Cybersecurity Threat Intelligence Augmentation and Embedding Improvement - A Healthcare Usecase

Matthew Sills[‡], Priyanka Ranade[*], Sudip Mittal[‡]
[‡]Dept. of Computer Science, University of North Carolina Wilmington, Email: {ms1491, mittals}@uncw.edu
[*]Dept. of Computer Science, University of Maryland Baltimore County, Email: priyankaranade@umbc.edu

***Abstract*—The implementation of Internet of Things (IoT) devices in medical environments, has introduced a growing list of security vulnerabilities and threats. The lack of an extensible big data resource that captures medical device vulnerabilities limits the use of Artificial Intelligence (AI) based cyber defense systems in capturing, detecting, and preventing known and future attacks. We describe a system that generates a repository of Cyber Threat Intelligence (CTI) about various medical devices and their known vulnerabilities from sources such as manufacturer and ICS-CERT vulnerability alerts. We augment the intelligence repository with data sources such as Wikidata and public medical databases. The combined resources are integrated with threat intelligence in our Cybersecurity Knowledge Graph (CKG) from previous research. The augmented graph embeddings are useful in querying relevant information and can help in various AI assisted cybersecurity tasks. Given the integration of multiple resources, we found the augmented CKG produced higher quality graph representations. The augmented CKG produced a 31% increase in the Mean Average Precision (MAP) value, computed over an information retrieval task.**

***Index Terms*—Artificial Intelligence, Cybersecurity, Knowledge Representation, Knowledge Graphs, Cyber Threat Intelligence**

I. Introduction

The medical industry actively adopts automated systems to assist with health data processing and sharing. These automated systems are considered medical Internet of Things (IoT) devices. The role of IoT in the healthcare sector has been especially useful in creating big data resources for patients, hospitals, and practitioners. Medical data is typically extracted from various sources, transformed into machine-readable formats, and fed into systems that use the transformations for various automated tasks. Automation through IoT Devices allows for convenient, fast, and larger data collection. In a traditional setting, data is typically collected manually or through hard-to-reach integrated systems. Examples like these make it difficult to provide practitioners with the latest information at all times. Through IoT devices, practitioners are able to continuously monitor patients, as well as receive the latest data from exterior sources. In addition, data collected from medical IoT devices can be remotely monitored which can bring many advantages, but also many serious security risks. For example, attackers can gain access to sensitive medical and financial data passing through hospital networks or even disable life-supporting assistive devices.

Medical and financial data is especially valuable to attackers as it sells well on the black market and can be used to commit targeted attacks. The Medjack attack is a well known attack allowing a breach on a secure hospital network, by using a compromised medical device as a backdoor. Once in the network, the attacker can deploy malicious software like ransomware to disrupt the ability of the hospital to function leaving all patients at risk, in order to steal sensitive data [17]. Another famous medical device vulnerability is the SweynTooth vulnerability, which affected bluetooth enabled devices that utilized Bluetooth Low Energy (BLE) for wireless communication. Allowing it to crash, deadlock, and bypass security on the impacted devices [31].

Medical IoT devices are especially prone to these attacks due to the lack of open big data resources specific to healthcare security. Our goal is to create a collection of medical device security vulnerabilities that can be used in knowledge augmentation tasks. Previously various systems have been built on top of cybersecurity specific natural language pipelines to create Cybersecurity Knowledge Graphs [19], [20]. These systems focus on collating scattered Cyber Threat Intelligence (CTI) mined from disparate sources to create a centralized repository of various threats and vulnerabilities [30]. Such analyst augmentation systems aid security operations center (SoC) workflows. We are motivated by the complicated nature of medical device vulnerabilities.

In this paper, we describe our data collection, processing, and augmentation methodologies for medical device vulnerabilities. We first parsed the web to gather information about medical devices with known security vulnerabilities. Using the collected data, we assert it in the Cybersecurity Knowledge Graph (CKG) and generated graph embeddings. Graph embeddings have been used to represent large graphical networks with the aim of improving tasks like: node classification, link prediction, community detection, network similarity and many others. Using our augmentation techniques we improve the quality of graph embeddings created.

