxy the "argumentative volume" of the cases (as detailed in Section 4: node & edge count, average degree, diameter, clustering coefficient and degree centrality.

Table 1 presents the median values for these metrics, while Fig. 5 presents their distribution based on violin plots as an alternative perspective.

The results reveal a convergence in topological complexity between the two jurisdictions, particularly in edge count and average degree. This suggests that the complexity of reasoning in pretrial detention cases is relatively agnostic to the specific national code. However, the Italian graphs display a larger median *Diameter* (8 vs. 6), suggesting a more linear, sequential chain of reasoning, whereas Brazilian graphs show higher *Clustering* and *Centrality*, indicating an argumentation style that revolves around central "hub" nodes.

Table 2
Semantic similarity of Knowledge Graph Nodes with Legal Factors Table.

Condition	Dataset	Levenshtein	Jaccard	LabSE
With Summary *(Proposed)*	IT	0.372	0.133	0.419
	BR	0.455	0.208	0.475
Raw Text *(Baseline)*	IT	0.332	0.049	0.288
	BR	0.329	0.042	0.284

Semantic similarity. Finally, to validate the utility of our proposed Factor-Guided Summarization method compared to direct LLM prompting with raw text, we measured the syntactic and semantic similarity between the generated graph nodes and the ground-truth *legal factors table*.

We compared our pipeline ("With Summary") against a baseline where KGs were generated directly from the raw judicial text ("Raw Text").[8] We utilized three metrics: *Levenshtein* (edit distance), *Jaccard* (lexical overlap), and *LabSE* (semantic similarity) to capture equivalence [26].

As shown in Table 2, our method significantly outperforms simpler prompting in all metrics, but specially in neural semantic similarity (scores of $\approx 0.42 - 0.47$ vs ≈ 0.28).

The divergence between the Jaccard (low) and LabSE (high) scores in the proposed method is of interest. It indicates that while the model does not reproduce the exact keywords (low lexical overlap), it successfully captures the *semantics* of the legal factors (high LabSE measures). The summarization step effectively functions as a filter: by priming the context window with the domain taxonomy, it forces the

[8] The knowledge graphs on which we applied the similarity metrics can be found in Appendices B and C.

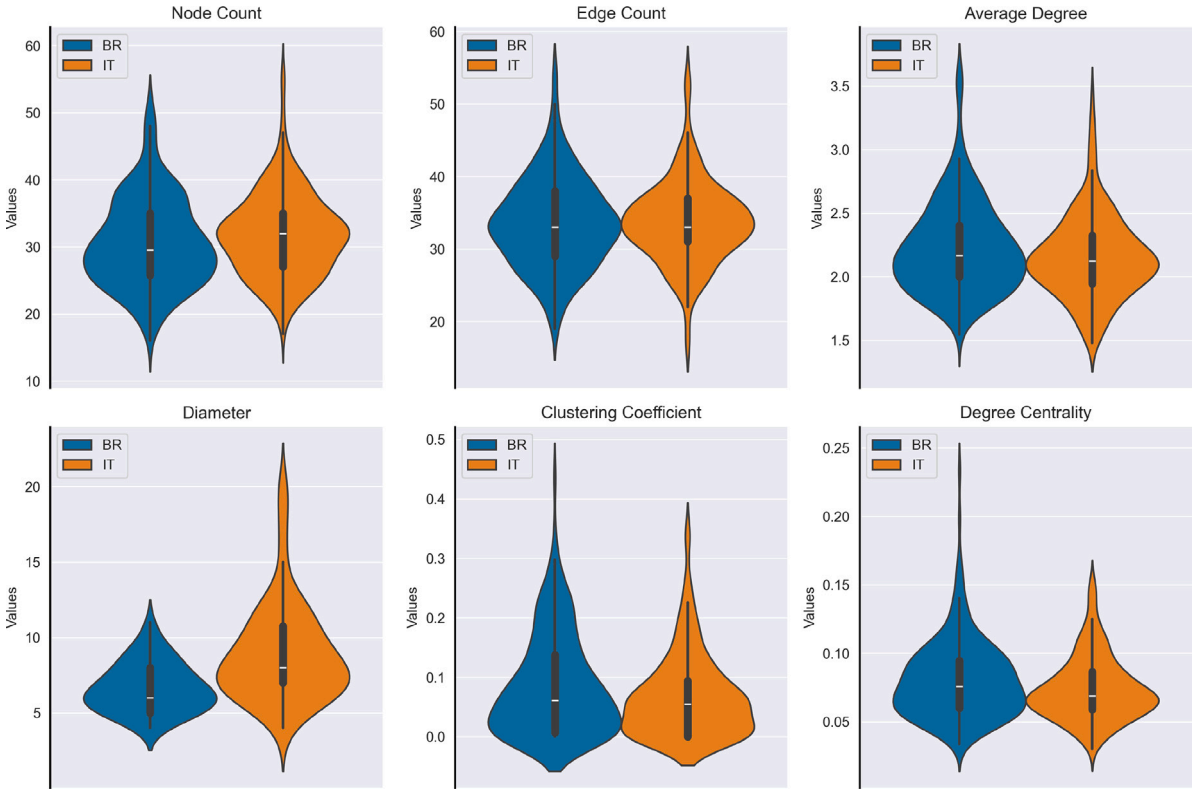

Fig. 5. Distribution of Graph Metrics (Brazil vs. Italy). The structural similarity suggests a convergent complexity in reasoning despite different legal traditions.

graph generation to converge on expert-defined legal factors, whereas the raw text baseline tends into procedural noise, as one can observe in Appendices B and C. These quantitative insights go in the hypothesis direction: that LLMs perform better with structured intermediate representations (summaries) for legal information extraction.

6. Conclusion

We presented a factor-guided framework that integrates LLMs with KGs to extract and visualize the key elements of judicial reasoning in pretrial detention. Our results confirm that while standard LLMs often struggle with legal nuance, the injection of expert-defined factors drastically reduces noise, enhancing the accuracy and semantic fidelity of the extraction.

This pipeline advances research at the intersection of Legal Tech and Visual Law by demonstrating how controlled extraction improves interpretability and traceability. Beyond assisting legal professionals in navigating case law, this methodology offers a scalable, replicable solution for constructing high-quality datasets for legal analytics. Furthermore, the resulting KGs provide structured, explainable inputs for predictive models, bridging the gap between raw text and factor-based reasoning.

However, limitations persist. We observed that while KGs effectively map entity relationships, they do not yet fully capture the hierarchical weight or causal dependencies of specific arguments. Future research will address this by incorporating rule extraction techniques to better model legal logic and by developing objective criteria to assess KG completeness.

