Houssam Razouk, M.Sc.

Integrating Domain Knowledge for the Analysis of Urban Blight Indicators Using NLP and Causal Data Science

Research Project
for Marshall Plan Scholarship
Doctoral programme in Engineering Sciences
Doctoral programme in Engineering Sciences: Computer Science

submitted to

Graz University of Technology

Supervisors

Ass.Prof. Dipl.-Ing. Dr.techn Roman Kern ,Institute of Interactive Systems and Data Science (TU Graz)

Mag. Dr. Michael Leitner M.A. Director, Graduate Certificate, Crime Mapping Geospatial Intelligence Analytics; Adjunct Professor, Department of Experimental Statistics (LSU)

Institute of Interactive Systems and Data Science

Villach, 1 2023

Abstract

Urban blight is estimated devising a set of indicators. These indicators are commonly divided into two main categories: physical and social. Theories about relationships between the urban blight indicators are often hypothesized. Many of these theories about relationships between the urban blight indicators involve causality. However, the lack of sufficient observations and the immorality of conducting interventional experiments increase the debate about these theories. Thus, domain knowledge is critical. To this end, this research facilitates the integration of domain knowledge in the analysis of urban blight indicators and their interaction. Hence, this paper proposes to model domain knowledge as causal diagrams to improve on the urban blight analysis and to achieve less biased results. In this research, natural language processing methods are leveraged for causal information extraction to automatically process urban blight related literature and documents. Moreover, this research proposes four rules to be used in the causal modeling phase. The results of this research guide future analysis of urban blight indicators by improving the understanding of their interactions.

Keywords: Urban Blight; Natural Language Processing; Causal Data Science; Causal Domain Knowledge.

1 Introduction

Urban blight is defined by Beers et al., 2011, as "the deteriorating property conditions that have deleterious effects on the community in which the property is situated". As such, scholars argued that blight might not be defined by how it looks, but it could be more determined by what it takes to reverse it (Schilling and Pinzón, 2016). To this end, urban blight is commonly estimated by devising a set of indicators. In many cases, these indicators are divided into two main categories: physical and social. Consequently, theories about relationships between the urban blight indicators are often hypothesized. Typically, these theories about relationships between the urban blight indicators involve causality. For example, the Broken Windows Theory introduced by Kelling, Wilson, et al., 1982, suggests that urban blight contributes to increased crime in a neighborhood. As such, "Blighted" areas are often associated with lack of security, increased crime rate and violence. The fear of crime has many negative effects on the economy, society, individuals, and their quality of life (Tandogan and Ilhan, 2016). Thus, those effects motivate residents of blighted areas to reallocate to other neighborhoods or cities (Costa et al., 2021). Therefore, urban blight's effects are seen through property deterioration, decreased property value, lower economic stability and environmental quality of surrounding neighborhoods (Costa et al., 2021). Generally, literature suggests that blight affects community welfare of the population in the affected neighborhoods (Larson et al., 2019).

In many cases, theories about relationships between the urban blight indicators (such as the Broken Windows Theory) are debated in the scientific community. Consequently, these theories are considered controversial. This controversy is mainly attributed to insufficient observations and the immorality of conducting interventional experiments. Keeping in mind that correlation does not imply causation, and the resources and technologies limitations to collect data at a scale, the available data sets are typically small and more likely to be biased. These biases severely affects the quality and generalizability of data analytics methods' results.

At the same time, researchers in the filed of causality demonstrated that data analytics methods benefit from integrating domain knowledge to address the quality and generalizability of these methods' results. As such, integrating domain knowledge showed its effectiveness in correcting for data biases and achieving more robust results (Pearl and Mackenzie, 2018). One great example of the effective use of domain knowledge is shown by Pearl, 2011 addressing the Simpson's paradox. In this example, Pearl, 2011 illustrate that the paradox resolution relies on the fact that causality is governed by its own logic. This logic is derived mainly from domain knowledge. Here, domain knowledge is acquired by the subject matter experts by observations and intervening on the environment under study.

Previous researchers have emphasized the complexity of the relationships between urban blight and its causes, effects, and countermeasures (Jones-Farmer and Hoerl, 2019; Shlay and Whitman, 2006). Therefore, particularly in the case of analyzing urban blight, domain knowledge is distributed among a vast number of subject matter experts from various disciplines. For example, Ristea et al., 2021 collected detailed information on instances of physical urban blight and people's perception of safety using a combination of methods. These methods include spatial video and geo narratives in the field with study participants similar to Mills et al., 2008, and the extraction of moments of stress from biosensing wristbands following the approach presented by Kyriakou et al., 2019.

For readers who are not familiar with the concept of geo narratives, it is worth noting that they were first introduced by Kwan and Ding, 2008. Geo narratives is a story telling method that accounts for geographical information and spatial data. Geo narrative might be used to explain the history and cultural significance of a place, to document changes in an area over time, or to provide context for current events or issues related to a particular location. In the case study presented by Ristea et al., 2021, the geo narrative method is employed to investigate themes or patterns that are relevant to urban blight, crime rate, and safety (Ristea et al., 2021).

As urban blight can be perceived as a tangible problem (Pinto et al., 2022), it is crucial to gather public feedback on indicators of urban blight. Ristea et al., 2021 also support these claims, as the study participants selected include not only specialized personnel, but also individuals from the local community who have had prolonged exposure to the study area and are not specifically specialized in urban blight related topics. Each study participant brings their own unique perspective and knowledge of urban blight based on their previous experiences.

In many cases, researchers use literature reviews to gain an understanding of the existing research in a specific field. This approach involves manually extracting and utilizing the domain knowledge of subject matter experts as recorded in scientific publications. For example, Schilling and Pinzón, 2016 examine urban blight's multiple dimensions by conducting an extensive literature review. However, the increasing number of publications in the field of urban blight's indicators analysis makes this approach increasingly challenging. As a prime example, a search on Google Scholar using the terms ("urban blight", "urban blight fear of crime", "urban blight and crime", "causes of urban blight", "effects of urban blight") yields more than 600 publications about urban blight. This illustrates the difficulty of manually extracting and utilizing the domain knowledge of urban blight's subject matter experts. Other methods, such as cognitive map based methods, are frequently used as an alternative approach to gather domain knowledge from a team of subject matter experts by collaborating and discussing a specific topic. Pinto et al., 2022 are among other researchers who have used cognitive map based methods to gather domain knowledge in case of urban blight. However, the authors recognized the limitations of the proposed methods. One of the major limitations is that choosing a different team of experts leads to vastly different results.

Geo narratives, literature reviews, and cognitive maps based methods represent an effective way to gather and organize domain knowledge. Namely, these methods allows for a more comprehensive understanding of the topic being studied (i.e. in the case of this research urban blight). However, these methods have three key limitations. First, geo narratives and cognitive maps based methods are time-consuming and can be influenced by biases introduced by the moderator. Additionally, such methods have limited coverage, particularly when researchers only have access to a limited number of participants. Second, the results of the geo narratives and scientific publications are in the form of texts, which then needs to be interpreted manually to extract the relevant knowledge. Finally, the criteria used in modeling domain knowledge with respect to urban blight using cognitive mapping are not clearly defined and might overlook some common pitfalls in modeling causal domain knowledge. Given the limitations in previous methods, the following research questions are raised:

- What are the challenges in the availability and accessibility of domain knowledge in the case of urban blight?
- How can collaboration between subject matter experts and computer scientists be leveraged to address these challenges?
- How can text-based documents, such as scientific publications and geo narratives interviews, be effectively processed to extract domain knowledge?
- How can domain knowledge be effectively integrated into future research on urban blight through causal modeling steps?
- What are the agreements and differences between the information extracted from literature and geo narratives interviews which could proxy the public's perception of urban blight?

The previous limitations greatly hinder the ability to access and utilize domain knowledge pertaining to urban blight. In response, this research aims to address the challenges of availability and accessibility of domain knowledge in the case of urban blight by utilizing natural language processing and fostering collaboration between subject matter experts and computer scientists. The goal is to facilitate the extraction of relevant information from text-based sources such as scientific publications and interviews. Additionally, this research provides guidance for integrating domain knowledge in causal modeling for future studies on urban blight. The contribution of this research can be summarized as follows:

- We propose a new approach for the extraction of domain knowledge in the case of urban blight using natural language processing techniques.
- We establish a close collaboration between subject matter experts and computer scientists to facilitate the processing of relevant text-based documents such as scientific publications and geo narratives interviews.
- We provide guidance for causal modeling steps to effectively integrate domain knowledge in future research on urban blight.
- Our approach addresses the limitations of previous methods, including the time-consuming nature of cognitive mapping and geo narratives methods, biases introduced by the moderator, and limited coverage when researchers only have access to a limited number of participants.
- We aim to increase the availability and accessibility of domain knowledge for the case of urban blight related analysis, thus providing a valuable resource for researchers and policy-makers.