The main *contributions* of this paper are -

- Creation of a knowledge graph that stores Cyber Threat Intelligence (CTI) about various medical devices. The CTI was collected using security alerts published by various manufacturers, CISA ICS-CERT, etc. (See Section II).
- We augment the available CTI using knowledge from sources like Wikidata, and FDA's AccessGUDID database (See Section III).

- We show that augmenting the CTI using these other sources improves the quality of graph embeddings generated. We test these different graph embeddings on information retrieval tasks (See Section IV).

The rest of the paper is organized as follows - Section II discusses some related work and background research. We describe our knowledge graph augmentation techniques and processes in Section III. We showcase improvement in the quality of graph embeddings as a result of our knowledge graph augmentation in Section IV. We conclude and discuss possible furture work in Section V.

II. Related Work

In this section we describe some related work in medical knowledge representations and relevant cybersecurity concerns. We also discuss cybersecurity threat intelligence, knowledge graphs and graph representational learning.

A. *Medical Knowledge Representations*

Medical professionals have developed various semantic languages like SNOMED CT [10], ICD-10 [9], and PubChem [24], etc. to communicate important diagnoses, medical procedures, and medications to each other. These languages serve as a back bone communication consensus among millions of physicians, nurses, researchers across various hospitals and countries. An important collection of medical documents that use these semantic languages is the PubMed[1] database maintained by the United States National Institutes of Health (NIH), National Center for BIoTechnology Information.

Xu et al. [35] using the PubMed database crated the PubMed knowledge graph (PKG) using the BioBERT Named Enntity Recognizer (NER). Other work by Wang et al. [33] and Muller et al. [22] further developed knowledge representation techniques for medical procedures, medicines, devices. U.S. Food and Drug Administration maintains a comprehensive database of medical devices [1]. The Global Unique Device Identification Database (GUDID)[2] lists device identification information and other details submitted to the FDA. Wikidata[3] [16] is a curated knowledge graph of the Wikimedia Foundation (WMF) and contains various details about medical device manufactures, which are accessible through it's SPARQL endpoint [32].

We have used some of these medical knowledge representation techniques to augment our Cybersecurity Knowledge Graph (See Section III).

B. *Medical Devices & Cybersecurity*

Medical devices are increasingly connected to the Internet, hospital networks, and other medical devices to provide features that improve health care. These features also increase the risk of cybersecurity threats. United States FDA is responsible for issuing cybersecurity guidance and safety communications. It also conducts multiple activities to inform medical professionals and patients about cybersecurity threats.

The Cybersecurity and Infrastructure Security Agency (CISA)[4] maintains the ICS-CERT Alert[5] which is intended to provide timely notification about critical infrastructure including some medical devices. For example, ICS Medical Advisory (ICSMA-18-123-01)[6] details various cybersecurity threats in the Philips Brilliance Computed Tomography (CT) System. Most medical device Manufacturers also maintain a repository of cyber threat intelligence. These are used to convey various technical details about cyber security threat. These alerts are generally available as HTML pages and need to be converted to raw files. These raw files can then be used as an input to a Natural Language Processing (NLP) Pipeline and output a knowledge graph.

Another popular method of representing various security vulnerabilities of a medical device is to create an attack trees and graphs for the device. Attack trees help security professionals determine how a device might be attacked and can show where stronger protection is needed [34] [15].

C. *Cybersecurity Knowledge Graphs*

A knowledge graph is a set of semantic triples, which are pairs of 'entities' with 'relationships' between them. Knowledge graphs allow for easy identification of related information. Having all of the data interconnected allows querying for related information to be done easily, specially for multiple cybersecurity applications [11], [12], [20], [23]. Knowledge graphs allow the user to find all of the entities that have a Uniform Resource Identifier.