Finally, we aim to extend this framework toward fully machine-readable and semantically interoperable graphs by leveraging Semantic Web technologies (RDF, OWL). Such developments will support cross-jurisdictional reasoning and integration with emerging knowledge-based legal AI systems [27].

Fundings

- Computable Law project. "CompuLaw" is a project funded by European Research Council (ERC) under the European Union's Horizon 2020 — Research and Innovation funding programme (2014/2020), Grant Agreement no. 833647. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them
- Brazilian National Council for Scientific and Technological Development (CNPq)
- Coordination for the Improvement of Higher Education Personnel - Institutional Program for Internationalisation (CAPES).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Full decisions and summaries

A.1. Italian case (full decision) - Sentenza n. 16939

```
SENTENZA
Cassazione penale sez. I - 01/04/2009, n. 16939
Intestazione
                    LA CORTE SUPREMA DI CASSAZIONE
                        SEZIONE PRIMA PENALE
Composta dagli Ill.mi Sigg.ri Magistrati:
Dott. FAZZIOLI   Edoardo       -  Presidente   -
Dott. ZAMPETTI   Umberto       -  Consigliere  -
Dott. DI TOMASSI Maria Stefani -  Consigliere  -
Dott. BRICCHETTI Renato        -  Consigliere  -
Dott. CASSANO    Margherita    -  Consigliere  -
ha pronunciato la seguente:
                       SENTENZA/ORDINANZA
sul ricorso proposto da:
difensori di             V.G., nato a (OMISSIS);
avverso l'ordinanza pronunciata in data 19 dicembre 2008 dal
Tribunale di Bologna;
udita la relazione del Consigliere Dott. BRICCHETTI Renato;
sentite le conclusioni del Pubblico Ministero, in persona del S.
Procuratore Generale Dott. CIAMPOLI Luigi, che ha chiesto rigettarsi
il ricorso;
udito il difensore di fiducia dell'imputato, avv. BONETTI Manrico di
Bologna, che ha chiesto accogliersi il ricorso.


SVOLGIMENTO DEL PROCESSO
1. Con il provvedimento indicato in epigrafe, il Tribunale di Bologna, costituito ai sensi
    ↪ dell'art. 310 c.p.p., rigettava l'appello presentato nell'interesse di V.G. avverso
    ↪ l'ordinanza con la quale il Giudice per le indagini preliminari della stessa città,
    ↪ richiamandosi all'art. 304 c.p.p., comma 1, lett. a) e c bis), aveva disposto la sospensione
    ↪ dei termini della custodia cautelare previsti dall'art. 303 c.p.p. perchè il processo
    ↪ (giudizio abbreviato) era stato sospeso, a norma dell'art. 71 c.p.p., comma 1, dopo che, a
    ↪ seguito degli accertamenti previsti dall'art. 70 c.p.p., era risultato che lo stato mentale
    ↪ dell'imputato era tale da impedirne la cosciente partecipazione.
2. Avverso l'anzidetta ordinanza hanno proposto ricorso per cassazione i difensori dell'imputato,
    ↪ chiedendone l'annullamento.
Deducono inosservanza della legge processuale penale, segnatamente dell'art. 304 c.p.p., comma 1,
    ↪ lett. a) e c bis), in relazione all'art. 70 c.p.p., e mancanza di motivazione.
Sostengono che, in presenza di un'incapacità accertata ex art. 70 c.p.p., che abbia determinato la
    ↪ sospensione del processo, il giudice non deve sospendere i termini di durata della custodia
    ↪ cautelare.

MOTIVI DELLA DECISIONE
3. Il ricorso non merita accoglimento.
a) A seguito degli accertamenti previsti dall'art. 70 c.p.p., essendo risultato che lo stato
    ↪ mentale dell'imputato era tale da impedirne la cosciente partecipazione al procedimento, il
    ↪ Giudice per le indagini preliminari del Tribunale di Bologna ha disposto, a norma dell'art.
    ↪ 71 c.p.p., comma 1, nel corso del giudizio abbreviato, la sospensione del processo, non
    ↪ ritenendo di dover pronunciare sentenza di proscioglimento, e, ai sensi dell'art. 304
    ↪ c.p.p., comma 1, lett. a) e c bis), la sospensione dei termini di custodia cautelare
    ↪ previsti dall'art. 303 c.p.p. "durante il tempo" di sospensione del processo "per
    ↪ impedimento dell'imputato".
Ha, in altre parole, ritenuto che nel concetto di impedimento rientri anche lo stato di incapacità
    ↪ dell'imputato che ha imposto la sospensione del processo. L'affermazione deve essere
    ↪ condivisa.
b) La questione era stata già affrontata nella vigenza del precedente codice di rito.
L'art. 272 c.p.p. 1930, prevedeva, al settimo comma, che i termini di durata della custodia
    ↪ cautelare rimanessero sospesi durante il tempo in cui il dibattimento era sospeso o rinviato
    ↪ per "legittimo impedimento" dell'imputato.
E questa Corte aveva avuto modo di stabilire (cfr. Cass. 5^ 12 agosto 1983, Messina, RV 10163;
    ↪ Cass. 1^ 16 febbraio 1977, Romeo, RV 136951) che nel concetto di "legittimo impedimento"
    ↪ rientravano sia le malattie fisiche, che impedivano all'imputato di essere presente al
    ↪ dibattimento, sia quelle psichiche le quali, pur non impedendogli di essere materialmente
    ↪ presente, gli precludevano un'utile assistenza al dibattimento medesimo.
```

c) Situazione normativa sostanzialmente identica è oggi riprodotta nell'art. 304 c.p.p., comma 1, che continua a prevedere quale causa di sospensione dei termini di custodia cautelare l'impedimento dell'imputato.

Nulla suggerisce, neppure oggi, un'interpretazione restrittiva di quest'ultima locuzione, esprimendo la norma l'esigenza di evitare che i termini di fase della custodia cautelare continuino a decorrere quando cause legate all'impossibilità dell'imputato di essere presente e di parteciparvi (e non necessariamente da lui determinate a fini meramente dilatori, come adombrato dal ricorrente) impediscano al processo di progredire, superando le varie fasi scandite nell'art. 303 c.p.p., comma 1, verso la propria meta finale.

Non vi è, pertanto, ragione alcuna, anche nel vigore dell'odierno art. 304 c.p.p., comma 1, di distinguere le ragioni dell'impedimento e, nel caso in cui il medesimo derivi da malattia, la natura della stessa.

La citata disposizione, d'altra parte, non contiene rinvii agli art. 420 ter e art. 484 c.p.p., comma 1 bis, vale a dire alle norme che si occupano in modo specifico della possibilità dell'imputato di essere fisicamente presente al processo.

