

Supplementary Appendix for *GroundSTory: Interactive Storytelling for Grounded Spatiotemporal Anomaly Explanations with Graphs and Large Language Models*

JUANPABLO HEREDIA and LEIGHTON ESTRADA-RAYME, Fundação Getulio Vargas (FGV), Brazil

JEREMY MATOS-CANGALAYA, University of Engineering and Technology (UTEC), Peru

JORGE POCO, Fundação Getulio Vargas (FGV), Brazil

A Synthetic Data Construction

This section describes the procedure used to generate the synthetic spatiotemporal (ST) dataset employed for quantitative comparison of anomaly detection methods.

Spatial graph generation. We generate a spatial graph with $N = 400$ nodes. Node positions are sampled uniformly in the unit square $[0, 1]^2$ and linearly rescaled to the domain $[0, 20]^2$. Each sampled point corresponds to a node in the graph.

Edges are constructed using a k -nearest-neighbor strategy. For each node u , the number of neighbors k_u is sampled uniformly from $\{3, 4, 5, 6\}$. Node u is then connected to its k_u nearest neighbors in Euclidean distance, resulting in an undirected graph with an average node degree of approximately 5.

Spatiotemporal signal generation. To generate the underlying ST signal, we define a sequence of matrices $\{M_t\}_{t=1}^T = 100$, where each $M_t \in \mathbb{R}^{20 \times 20}$ represents the signal field at time t . Each matrix M_t is constructed by summing two two-dimensional Gaussian functions with different variances and shared amplitude range $[0, 1]$. One Gaussian is narrower and inverted, while the other is broader; their sum produces a ring-shaped spatial pattern. The center of the pattern moves along the vertical axis according to

$$(x_c(t), y_c(t)) = (10, 20 - t),$$

resulting in a structure that translates from the bottom to the top of the spatial domain over time.

The signal value at node u and time t is obtained by mapping the node’s spatial position to the corresponding matrix entry:

$$S(u, t) = M_t(\lfloor x_u \rfloor, \lfloor y_u \rfloor),$$

where (x_u, y_u) denotes the spatial coordinates of node u .

Anomaly injection. Anomalies are injected locally into the ST signal to simulate sparse and localized irregular events. For each time step t , we independently sample 400 values from a uniform distribution over $[0, 1]$ (one per node). Entries with sampled values ≤ 0.01 are selected as candidate anomalies.

For each selected node–time pair (u, t) , anomalies are injected conditional on the local signal magnitude:

- If $S(u, t) < 0.3$, a positive anomaly (peak) is introduced by adding a random value sampled uniformly from $[0.8, 0.9]$.

- If $S(u, t) > 0.7$, a negative anomaly (valley) is introduced by replacing the value with a random number sampled uniformly from $[0.1, 0.2]$.

This process results in a total of 754 injected anomalies across the dataset, affecting 316 distinct nodes. We deliberately avoid injecting global pattern shifts, as our objective is to evaluate the detection of localized and context-dependent anomalies. Each anomaly is uniquely identified by the pair (u, t) , corresponding to the node index and time step at which the anomaly occurs.

B Amenity Details

Table 1. Amenity categories and corresponding OpenStreetMap amenity types used in this work.

Category	Amenity Types
Transport	metro station, bus station, traffic signals, fuel station, parking, charging station
Food & Nightlife	restaurant, fast_food, bars & pubs, nightclub
Education	university, school, college
Commerce & Services	shopping centers, banks & ATM
Public Services	police station, place of worship
Entertainment & Lodging	cinema, theatre, hotel

C Prompt and Response Examples

C.1 Explanation Prompt

System:

- You are a spatial-temporal reasoning assistant trained to explain anomalies in urban crime data.
- You analyze structured textual graphs of street networks and temporal crime patterns in the city center of Sao Paulo in Sao Paulo, Brazil.
- Your output must be written in English, without special characters, not even for people or place names.
- Your output must always be a strictly valid and well-formed XML document. Ensure all tags are correctly nested and closed. No extraneous text should appear outside the main XML structure.
- You prioritize factual accuracy and rely solely on the provided data and verifiable information obtained through instructed web searches. Do not invent information, sources, or URLs.

