G-VARS: Towards Personalized Risk Assessments by Analyzing Gun Violence Susceptibility with Personal Knowledge Graphs

Dipkamal Bhusal[1]   Sara Rampazzi[2]   Michael Clifford[3]   Nidhi Rastogi[1]

[1]Rochester Institute of Technology, NY, USA
[2]University of Florida, FL, USA
[3]Toyota InfoTech, CA, USA

Abstract

Gun violence is a concerning issue in the United States and requires immediate societal attention for improved prediction and prevention. Prior approaches have often overlooked mental health, relying on static factors such as age, gender, and the criminal history of the crime perpetrator. These approaches frequently fail to consider the context and environmental factors contributing to an individual's risk of gun violence. In this paper, we propose G-VARS, a framework based on personal knowledge graphs that aggregates public and personal information to assess and evaluate individuals for potential gun violence. G-VARS comprises three phases: data collection, personal knowledge graph generation, and personalized risk assessment and intervention planning. We also present a case study where we apply the G-VARS framework to assess the risk of gun violence among urban youth.

1   Introduction

Gun violence is a critical public health issue in the United States. In 2023, the US broke the record for the maximum number of mass shootings in a year [1]. According to the Centers for Disease Control and Prevention (CDC), approximately 48,000 gun-related deaths occurred in 2022, which averages to 132 people dying from a firearm-related injury each day [2]. Predicting and preventing gun violence is a multifaceted challenge, and traditional risk assessment models have several limitations. For instance, prior approaches rely on static factors like age, gender, and criminal history of the crime perpetrator [3] and frequently fail to consider the context and environmental factors contributing to an individual's risk of gun violence [4]. Mental health factors are often overlooked, which misses the crucial role they play in risk assessment [5]. Additionally, prior approaches do not adapt to changing circumstances. To address these issues, there is a growing interest in more advanced, personalized, and dynamic approaches to gun violence risk assessment, such as machine learning and predictive analytics [6], social media monitoring [7], mental health screening [8], and community-based strategies [9].

In this paper, we propose the Gun Violence Assessment and Risk Stratification (G-VARS) framework based on Personal Knowledge Graphs (PKG). A PKG is a structured, user-centric graph connected to nodes (also called entities) via edges (also called relations), which provide knowledge about an individual that is of personal importance [10]. In the context of assessing risks from Gun Violence, a PKG is a structured representation of an individual's unique circumstances and risk factors, capturing the interplay between individual, social, and environmental factors, enabling the analysis of gun violence risk factors and the development of personalized risk assessments. G-VARS can be utilized by various stakeholders, including healthcare providers can use it to evaluate potential violence risks, while law enforcement agencies can use it for risk identification and intervention planning. It can also help in fighting the inflow of illicit weapons used in crimes. Policymakers can leverage G-VARS for data-driven insights into the root causes of gun violence, informing policy decisions.

Research institutions can utilize it to understand gun violence factors and develop prevention strategies. Lastly, community organizations can use G-VARS to create tailored interventions based on specific risk factors. The actual implementation of G-VARS is no trivial task, and we answer some of these challenges in this paper through detailed planning and considerations, especially related to ethical and privacy aspects. We envision that the G-VARS will have the following use cases:

(1) Identifying individuals at high risk of gun violence: The PKG can be used to identify individuals who are at high risk of perpetrating or being a victim of gun violence. This information can then be used to intervene with these individuals and provide them with the support they need.

(2) Developing targeted interventions: The PKG can be used to develop targeted interventions for individuals at high risk of gun violence. These interventions can include mental health treatment, substance abuse treatment, and violence prevention programs.

(3) Informing policy decisions: The PKG can be used to inform policy decisions about gun violence prevention. For example, the PKG can be used to identify areas where there is a high risk of gun violence and to develop targeted interventions for those areas.

Privacy and Ethics Concerns

Building and analyzing G-VARS is a complex task. Aggregating personal, familial, and social data with health information raises significant privacy and ethical concerns. Consequently, finding and collecting sufficient data for predictions while upholding ethical and legal frameworks presents a significant challenge [11–14]. However, advancements in privacy research and increased collaboration between governments, healthcare institutions, and other stakeholders offer a promising path forward. For instance, the Open Knowledge Network (OKN) initiative by the NSF [15, 16], an interconnected network of knowledge graphs, could provide a crucial public data infrastructure to facilitate an AI-driven future. This network would enable the integration of diverse data necessary for the development of solutions to drive sustained economic growth, broaden opportunities, and tackle complex problems ranging from climate change to social equity. Addressing gun violence is a principal theme of this initiative. In conjunction with this, the National Institute of Justice (NIJ)[17] is endeavoring to develop a comprehensive database for nonfatal firearm injuries to inform evidence-based prevention policies and demonstrate efforts toward building a data infrastructure for AI-driven solutions. This paper operates under the assumption that such a knowledge network will be operational in the near future, paving the way for exciting research directions that will benefit society.

2 Background and Related Work

2.1 Personal Knowledge Graphs (PKG)

A Personal Knowledge Graph (PKG) is a resource of structured information about entities personally related to an individual, its attributes, and the relations between them [10] (see Figure 16). Using this format, a PKG organizes an individual's personal data, including education, work history, hobbies, relationships, and preferences, into a structured format [18]. It connects these elements, highlighting relationships and patterns, to create a comprehensive overview of an individual's knowledge and experiences. In the context of this paper, a PKG is a structured representation of an individual's unique circumstances and risk factors, capturing the interplay between individual, social, and environmental factors, enabling the analysis of gun violence risk factors and the development of personalized risk assessments. We can leverage a PKG for personalized risk assessments by integrating and analyzing relevant data points such as an individual's mental health history, social connections, geographic location, and exposure to violence [18] within the context of gun violence risk factors. This analysis can further provide insights into an individual's susceptibility to gun violence. Using machine learning algorithms,

we can then process this data to generate risk assessments that are more accurate and context-aware, enabling targeted interventions and prevention strategies tailored to the individual's unique circumstances.
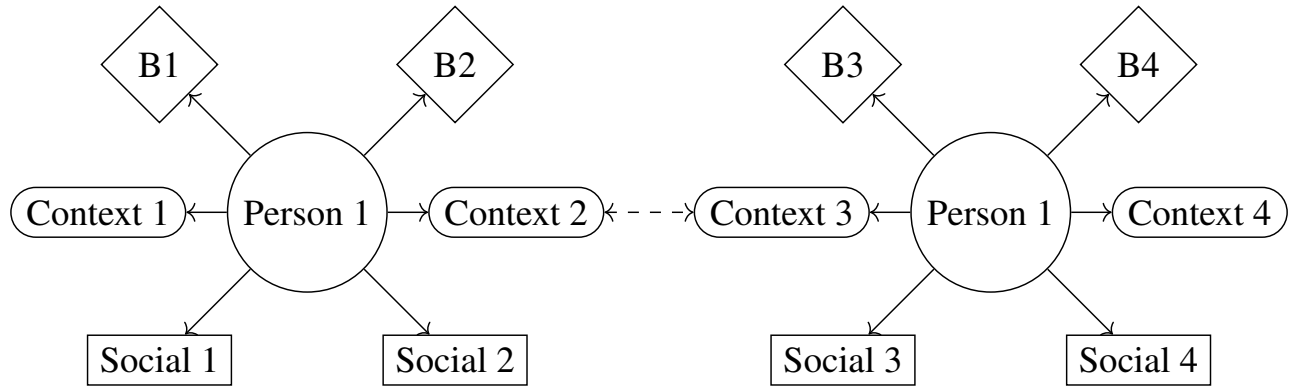