Sarebbe arbitrario, oltre che ingiustificato, pertanto, escludere dal concetto di impedimento di cui all'art. 304 c.p.p., comma 1, l'ipotesi in cui l'impossibilità di partecipare al processo dipenda dalla situazione mentale dell'imputato, contemplata dagli artt. 70, 71 e 72 c.p.p.; disposizioni che la Corte costituzionale (cfr.

sentenza 26 gennaio 2004, n. 39) ha, tra l'altro, interpretato come riferibili non soltanto ad una malattia definibile in senso clinico come psichica, ma anche a qualunque altro stato di infermità che renda non sufficienti o non utilizzabili le facoltà mentali dell'imputato.

Coerente con questa soluzione appare, tra l'altro, la disposizione contenuta nell'art. 305 c.p.p., comma 1, alla stregua della quale quando è disposta perizia sullo stato di mente dell'imputato, in ogni stato e grado del procedimento di merito, i termini di custodia cautelare sono prorogati per il periodo di tempo assegnato per l'espletamento della perizia.

d) E' il caso di aggiungere che nè la sospensione del processo, nè la sospensione dei termini di custodia cautelare sono assolute L'ordinamento contempla, invero, a garanzia dell'imputato, da una parte, il dovere del giudice (art. 70 c.p.p., comma 2) di assumere, a richiesta del difensore, le prove che possono condurre al proscioglimento dell'imputato, e, quando vi è pericolo nel ritardo, ogni altra prova richiesta dalle parti, dall'altra l'insuperabile tetto di durata massima della custodia cautelare di cui all'art. 304 c.p.p., comma 6. Se, dunque, è astrattamente ipotizzabile una pendenza sine die del procedimento (la realtà clinica prevede casi di infermità per le quali non sussiste nè possibilità di guarigione nè tanto meno di attenuazione del disturbo, ma è anzi possibile un progressivo peggioramento), fermo restando naturalmente il sistema delle verifiche periodiche dello stato di mente dell'imputato imposto dalla legge e funzionale all'eventuale ripresa del processo (positivamente scrutinato da Corte Cost. 28 maggio 2004 n. 157), è da escludere una prosecuzione sine die della custodia cautelare.

Si aggiunga che l'accertata (e del caso irreversibile) incapacità dell'imputato potrebbe, comunque, aprire scenari diversi, che possono andare dalla dovuta trasformazione della custodia cautelare in carcere in custodia cautelare in casa di cura (art. 73 c.p.p., comma 3) alla possibile revoca o sostituzione della misura (in caso di acclarata incompatibilità delle condizioni di salute con lo stato di detenzione oppure nell'ipotesi di cessazione o attenuazione delle concrete esigenze cautelari correlata allo stato mentale dell'imputato).

4. Al rigetto del ricorso consegue la condanna al pagamento delle spese processuali.

Deve disporsi, inoltre, che copia del presente provvedimento sia trasmessa al Direttore dell'istituto penitenziario di competenza ai sensi dell'art. 94 disp. att. c.p.p., comma 1 ter.

P.Q.M.

Rigetta il ricorso e condanna il ricorrente al pagamento delle spese processuali.

Dispone, inoltre, che copia del presente provvedimento sia trasmessa al Direttore dell'istituto penitenziario di competenza ai sensi dell'art. 94 disp. att. c.p.p., comma 1 ter.

Così deciso in Roma, il 1 aprile 2009.

Depositato in Cancelleria il 21 aprile 2009

A.2. Italian case (summary) - Sentenza n. 16939

```
- Nome del caso: Tribunale di Bologna, V.G.

- Riassunto:
Il Tribunale di Bologna ha rigettato l'appello presentato nell'interesse di V.G. contro l'ordinanza del Giudice per le indagini preliminari che aveva disposto la sospensione dei termini della custodia cautelare previsti dall'art. 303 c.p.p. a causa dell'incapacità dell'imputato di partecipare al processo a seguito di accertamenti previsti dall'art. 70 c.p.p. La Corte di Cassazione ha confermato la decisione del Tribunale di Bologna, ritenendo che la sospensione dei termini di custodia cautelare fosse giustificata dall'impedimento dell'imputato di partecipare al processo a causa della sua incapacità mentale.

- Fattori chiave:
F2: sospensione del tempo in carcere o sospensione del procedimento. La Corte di Cassazione ha ritenuto che la sospensione dei termini di custodia cautelare fosse giustificata dalla sospensione del processo a causa dell'incapacità dell'imputato di partecipare. Il ragionamento alla base di questo fattore è che l'impedimento dell'imputato di partecipare al processo, anche a causa di una incapacità mentale, giustifica la sospensione dei termini di custodia cautelare.

F7: complessità del procedimento o rischi esistenti come la fuga del detenuto. La Corte di Cassazione ha considerato che la sospensione dei termini di custodia cautelare non era assoluta e che l'ordinamento prevede garanzie per l'imputato, come il dovere del giudice di assumere prove che possono condurre al proscioglimento dell'imputato e la possibilità di verifiche periodiche dello stato di mente dell'imputato. Il ragionamento alla base di questo fattore è che la sospensione dei termini di custodia cautelare non può essere sine die e che devono essere considerati i rischi esistenti, come la fuga del detenuto.

F11: limitazione della difesa. La Corte di Cassazione ha considerato che la sospensione dei termini di custodia cautelare non limitava la difesa dell'imputato, poiché l'ordinamento prevede garanzie per l'imputato, come il dovere del giudice di assumere prove che possono condurre al proscioglimento dell'imputato. Il ragionamento alla base di questo fattore è che la sospensione dei termini di custodia cautelare non limita la difesa dell'imputato e che l'ordinamento prevede garanzie per l'imputato.
```