2 Background and related work

In the 20th century, many countries experienced both urbanization and suburbanization. Urbanization is the process of people moving from rural areas to urban areas (Curci and Masera, 2017). Urbanization can bring increased income, improved living standards, and access to better education and healthcare, but it also has numerous negative consequences. For example, crime and violence are mostly common and severe in urban areas (WHO, 2010). In contrast to urbanization, suburbanization is the movement of population from urban city centers to less densely populated areas, such as suburbs and rural towns (Curci and Masera, 2017). As people move from urban city centers to less densely populated areas during suburbanization, there can be a decline in population in the urban cities, which can lead to an increase in the number of vacant properties (Sugrue, 2014). In extreme cases, the decline in population urban area can lead to the abandonment of properties. For instance, in Detroit (USA), nearly 90,000 vacant properties were reported in 2013 (Sugrue, 2014). Consequently, urban areas can become predominantly populated by minority groups, who may face various social and economic challenges such as poverty, unemployment, and lack of access to education and healthcare (De Sousa, 2006). These social and economical issues together with vacant properties and brownfields (i.e. previously developed land that is not currently in use) can contribute to phenomena known as 'urban blight' (De Sousa, 2006).

"Urban blight" is a term that has multiple interpretations and evaluations due to its subjective nature and the variety of disciplines it encompasses. Thus, there is no universally accepted definition for this phenomenon (Pinto et al., 2022). Further, Schilling and Pinzón, 2016 argued that the definition of urban blight should not be based solely on its appearance, but rather on the actions and efforts required to reverse it. Thus, despite the ambiguity of the term, many scholars associate urban blight with high crime rates (Zhang and Song, 2014), vacant residential and commercial properties, reduced housing quality, a decline in local services (Kondo et al., 2018), and overall deterioration of urban areas (Rafiee and Mahesh, 2013). In some cases, urban blight as a widespread phenomenon is considered a potential cause of other drastic consequences such as drug trafficking, prostitution, and violence (Pinto et al., 2022). These cases align with the work of Kelling, Wilson, et al., 1982 who viewed urban blight through the "Broken Windows Theory". Here, urban blight can be identified by the presence of vacant properties and abandoned facilities that reflect the physical, social, and economic conditions of an area (Kelling, Wilson, et al., 1982). Moreover, urban blight can also be indicated by a lack of adequate housing, whereas in the mid-twentieth century the concept of urban blight included the depreciation of property and a decreased demand for services (Maghelal et al., 2014). More recently, urban blight is also characterized by urban poverty, ghettos, slums, high population density (overcrowded areas), high crime rates, and a concentration of minority households in blighted areas (Maghelal et al., 2014).

Additionally, much of the research on urban blight differentiates between physical and social indicators. For instance, the study on blight done by Maghelal et al., 2014 mapped physical indicators in the City of Dallas into seven sub-categories: abandoned, vacant residential, vacant commercial, mortgage foreclosed, tax foreclosed property, tax delinquent, and demolished. Additionally, Maghelal et al., 2014 examined the socio-economic blight indicators and mapped them into following categories: poverty, unemployment, ethnicity, race, renter household, population, and single parent household (Maghelal et al., 2014).

To account for both physical and social indicators, recent research on urban blight generally utilizes mixed method approaches, by combining both qualitative and quantitative factors. This is illustrated by Rafiee and Mahesh, 2013 where the authors leverage triangulation approach to assess the main causes of urban blight in the historical centre of Shiraz city (Rafiee and Mahesh, 2013). Another example is presented by Fabiyi, 2011 where they have analyzed the spatial and temporal pattern of urban blight in metropolitan Ibadan (Nigeria) through the data obtained from satellite images, questionnaire, and indicators identified by experts (Fabiyi, 2011). Furthermore, Hosseini and Hajilou, 2019 evaluated urban blight in Laleh-Zar neighborhood of Tehran by using literature review, statistical methods and questionnaires filled in by random samples of residents in the area. Moreover, De Tuya et al., 2017 aimed to identify data requirements to create value in the context of urban blight by including stakeholders feedback through workshops, meetings and focus groups (De Tuya et al., 2017).

In summary, most studies on urban blight have similar methods, leading to similar limitations. These include a focus on a single neighborhood or location and reliance on the subjective opinions of a small group of experts. Some of these limitations stem from the limited availability of domain knowledge and the difficulty of obtaining it.

2.1 Cognitive maps based approaches for urban blight analysis

To study the causes of urban blight, Pinto et al., 2022 devised multiple criteria decision analysis (MCDA) approach. Here, Pinto et al., 2022 leverage cognitive maps based method to identify, select, categorize and group potential causes of urban blight. Next, in order to perform quantitative analysis, Pinto et al., 2022 devised the decision-making trial and evaluation laboratory (DEMATEL) method. Consequently, several limitations of this study are noted by the authors. These limitations include: (i) if different group of domain experts are involved in the study, different causes of urban blight might have been identified; (ii) the limited number of panel members and the constrained study area could affect the representativeness of the study.

Similarly to Pinto et al., 2022, Ferreira et al., 2022 conducted a study on cause-and-effect relationships of urban blight prevention strategies. Ferreira et al., 2022 research methodology involve cognitive maps and multiple criteria decision analysis (MCDA). Moreover, Ferreira et al., 2022 opted for the MA-DEMATEL method aiming for more informed decision making. For readers who are interested in the MA-DEMATEL method, the method extends DEMATEL by incorporating moving averages (MAs). In Ferreira et al., 2022 study, cognitive maps are generated devising the strategic options development and analysis (SODA) approach. Here, various prevention initiatives are identified and organized into clusters (areas of concern) by the domain experts. The domain experts that were selected as decision-maker panel were urban architects, civil engineers, representatives of homeowner associations, and senior city planners within the Lisbon city council. Although Ferreira et al., 2022 opted for novel MA-DEMATEL approach to eliminate the limitations from previous studies on urban blight, concerns were raised by the domain experts. These concerns include: (i) simplified causal map, (ii) the model-building process is context dependent and includes subjective components, and (iii) limitations related to the comprehensibility and representativeness of the resulted cognitive map. These limitations are similar to the ones provided in a study done by Pinto et al., 2022.

Further, to create a conceptual model of the causes of blight, Lousada et al., 2021 used fuzzy cognitive maps (FCMs) which are graphical representations of causal reasoning. Specifically, to study the nonlinear behavior of complex systems over time, authors included a system dynamics approach (SD). Here, an aggregation of experts' opinions are presented to study the cause-and-effect relationships between urban blight causes. In Lousada et al., 2021 study, the expert panel is composed of 7 members with high level expertise in urban blight. The listed method limitations in Lousada et al., 2021 study include: (i) difficulty of finding experts specialized in urban blight; (ii) time consuming; (iii) differences in opinion among experts, (i.e. subjectivity); (iv) difficulty in consolidation of the results. In addition, limitations similar to the ones mentioned by Pinto et al., 2022 are noted.

In general, cognitive maps based approaches allow for the integration of multiple perspectives and sources of information, which can lead to a more comprehensive understanding of the problem. However, they have several key limitations. One of the main limitations is that they are time-consuming. Additionally, these methods are often vulnerable to biases introduced by the moderator. Furthermore, they have limited coverage, especially when researchers have access to only a limited number of subject matter experts. Finally, the criteria used in modeling domain knowledge regarding urban blight using cognitive mapping may not be clear, and this might cause overlooking of some common pitfalls in causal domain knowledge modeling.

2.2 Introduction to causal data science

Causal data science extends data science by explicitly considering the underlying data generative process, which can be used to improve the module prediction robustness and reliability. As such, causal data science is gaining more interest in parallel to the rapid development of data driven approaches, such as machine learning algorithms, and the era of big data (Guo et al., 2020). Machine learning algorithms are thought of as a grinder where the prediction quality is highly dependent on the quality of the validity of the assumptions made during the data analytic process. Greenland summarized the most common causal data science frameworks in his work (Greenland and Brumback, 2002). Hence, nowadays causal frameworks includes, but is not limited to, graphical causal models (causal diagrams), potential-outcome models, structural-equations models, and sufficient-component causal models. Causal diagrams, potential-outcome models, and structural-equations models are all tools that can be used to analyze cause and effect in a study population. Causal diagrams are useful for providing a clear, easy-to-understand visual representation of the qualitative assumptions behind a causal analysis. Potential-outcome and structural-equations models, on the other hand, can be used to articulate more detailed quantitative estimates about how different units in the population may respond to different factors. Sufficient-component cause models are distinct from these other models in that they depict more complex qualitative assumptions about the specific causal mechanisms by which different factors may be causing different effects within individual units of the population (Greenland and Brumback, 2002). In summary, all these models can be used to understand the relationships between causes and effects, but they provide different types of information and focus on different aspects of the study population.