Cybersecurity Knowledge Graphs (CKGs) have long been used to represent Cyber Threat Intelligence (CTI). To represent CTI in a CKG, the first step is to identify what entities and relationships need to be asserted. We also use an ontology called 'Unified Cybersecurity Ontology' (UCO 2.0) [25] to provide our system with cybersecurity domain knowledge. UCO 2.0 is based on Structured Threat Intelligence Language (STIX 2.0) [4] which provides a schema to represent cyber-threat intelligence. CKGs have also been developed from other open-source information by Mittal et al. [19], [20], [26]. In Section III we augment the CKG with external information and use it to generate rich graph embeddings in Section IV.

D. *Graph Representational Learning (GRL)*

Graphs are a powerful mathematical abstractions that can describe complex systems of relations and interactions. There are multiple types of graph representational techniques which are largely application dependent and differ if the application requires a static or a dynamic graph [2], [3]. Techniques usually either focus on individual nodes in the graph or on the entire graph and are similar to convolutional neural networks used in image analysis and computer vision. Popular software

[1]https://pubmed.ncbi.nlm.nih.gov/
[2]https://accessgudid.nlm.nih.gov/
[3]https://www.wikidata.org/wiki/Wikidata:Main_Page
[4]https://us-cert.cisa.gov/about-us
[5]https://us-cert.cisa.gov/ics/alerts
[6]https://us-cert.cisa.gov/ics/advisories/ICSMA-18-123-01

libraries such as node2vec [6], PyTorch Geometric[7] or Deep Graph Library[8] (DGL) are used to train and generate graph representations. In our work we encounter a static graph discussed in Section IV.

In cybersecurity, graph representational learning has been used for malware detection [7], intrusion detection [14], [36], event extraction [8], [13], [29], relationship extraction and threat intelligence [21], [25]–[28]. In this paper, we showcase the use of knowledge augmentation learning approaches to improve cybersecurity graph embeddings. We augment our Cybersecurity Knowledge Graph with other knowledge sources to train more robust embeddings (For more details see Section IV).

III. Cybersecurity Knowledge Graph Augmentation

In this section we will describe our knowledge graph augmentation techniques. Figure 1, showcases our knowledge augmentation architecture. The process starts with the mined Cyber Threat Intelligence, collected from various manufacturers and security bulletins. We use a cybersecurity named entity extractor to extract cybersecurity knowledge and threat intelligence. We augment this knowledge from other data sources. These are then collated and asserted in our Cybersecurity Knowledge Graph (CKG) [21], [25], [26].

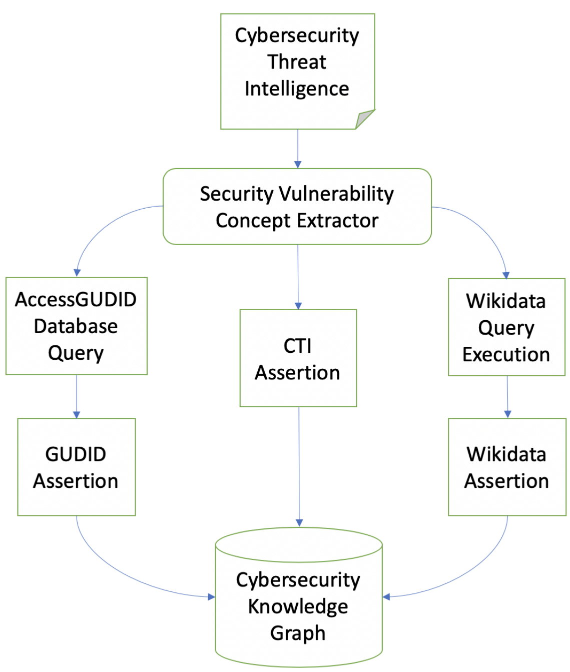

Fig. 1. Knowledge Augmentation Architecture Diagram.