A.3. Brazilian case (full decision) - Habeas Corpus n. 156667

```
EMENTA E ACÓRDÃO
04/12/2018
PRIMEIRA TURMA
HABEAS CORPUS 156.667 PARANÁ
RELATOR: MIN. MARCO AURÉLIO
REDATOR DO ACÓRDÃO: MIN. ROBERTO BARROSO
PACTE.(S): R. N. C.
IMPTE.(S): M. C.
COATOR(A/S)(ES): RELATOR DO HC No 447.695 DO SUPERIOR TRIBUNAL DE JUSTIÇA

EMENTA: HABEAS CORPUS. PRISÃO PREVENTIVA. ORGANIZAÇÃO CRIMINOSA. COMPLEXIDADE DA CAUSA. DECISÃO MONOCRÁTICA DE MINISTRO DO SUPERIOR TRIBUNAL DE JUSTIÇA. INADEQUAÇÃO DA VIA ELEITA. 1. Não cabe habeas corpus contra decisão monocrática de Ministro de Tribunal Superior. Precedentes. 2. Ausência de teratologia, ilegalidade flagrante ou abuso de poder que justifique a concessão da ordem de ofício, em especial pela complexidade da causa. Hipótese de paciente denunciado por crime de integrar organização criminosa (PCC), majorada por envolver emprego de arma de fogo e conexão com outras organizações criminosas independentes. Organização criminosa composta por mais de 700 acusados e responsável, entre outras infrações, por rebeliões em presídios, homicídios, roubos, extorsões e tráficos de drogas. 3. Habeas corpus não conhecido, revogada a liminar.
ACÓRDÃO
Vistos, relatados e discutidos estes autos, acordam os Ministros da Primeira Turma do Supremo Tribunal Federal, sob a presidência do Ministro Alexandre de Moraes, na conformidade da ata de julgamento, por maioria de votos, em não conhecer da impetração e revogar a liminar anteriormente deferida, nos termos do voto do Ministro Luís Roberto Barroso, Redator para o acórdão, vencido o Ministro Marco Aurélio, Relator.
Brasília, 4 de dezembro de 2018.
MINISTRO LUÍS ROBERTO BARROSO - REDATOR P/ O ACÓRDÃO
```

```
16
17 RELATÓRIO
18 O SENHOR MINISTRO MARCO AURÉLIO - O assessor Dr. Gustavo Mascarenhas Lacerda Pedrina assim revelou
       ↪ os contornos da impetração:
19
20 Eis o que informado quando da análise do pedido de liminar:
21 [...]
22 1. O assessor Dr. Gustavo Mascarenhas Lacerda Pedrina prestou as seguintes informações:
23 O Juízo da Oitava Vara Criminal da Comarca de Curitiba/PR, ao acolher manifestação do Ministério
       ↪ Público, no processo nº 0015539-58.2014.8.16.0013, determinou, em 25 de maio de 2015, a
       ↪ prisão preventiva do paciente, ocorrida em 17 de dezembro de 2015, e de outros investigados,
       ↪ ante o suposto cometimento da infração descrita no artigo 2º (integrar organização
       ↪ criminosa), parágrafos 2º e 4º, inciso IV (causas de aumento de pena alusivas ao emprego de
       ↪ arma de fogo e à conexão com outros grupos independentes), da Lei nº 12.850/2013.
       ↪ Reportou-se ao conteúdo de interceptações telefônicas, o qual, segundo assentou, revelou
       ↪ fortes indícios de participação do paciente na facção denominada Primeiro Comando da Capital
       ↪ (PCC). Destacou comprovada a expressa diretriz da organização de expansão das atividades
       ↪ ilícitas no Estado do Paraná, bem assim menção a eventos relacionados a rebeliões em
       ↪ presídios, prática de homicídios e tráfico de drogas.
24 Chegou-se ao Superior Tribunal de Justiça com o habeas corpus nº 447.695/PR, o qual teve a liminar
       ↪ indeferida. O impetrante ressalta inidôneo o ato por meio do qual determinada a preventiva,
       ↪ tendo-o como calcado em fundamentação abstrata. Sustenta o excesso de prazo da constrição,
       ↪ dizendo perdurar por mais de 2 anos. Frisa não atendidos os requisitos previstos no artigo
       ↪ 312 do Código de Processo Penal. Articula com o princípio da proporcionalidade. Assevera a
       ↪ viabilidade de substituição da medida por cautelar versada no artigo 319 do Código de
       ↪ Processo Penal.
25 [...]
26 Requereu, no campo precário e efêmero, a revogação da preventiva. No mérito, busca a confirmação
       ↪ da providência. Vossa Excelência, em 22 de maio de 2018, deferiu o pedido de medida
       ↪ acauteladora para afastar a prisão, considerado o excesso de prazo.
27 A Procuradoria-Geral da República opina pela inadmissão da impetração, dizendo-a formalizada
       ↪ contra decisão individual. Destaca ausente ilegalidade a ser reparada. Não foi possível
       ↪ acessar o andamento processual mediante consulta ao sítio do Tribunal estadual, uma vez
       ↪ submetido a segredo de justiça.
28 No Superior Tribunal de Justiça o processo revelador do habeas de nº 447.695/PR encontra-se
       ↪ concluso ao Relator.
29 A impetração foi distribuída a Vossa Excelência em virtude da vinculação com a de nº 151.610/PR, a
       ↪ qual está aparelhada para julgamento.
30
31 Lancei visto no processo em 24 de novembro de 2018, liberando-o para ser examinado na Turma a
       ↪ partir de 4 de dezembro seguinte, isso objetivando a ciência do impetrante.
32 É o relatório.
33
34 VOTO
35 O SENHOR MINISTRO MARCO AURÉLIO (RELATOR):
36
37 HABEAS CORPUS - ATO INDIVIDUAL - ADEQUAÇÃO. O habeas corpus é adequado em se tratando de
       ↪ impugnação a ato de Colegiado ou individual.
38 PRISÃO PREVENTIVA - ORGANIZAÇÃO CRIMINOSA. Uma vez decorrendo a custódia da prática do crime de
       ↪ integração a organização criminosa, relacionada a rebelião em presídios, cometimento de
       ↪ homicídios e tráfico de entorpecentes e armas, a teor de conteúdo obtido mediante
       ↪ interceptações telefônicas autorizadas e apreensão de documentos, tem-se dados a sinalizarem
       ↪ a periculosidade do envolvido, motivando, validamente, a prisão preventiva.
39 PRISÃO PREVENTIVA - PRAZO - EXCESSO. Configurado o excesso de prazo da custódia preventiva,
       ↪ impõe-se a devolução da liberdade ao acusado.
40
41 Improcede a preliminar veiculada pela Procuradoria-Geral da República. O habeas corpus é cabível
       ↪ contra ato de Colegiado ou individual. Importante é saber a existência de órgão com
       ↪ atribuição de examinar o merecimento da decisão atacada. Os integrantes do Superior Tribunal
       ↪ de Justiça estão submetidos, nos crimes comuns e de responsabilidade, à jurisdição do
       ↪ Supremo. Fora isso é esvaziar o alcance da impetração.
42 Reitero o que assentado, em 22 de maio de 2018, quando do implemento da medida de urgência:
43
44 [...]
```

```
2. O Juízo, ao determinar a prisão preventiva, apontou tratar-se de cidadão supostamente integrante de organização criminosa, aludindo a referências e anotações encontradas em caderno utilizado pela facção e apreendido em poder de corré. Destacou a existência de interceptações telefônicas nas quais revelada a diretriz do grupo de expansão das atividades ilícitas no Estado do Paraná, além da menção a eventos relacionados a rebeliões em presídios, prática de homicídios e tráfico de drogas, no que demonstrado estar em jogo a preservação da ordem pública. Sem prejuízo do princípio da não culpabilidade, a medida se impunha, ante a periculosidade do agente, ao menos sinalizada. Daí ter-se como razoável e conveniente o ato atacado. A inversão da ordem do processo-crime apurar para, selada a culpa, em verdadeira execução da pena, prender foi justificada, atendendo-se ao figurino legal.
Surge a problemática do excesso de prazo. Presente a data de cumprimento do mandado de prisão preventiva 17 de dezembro de 2015 , esta perdura por 2 anos, 1 mês e 20 dias. Privar da liberdade, por tempo desproporcional, pessoa cuja responsabilidade penal não veio a ser declarada em definitivo viola o princípio da não culpabilidade. Concluir pela manutenção da medida é autorizar a transmutação do pronunciamento mediante o qual implementada, em execução antecipada da pena, ignorando-se garantia constitucional.
[...]

Defiro a ordem, para afastar a prisão preventiva do paciente, formalizada no processo n° 0015539-58.2014.8.16.0013, da Oitava Vara Criminal da Comarca de Curitiba/PR. Advirtam-no da necessidade de permanecer com a residência indicada ao Juízo, atendendo aos chamamentos judiciais, de informar possível transferência e de adotar a postura que se aguarda do homem médio, integrado à sociedade. É como voto.

VOTO
O SENHOR MINISTRO LUÍS ROBERTO BARROSO - Presidente, é hipótese de Súmula n° 691, não conheço. Diante do exposto, não conheço do habeas corpus, revogada a liminar.

VOTO
O SENHOR MINISTRO ALEXANDRE DE MORAES (PRESIDENTE) - Também peço vênia ao eminente Ministro-Relator. É um caso semelhante ao anterior, o PCC no Paraná. Também não conheço.
```