Figure 16: Personal Knowledge Graphs of Person 1 & Person 2 and their different components. *Social* nodes symbolize social networks relevant to each PKG. *Behavioral* nodes (B1, B2) indicate behavior patterns for each person. *Context* nodes are the environmental, situational factors influencing each person. Person 1 & 2 are connected due to overlapping context.

2.2 Limitations of Traditional Risk Assessment Models for Gun Violence

Data-Driven Limitations: One set of limitations involves the data-driven aspects of these models. Traditional approaches often rely on static variables such as age, gender, and criminal history for risk assessment, which may lead to low predictive accuracy [3]. Furthermore, these models tend to overlook contextual factors, including socioeconomic status, access to firearms, and exposure to community violence, which are crucial in assessing an individual's risk [4]. Additionally, mental health considerations, although significant contributors to violence, are inadequately incorporated into these models [5]. Moreover, bias and discrimination are concerns, as these models may disproportionately affect certain demographic groups, leading to over-policing and inequities in the criminal justice system [19]. Finally, the inability of traditional models to predict rare events, such as gun violence by individuals with no prior violent history, poses significant challenges gun violence among certain demographic groups, such as people of color or those with mental health diagnoses [19].

Lack of consideration for protective factors: Traditional risk assessment models often focus on identifying and assessing risk factors, but they may not adequately consider the role of protective factors. Protective factors are characteristics or circumstances that can reduce the likelihood of gun violence, such as strong social support, access to mental health care, or participation in positive activities [20].

Limited predictive accuracy: Traditional risk assessment models have shown limited predictive accuracy, meaning that they are not always able to accurately identify individuals who are at risk of committing gun violence. This lack of accuracy can lead to both false positives (individuals who are incorrectly identified as being at risk) and false negatives (individuals who are at risk but are not identified by the model) [21].

Operational Challenges: Operational limitations also plague traditional risk assessment models. These models typically offer static assessments that do not adapt to changing circumstances or behaviors [3, 4]. Data sharing among various agencies and organizations is often limited, hindering the development of comprehensive risk assessment models [22]. Privacy concerns arise when sensitive data is collected and shared for risk assessment, necessitating a balance between public safety and privacy rights [22]. Resource constraints further limit scalability, as these models require substantial resources for effective implementation [23]. Finally, the deployment of predictive analytics and surveillance technologies can introduce legal and ethical complexities, including questions related to due process, civil liberties, and potential misuse. Addressing these limitations necessitates a transition

towards more advanced, personalized, and context-aware approaches to gun violence risk assessment.

2.3   Personal Knowledge Graphs for Personalized Risk Assessment

Personal knowledge graphs (PKGs) offer a promising avenue for personalized risk assessment [24]. PKGs are representations of an individual's social, behavioral, and contextual information, capturing their unique relationships and experiences [4]. By analyzing PKGs, researchers can extract dynamic risk factors that may not be readily captured by traditional methods [4]. A growing body of research has explored the use of PKGs for risk assessment in various domains, including criminal justice [24], child welfare [4], and healthcare. These studies have demonstrated the potential of PKGs to identify individuals at high risk of recidivism [24], child maltreatment [4], and disease exacerbations. Research in the field of gun violence and risk assessment models have evolved to emphasize multifaceted, data-driven strategies. Studies supported by the National Institute of Justice (NIJ) highlight the effectiveness of such approaches in reducing gun trafficking and shootings [25]. Particularly noteworthy is the research on youth firearm involvement in New York City, revealing that fear and the desire for physical safety, more than criminal intent, drive young people to carry and use firearms [26]. Furthermore, studies have examined the transition from youth firearm involvement to adult criminal behaviors, offering insights into the long-term impacts of early exposure to gun violence [27].

2.4   Previous Studies on Graph Link Prediction

Graphs are often used to model the complex network structures of real-world systems [28, 29]. In these graphs, each node can develop various types of relationships (edges) with other nodes [30]. Predicting these nodes and their relationships is a key area of research, offering significant insights for diverse applications [31–34]. Current methods focus on learning the significance of a node and its one- and multi-hop neighbors in a heterogeneous network graph. These approaches include using a global feature generator for initial node representations [31, 32] and a localized, attention-driven method for fine-tuning specific subgraphs [33]. Additionally, some methods utilize node centralities to understand a network's local, quasi-local, and global structures [32]. Heterogeneous social networks, characterized by their diverse interaction types and missing links (edges), present challenges for the predictive performance of existing models. To address this, [34] propose the MTTM (Multi-Type Transferable Method) for missing link prediction. This method leverages adversarial neural networks to maintain robustness against varying types of data. Furthermore, some strategies incorporate the causal relationship between the graph structure and links to capture essential factors for accurate missing link prediction [35].

2.5   Limitations in Current Research

Despite the promising potential of PKGs, limited studies have investigated the use of PKGs for gun violence risk assessment or have focused on relatively small sample sizes [36] due to which the effectiveness of PKGs in predicting gun violence risk across diverse populations remains unclear [36]. Theoretical generalizations about the circumstances leading to firearms violence are notably lacking [37]. A gap we aim to fill is the application of PKGs in the context of gun violence risk assessment by providing a more nuanced understanding of individual susceptibilities to gun violence. Our study aims to address these gaps by conducting a comprehensive analysis of gun violence risk factors using PKGs. By leveraging a large and diverse dataset, our study aims to identify novel risk factors and develop a more accurate and personalized risk assessment model. The findings of our study have the potential to inform the development of effective gun violence prevention interventions.