In the case of this research, the main goal is to improve the modeling of causal domain knowledge related to urban blight. To achieve this goal, this research has chosen to use the framework of causal diagrams. Causal diagrams are sets of random variables and the cause-effect relations between these variables. The variables in the causal diagrams are represented as nodes or vertices, while the cause-effect relations are represented as directed edges or arcs.

Causal diagrams have been shown to be an effective tool in helping scholars understand and address real-world problems. One example of where causal diagrams can be particularly helpful is in scenarios like the Simpson's paradox, where the relationship between different variables may not be immediately obvious. The Simpson's paradox occurs when a trend appears in different groups of data but disappears or reverses when these groups are combined Pearl, 2011. The main conclusion that can be drawn from Pearl's research is that when a stratification variable is found to introduce confounding bias, the results obtained from the stratified data set are more reliable. However, if the stratification variable does not introduce confounding bias, the correct results are found in the aggregated data set. This statement is emphasizing on the need to consider the potential sources of bias when interpreting research results and that the reliability of the results depends on the specific context. It's also important to note that the results obtained from the stratified data set or the aggregated data set do not necessarily mean the correct or the incorrect results, but the more reliable or less reliable results.

Causal diagrams are powerful tool for causal modeling that can be used to improve the robustness, reliability, and quality of predictions made by a model. However, in real-world scenarios, causal modeling using causal diagrams is vulnerable to pitfalls that can easily be overlooked (Suzuki et al., 2020). In the study "Tips for Using Causal Directed Acyclic Graphs in Causal Inference" by Suzuki et al. (2020), the authors provided a list of tips for scholars who are using causal diagrams, specifically causal directed acyclic graphs (DAGs). These tips include:

- Representing all the causal relations in the causal diagram, even if they only occurred once in the population. This ensures that the diagram is as comprehensive as possible, capturing all potential causal relationships.
- The nodes in the causal diagram should represent the causal variable and not its realized value. This ensures that the diagram represents the underlying causal mechanisms, rather than just the observable values of the variables.

In addition to the tips provided by Suzuki et al. (2020), VanderWeele and Robins, 2007, and VanderWeele and Robins, 2009 introduced the concept of sufficient structure causal diagrams as a way to link different causal modeling frameworks. Here, VanderWeele and Robins leveraged the concept of sufficient causes (as an added artificial node to the original causal diagrams) to allow the discovery of conditional independence between the parents (the causes) while conditioning on a certain strata of the outcome (effect) in a binary setting. In summary, VanderWeele and Robins introduced the concept of sufficient structure causal diagrams as a way to link different causal modeling frameworks, and to provide a more detailed understanding of the underlying causal mechanisms and the discovery of new causal relationships.

To ensure the validity of causal models, it is important to consider the principles of transitivity and proportionality in causation as outlined in Neil McDonnell's work McDonnell, 2018. Transitivity refers to the logical principle that if A causes B, and B causes C, then A must also cause C. This principle is often used as a checking criteria to make sure the relationship between variables is consistent. When transitivity is violated, it may indicate that the causal model needs to be revised. As an example, if it is assumed that "pollution causes respiratory disease" (A causes B) and "respiratory disease causes premature death" (B causes C), then it must be true that "pollution causes premature death" (A causes C).

Finally, in the study "Mechanisms, Modularity and Constitutive Explanation" by Kuorikoski, 2012, the author elaborates on the concept of modularity, which is an important assumption when using causal diagrams. Modularity refers to the way in which different parts of a causal diagram can be separated and considered independently. Here, Kuorikoski distinguishes between three main types of modularity in causal diagrams:

1. Causal model modularity in variables indicates the ability of surgically intervening on the variable (assigning a value to the variable) in the causal model (i.e. node in the causal diagrams) without affecting any other variable in the causal model.
2. The variable mechanism modularity indicates the independence between the variable (or at least a rang of the variable) and the causal effect mechanism[1] to ensure the smoothness of the outcome (i.e the causal mechanism is independent from the variable values).
3. Causal model modularity in parameters[2] indicates the ability of intervening on a parameter in the causal model would affect only one causal effect relation in the model.

In conclusion, the modeling of causal domain knowledge is preferably to be conducted in close collaboration between subject matter experts and

[1] the causal effect mechanism indicates the way the treatment affect the outcome

[2] parameters are the variables which are assumed to constant when developing the causal model for a certain context

causality experts. This is because causal models are powerful tools for understanding complex phenomena (e.g. phenomena of urban blight and its indicators), but they are also vulnerable to pitfalls that can easily be overlooked. Subject matter experts bring their domain-specific knowledge and understanding of the problem at hand, while causality experts bring their expertise in causal modeling, and the knowledge on how to avoid pitfalls and make the model robust. This collaboration ensures that the model is as accurate and reliable as possible, and that the results align with the real-world problems.

3 Methods

3.1 Improving the modeling of causal domain knowledge

In this research, some of the causal modeling's pitfalls in the cognitive maps presented by Pinto et al., 2022, Ferreira et al., 2022, and Lousada et al., 2021 are noted. For example, Pinto et al., 2022 applied DEMATAL on the cognitive map without taking the causal modeling perspective into account. Here, Pinto et al., 2022 gathered data through the collective negotiations and discussions with panel members. Next, the authors identify, select, categorize and cluster potential causes of urban blight. However, the clusters (the criteria in the cognitive map) are not independent. Namely, the identified clusters share one or more cause between each other. In many cases, the identified causes are articulated as events, which describe values of the causal variables not the variables themselves. Thus, events describing different or similar values of the same causal variable are listed in different clusters. Consequently, the results of the subsequent analysis (in this case DEMATAL) provide a great starting point, but are lacking from a causal perspective. Due to the fact that this research has limited access to the same team that participated in Pinto et al., 2022 study, this research cannot quantify the difference between the two approaches. Nevertheless, in the rest of this section, this research presents the recommendations for causal modeling by introducing set of four rules. These rules could be used in the modeling phase.

Causal Modeling Rule #1 If an event gives an information about another event, these events belong to the same causal variable

In general, events describe the value of the causal variable. For example, "vacant buildings" and "unoccupied housing (empties)" describe the same value of a causal variable "building occupation". Similarly, "vacant buildings" and "housing overcrowding" describe values of the same causal variable "building occupation". If left unaddressed, this would affect the modularity of the proposed causal model (e.g. knowing that a building is vacant, indicates that the building is not overcrowded). Therefore, it's important to check and discuss these values with domain experts during the causal modeling phase to extract the causal variable. To identify these cases, it's recommended to go through all the identified events and group the ones that are the same (or similar) or mutually exclusive. This rule is highlighted as one of the tips provided by Suzuki et al., 2020 and is in line with the modularity in variables case presented by Kuorikoski, 2012.

Causal Modeling Rule #2 If an event gives an information about more than one independent causal variable, it is the result of their interaction and should be modeled as an artificial node in the causal diagram

In some cases, an event can describe an interaction between two causal variables. For example, "abandoned buildings" are buildings that are unoccupied and in a state of severe disrepair. While "unoccupied building" is a value of the causal variable "building occupation" and "building in a state of severe disrepair" is a value of the causal variable "building conditions". "Abandoned buildings" cannot be modeled as a causal variable because it cannot be intervened on without affecting other causal variables (i.e. in this case "building occupation" and "building conditions"). To handle such cases, the value is represented in the causal diagram as an artificial node, which does not correspond to a causal variable. Additionally, the artificial node should be linked to the causal variables that created it through their interaction. This rule follows VanderWeele and Robins, 2009 approach in modeling the interactions between two or more causal variables.

Causal Modeling Rule #3 The presence of an edge in the causal diagram corresponds to the presence of at least one causal relation in the data set

After grouping the events under their designated variables, these variables are typically connected using directed edges representing the causal relation. The edge direction is from the cause to the effect. The presence of an edge corresponds to the presence of at least one event in the cause variable that acts as the cause for an event in the effect variable. This is also emphasized by Suzuki et al., 2020. However, it's important to note that these edges should be included only for direct effects, not distant effects, to facilitate mediation analysis. Similar recommendation is also given by Razouk and Kern, 2022. To this end, the absence of an edge between two variables means that the two variables are causally independent considering a specific group of other variables.

