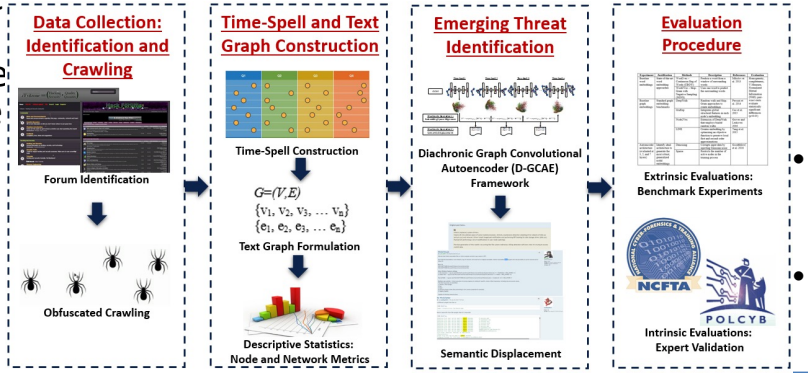
CRII: SaTC: Identifying Emerging Threats in the Online Hacker Community for Proactive Cyber Threat Intelligence

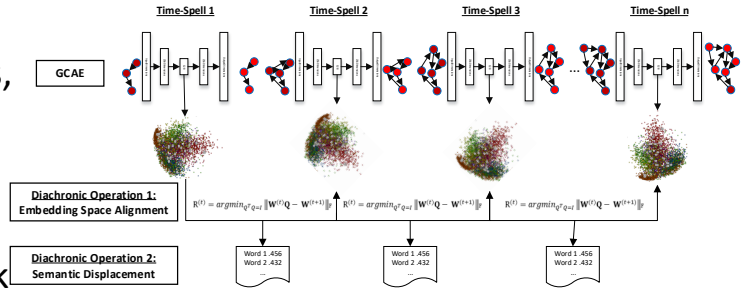
Challenge:

- The Dark Web is an emerging and viable CTI data source as it motivates millions of hackers from US, China, Russia to share malicious tools and knowledge
- Hackers rapidly develop new malware with novel functions.
- Unclear how semantics of hacker terms shift over time.



Solution:

- Draws upon and extends state-of-the-art in text graphs, diachronic linguistics, and unsupervised deep learning methodologies
- **Key Innovations:** Diachronic Graph Embedding Framework (D-GEF), Novel Graph of Words for Hacker Content, Identifying Emerging Threats



Scientific Impact:

- Novel CTI framework designed to collect and identify emerging threats from multi-million record hacker forums.
- Publicly accessible datasets and code for scientific reproducibility.
- Advancements for diachronic linguistics and lexical semantics.

Broader Impact:

- Dissemination and integration of research to two international information sharing entities → 800+ partners across academia, industry, and government.
- Integrating selected results and datasets into IU's AI4Cyber edX course

Project information:
Number: CNS-2041770
Institution: Indiana University, Bloomington

Contact: ssamtani@iu.edu

CENTER FOR APPLIED CYBERSECURITY RESEARCH
INDIANA UNIVERSITY
Pervasive Technology Institute