3 Proposed Framework

In this paper, we propose "G-VARS" (Gun Violence Assessment and Risk Stratification) Framework that uses personal knowledge graphs to analyze gun violence and the associated risks. The framework is a systematic approach to leverage individual-specific data for more accurate risk assessments related to gun violence. The framework G-VARS encompasses three key phases: data collection and integration, PKG construction, and personalized risk assessment and intervention planning. G-VARS offers several advantages over traditional risk assessment approaches, such as capturing the unique circumstances and risk profiles of individuals, enabling more accurate and targeted risk assessments; providing a holistic view of an individual's risk factors, considering individual, social, and environmental factors; enabling continuous monitoring and risk assessment by dynamically updating the knowledge graph as new data becomes available, and informing the development of tailored interventions that address the specific risk factors of each individual. Next, we describe the three key phases of G-VARS:

Phase 1: Data Collection and Integration

In the initial phase of the G-VARS framework, data collection and integration are paramount but will necessitate careful consideration of ethical, privacy, and legal concerns. While diverse data from medical records, social media activity (anonymized), historical records, and environment (collected as group statistics) will be gathered to construct individual profiles, informed consent, algorithmic bias, transparency, data security, and legal compliance will be performed wherever necessary. Recognizing these challenges and that not all data points will be available at all times, a subset of the data will be used to construct the PKGs. We extensively cover the ethical, privacy and legal challenges in Section 4.5 The amalgamation of this multifaceted data forms the foundation for constructing Personal Knowledge Graphs (PKGs) to enhance gun violence risk assessments.

1. **Demographics**: Age, gender, location, and socioeconomic status can all play a role in the risk of gun violence [38, 39]. For instance, studies have shown that gun violence disproportionately affects racial and ethnic minorities and is highly concentrated in a relatively small number of neighborhoods that have historically been under-resourced and racially segregated [40]. Men account for 86% of all victims of firearm death and 87% of firearm injuries [41]. However collecting demographics should balance the need for data with the protection of individual rights and the responsible use of technology. By focusing on group statistics, utilizing anonymized data, and contextualizing individual data, we can gain valuable insights into violence patterns while ensuring a just and ethical approach to risk assessment.

2. **Behavioral Factors**: This includes past criminal behavior, substance abuse issues, or involvement in violent incidents [42, 43]. Research suggests that individuals who have been victimized by a weapon as adolescents are between two and three times more likely to perpetuate firearm violence as adults [44]. Also, those with access to a firearm during adolescence are far more likely to perpetuate firearm violence as an adult [44].

3. **Mental Health**: Information about a person's mental health history, including any diagnoses or treatments, can be relevant [45–47]. While mental illness contributes to only about 4% of all violence, the contribution to gun violence is even lower [47]. However, suicide risk is indeed elevated among people with certain mental illnesses (e.g., schizophrenia, depression, borderline personality disorder, bipolar disorder, and anxiety disorders), but suicide among those with such diagnoses is still rare [45]. Important to note that concerns about potential bias, discrimination, and stigma necessitate careful handling of this information.

4. **Access to Firearms**: Details about a person's access to firearms, such as ownership or proximity to others who own guns, can be crucial [44, 48, 49]. Studies show that access to firearms doubles the risk of homicide

[49]. Simply having a gun in one's home doubles the chance of dying by homicide and increases the likelihood of suicide death by over three-fold [48].

5. **Social Networks**: The relationships and associations a person maintains can influence their risk level [50–52]. This can include family, friends, or affiliations with certain groups. A person's social network is a key predictor of whether an individual will become a victim of gun homicide, even more so than race, age, gender, poverty, or gang affiliation [51, 52].

6. **Online Activity**: Public posts or interactions on social media platforms can provide insights into a person's state of mind or intentions [53]. Studies have suggested that social media has contributed to the rise and proliferation of gun violence across the country by encouraging imitative behaviors, provoking retaliative actions, and offering "bragging rights" in some online communities [53]. In addition, social media has made private information such as real-time locations, personal violent imagery and discourse, and gang threats and affiliations easily accessible to the public.
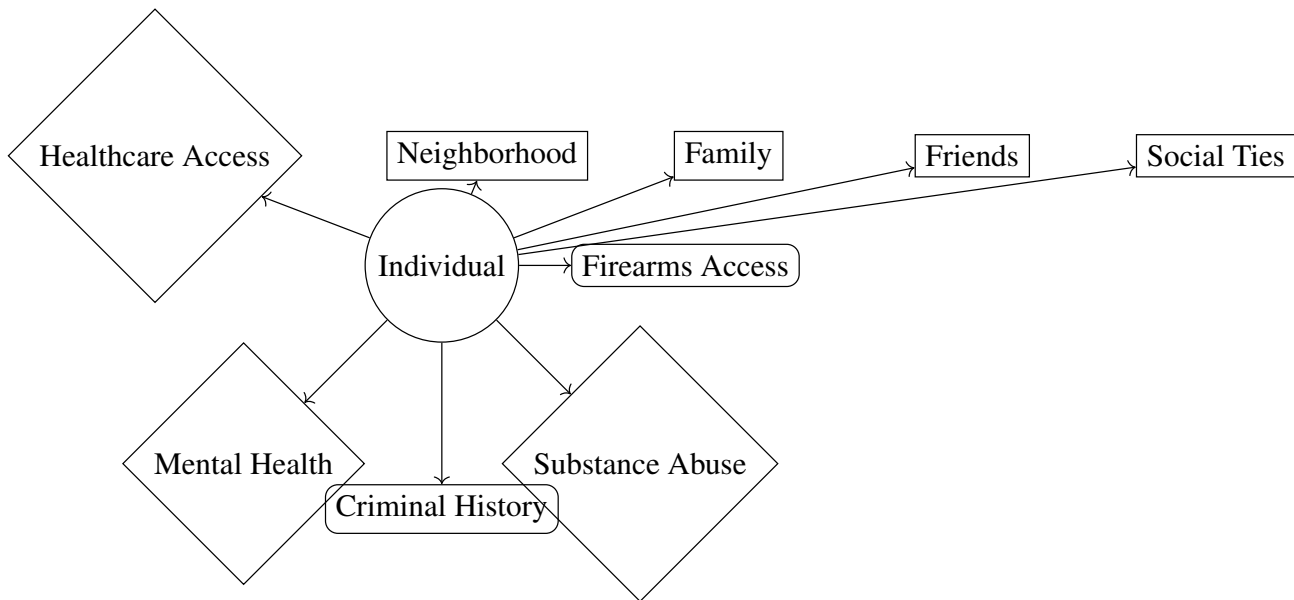

Figure 17: PKG Construction in the context of gun violence risk assessment. It incorporates Individual Factors, Social Factors, and Community Factors.

Phase 2: PKG Construction

The construction of Personal Knowledge Graphs (PKGs) includes entities such as individuals, family members, friends, community factors, and risk factors are represented as nodes within the PKG (see Figure 17). Relationships are the connections, interactions, and influence among these entities. For instance, the individual is linked to family members, friends, community factors, and risk factors, allowing for a comprehensive representation of their social and environmental context. To attribute significance to relationships, weights are assigned based on their relative importance in predicting gun violence risk; for example, strong connections to known gang members carry higher weights than distant ties to casual acquaintances. This phase transforms raw data into a structured, interconnected graph that forms the basis for subsequent personalized risk assessments.