Causal Modeling Rule #4 If a mediator variable violates the transitivity principles of causal relations this variable should be split into multiple variables

As a checking criterion, Rule #4 verifies the validity of the identified causal variables. Specifically, Rule #4 retrieves all the causal chains that are of length 3 variables. Next, the rule checks the validity of the transitive causal relation between the variables in the retrieved chains. Any violation of transitive causal relation validity could indicate an issue in the definition of the mediator causal variables, and these variables need to be revised and subsequently modified. This rule is in line with Neil McDonnell work in "Transitivity and proportionality in causation" McDonnell, 2018.

These four rules can be applied iteratively until the resulting causal models meet all the rules criteria. By applying these rules, the resulting cognitive map is more aligned with causal knowledge modeling. As such, the subsequent analysis results could be more reliable and less biased.

3.2 Improving the coverage of urban blight related analysis

This research aims to increase the availability and accessibility of domain knowledge for the case of urban blight related analysis. Thus, causal information extraction methods are proposed to be applied on urban blight relevant text. Causal information extraction, also known as causality extraction, is the process of identifying and extracting information about cause-and-effect relationships from text (Yang et al., 2022). This task is often performed using natural language processing (NLP) techniques and can be used to extract useful information from large amounts of text data (Yang et al., 2022). There are many applications for causal information extraction. These applications may include improving the quality of machine learning algorithms' results by providing a better understanding of the relationships between events in a given domain. One recent survey that provides an overview of the available methods and data sets for causal information extraction is presented by Liu et al., 2020.

There are several approaches to extract causal information from text, including manual annotation, rule-based methods, and machine learning based methods. Manual annotation involves manually identifying and labeling cause-and-effect relationships in a text, while rule-based methods use predefined rules to identify these relationships. Machine learning methods, on the other hand, use statistical techniques to learn patterns in the data and make predictions based on those patterns.

In this research, series of geo narratives conducted in Baton Rouge, Louisiana are manually analysed. Although, these series of geo narratives are preliminary analysed in previous research (Ristea et al., 2021). Nevertheless, this research further process them to extract causal information relevant to urban blight analysis.

Next, a data set of urban blight relevant publications is collected. These publication include valuable information that would improve the coverage of urban blight related analysis. This data set is collected by crawling publication search engine results based on a set of predefined search terms. The approach captures publication related meta data such as the year of publication and the name of the authors. Additionally, it removes duplicated search engine results the are caused by the overlapping of search results. The results of this approach are collected in a long list of publications relevant to urban blight. This list is manually checked to download the publication files.

Subsequently, different causal information extraction approaches including rule-based approach and transformers based approach have been applied on the collected data set. Next, to test and evaluate these approaches, a test data set have been annotated. The annotation process devises annotation guidelines to improve the interrater agreement.

The reset of this secession is organized as follows: more details of the geo narratives' series processing are presented. Next, the approach of collecting the relevant publications is elaborated. Finally, the different causal information extraction methods are described.

3.2.1 Geo narratives data analysis

In order to assess the understanding of the general population of the phenomenon of urban blight, a series of geo narratives are analysed. For instance, in the case of Baton Rouge, Louisiana, the geo narratives are devised by conducting interviews to examine the participant perceptions of urban blight, crime, and safety at specific locations (Ristea et al., 2021). The experiment presented by Ristea et al., 2021 included 46 test participants. These participants can be categorized into two groups based on their background. Namely, participants with background related to urban blight indicators are categorized as experts. Moreover, participants selected from general population are categorized as non-experts. The participants are driven following a predefined route around the study area. The selected route passes through different zones (i.e. types of land use). These zones include commercial zones and residential zones among others. Figure 3.1 provides information about the selected route, together with the information about the respective zones alongside the route.

At the same time, the participants were asked a number of questions based on a predefined set of topics depending on the surrounding environ-

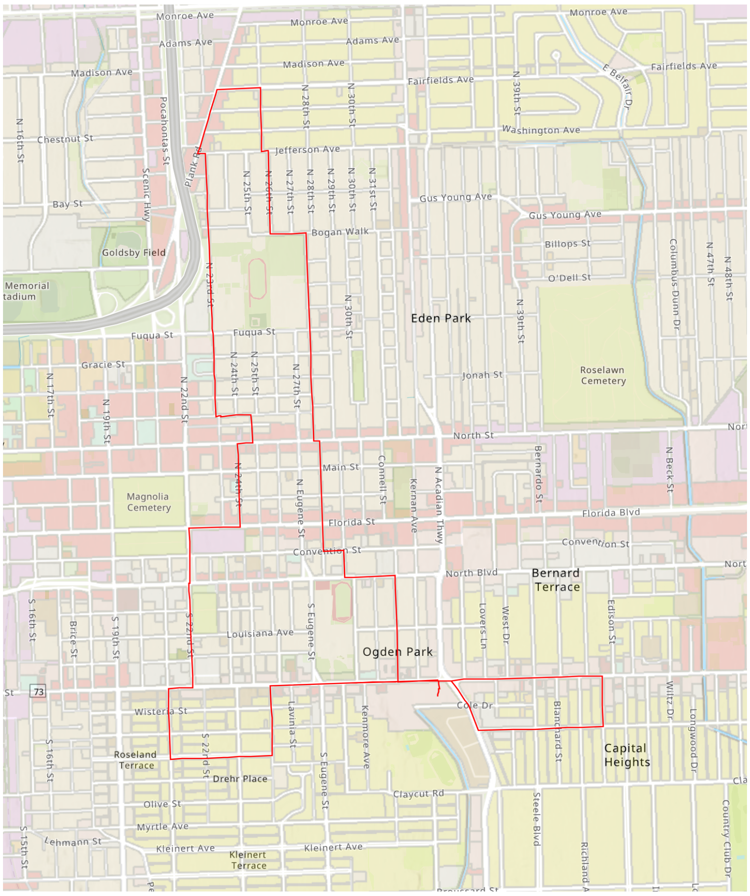

Figure 3.1: **Route taken for the geo narratives**. Each geo narrative was collected based on a predefined route, taking about 25 minutes with the car. The zones are represented as colours, for example in the upper-left corner the route follows Plank Road (diagonal), a commercial area represented as red, which has been associated with crime by the interviewees.

ment. The interviews are recorded together with the location information. The audio recordings are transcribed in textual form and saved together with a time stamp that could be used to retrieve the location where the text is stated in the interview. Since the geo narratives were collected in an interview style, many of the causal expressions span across multiple interactions and in many cases uses implicit language, i.e., inter-sentential causal expressions (Yang et al., 2022). As such, automatic causal relation extraction methods are not expected to perform well in such a setting. Consequently, close reading of the interviews for extracting causal information is conducted in a manual manner. It should be noted that here causal relationships are assumed based on linguistic traits and not on potential underlying philosophical and logical causal relationships. As a consequence, causal expressions reflect the sentiment of the interviewed participants and may also contradict each other.

3.2.2 Urban blight relevant publication data analysis

To increase the coverage of research conducted on urban blight analysis, many scholars opted for an extensive literature review. In many cases, researchers relay on publication search engines that retrieves lists of publications based on search queries. Next, these lists of publications are manually reviewed and summarized. It should be noted, that multiple search queries are typical in cases of multi disciplinary topics such as urban blight. Although the extensive literature review can increase the coverage of the conducted research, this process is labor intensive.

In this research, to simplify the literature review process, an approach that extracts the results of a search engine based on specific queries has been developed. The results from multiple queries are then combined to create a list of relevant publications. Duplicate publications are removed from this list. The publications are then manually downloaded in PDF format and their texts are extracted. Finally, these texts are analyzed using two causal information extraction methods.

The first causal information extraction method is a rule-based method that leverages predefined causal patterns to assess if a giving sentence contains causal language. Additionally, the same causal patterns are used to infer the direction of the causal relation. For example, '[B], because of [A]' and ' because [A], [B]' are two causal patterns related to the causal cue because. These causal patterns describe a causal relation between two events (event [A], and event [B]). These causal patterns are mainly inspired by the work presented by Luo et al., 2016. The direction of the causal relation in both of these patterns is from event [A] (i.e. the cause) to event [B] (i.e. the effects). These patterns are systematically leveraged to extract the sentences which contains causal language. Next, the causal cues and events are extracted from these sentences.

