NDSS The Network and Distributed System Security Symposium 2023

DO-I-TRUST: DISSECTING ON-CHAIN COMPROMISED DOMAINS VIA GRAPH LEARNING

Shuo Wang*, Mahathir Almashor, Sharif Abuadbba, Ruoxi Sun, Minhui Xue, Calvin Wang, Raj Gaire, Surya Nepal and Seyit Camtepe

The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

02-03-2023

02

PROBLEMS

PHISHING/URL

- 3 Billion phishing emails every single day according to Terranova Security (Dec 2021)
- One of the most common tools of cybercrime is the use of malicious URLs and websites to carry out phishing attacks, malware distribution, etc.

ALLOW/BLOCK LIST

- easy to use, but sometimes useless
- incompleted, passive, and time/labor intensive

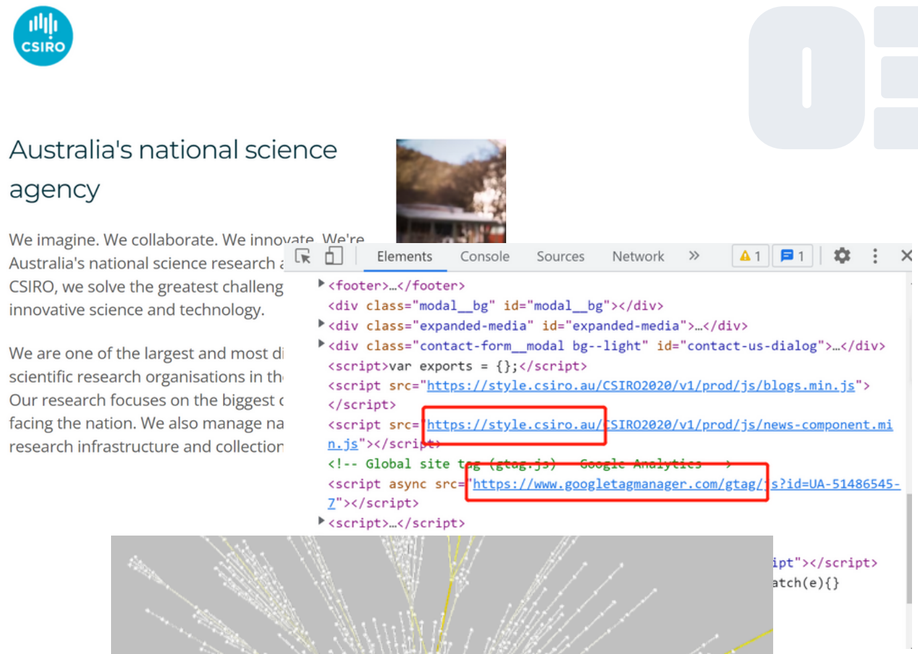
“how to develop a dynamic mechanism for expanding and monitoring the allow/block lists”

NDSS Symposium | 2023

03

EXPANSION

- Seed list, including gov, org, au, etc.
- Crawling the children urls in the html of main page of each domain in the seed list.
- Breadth-First Search layer by layer until reach 6 layer depth.
- 50K seeds to a graph with 1.7M node and 12M edges



TRANSITIVITY

- transitivity of trust, allowing the trust between two parties to be extended further.
- Good news: an extended list and introducing topology relations