Knowledge graphs use ontologies to describe various domain specific concepts through classes and properties. These properties include relationships between various classes and their attributes. These classes generally have sub-classes, and parent classes. Parent class relations are inherited by its children. Instances are individuals that are a type of a class. These have different data properties and can be associated with other instances by asserting object properties. These attributes are vital so as to differentiate between two different concepts. Knowledge graph augmentation is the process of adding from disparate sources, information to the knowledge graph to increase it's use fullness and adaptability.

For our CKG we used the Unified Cybersecurity Ontology (UCO) [25] to provide cybersecurity domain knowledge. An Intelligence ontology [19] was used to represent threat intelligence. We also create a medical device description ontology based on various medical domain knowledge and FDA's AccessGUDID.

Next we describe various medical data sources that help us with knowledge augmentation. We discuss how medical device security vulnerability data was gathered and augmented with knowledge from manufacturers, ICS-CERT, US FDA, and Wikidata (See Section II).

A. Data Sources

1) Manufacturer Cyber Threat Intelligence (CTI): To gather security vulnerability data for medical devices we built multiple crawlers for different medical device manufacturers. Crawlers were created for Phillips, GE, ICS-CERT, etc. using the python library Beautiful Soup. Once collected CTI was converted from HTML to raw text files. For example we collected the following CTI for a Phillips Ultrasound Machine (dated August 29, 2019)[9]:

```
Philips has become aware that if the Philips
    HDI 4000 Ultrasound system is running on
    outdated, unsupported operating systems,
    such as Windows 2000, an unauthorized
    user may be able to access ultrasound
    images or compromise image integrity.
```

The resulting CTI for each medical device was processed for assertion in the CKG. We extracted issues that consisted of terms related to various vulnerabilities using a *Security Vulnerability Concept Extractor* (SVCE) [19], [26]. The SVCE was able to tag each sentence with the following concepts: Means of an attack, Consequence of an attack, affected software, hardware and operating system, version numbers, network related terms, file names and other technical terms. The extracted concepts were used to generate an RDF [5] stored in the queryable CKG. RDF statements for the CTI about Phillips Ultrasound Machine can be seen in Figure 2.

2) Wikidata Knowledge Graph: To gather additional knowledge about various intelligence components extracted form the available CTI, we retrieve more information about these from the Wikidata Knowledge Graph [16]. Wikidata is an open source knowledge base that is machine processable. It contains all of the structured data for various Wikimedia

[7]https://pytorch-geometric.readthedocs.io/en/latest/
[8]https://www.dgl.ai/

[9]https://www.usa.philips.com/healthcare/about/customer-support/product-security

```
@prefix uco: <http://accl.umbc.edu/ns/ontology/uco#> .
@prefix intel: <http://accl.umbc.edu/ns/ontology/intelligence#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .

<Int24678359436> a intel:Intelligence ;
intel:hasVulnerability <unsupported_operating_systems> .

<Philips_HDI_4000_Ultrasound_system> a uco:Product ;
uco:hasVulnerability <unsupported_operating_systems> .

<unsupported_operating_systems> a uco:Vulnerability ;
uco:affectsProduct <Philips_HDI_4000_Ultrasound_system> ;
```

Fig. 2. RDF for the cyber threat intelligence about the Phillips ultrasound machine.

projects[10]. Wikidata API endpoints[11] allow users to issue complex SPARQL [32] queries. SPARQL is a RDF [5] query language. Wikidata has its own endpoint for SPARQL queries. SPARQL allows for a querying of Wikidata in a (*Subject*, *Predicate*, *Object*) format and returns various results as a JSON object.

To generate a Wikipedia SPARQL query we first select an entity extracted using the SVCE from CTI collected from manufacturer websites. This entity is then placed in either the subject, predicate, or object field of the SPARQL query. The query is then executed on the Wikidata Knowledge Graph. Each entity in Wikidata has its own universal identification number and is connected to other entities through the use of predicates. By specifying a predicate identification number and either a subject or an object tag in a query we can retrieve all of the Wikidata entities that have a specified relation to the imput entity.