A.4. Brazilian case (summary) - Habeas Corpus n. 156667

```
- Nome do caso: Habeas Corpus 156667 / PR
- Resumo:
O habeas corpus foi impetrado em favor de um paciente que foi preso preventivamente em 17 de dezembro de 2015, acusado de integrar a organização criminosa Primeiro Comando da Capital (PCC). A prisão foi determinada com base em interceptações telefônicas que revelaram indícios de participação do paciente na facção. O paciente alega que a prisão é indevida, pois foi baseada em fundamentação abstrata e que o prazo de prisão preventiva já ultrapassou os dois anos. O paciente também alega que não foram atendidos os requisitos previstos no artigo 312 do Código de Processo Penal e que a medida pode ser substituída por uma cautelar mais branda.

- Fatores chave:

F1: excesso de tempo no cárcere - O paciente alega que a prisão preventiva já ultrapassou os dois anos, o que é considerado excessivo e viola o princípio da proporcionalidade.

F6: presença de escuta telefônica - As interceptações telefônicas foram usadas como prova para determinar a prisão preventiva do paciente.

F7: complexidade do processo ou riscos existentes - O processo envolve uma organização criminosa complexa e há riscos de fuga ou de continuação de atividades ilícitas.

F11: limitação da defesa - O paciente alega que a prisão preventiva limita sua defesa e que não foram atendidos os requisitos previstos no artigo 312 do Código de Processo Penal.

Esses fatores chave são fundamentais para entender a decisão do habeas corpus e as razões pelas quais o paciente alega que a prisão preventiva é indevida.
```

Appendix B. Full knowledge graphs with summarization step

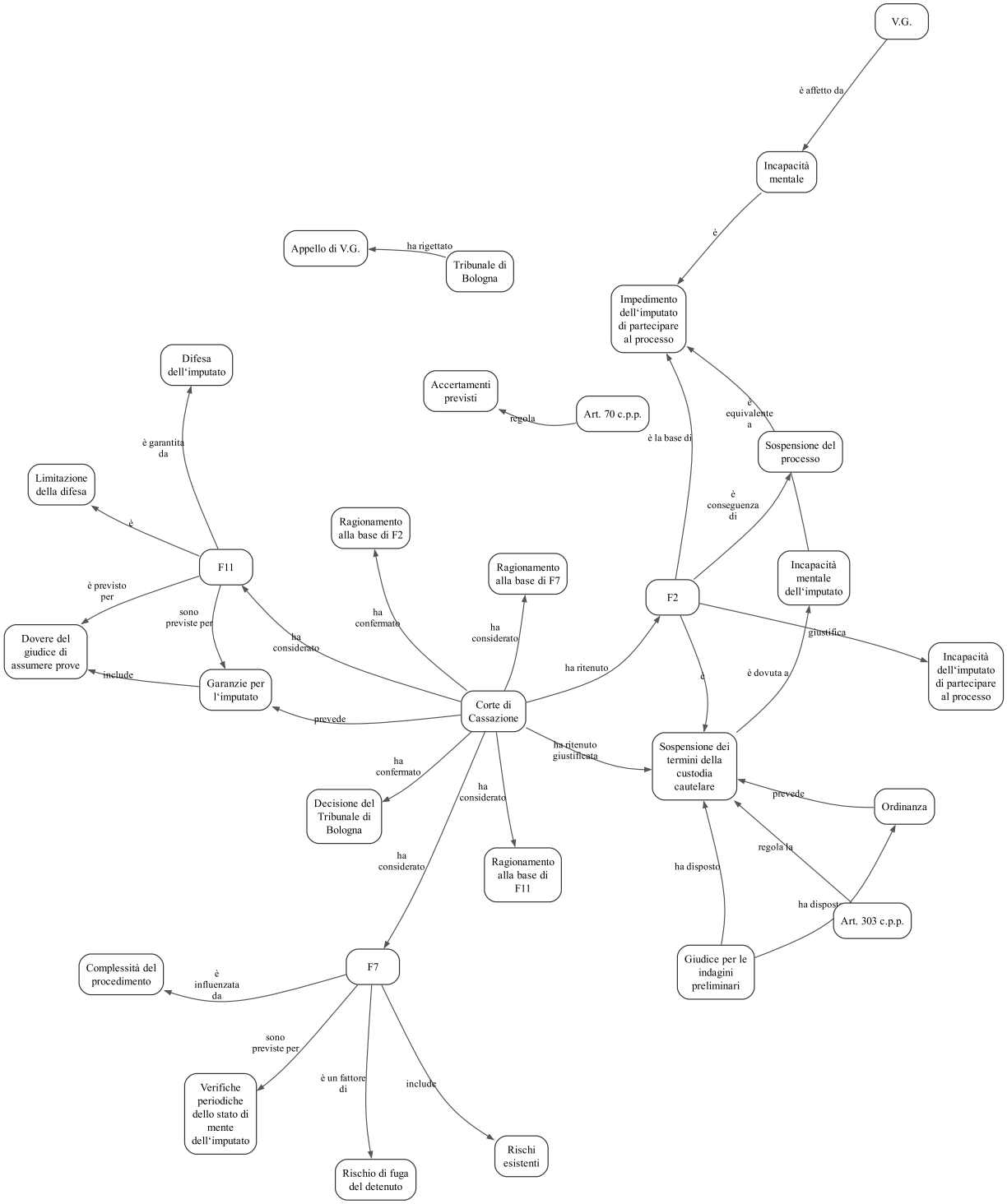

Fig. B.6. Visualisation of the KG for Sentenza n. 16939 using summary (IT).