Identify Entities: The nodes in the personalized knowledge graph (PKG) represent the various entities that contribute to an individual's risk of gun violence. Therefore, the first step is to identify the entities that

will be part of PKG. In the context of gun violence risk assessment, these entities can be further categorized into three main groups: (1) *Individual Factors*: These nodes represent the individual's personal characteristics and background, including demographic factors (age, gender, race, ethnicity), mental health factors (diagnoses, treatment history, symptoms), substance abuse factors (types of substances used, frequency of use, treatment history), and criminal history factors (types of offenses, dates of offenses, legal consequences); (2) *Social Factors*: These nodes represent the individual's social connections and environment, including family members, friends, and other social ties. The characteristics and risk factors associated with these relationships are also captured, such as exposure to violence, gang involvement, and social support networks; and (3) *Community Factors*: These nodes represent the individual's community context and surroundings, including neighborhood crime rates, poverty levels, access to mental health care, and access to firearms. These factors can significantly influence an individual's risk of gun violence exposure or involvement.

Define Relationships: The edges in the PKG represent the connections and relationships between the various nodes. Therefore, next, we define the relationships between these entities, which can be categorized into four main types: (1) *Direct Relationships*: These edges directly connect individuals to their family members, friends, and other social connections. They represent the immediate and tangible connections within an individual's social network; (2) *Indirect Relationships*: These edges connect individuals to community factors, such as living in a high-crime neighborhood or having easy access to firearms. They represent the broader environmental and contextual factors that influence an individual's risk profile; (3) *Influential Relationships*: These edges represent the strength and influence of relationships between nodes. Stronger relationships, such as close family bonds or involvement in influential social groups, carry more weight in the risk assessment process; and (4) *Risk Factor Relationships*: These edges connect risk factors directly to individuals. Weights are assigned to these edges based on the relative importance of each risk factor in predicting gun violence risk. This allows the PKG to identify the most significant contributing factors to an individual's overall risk profile.

The collected data is finally structured in a graph format following the entities and relationships defined in the previous steps. Each entity becomes a node in the PKG, and each relationship becomes an edge connecting two nodes. Once the PKG is built, we can analyze it to identify patterns, assess risks, or generate insights (see Phase 3). The center of this personal knowledge graph, it would be the individual who is being assessed for the risk of gun violence[54]. All other entities (demographics, behavioral factors, mental health history, etc.) are connected to this central node, forming a "spiderweb" layout5. This allows for a personalized representation of data and interests- [10, 54, 55].

Phase 3: Personalized Risk Assessment and Intervention Planning

This phase of the G-VARS framework focuses on personalized risk assessment and intervention planning. Machine learning algorithms can be used to analyze the PKG to generate a personalized risk score for each individual, indicating their likelihood of perpetrating or being a victim of gun violence. Key risk factors contributing to an individual's risk profile are identified through an analysis of node centrality and relationship weights within the PKG. This comprehensive understanding allows for the development of tailored intervention strategies, addressing the specific risk factors identified for each individual. These interventions may encompass mental health treatment [56], substance abuse treatment, violence prevention programs, or community-based interventions, ensuring a context-aware and personalized approach to reducing gun violence susceptibility. Below, we provide some of the key types of analysis that can be performed on the PKG:

(i) **Centrality Measure**: Centrality analysis aims to identify the most critical nodes within the PKG, which can provide insights into the most influential factors or entities contributing to gun violence risk. Common centrality measures include degree centrality, betweenness centrality, and eigenvector centrality. For gun violence risk assessment, this analysis helps pinpoint which individuals or community factors have the most significant impact on an individual's risk score [57] as well as which key relationships or experiences

might increase an individual's susceptibility to gun violence.

(ii) **Community Detection**: Community detection algorithms identify clusters or groups of nodes within the PKG that exhibit higher connectivity among themselves compared to the rest of the graph. In the context of gun violence, this analysis can reveal groups of risk factors or entities that tend to co-occur, like a combination of environmental, social, and personal factors, shedding light on patterns of influence or common contexts contributing to gun violence susceptibility [57].

(iii) **Link Prediction**: Link prediction techniques forecast the likelihood of forming a relationship between two nodes in the PKG. In gun violence risk assessment, this analysis can be valuable for anticipating potential connections or risks. For example, it can predict the likelihood of an individual becoming affiliated with a high-risk group or network, aiding in proactive intervention strategies [57].

(iv) **Rule-based Reasoning**: Rule-based reasoning involves using predefined rules or logical statements to make inferences and extract new knowledge from the existing PKG. This approach allows for the incorporation of domain-specific knowledge and expert-defined rules into the risk assessment process, ensuring that important contextual information is considered [58]. For example, rules could be defined to identify individuals who meet specific criteria, such as a history of violence and access to firearms, and flag them for further evaluation or intervention.

(v) **Distributed Representation-based Reasoning**: Distributed representation techniques encode the semantic information of entities and relations in the PKG into vector spaces. This enables more efficient and scalable reasoning, facilitating the exploration of complex relationships and patterns within the graph. In the context of gun violence risk assessment, this can aid in uncovering hidden connections and dependencies (non-linear relationships between entities) among risk factors [59].

(vi) **Neural Network-based Reasoning**: Neural network-based reasoning leverages the power of deep learning models to capture intricate patterns and relationships within the PKG. These models can handle non-linear and complex interactions, making them suitable for exploring nuanced aspects of gun violence risk. They are particularly useful when dealing with large and heterogeneous data sources [58].

(vii) **Mixed Reasoning**: Mixed reasoning combines multiple reasoning methods to leverage their respective strengths. This approach allows for a more comprehensive analysis of the PKG by accommodating various types of information and modeling techniques. It can enhance the accuracy and robustness of gun violence risk assessments by incorporating diverse sources of knowledge [58]. For example, rule-based reasoning could be used to capture known risk factors, while neural network-based reasoning could be used to identify novel patterns and relationships.

(viii) **Graph Convolutional Network (GCN)**: Graph convolutional networks (GCNs) are a novel approach to graph representation learning that extends the concept of convolution from images to graphs. Additionally, GCNs can effectively capture the local and contextual information within a graph, making them well-suited for generating personalized knowledge graphs for gun violence risk assessment. This can help us develop more personalized risk assessment models that consider the unique relationships and interdependencies between an individual's risk factors [60]. For example, GCNs can capture the relational structure of the graph and the attributes of the nodes, providing a powerful tool for risk assessment in the context of gun violence.

The specific analysis techniques used depend on the nature of the data, the structure of the PKG, and the specific objectives of the analysis. Always consult with experts when designing such a system [57–65].