The second causal information method is a data driven based method. This method leverages transformer based language models to extract meaningful representation of the tokens in the sentences extracted from the publication. These embeddings are passed to a trigger detection model (i.e a model that detects the presence of causal language in a sentence). This model's results indicate if a token is part of a causal trigger. Next, the tokens embedding and the results of the trigger detection model are passed to a trigger grouping model. This model addresses cases where a sentence contains multiple triggers. Each trigger could be composed of multiple tokens. To this end, the trigger grouping model classifies the embedding of each pair of tokens that are labeled with trigger class from the trigger detection model. Finally, each grouped trigger embedding (a weighted average of the tokens embeddings that construct the same trigger) are passed together with the tokens embedding to an argument detection model. This model aims to detect the cause and the effect of each causal relations present in the sentence. This method is firstly introduced by Daniel Gerber in his master thesis work Gärber, 2022 [1] [2]. These models are pre-trained on the Because corpus Dunietz et al., 2017 and directly applied on the collected publication data set. Similar to the rule-based methods, these three models are systematically applied to the sentences extracted from urban blight related publications. Hence, if the trigger detection model detects a trigger in a sentence, the trigger grouping model is applied on the same sentence followed by applying the argument detection model.

The two methods are developed to address causal information extraction on text from the general domain. As such, the performance of these methods on urban blight related text is not known. Therefore, in this research, a test data set which is used to evaluate and compare the two methods has been annotated. The annotation process is conducted by three participants from different scientific domains. Based on preliminary analysis of the annotated data, high disagreement between the annotators is noted. Thus annotation guidelines are articulated in the next subsection.

3.2.3 Annotation Guidelines

To estimate and compare the performance of the proposed causal information extraction methods on urban blight related text, 100 sentences were extracted from the scientific publications. Subsequently, these sentences are annotated by three annotators. The annotators are asked to annotate: (i) If the sentences contain causal language? (ii) What is the causal trigger? (iii) What are the causes? (iv) And what are the effects?

[1] An online demonstrator is provided by Daniel Gerber

[2] https://www.danielgaerber.at/cbert/

Here, Fleiss' kappa is devised to estimate the interrater agreement between the three annotators. In the first round, as expected, the resulting interrater agreement had been 16%, which translates to a slight interrater agreement according to Landis and Koch, 1977. The slight interrater agreement is mainly attributed to the limited annotation guidelines and training given to the annotators. As such, in the second round, we follow the annotation guideline suggested by Dunietz et al., 2015. Here, Dunietz defined causal language as a "term refers to clauses or phrases in which one event, state, action, or entity (the cause) is explicitly presented as promoting or hindering another (the effect)." Thus, Dunietz advises the annotators to the language used to appeal notions of causality not the relationships that are causal in some actual metaphysical sense. As such, Dunietz advises to exclude four main types of causal language: (i) Causal relationships with no lexical trigger. These cases are referred to as implicit causal relations. (ii) Connectives that lexicalize the means or the result of the causation. In such cases, the connective and effect are included in the same verb with no clear separation between the two. Fore instance, the verb heal encodes the "cause of recovery", but it includes the effect as well, i.e. the person had recovered. (iii) Temporal language, in such cases, temporal indicators (e.g. "after") can be interpreted as causal relations not due to the language used in the sentences but because of knowledge of the domain. If this knowledge is absent for some annotators, it will create disagreement in the annotation. After all, causal information extraction aims to extract causal information from the text without knowledge about the domain. Finally, (iv) connectives that assert an unspecified causal relationship. Here, Dunietz strives to enrich the extracted causal relation with additional information such as consequence or motivation, etc. However, for the case of this research, the annotation guidelines are only limited to the first three types suggested by Dunietz.

4 Results

4.1 Improving the modeling of causal domain knowledge in cognitive maps based approaches

To test the validity of the proposed causal modeling rules, the cognitive map presented by Pinto et al., 2022 have been digitized. Based on preliminary examination, the clusters proposed by the authors share numerous values. As a result, these clusters overlap. To address this observation, in this research a team of three experts have attempted to reassign the described events in the cognitive maps to their causal variables by applying the first rule form the suggested rules. Here, we report that, in some cases, this process has been successful. For example, the event ″little inspection″ and ″lack of inspection″ are two values of the same variable ″inspection″. However, in some other cases, this process is limited due to the ambiguity of the described event and the limited access to the original experts who participated in Pinto et al., 2022 study.

Next, the second rule is applied. This rule searches for events that give information about more than one causal variable. In this study, the event ″abandoned buildings″ have been identified as a prime example of this case. As such, the event is represented in the causal diagram devising a different color to emphasize that it is an event rather than a causal variable. As mentioned in 3.1, an event in this research refers to the realized value of a variable. The relations between the events, that give information about more than one causal variable, and thier corresponding causal variables are also represented in the causal diagram using a different type of arrows to indicate that these relations are different from the relations that connect causal variables.

To identify the causal relations between the proposed causal variables, the relationships between events within the variables are analyzed. The presence of at least one cause-and-effect pair where the cause event belongs to one variable and the effect belongs to the other variable indicates the presence of a causal relation directed from the first variable to the second one. For example, the event ″lack of urban policies″ from the causal variable ″urban policies″ could lead to the event ″lack of inspections″ from the causal variable ″inspections″. Thus, a causal relation from the causal variable ″urban policies″ to the causal variable ″inspections″ is added in the diagram.

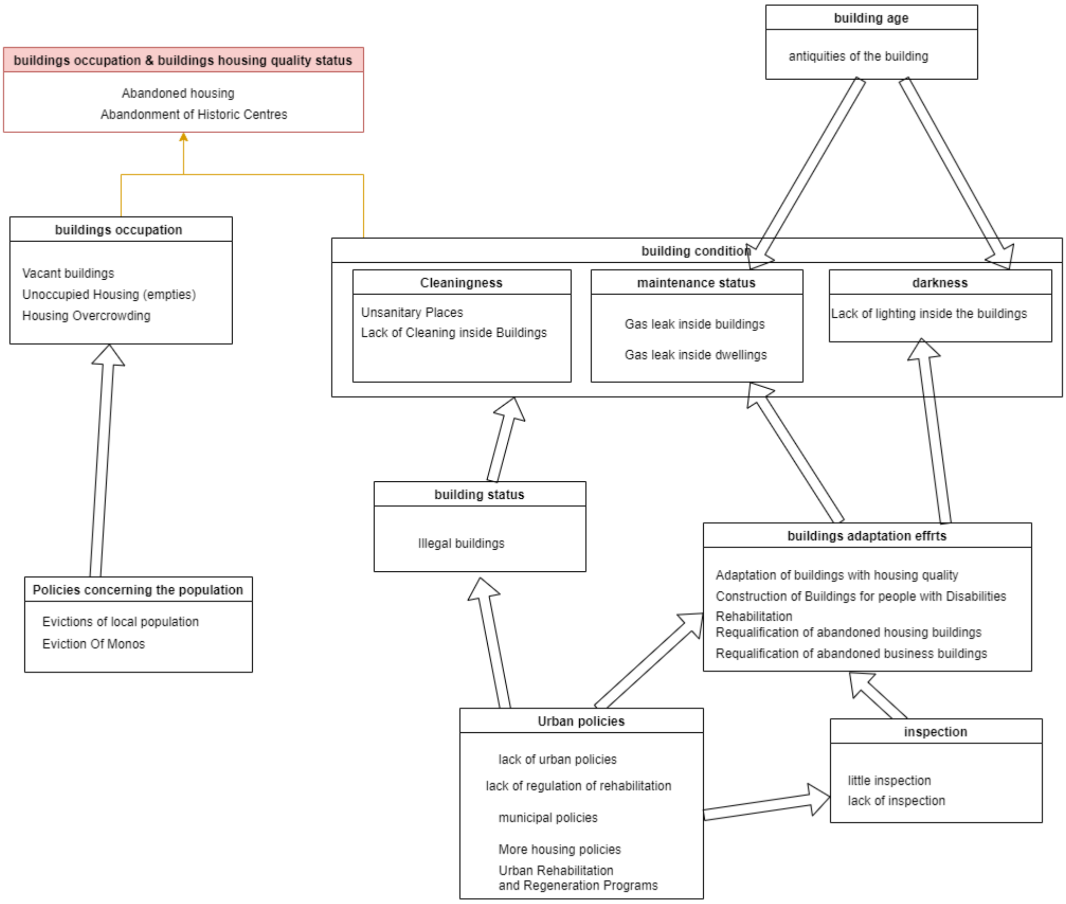

Figure 4.1: **The processing results of part of the cognitive map**. This processing includes applying the four proposed rules to an already existing cognitive map. Limitations concerning the ambiguity of the described events are reported.

Finally, experts are strongly encouraged to use the fourth rule to verify the validity of the developed causal models. A violation of the transitive property of the causal relation could indicate the need to adapt the causal variables' definitions. For example, while "building conditions" is the causal variable with respect to the event of abandoned buildings, it contains three criteria, each of which could be considered as a causal variable on its own. Each criterion could have its own root causes and effects that may not be related to the overall variable "building condition". To represent these cases, these causal variables are depicted independently within the general causal variable (in this case "building condition"). Additionally, the causal relations which link the general causal variable are differentiated from the ones related only to one criterion. In Figure 4.1 the processing results of part of the cognitive map presented by Pinto et al., 2022 is depicted.