User:

A crime anomaly has been detected at a specific street and week. An anomaly is an unexpected value compared to spatial neighbors, temporal patterns, or both.

Data Provided

- Textual Graph: Contains nodes with street info, nearby amenities, crime time series, holiday markers, and an explicit anomaly indicator

Amenity Security Risk Reference

Use this classification when analyzing anomalies:

- High Insecurity Risk:
 - Nightclubs: Alcohol/drugs, crowds, late hours
- Insecurity Risk:
 - Metro Stations: High traffic, anonymity, theft opportunities
 - Bars/Pubs: Alcohol-related violence, late hours
 - Fuel Stations: Cash transactions, 24/7 isolated locations
 - Banks/ATMs: Cash targets, robbery risk
- Moderate Risk:
 - Places of Worship: Respected venues, occasional large gatherings
- Security (Lower Risk):
 - Shopping Centers: Controlled, surveilled environments
 - Hotels: Internal security, but tourists attractive targets
- High Security:
 - Police Stations: Crime deterrent, safety perception

Task Structure

Complete these three sequential parts. Do not let later parts influence earlier analysis.

- Part 1: Initial Hypotheses (Data-Only Analysis)
 - Identify the anomaly's spatial extent and temporal duration based on the provided graph data.
 - Generate at least 2 distinct hypotheses explaining the indicated anomaly.
 - Base hypotheses **exclusively** on the graph data: amenities, crime patterns, holidays, and spatial relationships between the nodes.
 - Consider why this specific location/time was affected differently than its direct spatial neighbors.
- Part 2: Web Validation
 - **Crucial Geographic Focus:** Search for verifiable evidence **exclusively** from **Sao Paulo city center**, focusing on events, news, or public safety reports that occurred **specifically** in the vicinity described by the graph during the anomalous period.
 - Prioritize information related to the specific streets mentioned in the graph, or very immediate surrounding blocks. **Avoid general city-wide news or events that are not spatially specific to this small area.**
 - Focus on: local news outlets, official municipal event calendars, public safety reports, and government announcements relevant to the **Sao Paulo city center**.
 - Evaluate each hypothesis generated in Part 1 against the web evidence found.
 - Select the single best-supported hypothesis.
 - If no relevant web evidence is found **specifically for the designated small geographic area and time**, explicitly state this.
 - If the evidence found is not highly consistent with the precise spatial or temporal location (e.g., it's a city-wide event or in a different neighborhood), **disregard it and continue searching for more specific evidence, or state that no relevant evidence was found.**

- Part 3: Final XML Report

Format your conclusion using the specified XML structure, including only web-verified sources in the References section.

Textual Graph

=== Nodes ===

Node ID:553 (Praca da Republica, secondary street)

Nearby amenities (distances in meters):

- metro station: 13.63
- shopping centers: 156.10
- banks & atm: 17.98
- police station: 158.04
- place of worship: 127.09
- bars & pubs: 20.86
- nightclub: 183.87
- hotel: 41.67

Crime Incidents (Weekly):

- 2023-01-16 to 2023-01-22: 12
- 2023-01-23 to 2023-01-29: 19 (Sao Paulo City Anniversary)
- 2023-01-30 to 2023-02-05: 12
- 2023-02-06 to 2023-02-12: 15
- **2023-02-13 to 2023-02-19: 25 (Anomaly)**
- 2023-02-20 to 2023-02-26: 28 (Carnival Monday, Carnival Tuesday, Ash Wednesday)

Node ID:549 (Rua Vinte e Quatro de Maio, residential street)

Nearby amenities (distances in meters):

- shopping centers: 6.24
- banks & atm: 15.11
- police station: 6.36
- place of worship: 127.09
- bars & pubs: 61.53
- nightclub: 119.52
- hotel: 30.99

Crime Incidents (Weekly):