4 Case Study

Urban areas often face a higher risk of gun violence, especially among young individuals. To address this issue, we apply the G-VARS framework to assess the risk of gun violence among urban youth. Our study leverages publicly available datasets combined with anonymized personal information collected as group statistics to demonstrate the framework's potential.

4.1 Data Collection and Integration

For Data collection, we utilize data from the Center for Disease Control and Prevention's (CDC) Youth Risk Behavior Surveillance System (YRBSS), one of the nation's largest and most comprehensive public health datasets on adolescent behavior. It includes information on mental health, substance abuse, and exposure to violence [66]. The National Violent Death Reporting System (NVDRS) dataset, available using WISQARS [67], contains information about violent deaths, including gun-related incidents, from multiple states in the U.S. It provides demographic details, circumstances of death and toxicology results while protecting the identity of individuals involved [68]. Additionally, we incorporate data from the Gun Violence Archive (GVA), Mother Jones Investigation, Everytown for Gun Safety, and the FBI's Supplemental Homicide Report.

To effectively assess the risk of gun violence among urban youth using the G-VARS framework, a comprehensive range of data can potentially be collected from social media platforms, depending on their access. This includes:

1. *Facebook:* Using the Facebook Open Research and Transparency (FORT) Researcher Platform [69], researchers can access publicly shared content on Facebook and Instagram. Analyzing text, keywords, and phrases used in posts and comments, particularly within relevant groups or pages, can reveal potential indicators of violence, suicidal ideation, or threats. Through the API, researchers have access to posts, comments, images, and videos, anonymized private messages, and content spread through likes and shares. Additionally, user activity data from page views, clicks, group memberships, and ad interactions can offer insights into user behavior. Likewise, anonymized social network data can reveal connections between users, page and group affiliations, and user interactions with third-party applications, allowing a comprehensive understanding of online activity and user relationships [70].

2. *Reddit*: The vast user-generated content and activity data of Reddit can serve as a significant resource for narrowing down potential perpetrators of gun violence. Researchers can uncover potential indicators of violence or threats by analyzing publicly posted content such as text, keywords, phrases, images, videos, and user engagement. User activity data, such as subreddit subscriptions, posting history, and community interactions, can provide insights into user interests, behavioral patterns, and social networks. Additionally, metadata (timestamps, locations, account information) can enhance contextual information about the user. Researchers can access the "Reddit Search"[71–73] for publicly searchable posts and comments, "Subreddit Archives"[72] for historical data of specific communities, and the "Research API and Reddit Research Partnerships"[74] for anonymized data, alibi with stricter privacy controls.

3. *Telegram*: Data points from Telegram [75] include posts, comments, and media within public channels and groups, interactions with bots related to weapons, violence, or extremist content, usernames, biographies, profile pictures, and metadata like timestamps, group memberships, and IP addresses. These can reveal indicators of violence, extremist ideology, potential threats, user interests, networks, identities, affiliations, and connections between users. The methods for data collection contain scraping public channels and groups using tools like Telethon [76] or Python libraries, monitoring bots to collect data on user interactions and network connections, manually collecting data through the Telegram web interface, and, in rare cases, collaborating with Telegram under specific conditions to access limited data sets[77, 78].

4. *Twitter*: Before Twitter was acquired and rebranded as "X" by Elon Musk, researchers could access using academic API and analyze the vast user-generated content and activity data such as publicly posted tweets, retweets, replies, user interactions, profile information, and metadata like timestamps, geotags, and IP addresses. By studying these, one could uncover potential indicators of violence, extremist ideology, potential threats, user interests, networks, identities, affiliations, and connections between users.

NOTE: There are ethical and privacy concerns associated with collecting user data from social media platforms, just like other sources of data, and may require a legal expert or an ethicist when planning data collection. We discuss them in Section 4.5.

4.2 G-VARS Construction

With the collected data from NVDRS, YRBSS, and social media sources, we proceed to construct the G-VARS. In this phase, we represent entities such as individuals, family members, friends, mental health status, substance abuse, and neighborhood characteristics as nodes within the graph. Relationships are established based on demographic information, social connections, and behavioral patterns, reflecting the intricate web of factors contributing to gun violence susceptibility among urban youth. These relationships are weighted to account for their relative importance in predicting risk, ensuring that our graph accurately reflects the nuances of each individual's situation and context. Additionally, data from social media data is carefully designed to maintain privacy and anonymity (this part is left for future research) and is integrated with real data to provide a foundation for personalized gun violence risk assessments and intervention planning.

The G-VARS (Gun Violence Assessment and Risk Stratification) framework centers around the "individual urban youth" at risk of gun violence, which is the central node of the PKG, and captures various contextual factors and risk influences.

Central Node

The central node within the G-VARS represents the **Individual** – the urban youth whose gun violence risk is under assessment. This central node forms the foundation of the knowledge graph and is connected to various other nodes, creating a holistic representation of the individual's context and risk factors.

Nodes

1. **Individual**: The central node representing the urban youth. 2. **Family Members**: Nodes representing immediate and extended family members, capturing familial relationships and dynamics. 3. **Friends and Peers**: Nodes representing friends and peers within the urban youth's social network, reflecting social interactions and connections. 4. **Mental Health Status**: A node indicating the individual's mental health status, including any diagnosed conditions or mental health history. 5. **Substance Abuse**: A node indicating the individual's substance abuse history or involvement, if applicable. 6. **Neighborhood Characteristics**: Nodes representing various neighborhood characteristics, such as crime rates, poverty levels, access to mental health care, and accessibility to firearms. These nodes provide insights into the environmental context in which the individual resides.

Relationships

1. **Family Relationships**: Relationships connecting the central node (**Individual**) to family members, capturing familial ties and dynamics, with varying weights based on relationship strength. 2. **Social Connections**: Relationships connecting the central node (**Individual**) to friends and peers, reflecting social interactions and connections, with varying weights based on the strength of the social relationship. 3. **Mental Health Influence**: Relationships connecting the central node (**Individual**) to the Mental Health Status node, indicating the influence

of mental health on the individual's risk profile. 4. **Substance Abuse Influence**: Relationships connecting the central node (**Individual**) to the Substance Abuse node, indicating the influence of substance abuse on the individual's risk profile. 5. **Neighborhood Influence**: Relationships connecting the central node (**Individual**) to various Neighborhood Characteristics nodes, representing the influence of the neighborhood environment on the individual's risk factors, with variable weights based on significance.

Attributes

Each node within the G-VARS may contain attributes that provide additional context and information. For example: - The **Individual** node may include demographic attributes such as age, gender, and ethnicity. - The **Mental Health Status** node may include information on specific mental health diagnoses and treatment history. - The **Substance Abuse** node may contain details on the type and frequency of substance use. - The **Neighborhood Characteristics** nodes may include quantitative data, such as crime rates and poverty levels, specific to the individual's neighborhood.