4.2 Causal information extraction for increasing the coverage of urban blight related cognitive maps

4.2.1 Geo narratives data analysis

The city of Baton Rouge has a high crime rate, with a higher than Louisiana statewide and US national average number of violent crime and property crime incidence. It is considered a dangerous city to live in and residents should take precautions to protect themselves and their property. This is acknowledged by local leaders who are working to address the problem through various initiatives and programs. Hence, geo narratives are devised by conducting interviews to examine the participant perceptions of urban blight, crime, and safety at specific locations (Ristea et al., 2021). The participants in the geo narratives can be categorized into two groups based on their background. Namely, participants with background related to urban blight indicators are categorized as experts. Otherwise they are categorized as non-experts. This categorisation can allow for examining the agreements and differences of urban blight experts' perception of the surroundings compared to the general population (i.e. non-expert). Based on the analysis of these geo narratives we report the following.

Non-experts

One of the biggest influencing factors why participants feel more safe is familiarity with the study area. Participants, who grew up or worked in an area with high urban blight are able to make a clearer distinction between indicators of poor areas and the impression of higher crime levels.

In Table 4.1 an overview is given of the main influencing factors as mentioned in the interviews with non-experts. Since the causal relations spawn a complex network, the most common relationships were further analysed and grouped according to four categories.

First, the root causes and direct causes are listed, which have been mentioned to result in the intermediary causes, consequences and the perceptions. The intermediary causes list mediators, such as lighting conditions on the street, which may result in people feeling unsafe during the night. Consequences are result of the causes, for example the perception of urban blight has been reported to be an indicator of crime. Further, the news reporting of crime also heavily influence the perception of crime. The two mentioned influencing factors for the perception of crime differed also on level of familiarity. People, who never visited the areas before rather relied on news reports and superficial perception of urban blight, such as graffiti.

Finally, the last category represents the subjective perception of the interviewees, including the perception of urban blight, crime, and their feeling of safety.

Participants also disagreed on the presence of other people on the streets. Some reported that people on the streets in general improve the subjective perception of feeling safe, with exceptions like homeless and groups of people. Others reported that even situations like people close to convenience stores contribute to a feeling of being unsafe. Corner stores and convenience stores are generally assumed to at least indicate higher crime rates. Interestingly, in multiple interviews the presence of certain stores and the absence of others are mentioned as indicator for areas to avoid.

Additionally, some interviewees separate property crime from other types of crime, e.g., violent crime. For example, visible alarm systems are being mentioned as being an indicator for property crime.

In most of the interviews little differences are being made between poorer neighborhoods, urban blight, and crime. For example, overgrown vegetation and untrimmed lawns are generally mentioned as indicators for urban blight, and contribute to the perception of feeling unsafe.

Finally, cars appear to be the main mode of transportation, as in the city of Baton Rouge public transport is not well received. Public transportation is reported to be dark and associated with bad smell, according to the interviews.

Experts

In contrast to the non-experts, the experts made a clear distinction between poor neighborhoods, urban blight, and crime, highlighting the difference in the underlying (social) mechanism. The experts acknowledge the complex interactions between poorness and its implications, e.g., despair and hopelessness to change the situation. For example, rented houses are expected to receive less care than self home ownership. Especially, in cases where the owners are interested in rising prices of their investment. In Figure 4.2 the complete causal diagram extracted from the interviews with experts is shown.

4.2.2 Urban blight relevant publication analysis

In order to streamline the literature review process, this research utilizes advanced search engines to extract relevant results based on a predefined queries. This improves upon traditional literature review methods by providing more targeted and efficient results. Specifically, the following keywords were utilized in our research: (i) "urban blight" (ii) "urban blight fear of crime" (iii) "urban blight and crime" (iv)"causes of urban blight" (v) "effects

Root, direct causes	Intermediary causes	Consequences	Perception
Lack of investment	Bad road conditions; abandoned houses; empty lots	Unmaintained neighborhoods	Urban blight
Poor residents	Lack of care; trash; overgrown vegetation; blocked, foiled, and broken windows	Unmaintained neighborhoods	Urban blight
Small and dense houses; no green area; small spaces;	Foiled windows; abandoned cars and furniture	Unmaintained neighborhoods	Poor area
Lack of community	'No trespassing' signs; unkept neighborhoods	Perception of urban blight; crime news reporting	Crime
Drugs, gangs, lack of money		Bars, fences, and cameras	Crime
Known area (worked or lived)			Feeling safe
Proximity of schools; police presence; cameras	Lit up during night; weekend; residential area; walkability		Feeling safe
People walking the dog; people on the streets			Feeling safe
Abandoned houses; empty lots; dense housing; bad road conditions	Not getting out of the car;		Feeling unsafe
People with guns; walking in the middle of the street; homeless; lurking; man on the street; self-walking dogs; being a girl	drive instead of walk;		Feeling unsafe

Table 4.1: **Most common causal expressions by non-experts**.

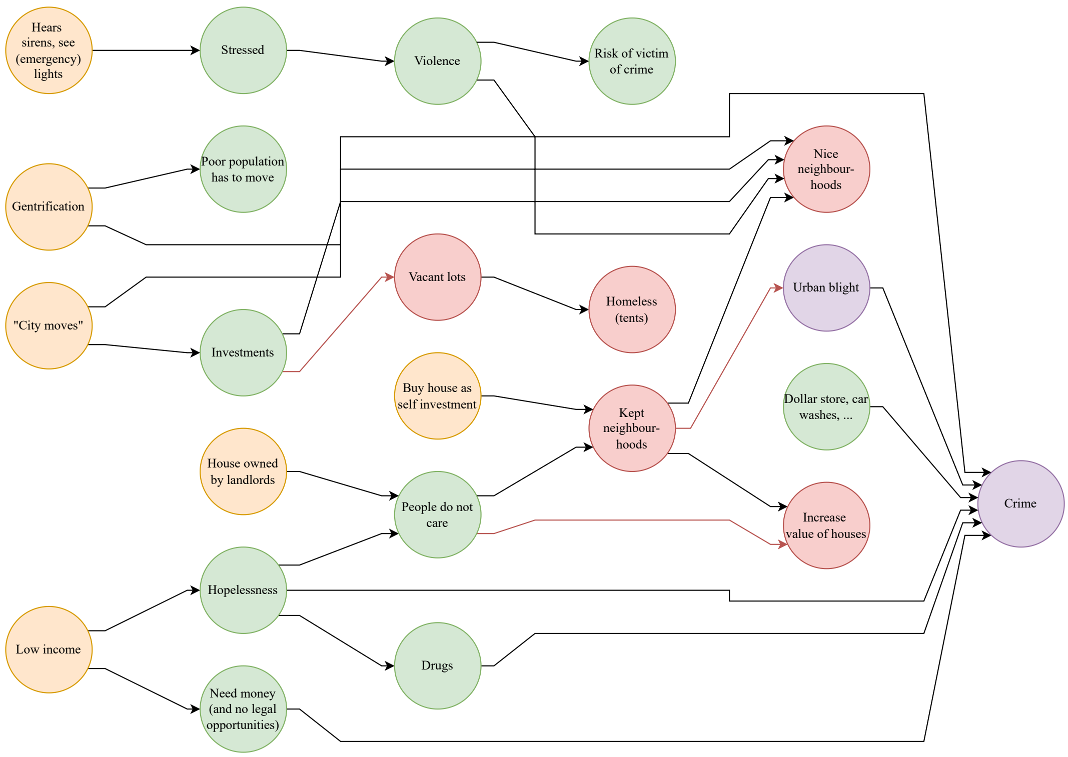

Figure 4.2: **Causal diagram extracted from experts interviews.** .

of urban blight". These keywords were selected to ensure that our data set encompasses a wide range of perspectives and information on the topic of urban blight. In total, a list of over 600 papers related to urban blight is obtained. This list has been manually checked to ensure that it includes only relevant and high-quality papers. In Figure 4.3 an overview of papers' count related to urban blight published each year. Here, an increase trend can be clearly observed. This trend could indicate the increase interest in the topic of urban blight and highlights the urge of the problem. Next, due to the fact these papers are coming from different journals with different access rights, the papers are manually downloaded by visiting each website and downloading them one by one. Moreover, texts are extracted from the downloaded documents. These texts are then processed by a pre-processing function that addresses systematic errors introduced by the text extraction. Finally, sentence tokenization is used to extract sentences from the body of texts. These sentences are used as input for causal information extraction methods.