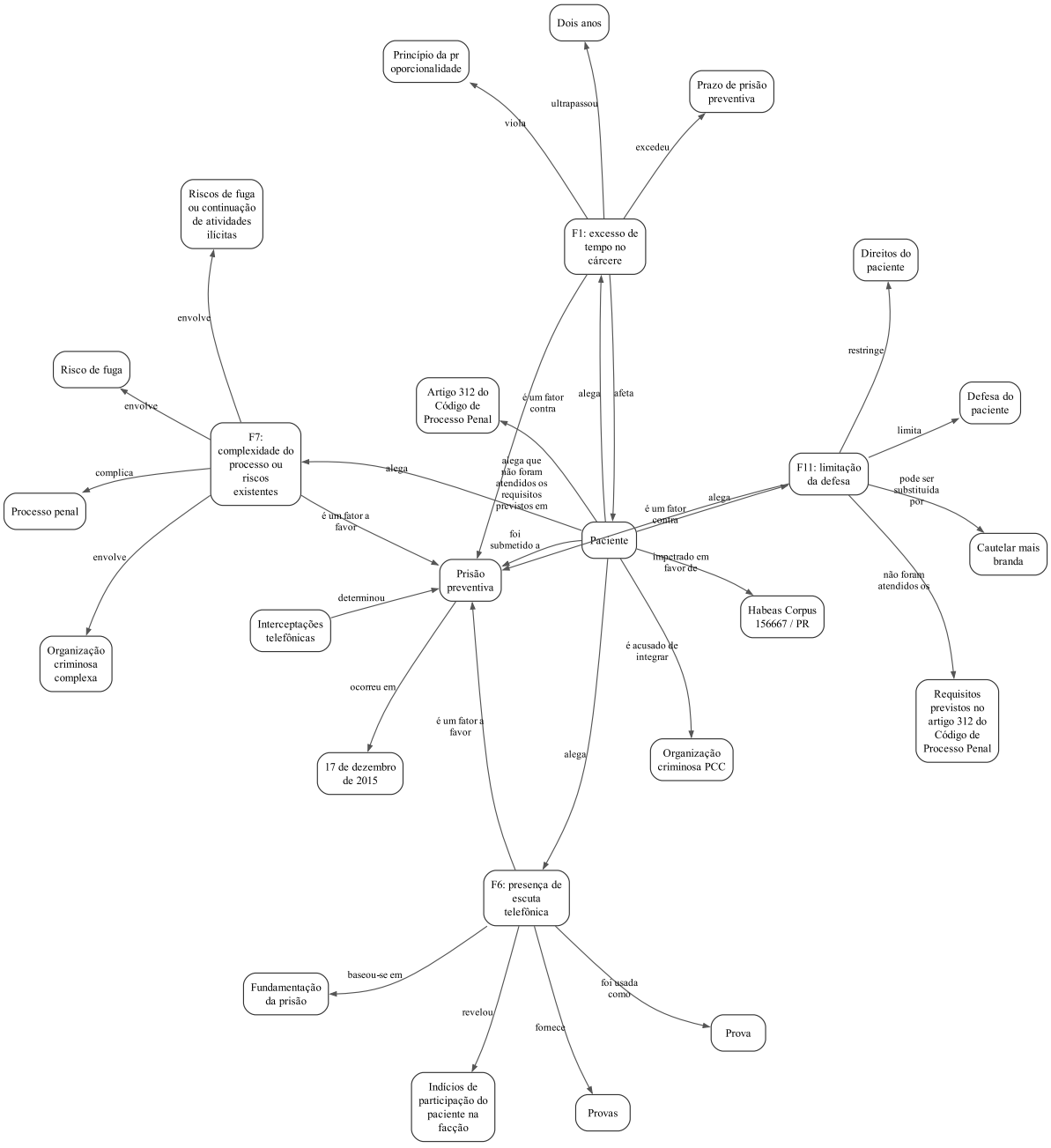

Fig. B.7. Visualisation of the KG for HC n. 156667 using summary (BR).

Appendix C. Full knowledge graphs without summarization step

See Figs. C.8 and C.9.