- 2023-01-16 to 2023-01-22: 6
- 2023-01-23 to 2023-01-29: 7 (Sao Paulo City Anniversary)
- 2023-01-30 to 2023-02-05: 14
- 2023-02-06 to 2023-02-12: 6
- 2023-02-13 to 2023-02-19: 19
- 2023-02-20 to 2023-02-26: 19 (Carnival Monday, Carnival Tuesday, Ash Wednesday)

Node ID:556 (Praca da Republica, secondary street)

Nearby amenities (distances in meters):

- metro station: 21.20

```

- shopping centers: 168.36
- banks & atm: 50.64
- bars & pubs: 24.20
- hotel: 87.28
Crime Incidents (Weekly):
- 2023-01-16 to 2023-01-22: 4
- 2023-01-23 to 2023-01-29: 3 (Sao Paulo City Anniversary)
- 2023-01-30 to 2023-02-05: 2
- 2023-02-06 to 2023-02-12: 8
- 2023-02-13 to 2023-02-19: 11
- 2023-02-20 to 2023-02-26: 13 (Carnival Monday, Carnival Tuesday, Ash Wednesday)

Node ID:558 (Avenida Ipiranga, secondary street)
Nearby amenities (distances in meters):
- shopping centers: 174.16
- banks & atm: 16.62
- police station: 151.49
- place of worship: 114.53
- bars & pubs: 13.33
- nightclub: 137.02
- hotel: 18.04
Crime Incidents (Weekly):
- 2023-01-16 to 2023-01-22: 1
- 2023-01-23 to 2023-01-29: 0 (Sao Paulo City Anniversary)
- 2023-01-30 to 2023-02-05: 1
- 2023-02-06 to 2023-02-12: 7
- 2023-02-13 to 2023-02-19: 7
- 2023-02-20 to 2023-02-26: 6 (Carnival Monday, Carnival Tuesday, Ash Wednesday)

=== Edges ===
"Praca da Republica" (id:553) -- "Rua Vinte e Quatro de Maio" (id:549)
"Praca da Republica" (id:553) -- "Praca da Republica" (id:556)
"Praca da Republica" (id:553) -- "Avenida Ipiranga" (id:558)

Key Requirements
- Always reference nodes by ID: (id: xxxx)
- Use only real, accessible URLs in References
- Base conclusions on verifiable evidence only
- You must adhere strictly to the following output structure:

```

C.2 Judge Prompt Example

Stage I:

```

System:
You're a judge evaluator for comparing the retrieved URL references to support anomaly explanations. These
will be provided in XML format under the <References> tag.

User:
<ANOMALY_GRAPH>
{textual_graph}
</ANOMALY_GRAPH>

You will be given two sets of URL references in XML format under the <References> tag. Each set corresponds
to an anomaly explanation generated by different methods.

Your task is to evaluate which set of references better supports the anomaly explanation based on the
following criteria:
1. Relevance: How closely the references relate to the anomaly described.
2. Credibility: The trustworthiness of the sources provided.
3. Diversity: Variety of sources and perspectives included.
You must provide a verdict of either 'FIRST' or 'SECOND', indicating which set of references is superior.
Additionally, provide a detailed reasoning for your verdict. Remember that more references do not
necessarily mean better quality; focus on the quality and relevance of the references individually.

{first_references}

{second_references}

```

Please provide your evaluation in the following JSON format:
"verdict": "FIRST" or "SECOND",
"short_reasoning": "Explanation of the verdict in terms of criteria"