The three annotators each annotated 100 sentences that contained causal language in the first round. To estimate the agreement between the annotators, Fleiss' Kappa was used. The resulting interrater agreement was 16%, indicating a slight level of agreement between the annotators. In the second round, after the annotation guidelines were shared, an interrater agreement

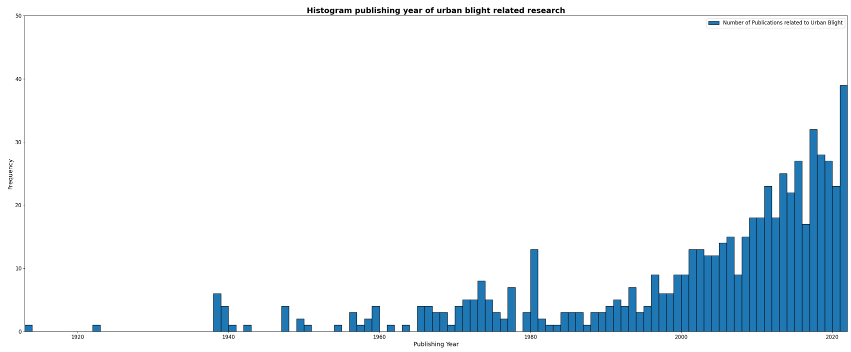

Figure 4.3: **Number of publications related to urban blight in each year.** An increase trend can be clearly observed. This trend could indicate the increase interest in the topic of urban blight.

of 46% was achieved, indicating a moderate level of agreement between the annotators. It should be noted that the two experts with similar background in dealing with causality achieved an interrater agreement of 67%, which is considered substantial level of agreement between the annotators. Based on this, the annotated data was used, and the max voting criteria was applied to evaluate the causal information extraction methods.

Both causal information extraction methods have been evaluated on the annotated data set. Here, the evaluation metrics are the precision, recall, the harmonic mean of precision and recall (i.e.F1 score), and Matthew's correlation coefficient (i.e.MCC). These metrics are calculated to evaluate the models ability to detect causal language in a sentence. By applying these methods to the annotated sentences, the following results were obtained:

Table 4.2: **Comparison between rule-based and BERT-based causal information extraction methods performance on causal language detection**. The rule-based approach delivers higher precision, but its recall is limited. On the other hand, the BERT-based method exhibits higher recall and overall superior performance, as demonstrated by its higher F1 score and Matthew's Correlation Coefficient (MCC).

Method	**Precision**	**Recall**	**F1 score**	**MCC**
Rule-based	**66**%	16%	26%	22%
Bert-based	44%	**68**%	**53**%	**26**%

5 Discussion

Urban blight is a complex issue that requires an interdisciplinary approach to address its various challenges. One of the significant challenges in this domain is the availability and accessibility of domain knowledge. This challenge can be attributed to several factors such as the lack of data collection and standardization, the dispersed and fragmented nature of relevant information, and the difficulty in extracting meaningful insights from the large amounts of data. This is highlighted by the results on the causal modeling of the knowledge documented in the cognitive maps.

To address these challenges, collaboration between subject matter experts and computer scientists is crucial. Subject matter experts have deep knowledge and experience in the domain, while computer scientists have the technical skills and tools to model this knowledge and include it in the processing and analysis of the data. By working together, they can leverage their strengths to extract and integrate domain knowledge more effectively. Here, we provided and example of this collaboration by articulating and testing four rules for causal modeling of domain knowledge.

Text-based documents, such as scientific publications and geo narratives interviews, are an essential source of domain knowledge in the case of urban blight. However, effectively processing these documents to extract meaningful insights is a challenge. Natural language processing (NLP) techniques, such as causal information extraction, can be leveraged to extract relevant information from these documents. To improve its performance, NLP algorithms can be further trained to recognize patterns and relationships in domain specific text by providing adequate labeled data set, which can then be used to extract domain knowledge.

Finally, comparing and contrasting the information extracted from geo narratives interviews can provide valuable insights into the public's perception of urban blight. Experts provide a more objective and scientific perspective on the issue, while geo narratives interviews with people from the general population provide a more subjective and personal perspective. By comparing and contrasting these perspectives, researchers can gain a better understanding of the complex and nuanced nature of urban blight.

As future work, once domain knowledge has been extracted, it can be effectively integrated into future research on urban blight through causal modelling steps. Causal models can help to identify and clarify the relationships between different factors and variables that contribute to urban blight. This information can be used to design and implement more effective policies and interventions aimed at reducing urban blight.

6 Conclusion

In conclusion, the challenges of availability and accessibility of domain knowledge in the case of urban blight are significant issues that requires an interdisciplinary approach to address. Collaboration between subject matter experts and computer scientists is crucial to overcome this challenge, as it leverages the strengths of both disciplines to extract and integrate domain knowledge more effectively. Text-based documents, such as scientific publications and geo narratives interviews, are a valuable source of domain knowledge, and NLP techniques can be leveraged to extract relevant information from these documents. Causal modeling can help to integrate this knowledge into future research on urban blight, and comparing and contrasting the information extracted from literature and geo narratives interviews can provide valuable insights into the public's perception of urban blight and hence help to alleviate this societal problem.

Bibliography

Beers, A., Daley, C., McLaughlin, I., & Pavlek, G. (2011). Quick guide: New tools to address blight and abandonment. *Pennsylvania: The Housing Alliance of Pennsylvania* (cit. on p. 1).

Costa, J. B., Ferreira, F. A., Spahr, R. W., Sunderman, M. A., & Pereira, L. F. (2021). Intervention strategies for urban blight: A participatory approach. *Sustainable Cities and Society*, *70*, 102901 (cit. on p. 1).

Curci, F., & Masera, F. (2017). *Flight from urban blight: Lead poisoning, crime and suburbanization* (tech. rep.). Working Paper. (Cit. on p. 5).

De Sousa, C. (2006). Residential development activity on urban brownfields in milwaukee and chicago: An examination of redevelopment trends, developer perceptions and future prospects (cit. on p. 5).

De Tuya, M., Cook, M., Sutherland, M. K., & Luna-Reyes, L. F. (2017). Information requirements to create public value: Sharing and opening data to address urban blight. *Transforming Government: People, Process and Policy* (cit. on p. 6).

Dunietz, J., Levin, L., & Carbonell, J. G. (2015). Annotating causal language using corpus lexicography of constructions. *Proceedings of The 9th Linguistic Annotation Workshop*, 188–196 (cit. on p. 19).

Dunietz, J., Levin, L., & Carbonell, J. G. (2017). The because corpus 2.0: Annotating causality and overlapping relations. *Proceedings of the 11th Linguistic Annotation Workshop*, 95–104 (cit. on p. 18).

Fabiyi, O. O. (2011). Analysis of urban decay from low resolution satellite remote sensing data: Example from organic city in nigeria. *International Journal of Development and Management review*, *6*(1) (cit. on p. 6).

Ferreira, F. A., Spahr, R. W., Sunderman, M. A., Govindan, K., & Meidutė-Kavaliauskienė, I. (2022). Urban blight remediation strategies subject to seasonal constraints. *European Journal of Operational Research*, *296*(1), 277–288 (cit. on pp. 7, 12).

Gärber, D. (2022). *Causal Relationship Extraction from Historical Texts using BERT* (Master's thesis). Graz University of Technology. (Cit. on p. 18).

Greenland, S., & Brumback, B. (2002). An overview of relations among causal modelling methods. *International journal of epidemiology*, *31*(5), 1030–1037 (cit. on p. 8).

Guo, R., Cheng, L., Li, J., Hahn, P. R., & Liu, H. (2020). A survey of learning causality with data. *ACM Computing Surveys*, *53*(4), 1–37. https://doi.org/10.1145/3397269 (cit. on p. 8)

Hosseini, S. H., & Hajilou, M. (2019). Drivers of urban sprawl in urban areas of iran. *Papers in Regional Science*, *98*(2), 1137–1158 (cit. on p. 6).

Jones-Farmer, L. A., & Hoerl, R. (2019). A unified approach. *Quality Progress*, *52*(5), 48–51 (cit. on p. 2).

Kelling, G. L., Wilson, J. Q., et al. (1982). Broken windows. *Atlantic monthly*, *249*(3), 29–38 (cit. on pp. 1, 5, 6).

Kondo, M. C., Morrison, C., Jacoby, S. F., Elliott, L., Poche, A., Theall, K. P., & Branas, C. C. (2018). Blight abatement of vacant land and crime in new orleans. *Public Health Reports*, *133*(6), 650–657 (cit. on p. 5).