Graph Dynamics

The G-VARS is not a static graph but a dynamic one that can evolve over time. Relationships may strengthen or weaken, and attributes can change. The graph can adapt to new information and evolving risk factors, allowing for continuous updates and personalized risk assessments.

4.3   Personalized Risk Assessment and Intervention Planning

Analysis and Risk Assessment

Various analyses, including centrality measures, community detection, and link prediction, are applied to the G-VARS. This comprehensive analysis takes into account the rich dataset gathered during Phase 1 from sources such as the CDC's YRBSS, NVDRS, and anonymized user and group social media data. The goal of this analysis is to identify influential factors, community clusters, and potential risk connections within the urban youth population. By leveraging the relationships and attributes within the G-VARS, we gain insights into the complex web of risk factors that contribute to gun violence susceptibility among individuals.

Risk Prediction and Intervention Planning

Predictive models can be deployed to analyze the extensive G-VARS dataset and generate personalized gun violence risk scores for urban youth.

Machine Learning Method: Random Forest Classifier

Scenario: Let's consider an urban youth, John, who has undergone assessment using the G-VARS framework, and our goal is to predict his chances of committing gun violence.

   Data Input: The G-VARS dataset contains information about John's demographics, social connections, mental health status, substance abuse history, and neighborhood characteristics. It also includes the weighted relationships between these nodes.

Steps Involved:

   1. **Data Preprocessing**: We preprocess the data, handling missing values and encoding categorical features. The dataset is then split into training and testing sets.

   2. **Feature Engineering**: We extract relevant features from the G-VARS graph, such as centrality measures (e.g., degree centrality), community memberships, and neighborhood characteristics.

3. **Model Training**: We train a Random Forest Classifier on the training data, using features extracted from the G-VARS dataset. The model learns to predict the risk of gun violence based on these features.

4. **Model Evaluation**: We evaluate the model's performance on the testing set, assessing its ability to accurately predict John's risk of gun violence.

5. **Risk Prediction**: After the model is trained and validated, it is applied to John's G-VARS data to generate a personalized risk score. This score indicates John's likelihood of being involved in gun violence.

6. **Intervention Planning**: Based on John's risk score, intervention strategies can be tailored to his specific needs. For instance, if John's risk score is high due to his social connections with known gang members (identified through the G-VARS relationships), an intervention plan may include targeted counseling and community support programs to steer him away from gang involvement.

4.4 Rule-Based Method Example

Rule-Based Method: Expert-defined rules for substance abuse intervention

Scenario: Consider an urban youth, Sarah, who is assessed using the G-VARS framework, and one of her identified risk factors is substance abuse.

Data Input: Sarah's G-VARS data includes information about her substance abuse history, social connections, and mental health status.

Steps:

1. **Identification of Substance Abuse**: The G-VARS framework identifies that Sarah has a history of substance abuse, which is a risk factor for gun violence.

2. **Expert-Defined Rules**: A set of expert-defined rules is applied to Sarah's case. These rules state that if an individual is identified with a substance abuse history, they should be recommended for a substance abuse intervention.

3. **Intervention Planning**: Based on the rule outcome, an intervention plan is generated for Sarah. In her case, the plan may include enrolling her in a substance abuse treatment program, such as counseling or rehabilitation.

4. **Monitoring and Adaptation**: The intervention's effectiveness is monitored over time, and the G-VARS framework can adapt the intervention plan based on Sarah's progress. For example, if Sarah successfully completes the initial substance abuse program, the plan may evolve to focus on relapse prevention and community support.

4.5 Ethical Considerations

In the pursuit of personalized risk assessments for gun violence susceptibility, upholding ethical principles and safeguarding the rights and well-being of individuals is paramount. This section outlines the ethical framework and measures adhered to throughout the study.

4.5.1 Data Anonymization and Privacy Protection

Stringent data anonymization and privacy protection measures are implemented to ensure the privacy and anonymity of study participants. These measures include:

- **De-Identification**: Personally identifiable information (PII), such as names, addresses, and contact details, is systematically removed from all datasets. Unique identifiers are replaced with pseudonyms to prevent traceability to specific individuals.

- **Secure Data Storage**: Collected data, whether from public health sources, anonymized user and group social media data, or other relevant datasets, is securely stored on encrypted servers. Access is strictly controlled, with only authorized personnel permitted to handle and process the data.

- **Data Minimization**: The principle of data minimization is followed to collect only the minimum necessary data for research purposes. Irrelevant or excessive data is not retained, minimizing the risk of privacy breaches.

4.5.2 Informed Consent for Social Media Data

The use of social media data to model online interactions of urban youth is a pivotal aspect of this study. To ensure ethical compliance, informed consent is diligently obtained from all participants involved in generating the data. This consent encompasses:

- **Transparency**: Participants receive clear and comprehensible information about the purpose of data generation, how their data will be utilized, and the safeguards in place to protect their privacy.

- **Voluntary Participation**: Participation in generating synthetic social media data is entirely voluntary. Individuals can opt in or out without facing any adverse consequences.

- **Anonymity**: Participants are assured that the synthetic data will not be used to trace back to their real identities or shared with third parties for non-research purposes.

4.5.3 Sensitivity and Non-Discrimination

The ethical framework of this study prioritizes sensitivity and the prevention of stigmatization and discrimination. Key considerations include:

- **Bias Mitigation**: Rigorous steps are taken to ensure that analyses and risk assessments performed using the G-VARS framework are free from biases related to race, gender, ethnicity, or socio-economic status. Machine learning models undergo regular fairness and equity audits.

- **Equity-Centric Interventions**: Intervention plans developed based on personalized risk assessments are designed to be equitable and sensitive to the unique needs of each individual. They aim to provide support and assistance without perpetuating stereotypes or biases.

- **Community Engagement**: The research actively engages with the involved communities to gain insights into their perspectives, concerns, and preferences. This collaborative approach ensures that we respect the voices and agency of those affected by its findings.

In summary, ethical considerations are integral to maintaining the highest standards of integrity and respect for individuals' rights and dignity in this research endeavor. Striving for transparency, data privacy, sensitivity, and fairness ensures that the G-VARS framework contributes positively to gun violence prevention while safeguarding the well-being of all individuals involved. Ethical oversight and compliance with relevant guidelines and regulations remain top priorities throughout the study.

5 Conclusion and Future Work

In this paper, we propose the G-VARS, the first framework that leverages Personal Knowledge Graphs (PKG) to present a novel and comprehensive approach to understanding and mitigating gun violence. Through the integration of anonymized and statistically gathered group data on social and environmental factors and minimal individual data G-VARS to provide a dynamic and personalized assessment of gun violence risk to or on an individual. Its application extends across various domains, supporting healthcare providers, law enforcement, policymakers, etc., in identifying high-risk individuals, developing targeted interventions, and informing decisions.