EXPANSION GRAPH

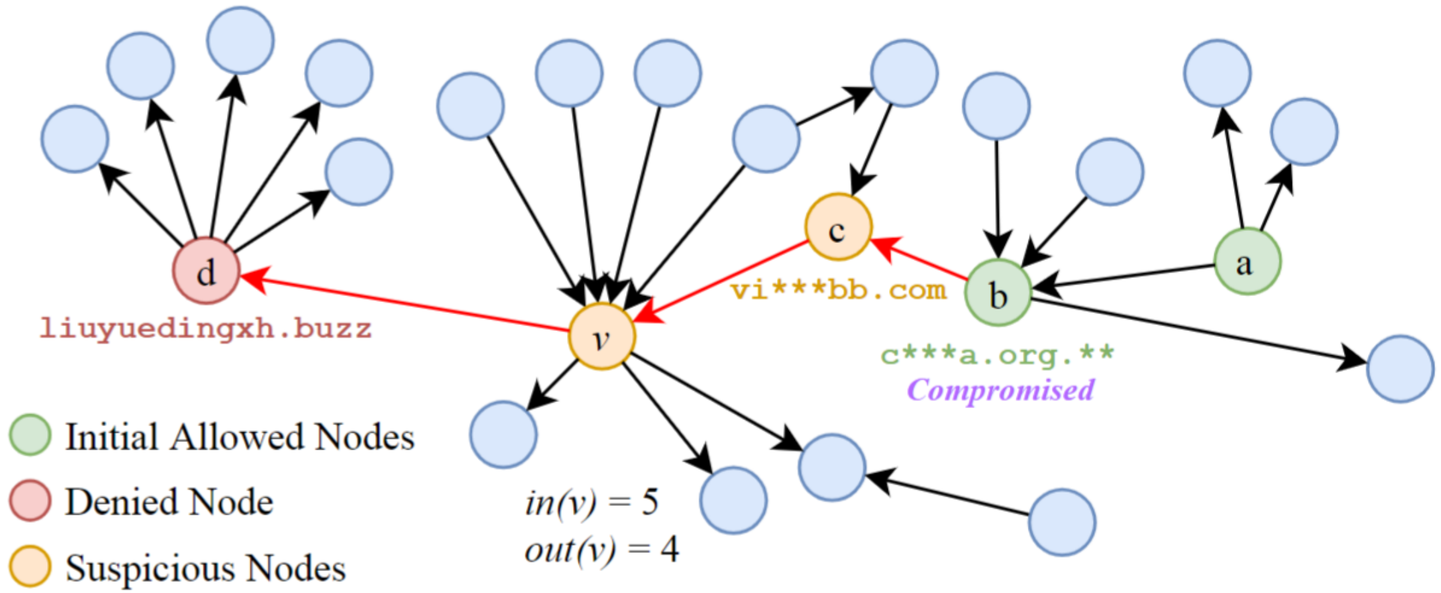
PROBLEM STATEMENT

MALICIOUS V.S. ON-CHAIN COMPROMISE

- malicious patterns
- evasion of detection
- content stealthiness
- constantly updated
- topology (intent) stealthiness

BENEFITS OF GRAPH

- build an extensive and dynamic allow list;
- wide view to further monitor the trust of allow-list.
- modeling the suspicious behavior pattern and analyzing how benign webpages are compromised and eventually linked to malicious websites.



TECHNICAL CHALLENGES

LABELS AND SUPERVISION

- No supervision information available for extracting malicious patterns.
- Only a small percentage of nodes (180/1.7M) have labels
- insufficient supervision information.
- Limited ground truth

EFFICACY

- either based on individual node features or only focusing on the topology
- High false positive rate and low accuracy
- Lightweight feature extraction for fast inference.

SCALABILITY AND PORTABILITY

- GNNs excel by considering both individual and global structural information.
- performance and scalability in large graphs
- web-compromising behaviors have a typical long chain of influence, beyond the 2-hops scope of ordinary graph learning schemes.
- always from 4 to 6
- overfitting and exponential explosion for the number of parameters

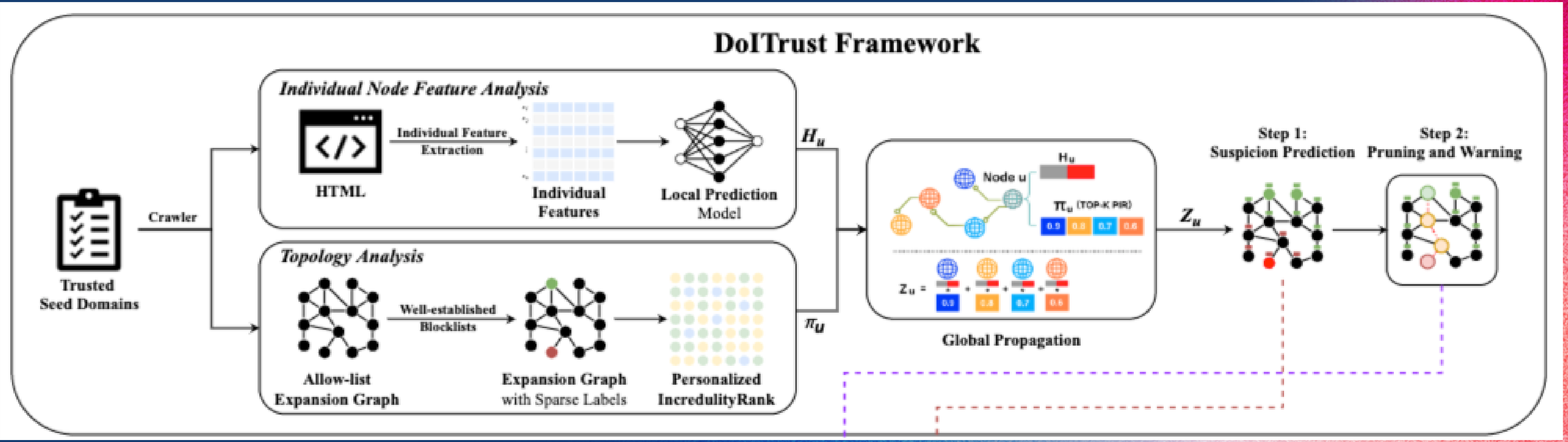
OUR INTEGRATED SCHEME 07

SUSPICION PREDICTION

target or intent, customized
small fraction of labels
semi-supervised node classification
scalability: local+global
node feature from HTML and prediction
ranking as message passing scheme

