THE UNIVERSITY OF TEXAS, ARLINGTON

A7143

CREDIBLE OPEN KNOWLEDGE NETWORK (COKN)

Chengkai Li
cli@uta.edu

Dolores Albarracin
Xiaojing Liao
Yinghui Wu
Jun Yang

OVERVIEW

In a world where decision making is heavily driven by data, lack of information credibility leads to a populace misled and divided by widespread misinformation and algorithmic decisions that are irrelevant or even catastrophic. Our team is building framework and tools to assist software developers and domain experts to ensure the credibility of their decision-making software powered by open knowledge networks (OKN). Applying our technologies and tools to practical domains, we construct knowledge graphs and software to (a) assist healthcare professionals and app developers in mitigating health misinformation and (b) help cybersecurity professionals in accurately assessing and mitigating software vulnerability.

DESCRIPTION

Credibility as perceived by information recipients stems from both "objective" and "subjective" factors. Objective credibility involves the factual accuracy of information, and subjective credibility involves perceptions that make information convincing to recipients. To improve objective credibility, we will develop data veracity tools for repairing incorrect and missing data, and we will develop data provenance tools that explain query and analytic results. To improve subjective credibility, we will develop strategies to contextualize results from queries and analytics according to information recipients' needs and profiles, and we will develop communication strategies to present such results in a way that is likely to be judged credible by target recipients.

In addition to providing a novel framework of credibility and a suite of tools that can be instantiated in a variety of domains, we directly apply our tools to produce credibility-aware knowledge graphs and software in two highly important domains: cyber threat intelligence and mitigation of health misinformation related to vaccination hesitancy and to the COVID-19 pandemic. Furthermore, we will collaborate with several other teams (A6677 GIS knowledge, A7017 chemicals and materials systems, and A7115 civil infrastructure) to apply our framework and tools in their domains.

Without being addressed, credibility challenges will become a road block to the ambition of OKN. Opening up the knowledge network also opens the door to low-credibility decisions, since an OKN without credibility assurance may lead to inaccurate or even disastrous decisions. Outcomes from our ongoing project revealed critical misinformation in high-profile security vulnerability repositories such as the National Vulnerability Database, which is used for assessing the security of applications deployed on Amazon Web Services.

DIFFERENTIATORS

Decades of research and development in computing technologies such as data extraction, integration and cleaning, semantic web, database query, and security has resulted in vast experience in tackling important challenges in creating an OKN. However, little has been done to systematically ensure the credibility of software powered by knowledge graphs. This project will complement the aforementioned advancements and together will help realize the vision of OKN.

Ensuring credibility is imperative for all the convergence accelerator projects. While other teams are developing technologies for intelligent decision making based on OKNs in various application domains, our cross-cutting technology is an important enablement of the vision of OKN.

There have been many efforts in studying information credibility and mitigating harmful misinformation, especially in the context of news and social media,

including highly influential works by members of this team. However, except for methods of repairing erroneous data and augmenting incomplete data, there is no prior state-of-the-art for ensuring the credibility of knowledge graphs and software that use such data.

ROAD MAP

The project so far has produced initial versions of credible knowledge graphs and decision-making tools for health misinformation mitigation and cybersecurity. We have also created a public dashboard and a browser plugin for mitigating COVID-19 related misinformation.

The Phase II tasks will be carried out in parallel from onset. The anticipated completion dates of key milestones and deliverables are as follows:

Month 6: Framework of credibility and data model.

Month 9: Knowledge object collection on cyber threat intelligence and health misinformation.

Month 13: Tools for objective credibility.

Month 15: Tools for subjective credibility.

Month 15: Software for health misinformation mitigation and software vulnerability prioritization.

Month 15: Adapted tools to assist other teams in producing credible decision-making systems.

Month 18: Evaluation of deliverables with partners.

Month 19: Refined datasets, tools and software based on the evaluation results.

Month 21: Final evaluation with partners.

Month 24: Releasing datasets, software and website.

PARTNERSHIPS

Our interdisciplinary team has significant critical mass: 1) leading computer scientists in data-related areas; 2) renowned social scientists in credibility, persuasion, psychology, and communication; 3) domain experts and partners in healthcare, vaccine, health misinformation, and cybersecurity; 4) collaborators from the Army Research Lab, Qatar Computing Research Institute, and collaborators with major contributions to the largest knowledge graph products at Amazon, Google and IBM, and to Google's fact-checking products.

The project is informed by interviews, datasets and use cases from partners such as Fortiphyd Logic, Inc., Siemens, Metafact, FactMata, and Duke University Health System, as well as the aforementioned Track A teams. The deliverables will be evaluated and deployed by these partners. A preliminary version of our vulnerability severity prioritization and assessment tool has been deployed by Fortiphyd Logic to mitigate security risks in industrial control systems.

INTELLECTUAL PROPERTY

This project is not using any proprietary technology. We are mostly using publicly available data. For development and evaluation in industrial environments at our partners, steps involving their proprietary data will be directly performed by them. With regard to intellectual property produced in the project, we aim to open source all tools and software and make all datasets produced in the project available to the public.

ADDITIONAL INFORMATION

https://idir.uta.edu/cokn/