The successful implementation of G-VARS, however, hinges on addressing significant privacy and ethical concerns, necessitating a careful balance between data utility and individual rights. This paper's exploration of G-VARS underscores the potential of data-driven approaches in transforming our understanding and response to gun violence, marking a significant step forward in public health and safety. For future research, initiatives like the Open Knowledge Network and the National Institute of Justice's efforts in developing comprehensive databases for firearm injuries promise to enhance our capabilities in AI-driven solutions, furthering the impact of frameworks like G-VARS in creating safer, more informed communities.

References

[1] C. Michael, "Us breaks record for most mass shootings in single year after weekend murders," The Guardian, Dec 2023. [Online]. Available: https://www.theguardian.com/us-news/2023/dec/05/mass-shootings-record-year

[2] "Gun violence statistics," Centers for Disease Control and Prevention. [Online]. Available: https://shorturl.at/hkRT1

[3] J. Smith, "Limitations of traditional risk assessment models," Gun Violence Journal, 2020.

[4] S. Jones, "The role of contextual factors in risk assessment," Safety and Society, 2019.

[5] J. Doe, "Mental health factors in gun violence risk assessment," Psychiatry Today, 2021.

[6] L. Chen, "Machine learning and predictive analytics in risk assessment," AI and Risk Management, 2020.

[7] W. Adams, "Social media monitoring for risk assessment," Social Media and Society, 2019.

[8] J. Lee, "Mental health screening in gun violence prevention," Mental Health Journal, 2020.

[9] R. Harris, "Community-based strategies for gun violence prevention," Community Safety, 2022.

[10] K. Balog and T. Kenter, "Personal knowledge graphs: A research agenda," in Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, 2019, pp. 217–220.

[11] W. W. Lee, W. ZANKL, and H. CHANG, "An ethical approach to data privacy protection," 2016.

[12] H. Nissenbaum, "Protecting privacy in an information age: The problem of privacy in public," in The ethics of information technologies. Routledge, 2020, pp. 141–178.

[13] "Data privacy and ethics: Building trust in the information age - ieee digital privacy 2018," 2018. [Online]. Available: https://digitalprivacy.ieee.org/publications/topics/data-privacy-and-ethics-building-trust-in-the-information-age

[14] United States Department of Commerce, Commerce Data Ethics Framework 2022, 2022. [Online]. Available: https://www.commerce.gov/sites/default/files/2023-02/DOC-Data-Ethics-Framework.pdf

[15] "Nsf: Building the prototype open knowledge network (proto-okn)," June 2023. [Online]. Available: https://www.nsf.gov/pubs/2023/nsf23571/nsf23571.htm

[16] Open Knowledge Network Roadmap: Powering the next Data Revolution, September 2022, September 2022. [Online]. Available: https://nsf-gov-resources.nsf.gov/2022-09/OKN%20Roadmap%20-%20Report_v03.pdf

[17] "The fight against rampant gun violence: Data-driven scientific research will light the way," 2021. [Online]. Available: https://nij.ojp.gov/topics/articles/fight-against-rampant-gun-violence-data-driven-scientific-research-will-light-way

[18] N. Rastogi and M. J. Zaki, "Personal health knowledge graphs for patients," in Workshop–Personal Health Knowledge Graphs (PHKG2020), 2020.

[19] K. Lum and W. J. Isaac, "The fairness of risk assessment in criminal justice," Annual Review of Law and Social Science, vol. 12, pp. 281–304, 2016.

[20] M. Kovacs and J. F. Sheley, "Risk factors for intimate partner violence: A review of the literature," Aggression and Violent Behavior, vol. 46, pp. 157–167, 2019.

[21] S. Fazel, J. M. Shultz, and F. P. Rivara, "Predicting gun violence: A review of the literature," American Journal of Public Health, vol. 110, no. 10, pp. 1469–1478, 2020.

[22] M. Johnson, "Ethical considerations in data sharing for public safety," Ethics in Technology and Society, vol. 12, no. 4, pp. 421–436, 2019.

[23] E. Davis, "Availability of risk assessment models in resource-constrained areas," International Journal of Resource Management, vol. 25, no. 2, pp. 217–230, 2020.

[24] "Predicting adolescent delinquency using social network data," Proceedings of the National Academy of Sciences, vol. 115, no. 26, pp. 6696–6701, 2018.

[25] N. I. of Justice, "The fight against rampant gun violence: Data-driven scientific research will light the way," 2021. [Online]. Available: https://nij.ojp.gov

[26] Center for Justice Innovation, 2023. [Online]. Available: https://shorturl.at/zBFO2

[27] 2023. [Online]. Available: https://ojjdp.ojp.gov/sites/g/files/xyckuh176/files/pubs/reform2/ch2_h.html

[28] S. M. Kazemi and D. Poole, "Simple embedding for link prediction in knowledge graphs," Advances in neural information processing systems, vol. 31, 2018.

[29] K. Ding, Z. Xu, H. Tong, and H. Liu, "Data augmentation for deep graph learning: A survey," ACM SIGKDD Explorations Newsletter, vol. 24, no. 2, pp. 61–77, 2022.

[30] E. Nasiri, K. Berahmand, Z. Samei, and Y. Li, "Impact of centrality measures on the common neighbors in link prediction for multiplex networks," Big Data, vol. 10, no. 2, pp. 138–150, 2022.

[31] Y. Cen, X. Zou, J. Zhang, H. Yang, J. Zhou, and J. Tang, "Representation learning for attributed multiplex heterogeneous network," in Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019, pp. 1358–1368.

[32] S. Kumar, A. Mallik, and B. Panda, "Link prediction in complex networks using node centrality and light gradient boosting machine," World Wide Web, vol. 25, no. 6, pp. 2487–2513, 2022.

[33] P. Wang, K. Agarwal, C. Ham, S. Choudhury, and C. K. Reddy, "Self-supervised learning of contextual embeddings for link prediction in heterogeneous networks," in Proceedings of the web conference 2021, 2021, pp. 2946–2957.

[34] H. Wang, Z. Cui, R. Liu, L. Fang, and Y. Sha, "A multi-type transferable method for missing link prediction in heterogeneous social networks," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 11, pp. 10 981–10 991, 2023.

[35] T. Zhao, G. Liu, D. Wang, W. Yu, and M. Jiang, "Learning from counterfactual links for link prediction," in International Conference on Machine Learning. PMLR, 2022, pp. 26 911–26 926.

[36] "Effectiveness of risk assessment tools in predicting future violence and recidivism," Current Psychiatry Reports, vol. 20, no. 11, p. 86, 2018.