3) US FDA AccessGUDID Database: To gather more knowledge about various insecure medical devices we also used the Global Unique Device Identification Database (GUDID). The GUDID contains all of the devices that have Unique Device Identifiers (UDI) that have been submitted to the FDA and is available publicly. To gather the required data we wrote a program to parse through each device in the database and turn it into a device object. The information that we considered useful from the available data was the unique device id, device's manufacturer name, brand name, description of the device, type of device, and purpose of the device. Here is the data mined about GE's ultrasound machine:

```
UID : 00840682146944;
Manufacturer : GE Medical Systems (China)
    Co.,Ltd.;
Brand Name : LOGIQ;
Type : General-purpose ultrasound imaging
    system;
Description : A stationary or mobile (e.g.,
    on wheels) assembly of devices designed
    to collect, display, and analyse
    ultrasound images during a variety of
    extracorporeal and/or intracorporeal
    (endosonography or endoscopic) ultrasound
    imaging procedures (e.g., cardiac,
    OB/GYN, endoscopy, breast, prostate,
    vascular, and intra-surgical imaging). It
    consists of a mains (AC-powered) data
    processing unit with integrated software
    and a monitor. It is typically presented
    as a mobile assembly which may support a
    wide variety of transducers and related
    application software packages; an
    ultrasound transducer(s) may be included.
```

The manufacturer and brand name data were used to collate known Cyber Threat Intelligence (CTI) about various devices. We were able to link CTI from manufacturers website and devices on FDA's AccessGUDID database. Table I, lists the number of known vulnerabilities in popular medical devices.

Next we collate all these data sources and assert them in our Cybersecurity Knowledge Graph.

B. Cybersecurity Knowledge Graph Assertions

After we have mined the knowledge from various data sources discussed above, we assert it in our cybersecurity knowledge graph. We then associated the extracted entities and concepts with Uniform Resource Identifiers (URIs). These URIs are then converted to nodes in our CKG.

Using Wikidata SPARQL queries (See Section III-A2), we fetch the sub-graph for each URI. This helps in including more global knowledge about an entity in our CKG. For example we can use Wikidata to map the URI for "GE Healthcare" to *wiki:Q1152374*[12]. This external knowledge graph help us map our entities to real world conceptual instances.

We also created a light medical device description ontology based on FDA's AccessGUDID database fields (See Section III-A3). We stored the linked data as RDF triples in our CKG.

We collected 5,843 CTI from manufacturers like Phillips, GE, Medtronic, CISA ICS-CERT Alerts, etc. The *Security Vulnerability Concept Extractor* (SVCE) [19]–[21] was used to convert these into RDF [5] linked data format and asserted it in our broad Cybersecurity Knowledge Graph (CKG) [21], [25], [26]. The CTI linked data was augmented with 1739 Wikidata objects and information about 163 medical devices listed on the AccessGUDID database. To evaluate the impact of knowledge augmentation on embedding quality improvement we first evaluate our knowledge augmentation process and then it's impact on embedding generation.

We used the '*owl:SameAs*' assertion to 'connect' different knowledge obtained from manufacturer CTI, Wikidata, and AccessGUDID. Assertions were dependent on the match between manufacturer's/brand name and the device name. If there was a complete match the entities were linked directly. However, in case, there is no exact match, we calculate the term frequency inverse document frequency (tf-idf) scores to calculate similarity and connect nodes. A similar technique

[10]https://www.wikidata.org/wiki/Wikidata:Main_Page
[11]https://www.wikidata.org/wiki/Wikidata:Data_access
[12]https://www.wikidata.org/wiki/Q1152374

was used by Piplai et al. [26]. The augmented CKG can handle complex queries using the SPARQL query language [20].