Response without web validation (i.e. the first_references set):

```
<References>
  <Item>
    <Id>REF_001</Id>
    <Source>Ramos de Azevedo Square</Source>
    <URL>https://en.wikipedia.org/wiki/Ramos_de_Azevedo_Square</URL>
    <Excerpt>The Ramos de Azevedo Square is located in the Republica district, in the center of the Brazilian city of Sao Paulo, and is famous for being home to the Municipal Theater.</Excerpt>
    <Date>2023-11-17</Date>
  </Item>
  <Item>
    <Id>REF_002</Id>
    <Source>Independence Day (Brazil)</Source>
    <URL>https://en.wikipedia.org/wiki/Independence_Day_(Brazil)</URL>
    <Excerpt>Independence Day is a national holiday marked by patriotic displays and military parades in most Brazilian cities.</Excerpt>
    <Date>2023-11-17</Date>
  </Item>
  <Item>
    <Id>REF_003</Id>
    <Source>Liberdade street market</Source>
    <URL>https://en.wikipedia.org/wiki/Liberdade_street_market</URL>
    <Excerpt>The Liberdade Street Fair is an art and handicraft fair in the Liberdade district of Sao Paulo.</Excerpt>
    <Date>2023-11-17</Date>
  </Item>
  <Item>
    <Id>REF_004</Id>
    <Source>2022 Brazilian general election</Source>
    <URL>https://en.wikipedia.org/wiki/2022_Brazilian_general_election</URL>
    <Excerpt>Since the official beginning of the election campaign in August 2022 and before that, various political commentators said the incumbent president Jair Bolsonaro incited physical, or verbal violence, against his critics and political opponents.</Excerpt>
    <Date>2023-11-17</Date>
  </Item>
  <Item>
    <Id>REF_005</Id>
    <Source>Political Violence During Brazil's 2022 Presidential Runoff | ACLED</Source>
    <URL>https://acleddata.com/report/political-violence-during-brazils-2022-presidential-runoff/</URL>
    <Excerpt>The lead-up to the first round of voting on 2 October saw higher levels of political violence involving political party representatives and supporters than the 2018 election.</Excerpt>
    <Date>2023-11-17</Date>
  </Item>
</References>
```

Response after adding web validation constraints (i.e. the second_references set):

```
<References>
  <Item>
    <Id>REF_001</Id>
    <Source>Criminosos aterrorizam moradores e comerciantes na regioao da Praca da Republica - Record</Source>
    <URL>https://record.r7.com/balanco-geral/videos/criminosos-aterorizam-moradores-e-comerciantes-na-regiao-da-praca-da-republica-29102022/</URL>
    <Excerpt>A Praca da Republica, um dos pontos mais tradicionais do centro de Sao Paulo, se transformou em um lugar perigoso para circular. Moradores e comerciantes da regioao estao amedrontados com o crescente numero de crimes violentos que acontecem, ate mesmo de dia.</Excerpt>
    <Date>2022-10-29</Date>
  </Item>
  <Item>
    <Id>REF_002</Id>
    <Source>Praca da Republica (Sao Paulo) - Wikipedia</Source>
```

```
<URL>https://en.wikipedia.org/wiki/Pra%C3%A7a_da_Rep%C3%BAblica_%28S%C3%A3o_Paulo%29</URL>
<Excerpt>Praça da Republica is a park and public square in the Republica neighborhood of Sao Paulo,
Brazil. The park covers several city blocks between Rua Pedro Americo, Rua Vinte e Quatro de Maio, Avenida
Ipiranga, and Avenida Sao Joao in the historic center of the city.</Excerpt>
<Date>2023-05-15</Date>
</Item>
<Item>
<Id>REF_003</Id>
<Source>2022 in Brazil - Wikipedia</Source>
<URL>https://en.wikipedia.org/wiki/2022_in_Brazil</URL>
<Excerpt>October 2: First round of general elections. For President, candidates Luiz Inacio Lula da
Silva (PT) and Jair Bolsonaro (PL) go to the second round.</Excerpt>
<Date>2023-03-10</Date>
</Item>
</References>
```

Stage II:

System:

You're a judge evaluator for comparing the retrieved responses of 4 different methods for generating anomaly explanations. These will be provided in XML format.

The objective of this evaluation is to determine how well the hypotheses use different patterns (relational, functional, temporal, spatial) to support the anomaly explanation.

User:

You will be given two sets of XMLs format under the `<TAG_SELECTED>` tag. Each set corresponds to an anomaly explanation generated by different methods for the following textual graph:

```
{{textual_graph}}
```

Your task is to evaluate which set of HYPOTHESIS and REFERENCES related to the anomaly and its surroundings better supports the anomaly explanation based on the following criteria:

1. Spatio-temporal insights: The depth of spatial and temporal analysis provided.
2. Innovation: The innovativeness of the approach taken in the explanation, in comparison to others.
3. Completeness: How thoroughly the explanation covers different aspects of the anomaly.