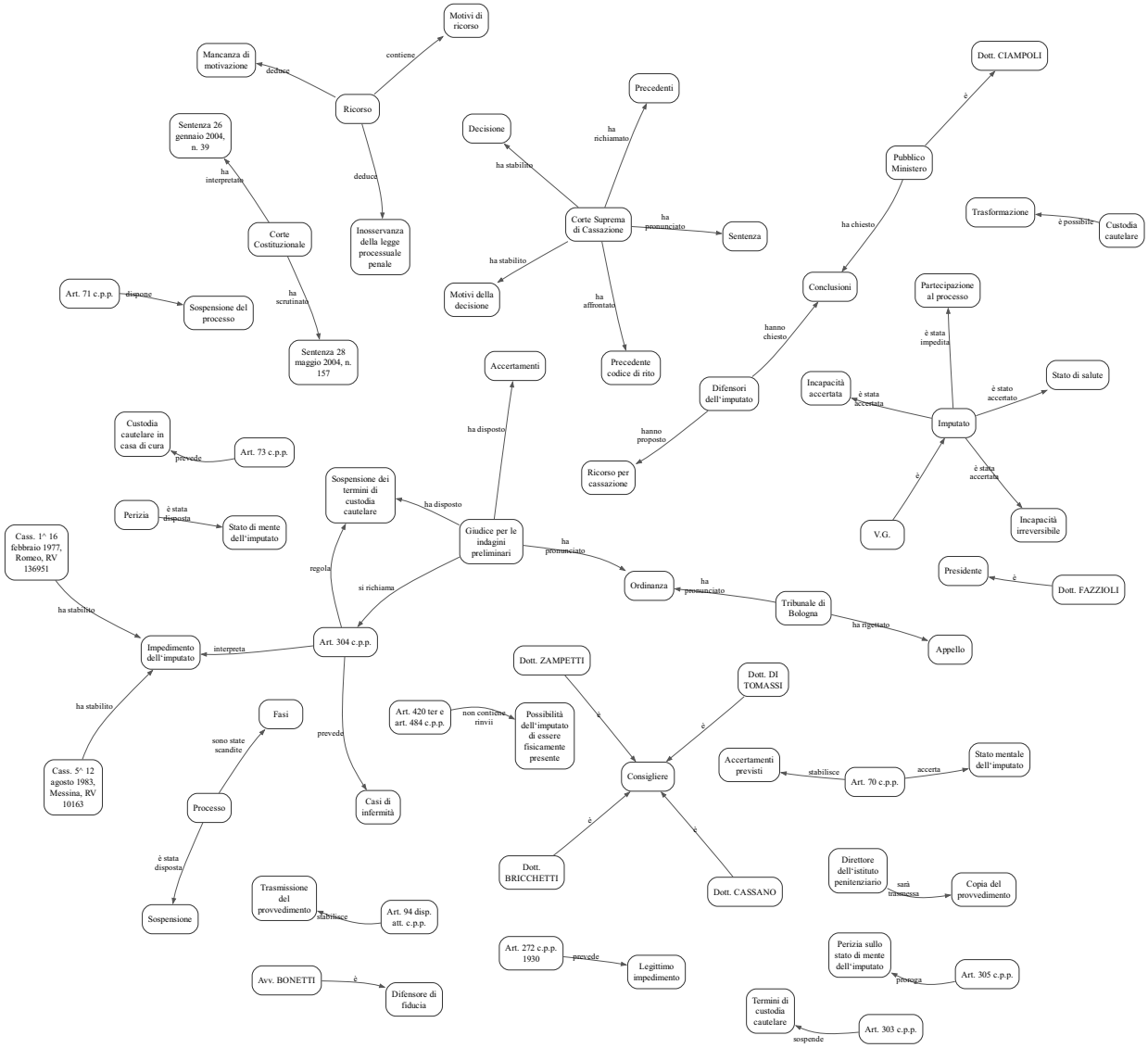

Fig. C.8. Visualisation of the KG for Sentenza n. 16939 without summary (IT).

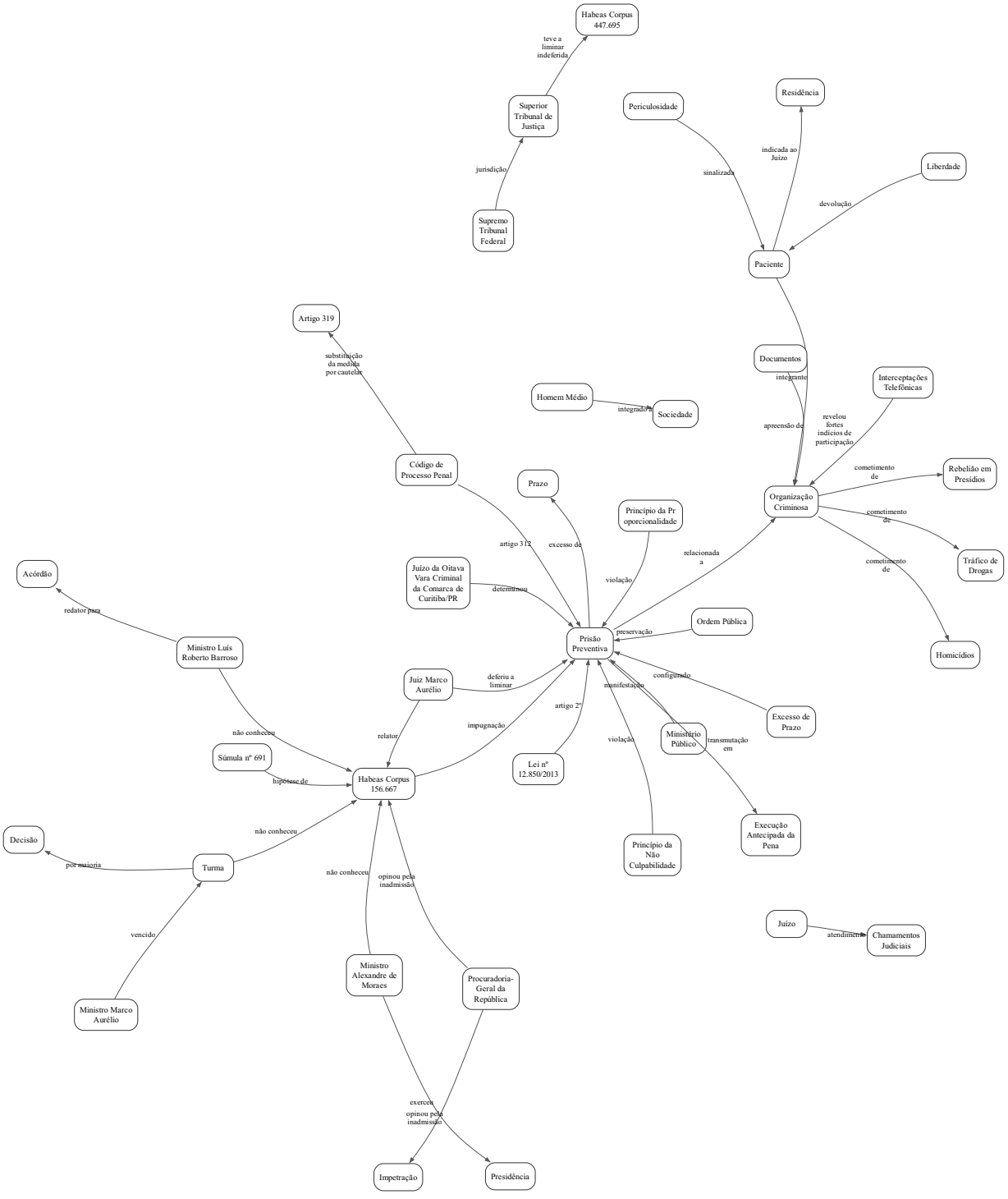

Fig. C.9. Visualisation of the KG for HC n. 156667 without summary (BR).

References

[1] Martineau K. What is generative AI? 2024, Online; https://research.ibm.com/blog/what-is-generative-AI. [Accessed 01 September 2024].

[2] Khennouche F, Elmir Y, Himeur Y, Djebari N, Amira A. Revolutionizing generative pre-traineds: Insights and challenges in deploying ChatGPT and generative chatbots for FAQs. Expert Syst Appl 2024;246:123224. http://dx.doi.org/10.1016/j.eswa.2024.123224.

[3] Tang L, Sun Z, Idnay B, Nestor JG, Soroush A, Elias PA, Xu Z, Ding Y, Durrett G, Rousseau JF, Weng C, Peng Y. Evaluating large language models on medical evidence summarization. Npj Digit Med 2023;6(1). http://dx.doi.org/10.1038/s41746-023-00896-7.