Evaluating the knowledge augmentation process: In order to ensure that knowledge from different sources was connected correctly, we used a group of 3 annotators to manually check 150 randomly selected connections. Out of the 98 (these were the ones with and inter-annotator agreement higher than 0.66), 68 were marked correct, 15 were marked somewhat correct and the rest were marked incorrect.

IV. CKG Embedding Generation

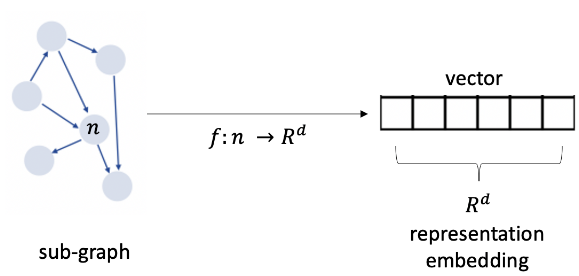

Fig. 3. Graph Embedding generation process.

Graph embeddings have been used to represent large graphical networks with the aim of improving tasks like, node classification, link prediction, community detection, network similarity and many others. The goal is to map each node into a low-dimensional space, while preserving most of the network information. Given a graph G, with nodes V, and relations E, the task is to learn -

$$f : n \longrightarrow R^d, n \in V$$

We want to learn the feature representation that is predictive of nodes in n neighborhood $N_s(n)$. The neighbourhood is generally defined depending on the use-case where the representations are being utilized. For a global macroscopic view, Depth First Search (DFS) can be used to define the neighbourhood. In tasks where a local microscopic view is needed Breadth First Search (BFS) can be used to describe the neighbourhood.

To generate the graph embeddings for our CKG, we use the Breadth First Search to define a local neighbourhood for the node. Once the neighbourhood has been defined we generate the embeddings using the node2vec algorithm [6]. We next evaluate the impact of our knowledge augmentation on the process of embedding improvement.

Brand	Number of CTI
EchoPAC	7
Versana (all except for Versana Essential)	5
ViewPoint product line	6
Vivid product line	32
LOGIQ (all except for LOGIQ 100 Pro)	45
Voluson product line	38
Invenia ABUS Scan station	6
Venue (all except for 40 R1-3, 50 R4-5)	7

TABLE I
Some popular medical device brands and number of known threat intelligence about their products.

Knowledge Augmentation Level	MAP score
CTI	0.54
CTI + Wikidata	0.66
CTI + Wikidata + AccessGUDID	0.71

TABLE II
MAP values for embedding models created using different levels of knowledge augmentation.

Evaluating impact of augmentation on embedding quality: We used node2vec [6], to generate our embeddings. In our generation process, we define *empirically* the neighbourhood size of a node to be at-most 4 degrees of separation and embedding size of 200 dimensions. Each node in the CKG was represented as a vector of 200 dimensions. This evaluation task was converted to an information retrieval task, where similar nodes to an input vector are compared to a predefined set of known similar entities. This process has been used in evaluating the word2vec model suggested by Mikolov et al. [18] and Mittal et al. [21]. OWASP[13] maintains groups of similar vulnerabilities[14] and attacks[15]. We created 14 groups of similar vulnerabilities, 11 groups of similar attacks, and 15 groups of similar medical products. For the experiment one entity from these group was set up as an input to the embedding model, and similar entities to this input were computed. To evaluate the impact of knowledge augmentation on the quality of embeddings created as an information retrieval task, we used the Mean Average Precision (MAP) metric. MAP is a popular metric used to measure the performance of models doing document/information retrieval. MAP values are between 0 and 1 and higher is better. In our case, we created 3 levels of knowledge augmentation which are, Just the CTI, CTI augmented with Wikidata, CTI and Wikidata augmented with AccessGUDID. Table II, shows the different MAP values obtained. CTI and Wikidata augmented with AccessGUDID performed the best in the experiment by about 31% (over the base MAP score of 0.54 with CTI only).