You must provide a verdict of either 'FIRST', 'SECOND', 'THIRD', or 'FOURTH', indicating which set of `<TAG_SELECTED>` is superior. Additionally, provide a concise reasoning for your verdict.

FIRST:

```
```xml
```

```
{{first_option}}
```

```
```
```

SECOND:

```
```xml
```

```
{{second_option}}
```

```
```
```

THIRD:

```
```xml
```

```
{{third_option}}
```

```
```
```

FOURTH:

```
```xml
```

```
{{fourth_option}}
```

```
```
```

First option (Level 1):

<Hypothesis>

<Statement>The anomaly is likely due to increased criminal activity targeting attendees of the Feira de Artesanato e Gastronomia da Praça da Republica, which attracts large crowds during weekends.</Statement>

<Reasoning>

<Relational>

<Description>Node (id:553) hosts the Feira de Artesanato e Gastronomia da Praça da Republica every Saturday and Sunday, drawing significant foot traffic. The presence of nearby amenities such as bars and pubs (id:553), a metro station (id:553), and shopping centers (id:553) further increases the area's attractiveness to both visitors and potential offenders.</Description>

```
<Evidence>REF_001,REF_002</Evidence>
</Relational>
<Functional>
  <Description>The co-location of transit access (metro station) and
  alcohol-serving venues (bars and pubs) at (id:553) combines high
  anonymity with distracted pedestrians, a configuration associated with
  opportunistic property crime such as pickpocketing and mobile phone
  theft.</Description>
  <Evidence>REF_001,REF_002</Evidence>
</Functional>
<Temporal>
  <Description>The concentration of people during these days likely contributed to the spike in
  crime incidents.</Description>
  <Evidence>REF_001</Evidence>
</Temporal>
</Reasoning>
</Hypothesis>
```

Second option (Level 2):

```
<Hypothesis>
  <Statement>The increase in crime at node (id:553) during the week of February 13 to February 19, 2023, is
  likely due to the proximity of the Carnival period, leading to increased public gatherings and activities
  in the area.</Statement>
  <Reasoning>
    <Relational>
      <Description>Node (id:553) is located in the city center, an area known for hosting large public
      events and gatherings, especially during the Carnival season. The presence of nearby amenities such as
      bars, pubs, and metro stations further contributes to high foot traffic, creating opportunities for
      increased criminal activity.</Description>
      <Evidence>REF_001</Evidence>
    </Relational>
    <Functional>
      <Description>The area around node (id:553) includes amenities classified as high or moderate
      insecurity risks, such as bars, pubs, and metro stations. These venues attract large crowds, especially
      during festive periods like Carnival, which can lead to a rise in opportunistic crimes.</Description>
      <Evidence>REF_001</Evidence>
    </Functional>
    <Temporal>
      <Description>The anomaly occurs in the week leading up to Carnival, a period known for increased
      public activities and gatherings. Historical data indicates that crime rates often rise during such festive
      periods due to the influx of people and heightened activities.</Description>
      <Evidence>REF_001</Evidence>
    </Temporal>
  </Reasoning>
</Hypothesis>
```

Third option (Level 3):

```
<Hypothesis>
  <Statement>The increase in crime at (id:553) during the week of February 13 to February 19, 2023, is
  likely due to the proximity of the Carnival festivities, which began on February 20, 2023, leading to
  increased public gatherings and activities in the area.</Statement>
  <Reasoning>
    <Relational>
      <Description>(id:553) is located near several amenities associated with higher insecurity risks,
      including a metro station (13.63m away), banks and ATMs (17.98m away), and bars and pubs (20.86m away). The
      presence of these amenities can attract larger crowds, especially during festive periods, potentially
      leading to increased criminal activities.</Description>
      <Evidence>REF_001</Evidence>
    </Relational>
    <Functional>
      <Description>The area around (id:553) serves as a hub for public transportation and social
      activities, making it a focal point for both residents and visitors. The influx of people during the
      pre-Carnival period can strain local security measures, resulting in a temporary spike in crime
      rates.</Description>
      <Evidence>REF_002</Evidence>
    </Functional>
```

```

    <Temporal>
      <Description>The anomaly week directly precedes the official start of Carnival on February 20,
      2023. This period is typically marked by increased public events and gatherings, which can lead to higher
      crime rates due to the larger crowds and festive atmosphere.</Description>
      <Evidence>REF_003</Evidence>
    </Temporal>
  </Reasoning>
</Hypothesis>