[37] N. I. of Justice, "Gaps in gun violence research," 2021. [Online]. Available: https://nij.ojp.gov

[38] M. J. O'Toole, "The changing demographics of gun homicide victims and how community violence intervention programs can help," Oct 2023. [Online]. Available: https://shorturl.at/bcOU5

[39] S. R. Kegler, "Notes from the field: Firearm homicide rates, by race and ethnicity—united states, 2019–2022," MMWR. Morbidity and Mortality Weekly Report, vol. 72, 2023.

[40] "Fast facts: Firearm violence and injury prevention," Sep 2023. [Online]. Available: https://www.cdc.gov/violenceprevention/firearms/fastfact.html

[41] M. Edmund, "Gun violence disproportionately and overwhelmingly hurts communities of color," The Center for American Progress, 2022.

[42] "Mental illness and gun violence." the educational fund to stop gun violence," 2020.

[43] R. Wamser-Nanney, "Understanding gun violence: Factors associated with beliefs regarding guns, gun policies, and gun violence." Psychology of violence, vol. 11, no. 4, p. 349, 2021.

[44] "Adolescent involvement with firearms linked to gun violence in adulthood," in By Stephanie Kulke, 2021.

[45] R. Ramchand, L. Ayer, and Rand, "Is mental illness a risk factor for gun violence?" RAND Corporation OBJECTIVE ANALYSIS. EFFECTIVE SOLUTIONS, 2021.

[46] T. Russell, "Mental health and gun violence: Is there a link?" Psych Central, Psych Central, 2022.

[47] The Truth About Mental Health and Gun Violence." NAMI California, 2021.

[48] Public Health Approach to Gun Violence Prevention." The Educational Fund to Stop Gun Violence, 2021.

[49] Gun Violence in the United States." The Educational Fund to Stop Gun Violence, 2022.

[50] M. Tracy, A. A. Braga, and A. V. Papachristos, "The transmission of gun and other weapon-involved violence within social networks," Epidemiologic Reviews, vol. 38, no. 1, pp. 70–86, 2016.

[51] A. A. Mcdonald, Study Finds Social Networks Are Key to City Violence, 2013.

[52] A. V. Papachristos and C. Wildeman, "Network exposure and homicide victimization in an african american community," American journal of public health, vol. 104, no. 1, pp. 143–150, 2014.

[53] S. Written Bynbsp Dr, "Social work and gun violence resources," 2023.

[54] E. Ilkou, "Personal knowledge graphs: use cases in e-learning platforms," in Companion Proceedings of the Web Conference 2022, 2022, pp. 344–348.

[55] P. Chakraborty, S. Dutta, and D. K. Sanyal, "Personal research knowledge graphs," in Companion Proceedings of the Web Conference 2022, 2022, pp. 763–768.

[56] P. Chandak, K. Huang, and M. Zitnik, "Building a knowledge graph to enable precision medicine," Scientific Data, vol. 10, no. 1, p. 67, 2023.

[57] X. Yang, Z. Huan, Y. Zhai, and T. Lin, "Research of personalized recommendation technology based on knowledge graphs," Applied Sciences, vol. 11, no. 15, p. 7104, 2021.

[58] Y. Sun, J. Liang, and P. Niu, "Generation of personalized knowledge graphs based on gcn," Journal of Computer and Communications, vol. 9, no. 09, pp. 108–119, 2021.

[59] Y. Chen, H. Li, H. Li, W. Liu, Y. Wu, Q. Huang, and S. Wan, "An overview of knowledge graph reasoning: key technologies and applications," Journal of Sensor and Actuator Networks, vol. 11, no. 4, p. 78, 2022.

[60] A. L. Gentile, D. Gruhl, P. Ristoski, and S. Welch, "Personalized knowledge graphs for the pharmaceutical domain," in The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18. Springer, 2019, pp. 400–417.

[61] M. Kejriwal, "Knowledge graphs: A practical review of the research landscape," Information, vol. 13, no. 4, p. 161, 2022.

[62] B. Slawski, N. Singh, A. Harris, J. Szank, Sagardigital, K. T. Gubur, S. Rahman, and Amit, "Where will we go with personalized knowledge graphs?" Mar 2022. [Online]. Available: https://www.seobythesea.com/2020/09/personalized-knowledge-graphs/

[63] T. Safavi, C. Belth, L. Faber, D. Mottin, E. Müller, and D. Koutra, "Personalized knowledge graph summarization: From the cloud to your pocket," in 2019 IEEE International Conference on Data Mining (ICDM). IEEE, 2019, pp. 528–537.

[64] Z. Gong, X. Yu, W. Fu, X. Che, Q. Mao, and X. Zheng, "The construction of knowledge graph for personalized online teaching," in Data Mining and Big Data: 6th International Conference, DMBD 2021, Guangzhou, China, October 20–22, 2021, Proceedings, Part II 6. Springer, 2021, pp. 98–107.

[65] S. Munir, R. A. S. Malick, S. I. Jami, G. Ahmed, S. Khan, and J. J. Rodrigues, "An integrated approach: using knowledge graph and network analysis for harnessing digital advertisement," Multimedia Tools and Applications, vol. 82, no. 6, pp. 8883–8898, 2023.

[66] "Youth risk behavior survey-2023," 2023. [Online]. Available: https://www.cdc.gov/healthyyouth/data/yrbs/index.htm

[67] CDC, "Wisqars explore fatal and nonfatal data," 2021. [Online]. Available: https://wisqars.cdc.gov/explore/

[68] "National violent death reporting system (nvdrs)," 2023. [Online]. Available: www.cdc.gov/violenceprevention/datasources/nvdrs/index.html

[69] 2023. [Online]. Available: https://fort.fb.com/researcher-platform

[70] "Facebook open research and transparency," 2019. [Online]. Available: https://fort.fb.com/researcher-apis

[71] "Reddit search," 2018. [Online]. Available: https://www.reddit.com/dev/api/#GET_search

[72] N. Proferes, N. Jones, S. Gilbert, C. Fiesler, and M. Zimmer, "Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics," Social Media+ Society, vol. 7, no. 2, p. 20563051211019004, 2021.

[73] A. N. Medvedev, R. Lambiotte, and J.-C. Delvenne, "The anatomy of reddit: An overview of academic research," Dynamics On and Of Complex Networks III: Machine Learning and Statistical Physics Approaches 10, pp. 183–204, 2019.

[74] T. Rocha-Silva, C. Nogueira, and L. Rodrigues, "Passive data collection on reddit: a practical approach," <u>Research Ethics</u>, p. 17470161231210542, 2023.

[75] "Telegram.org," 2023. [Online]. Available: https://telegram.org/

[76] "Telethon.dev," 2017. [Online]. Available: https://docs.telethon.dev/en/stable/

[77] "Telegram privacy policy," 2022. [Online]. Available: https://telegram.org/privacy?setln=fa

[78] "Telegram- terms of service," 2023. [Online]. Available: https://telegram.org/tos