PRUNING AND WARNING

prediction indicates the compromise intent value.
pruning nodes of the graph according to the threshold of the suspicion value.
two pruning strategies to further achieve a clean extended allow-list



03

NDSS Symposium | 2023

LOCAL

STATISTICAL FEATURES

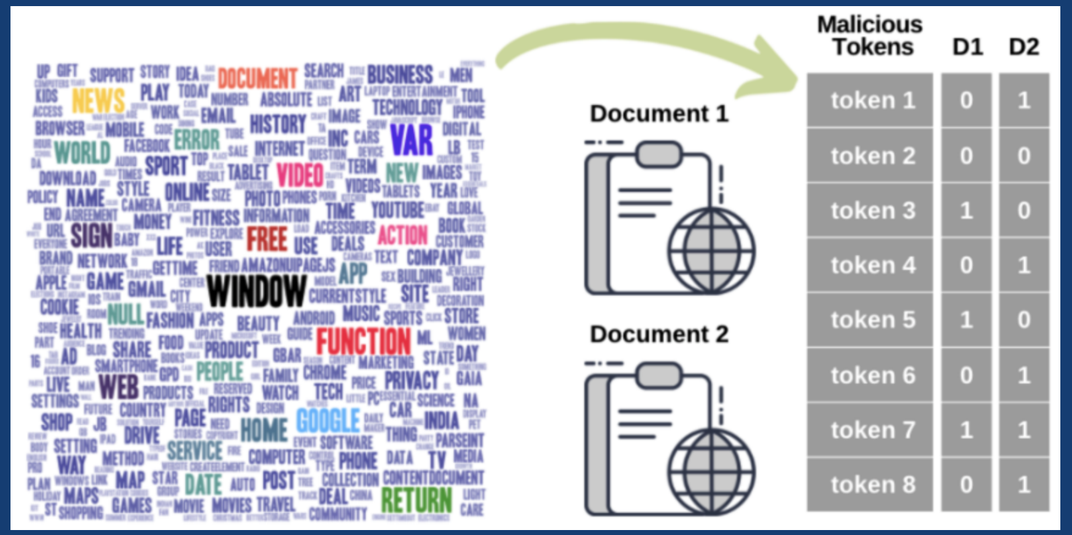
- length of a URL,
- number of special characters
- whether an IP address or re-direction

LEXICAL FEATURES

- vectorize HTML contents with BoM
- 2400 most frequent tokens (words or symbols)
- HTML --> frequency distribution matrix
- worse case

LOCAL PREDICTION

- a binary classification task
- 2-layer fully connected network (FCN).



GLOBAL

INCREDULITY RANK

- totally “benign” nodes are hard to define
- given the identified untrustworthy recommendation, find a recommender who strongly supports the recommendation after a few iterations of distrust backward propagation
- splitting and dampening distrust upwards

APPROXIMATION

- sparse initialization
- low efficiency on a large-scale
- adaptive initialization
- top-K approxiamtion

GLOBAL PROPAGATION & INCORPORATION

- Real-time and efficient inference is important for practical applications.
- trained once providing continuous inference.
- the impact of incorporation on the performance

COMPROMISE PRUNING

two pruning strategies

EXPERIMENT SETUPS

DATA AND SETTINGS

- Ground truth and scope: Suspicious prediction, denied nodes using GSB, Phishtank, blocklist; allowed nodes using Alexa with dropping
- exclude popular interactive websites
- top-level domains and default homepage.
- Small and large datasets: 10K and 1M
- network not in the training and validation

BASELINES

Individual Machine Learning Baselines (IML)
Graph Neural Network Baselines (GNN)
Scale Graph Neural Network Baselines
Structural Processing Only Baselines (SPO)