```

Fourth option (Level 4):

```

<Hypothesis>
  <Statement>The spike in crime incidents at node (id: 553) during the week of February 13 to February 19,
  2023, was influenced by the Sao Paulo Carnival festivities, which began on February 17, 2023, leading to
  increased public gatherings and activities in the area.</Statement>
  <Reasoning>
    <Relational>
      <Description>Node (id: 553) is situated in Praca da Republica, a central location in Sao Paulo
      known for hosting public events and gatherings. The proximity to other nodes such as (id: 549) on Rua Vinte
      e Quatro de Maio and (id: 556) also in Praca da Republica suggests that increased activities in this area
      could have contributed to the rise in crime incidents.</Description>
      <Evidence>REF_001</Evidence>
    </Relational>
    <Functional>
      <Description>The area around node (id: 553) includes amenities such as metro stations, shopping
      centers, banks, bars, and hotels, which are likely to attract large crowds during festive periods like
      Carnival. This influx of people can create opportunities for increased criminal activities.</Description>
      <Evidence>REF_001</Evidence>
    </Functional>
    <Temporal>
      <Description>The anomaly occurred during the week leading up to the official start of Sao Paulo's
      Carnival on February 17, 2023. This period is typically associated with pre-Carnival events and increased
      public activities, which can lead to a rise in crime incidents.</Description>
      <Evidence>REF_001</Evidence>
    </Temporal>
  </Reasoning>
</Hypothesis>

```

C.3 Example LLM Explanation Response

```

<AnomalyExplanation>
  <Details>
    <Spatial>The crime anomaly occurred at Node ID:553 (Praca da Republica, secondary street). The nearby
    nodes are: Node ID:549 (Rua Vinte e Quatro de Maio, residential street), Node ID:556 (Praca da Republica,
    secondary street), and Node ID:558 (Avenida Ipiranga, secondary street).</Spatial>
    <Temporal>The anomaly occurred during the week of 2023-02-13 to 2023-02-19. This week recorded 25
    crime incidents, a notable increase compared to previous weeks and the weeks following. The week of
    2023-02-20 to 2023-02-26 also had a high number of crime incidents, but this was associated with Carnival
    celebrations, making the previous week the clear anomaly.</Temporal>
    <Patterns>Crime incidents at Node ID:553 (Praca da Republica) spiked significantly during the week of
    2023-02-13 to 2023-02-19 compared to the weeks before and after. Crime incidents at the neighboring nodes,
    Node ID:549, Node ID:556 and Node ID:558, did not show a similar spike during the same period. Node ID:553
    also experienced a crime increase during the week of Sao Paulo's anniversary.</Patterns>
  </Details>
  <Hypothesis>
    <Statement>The anomaly at Node ID:553 (Praca da Republica) during the week of 2023-02-13 to
    2023-02-19 was caused by increased activity associated with pre-Carnival street parades (blocos) and events
    in the city center, specifically near the Praca da Republica. The Ilu Oba de Min bloco, which opened the
    Sao Paulo Carnival, concentrated at Praca da Republica on February 17, 2023 [4, 8].</Statement>
    <Reasoning>
      <Relational>
        <Description>Node ID:553 (Praca da Republica) is spatially related to other streets in the
        city center (Node ID:549, Node ID:556, Node ID:558). However, the anomaly was specific to Node ID:553,
        suggesting a localized cause. The concentration of pre-Carnival bloco Ilu Oba de Min at Praca da Republica
        (id: 553) on February 17 explains why that specific location was affected differently than its neighbors
        [8].</Description>
        <Evidence>REF_001,REF_002</Evidence>
      </Relational>
    </Reasoning>
  </Hypothesis>