V. Conclusion and Future Work

In order to better protect users of medical devices and those with information in the hospital ecosystem it is necessary to create a repository of known security vulnerabilities for medical devices. For medical devices to become more secure there needs to be a well developed machine understandable knowledge repository for current and past vulnerabilities. In this paper, we collect Cyber Threat Intelligence (CTI) about various medical devices from sources like manufacturers, CISA ICS-CERT, etc. We augment this intelligence with data from the Wikidata knowledge graph and medical databases like FDA's AccessGUDID. These data sources are integrated along with the threat intelligence in our Cybersecurity Knowledge Graph (CKG). The augmented CKG helps produce better

[13]https://www.owasp.org/index.php/Main_Page
[14]https://www.owasp.org/index.php/Category:Vulnerability
[15]https://www.owasp.org/index.php/Category:Attack

quality node graph representations. We were able to get a 31% increase in the Mean Average Precision (MAP) value as computed over a information retrieval task. These graph embeddings have been used to represent large graphical networks with the aim of improving tasks like, node classification, link prediction, community detection, network similarity and many others. In the future, we will be able to augment the CKG with other data sources further improving the quality of the graph embeddings. These augmented graph embeddings will help further improve various natural language processing tasks on cybersecurity text data.

ACKNOWLEDGEMENT

This work was supported by a National Science Foundation (NSF) grant, award number 2025685. The authors will also like to thank members of the Ebiquity research group at University of Maryland Baltimore County.

REFERENCES

[1] Medical Device Databases. https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/medical-device-databases.

[2] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

[3] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[4] Oasis group. Stix 2.0 documentation. https://oasis-open.github.io/cti-documentation/stix/examples.html, May 2013.

[5] RDF Working Group. Resource description framework (rdf). https://www.w3.org/RDF/, 2020. [Online].

[6] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[7] Hashem Hashemi, Amin Azmoodeh, Ali Hamzeh, and Sattar Hashemi. Graph embedding as a new approach for unknown malware detection. *Journal of Computer Virology and Hacking Techniques*, 13(3):153–166, 2017.

[8] Shin-Ying Huang, Yen-Wen Huang, and Ching-Hao Mao. A multi-channel cybersecurity news and threat intelligent engine-secbuzzer. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 691–695, 2019.

[9] ICD10. Icd-10-cm/pcs medical coding reference. https://www.icd10data.com/.

[10] SNOMED International. The value of snomed ct. http://www.snomed.org/snomed-ct/why-snomed-ct.

[11] Karuna P Joshi, Aditi Gupta, Sudip Mittal, Claudia Pearce, Tim Finin, et al. Alda: Cognitive assistant for legal document analytics. In *2016 AAAI Fall Symposium Series*, 2016.

[12] Maithilee Joshi, Sudip Mittal, Karuna P Joshi, and Tim Finin. Semantically rich, oblivious access control using abac for secure cloud storage. In *2017 IEEE international conference on edge computing (EDGE)*, pages 142–149. IEEE, 2017.

[13] Nitika Khurana, Sudip Mittal, Aritran Piplai, and Anupam Joshi. Preventing poisoning attacks on ai based threat intelligence systems. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2019.

[14] Fucheng Liu, Yu Wen, Dongxue Zhang, Xihe Jiang, Xinyu Xing, and Dan Meng. Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 1777–1794, 2019.

[15] Patrick Luckett, J Todd McDonald, and William Bradley Glisson. Attack-graph threat modeling assessment of ambulatory medical devices. *arXiv preprint arXiv:1709.05026*, 2017.

[16] Stanislav Malyshev, Markus Krötzsch, Larry González, Julius Gonsior, and Adrian Bielefeldt. Getting the most out of wikidata: Semantic technology usage in wikipedia's knowledge graph. In Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, and Elena Simperl, editors, *The Semantic Web – ISWC 2018*, pages 376–394, Cham, 2018. Springer International Publishing.