[4] Ji S, Pan S, Cambria E, Marttinen P, Yu PS. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Trans Neural Netw Learn Syst 2022;33(2):494–514. http://dx.doi.org/10.1109/tnnls.2021.3070843.

[5] Horty JF, Bench-Capon TJ. A factor-based definition of precedential constraint. Artif Intell Law 2012;20(2):181–214.

[6] Horty J. Reasoning with dimensions and magnitudes. Artif Intell Law 2019;27(3):309–45.

[7] Ashley KD. Prospects for legal analytics: Some approaches to extracting more meaning from legal texts. Univ Cincinnati Law Rev 2022;90(4):5.

[8] Dal Pont TR, et al. Classification and association rules in Brazilian supreme court judgments on pre-trial detention. In: International conference on electronic government and the information systems perspective. Springer; 2021, p. 131–42.

[9] Sabo IC, Billi M, Lagioia F, Sartor G, Rover AJ. Unsupervised factor extraction from pretrial detention decisions by Italian and Brazilian supreme courts. In: Guizzardi R, Neumayr B, editors. Advances in conceptual modeling. Cham: Springer International Publishing; 2022, p. 69–80. http://dx.doi.org/10.1007/978-3-031-22036-4_7.

[10] Billi M, Dal Pont TR, Sabo IC, Lagioia F, Sartor G, Rover AJ. Supervised learning, explanation and interpretation from pretrial detention decisions by Italian and Brazilian supreme courts. In: Sales TP, Araújo J, Borbinha J, Guizzardi G, editors. Advances in conceptual modeling. Cham: Springer Nature Switzerland; 2023, p. 131–40. http://dx.doi.org/10.1007/978-3-031-47112-4_12.

[11] Calvin Barron R, Eren M, Serafimova O, Matuszek C, Alexandrov B. Bridging legal knowledge and AI: Retrieval-augmented generation with vector stores, knowledge graphs, and hierarchical non-negative matrix factorization. In: Proceedings of the twentieth international conference on artificial intelligence and law. New York, NY, USA: Association for Computing Machinery; 2026, p. 51–60. http://dx.doi.org/10.1145/3769126.3769215, URL https://doi.org/10.1145/3769126.3769215.

[12] Bellan P, Dragoni M, Ghidini C. Process knowledge extraction and knowledge graph construction through prompting: A quantitative analysis. In: Proceedings of the 39th ACM/SIGAPP symposium on applied computing, vol. 1, ACM; 2024, p. 1634–41. http://dx.doi.org/10.1145/3605098.3635957.

[13] Zhao W, Chen Q, You J. LlmRe: A zero-shot entity relation extraction method based on the large language model. In: Proceedings of the 2023 7th international conference on electronic information technology and computer engineering, vol. 35, ACM; 2023, p. 475–80. http://dx.doi.org/10.1145/3650400.3650478.

[14] Dong H, Wu B. Enhancing named entity recognition in safety hazard analysis through GBD and LLMs. In: 2024 7th international conference on information and computer technologies. IEEE; 2024, p. 13–9. http://dx.doi.org/10.1109/icict62343.2024.00009.

[15] De Bellis A. Structuring the unstructured: an LLM-guided transition. In: d'Amato C, Pan J, editors. Proceedings of the doctoral consortium at ISWC 2023 co-located with 22nd international semantic web conference. CEUR workshop proceedings, vol. 3678, Athens, Greece; 2023-11-07, URL https://ceur-ws.org/Vol-3678/paper12.pdf.

[16] Yang X, Wang Z, Wang Q, Wei K, Zhang K, Shi J. Large language models for automated Q&A involving legal documents: a survey on algorithms, frameworks and applications. Int J Web Inf Syst 2024;20(4):413–35. http://dx.doi.org/10.1108/ijwis-12-2023-0256.

[17] Liu S, Cao J, Li Y, Yang R, Wen Z. Low-resource court judgment summarization for common law systems. Inf Process Manage 2024;61(5):103796. http://dx.doi.org/10.1016/j.ipm.2024.103796.

[18] Nay JJ, Karamardian D, Lawsky SB, Tao W, Bhat M, Jain R, Lee AT, Choi JH, Kasai J. Large language models as tax attorneys: a case study in legal capabilities emergence. Philos Trans R Soc: Math Phys Eng Sci 2024;382(2270). http://dx.doi.org/10.1098/rsta.2023.0159.

[19] Dal Pont T, Galli F, Loreggia A, Pisano G, Rovatti R, Sartor G. Legal Summarisation through LLMs: The PRODIGIT Project. 2023, arXiv preprint 2308.04416, https://arxiv.org/abs/2308.04416.

[20] Gray MA, Savelka J, Oliver WM, Ashley KD. Empirical legal analysis simplified: reducing complexity through automatic identification and evaluation of legally relevant factors. Philos Trans R Soc A: Math Phys Eng Sci 2024;382(2270). http://dx.doi.org/10.1098/rsta.2023.0155.

[21] Shi J, Guo Q, Liao Y, Wang Y, Chen S, Liang S. Legal-LM: Knowledge graph enhanced large language models for law consulting. In: Advanced intelligent computing technology and applications. Springer Nature Singapore; 2024, p. 175–86. http://dx.doi.org/10.1007/978-981-97-5672-8_15.

[22] Rose DE, Belew RK. A connectionist and symbolic hybrid for improving legal research. Int J Man-Mach Stud 1991;35(1):1–33. http://dx.doi.org/10.1016/s0020-7373(07)80006-x.

[23] Reimer U, Margelisch A. A hybrid knowledge representation approach to reusability of legal knowledge bases. In: Proceedings of the fifth international conference on artificial intelligence and law. ACM Press; 1995, p. 246–55. http://dx.doi.org/10.1145/222092.222249.

[24] Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. Sov Phys Dokl 1965;10:707–10.

[25] Jaccard P. The distribution of the flora in the alpine zone. New Phytol 1912.

[26] Feng F, Yang Y, Cer D, Arivazhagan N, Wang W. Language-agnostic BERT sentence embedding. In: Muresan S, Nakov P, Villavicencio A, editors. Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers). Dublin, Ireland: Association for Computational Linguistics; 2022, p. 878–91. http://dx.doi.org/10.18653/v1/2022.acl-long.62, URL https://aclanthology.org/2022.acl-long.62/.

[27] Costa MZ, Robson DF, Vieira TBP, Bourguet J-R, Guizzardi G, Almeida JPA. Automated semantic annotation pipeline for brazilian judicial decisions. In: 37th annual conference on legal knowledge and information systems. IOS; 2024, p. 226–38.