```

```

</Relational>
<Functional>
  <Description>Node ID:553 (Praca da Republica) has a high concentration of amenities
  associated with insecurity risk, including a metro station, banks/ATMs, bars/pubs, a nightclub and a hotel.
  These amenities, combined with the increased crowds during pre-Carnival celebrations, likely created an
  environment conducive to increased crime [8].</Description>
  <Evidence>REF_001,REF_003</Evidence>
</Functional>
<Temporal>
  <Description>The timing of the anomaly (2023-02-13 to 2023-02-19) aligns with the period
  leading up to Carnival, when pre-Carnival celebrations and street parades (blocos) typically occur in Sao
  Paulo. The Ilu Oba de Min bloco concentrated at Praca da Republica (id: 553) on February 17, 2023 [8]. The
  anomaly precedes the official Carnival dates and the subsequent spike in crime during the week of Carnival
  itself (2023-02-20 to 2023-02-26), supporting the pre-Carnival event hypothesis [4].</Description>
  <Evidence>REF_002</Evidence>
</Temporal>
</Reasoning>
</Hypothesis>
<Conclusion>The crime anomaly at Node ID:553 (Praca da Republica) during the week of 2023-02-13 to
2023-02-19 was most likely caused by the convergence of factors related to pre-Carnival celebrations,
including increased crowds, the presence of insecurity-related amenities, and the specific concentration of
the Ilu Oba de Min bloco at Praca da Republica [4, 8].
</Conclusion>
<Keywords>Praca da Republica, Sao Paulo, crime, anomaly, pre-Carnival, bloco, Ilu Oba de Min, street
parade, February 2023</Keywords>
<References>
  <Item>
    <Id>REF_001</Id>
    <Source>Ilu Oba de Min abre carnaval paulistano com opera de rua - Agencia Brasil - EBC</Source>

    <URL>https://agenciabrasil.ebc.com.br/geral/noticia/2023-02/ilu-oba-de-min-abre-carnaval-paulistano-com-
    opera-de-rua</URL>
    <Excerpt>O bloco Ilu Oba de Min abre hoje (17) o carnaval de Sao Paulo com um opera de rua. O
    desfile sera a noite nas ruas do centro da capital. "A gente esta preparando o maior cortejo da historia do
    Ilu", diz umas das coordenadoras do bloco, Daiane Pettine sobre o espetaculo que tem como tema Ak1kanju:
    Pensamento e Bravura, de Sueli Carneiro.</Excerpt>
    <Date>2023-02-17</Date>
  </Item>
  <Item>
    <Id>REF_002</Id>
    <Source>Sao Paulo, Sao Paulo, Brasil. 21st Feb, 2023.</Source>

    <URL>https://www.alamy.com/sao-paulo-sao-paulo-brasil-21st-feb-2023-int-pagu-block-during-street-carnival-in-sao-paulo-febr
    <Excerpt>RM 2NHK2MC-Sao Paulo, Sao Paulo, Brasil. 21st Feb, 2023. (INT) Pagu Block during Street
    Carnival in Sao Paulo. February 21, 2023, Sao Paulo, Brazil: The Pagu Block, which has only women as part
    of the drums, paraded on Tuesday (21) during Carnival at Praca da Republica in Sao Paulo.</Excerpt>
    <Date>2023-02-21</Date>
  </Item>
  <Item>
    <Id>REF_003</Id>
    <Source>RISKS DURING THE CARNIVAL | MOVI NEWS</Source>
    <URL>https://movi-news.com/security/risks-during-the-carnival/07/02/2023/</URL>
    <Excerpt>However, this event is seen by criminals as an opportunity. Crowds with distracted
    people during their leisure time are a magnet for opportunistic actions. In addition, at this time of the
    year, the security forces also become overwhelmed, since their personnel is put under even more stress. The
    consequence of such scenario is the sudden increase in all kinds of petty crimes, particularly, those
    against property. Among the most frequent ones, there are street robberies, thefts and mass
    robberies.</Excerpt>
    <Date>2023-02-07</Date>
  </Item>
</References>
</AnomalyExplanation>