[17] Sinclair Meggitt. Medjack attacks: The scariest part of the hospital. 2018.

[18] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[19] Sudip Mittal, Prajit Das, Varish Mulwad, Anupam Joshi, and Tim Finin. Cybertwitter: Using twitter to generate alerts for cybersecurity threats and vulnerabilities. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2016.

[20] Sudip Mittal, Anupam Joshi, and Tim Finin. Thinking, fast and slow: Combining vector spaces and knowledge graphs. *arXiv preprint arXiv:1708.03310*, 2017.

[21] Sudip Mittal, Anupam Joshi, and Tim Finin. Cyber-all-intel: An ai for security related threat intelligence. *UMBC Faculty Collection*, 2019.

[22] Lars Müller, Rashmi Gangadharaiah, Simone C Klein, James Perry, Greg Bernstein, David Nurkse, Dustin Wailes, Rishi Graham, Robert El-Kareh, Sanjay Mehta, et al. An open access medical knowledge base for community driven diagnostic decision support system development. *BMC medical informatics and decision making*, 19(1):93, 2019.

[23] Lorenzo Neil, Sudip Mittal, Anupam Joshi, et al. Mining threat intelligence about open-source projects and libraries from code repository issues and bug reports. *IEEE Intelligence and Security Informatics (IEEE ISI) 2018*, 2018.

[24] National Institutes of Health. Pubchem. https://pubchem.ncbi.nlm.nih.gov/.

[25] Aditya Pingle, Aritran Piplai, Sudip Mittal, Anupam Joshi, James Holt, and Richard Zak. Relext: Relation extraction using deep learning approaches for cybersecurity knowledge graph improvement. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2019.

[26] Aritran Piplai, Sudip Mittal, Anupam Joshi, Tim Finin, James Holt, and Richard Zak. Creating cybersecurity knowledge graphs from malware after action reports. *UMBC Faculty Collection*, 2019.

[27] Priyanka Ranade, Sudip Mittal, Anupam Joshi, and Karuna Joshi. Using deep neural networks to translate multi-lingual threat intelligence. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 238–243. IEEE, 2018.

[28] Arpita Roy, Youngja Park, and SHimei Pan. Learning domain-specific word embeddings from sparse cybersecurity texts. *arXiv preprint arXiv:1709.07470*, 2017.

[29] Taneeya Satyapanich, Francis Ferraro, and Tim Finin. Casie: Extracting cybersecurity event information from text. *UMBC Faculty Collection*, 2020.

[30] Robert David Steele. Open source intelligence: What is it? why is it important to the military. *American Intelligence Journal*, 17(1):35–41, 1996.

[31] US-CERT. Sweyntooth cybersecurity vulnerabilities. https://us-cert.cisa.gov/ics/alerts/ics-alert-20-063-01.

[32] W3. Sparql query language. https://www.w3.org/TR/rdf-sparql-query/.

[33] Meng Wang, Mengyue Liu, Jun Liu, Sen Wang, Guodong Long, and Buyue Qian. Safe medicine recommendation via medical knowledge graph embedding. *arXiv preprint arXiv:1710.05980*, 2017.

[34] Jian Xu. Systematic vulnerability evaluation of interoperable medical device system using attack trees. 2015.

[35] Jian Xu, Sunkyu Kim, Min Song, Minbyul Jeong, Donghyeon Kim, Jaewoo Kang, Justin F Rousseau, Xin Li, Weijia Xu, Vetle I Torvik, et al. Building a pubmed knowledge graph. *Nature Sci Data 7, 205 (2020)*, 2020.

[36] Suya Zhao, Renzheng Wei, Lijun Cai, Aimin Yu, and Dan Meng. Ctlmd: Continuous-temporal lateral movement detection using graph embedding. In *International Conference on Information and Communications Security*, pages 181–196. Springer, 2019.