Kuorikoski, J. (2012). Mechanisms, modularity and constitutive explanation. *Erkenntnis*, *77*. https://doi.org/10.1007/s10670-012-9389-0 (cit. on pp. 10, 13)

Kwan, M.-P., & Ding, G. (2008). Geo-narrative: Extending geographic information systems for narrative analysis in qualitative and mixed-method research. *The Professional Geographer*, *60*(4), 443–465 (cit. on p. 2).

Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A., Werner, C., Niederseer, D., Liedlgruber, M., Wilhelm, F., Osborne, T., & Pykett, J. (2019). Detecting moments of stress from measurements of wearable physiological sensors. *Sensors*, *19*(17), 3805 (cit. on p. 2).

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174 (cit. on p. 19).

Larson, M., Xu, Y., Ouellet, L., & Klahm, C. F. (2019). Exploring the impact of 9398 demolitions on neighborhood-level crime in detroit, michigan. *Journal of Criminal Justice*, *60*, 57–63. https://doi.org/https://doi.org/10.1016/j.jcrimjus.2018.11.002 (cit. on p. 1)

Liu, K., Chen, Y., Liu, J., Zuo, X., & Zhao, J. (2020). Extracting events and their relations from texts: A survey on recent research progress and challenges. *AI Open*, *1*, 22–39 (cit. on p. 14).

Lousada, A. L., Ferreira, F. A., Meidutė-Kavaliauskienė, I., Spahr, R. W., Sunderman, M. A., & Pereira, L. F. (2021). A sociotechnical approach to causes of urban blight using fuzzy cognitive mapping and system dynamics. *Cities*, *108*, 102963 (cit. on pp. 7, 12).

Luo, Z., Sha, Y., Zhu, K. Q., Hwang, S.-w., & Wang, Z. (2016). Commonsense causal reasoning between short texts. *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning* (cit. on p. 17).

Maghelal, P., Andrew, S., Arlikatti, S., & Jang, H. (2014). Assessing blight and its economic impacts: A case study of dallas, tx. *WIT Transactions on Ecology and the Environment*, *181*, 187–197 (cit. on p. 6).

McDonnell, N. (2018). Transitivity and proportionality in causation. *Synthese*, *195*(3), 1211–1229 (cit. on pp. 10, 14).

Mills, J. W., Curtis, A., Fagan, W., & Core, C. (2008). *The spatial video acquisition system as an approach to capturing damage and recovery data after a disaster: A case study from the super tuesday tornadoes*. Citeseer. (Cit. on p. 2).

Pearl, J. (2011). Simpson's paradox: An anatomy (cit. on pp. 2, 9).

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic books. (Cit. on p. 1).

Pinto, B., Ferreira, F. A., Spahr, R. W., Sunderman, M. A., & Pereira, L. F. (2022). Analyzing causes of urban blight using cognitive mapping and dematel. *Annals of Operations Research*, 1–28 (cit. on pp. 2, 3, 5, 7, 8, 12, 20, 21).

Rafiee, N., & Mahesh, T. (2013). Urban blight in historical centre of shiraz city. *Global Journal of Current Research Vol*, *1*(2), 77–84 (cit. on pp. 5, 6).

Razouk, H., & Kern, R. (2022). Improving the consistency of the failure mode effect analysis (fmea) documents in semiconductor manufacturing. *Applied Sciences*, *12*(4), 1840 (cit. on p. 13).

Ristea, A., Leitner, M., Resch, B., & Stratmann, J. (2021). Applying spatial video geonarratives and physiological measurements to explore perceived safety in baton rouge, louisiana. *International journal of environmental research and public health*, *18*(3), 1284 (cit. on pp. 2, 15, 22).

Schilling, J., & Pinzón, J. (2016). The basics of blight. recent research on its drivers, impacts, and interventions. *VPRN Research & Policy Brief*, (2) (cit. on pp. 1, 2, 5).

Shlay, A. B., & Whitman, G. (2006). Research for democracy: Linking community organizing and research to leverage blight policy. *City & Community*, *5*(2), 153–171 (cit. on p. 2).

Sugrue, T. J. (2014). The origins of the urban crisis. In *The origins of the urban crisis*. Princeton University Press. (Cit. on p. 5).

Suzuki, E., Shinozaki, T., & Yamamoto, E. (2020). Causal diagrams: Pitfalls and tips. *Journal of epidemiology*, JE20190192 (cit. on pp. 9, 13).

Tandogan, O., & Ilhan, B. S. (2016). Fear of crime in public spaces: From the view of women living in cities. *Procedia engineering*, *161*, 2011–2018 (cit. on p. 1).

VanderWeele, T. J., & Robins, J. M. (2007). Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *American journal of epidemiology*, *166*(9), 1096–1104. https://doi.org/10.1093/aje/kwm179 (cit. on p. 9)

VanderWeele, T. J., & Robins, J. M. (2009). Minimal sufficient causation and directed acyclic graphs. *The Annals of Statistics*, *37*(3). https://doi.org/10.1214/08-aos613 (cit. on pp. 9, 13)

WHO. (2010). *Hidden cities: Unmasking and overcoming health inequities in urban settings*. World Health Organization. (Cit. on p. 5).

Yang, J., Han, S. C., & Poon, J. (2022). A survey on extraction of causal relations from natural language text. *Knowledge and Information Systems*, 1–26 (cit. on pp. 14, 17).

Zhang, H., & Song, W. (2014). Addressing issues of spatial spillover effects and non-stationarity in analysis of residential burglary crime. *GeoJournal*, *79*(1), 89–102 (cit. on p. 5).

Appendix

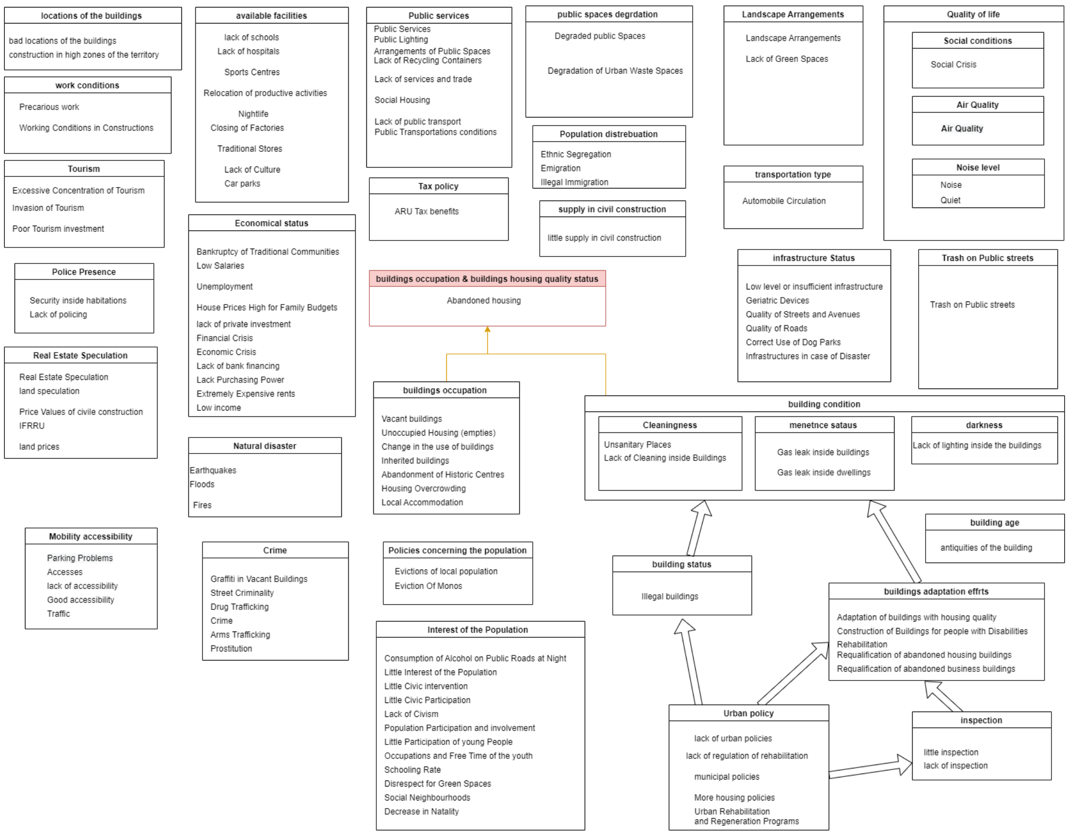

Figure 1: **The processing results of part of the cognitive map**. This processing include applying the fore proposed rules to an already existing cognitive map. Limitations concerning the ambiguity of the descried events are reported.