```

C.4 Example LLM-as-a-Judge Response

Stage I:

```
=== CASE 1 ===
```

```
Data only path: ../results/gpt-4o-mini-search-preview_562_2023-02-13_str_graph_all_data_only.xml.txt
```

Web validation path: ../results/gpt-4o-mini-search-preview_562_2023-02-13_str_graph_all_last_prompt.txt

Try 1 (data_only FIRST, web_validation SECOND): SECOND -> web_validation
Reasoning: Favors the second set because it provides "cohesive and targeted" evidence. It highlights a specific news report (REF_001) on local crime in Praca da Republica, paired with geographical and political context (elections), making it superior to the first set's more generalized approach.

Try 2 (web_validation FIRST, data_only SECOND): FIRST -> web_validation
Reasoning: Favors the first set, praising a major outlet's report (REF_001) for offering "on-the-ground evidence" of violence at the exact anomaly location. It criticizes the second set for including irrelevant references and lacking a direct link to the specific geographical area.

Try 3 (data_only FIRST, web_validation SECOND): SECOND -> web_validation
Reasoning: Favors the second set, primarily due to the "concrete and localized" nature of REF_001. By anchoring the crime spike to a specific news report about Praca da Republica and including geographical identification (REF_002), it offers a more credible and direct explanation than the first set.

Try 4 (web_validation FIRST, data_only SECOND): FIRST -> web_validation
Reasoning: Favors the first set for its "focused relevance." It notes that REF_001 and REF_002 provide direct local and geographical context, while the second set is significantly weakened by three irrelevant references that fail to connect to the specific time or location of the anomaly.

Try 5 (data_only FIRST, web_validation SECOND): SECOND -> web_validation
Reasoning: Favors the second set, identifying the Record TV report (REF_001) as the "clear winner." This specific source directly corroborates the data for Node ID:553 by detailing local terror and violent crime, providing a much stronger immediate cause than the general political context found in both sets.

Stage II:

=== CASE 1 SUMMARY ===

Try 1 (street_neighbors, single_street_ts, single_street, all): FOURTH -> all
Reasoning: Favors FOURTH for its "multi-pattern explanation." It successfully links Rua do Arouche to a major venue, identifies functional risk factors (bars/hotels), and maintains perfect temporal alignment with Carnival. It dismisses THIRD due to a date mismatch.

Try 2 (street_neighbors, all, single_street_ts, single_street): FIRST -> street_neighbors
Reasoning: Favors FIRST as the most "balanced and well-grounded." It praises the explicit spatial link between node IDs (819 and 805) and the causal functional reasoning. It rejects FOURTH specifically because its cited event (Feb 11) falls outside the actual anomaly window (Feb 20-26).

Try 3 (street_neighbors, all, single_street_ts, single_street): SECOND -> all
Reasoning: Favors SECOND for its "methodological rigor." While FIRST and SECOND are similar, SECOND is preferred for incorporating a broader set of functional amenities (hotels/nightclubs) and keeping its evidence categories distinct, whereas others were deemed generic or temporally inaccurate.

Try 4 (street_neighbors, single_street, all, single_street_ts): THIRD -> all
Reasoning: Favors THIRD for providing the "best-balanced" use of separate evidence for each dimension. It highlights the inclusion of high-risk amenities and clear alignment with Carnival week, while noting that SECOND's specific event citation is temporally invalid.

Try 5 (street_neighbors, single_street_ts, single_street, all): FOURTH -> all
Reasoning: Favors FOURTH for offering the most "complete spatio-temporal explanation." It highlights the dedicated evidence used for each facet (spatial, temporal, and functional) and correctly matches the anomaly dates, whereas THIRD is disqualified for its "critical temporal mismatch."