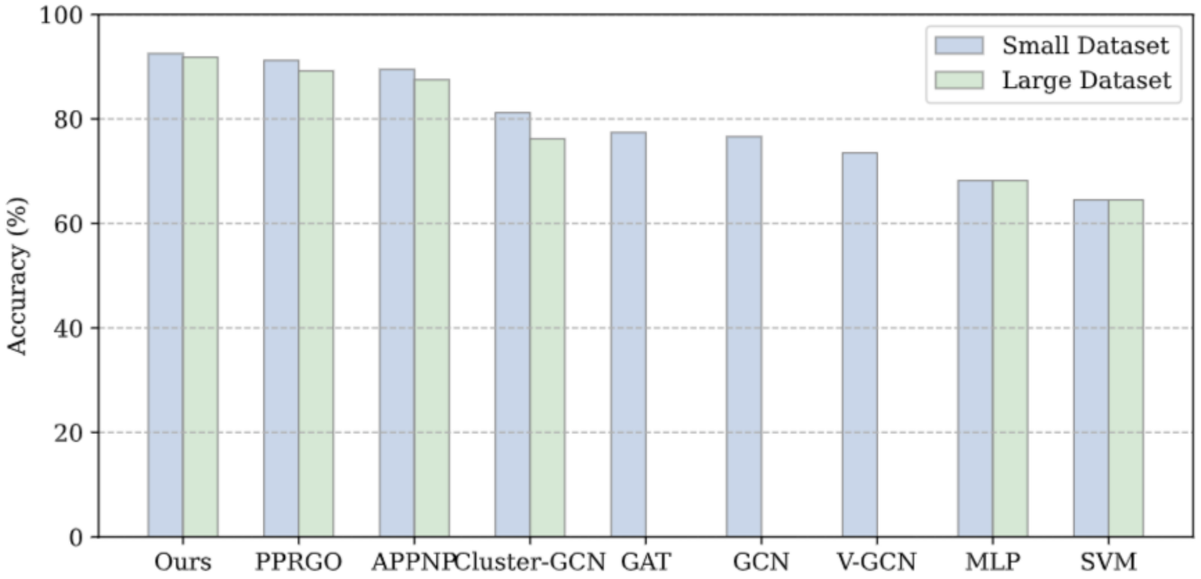
TWO-FOLD VALIDATION

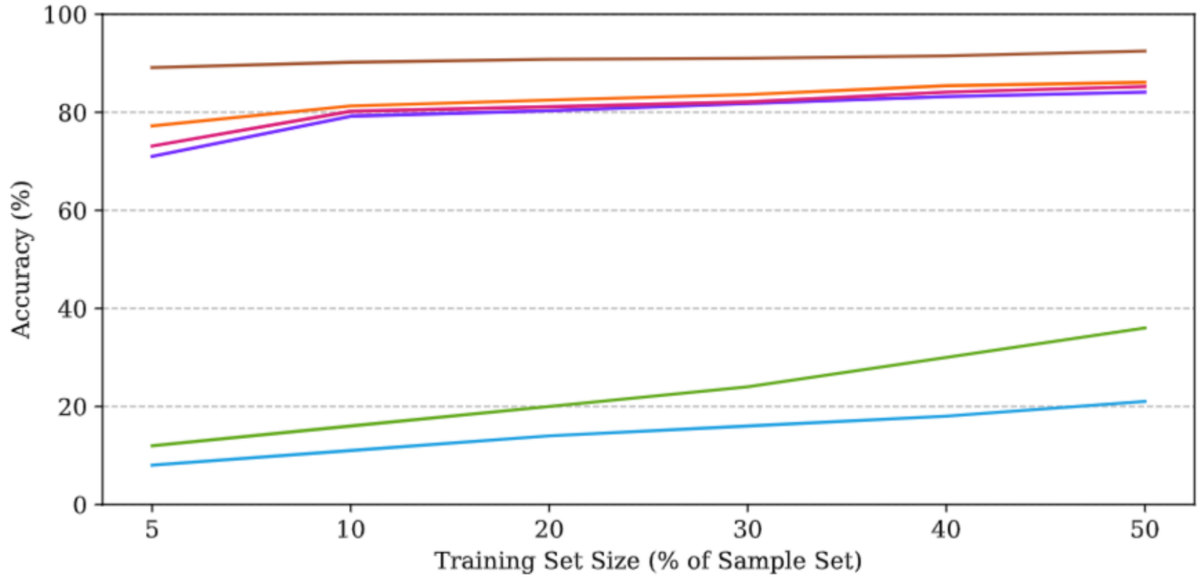
- denied node prediction validation with full ground truth (automatically quantitative evaluation)
- prediction accuracy and false positive rates;
- compromise analysis without ground truth (post-processing and analysis)
- confirm the correlation and to analyze the underlying indicators of compromising.
- accuracy of obvious-positive alarms (highly suspicious nodes with obviously suspicious information found)
subtle-positive alarms (highly suspicious nodes without obviously suspicious information found)

Method	Threshold	Precision	Recall	F1-score	Accuracy
PageRank	0.1	53.06%	66.67%	59.09%	54.43%
	0.2	50.41%	78.21%	61.31%	51.27%
	0.3	50.40%	80.77%	62.07%	51.27%
	0.4	50.71%	91.03%	65.14%	51.90%
	1.0	49.67%	97.44%	65.80%	50.00%
TrustRank	0.001	52.58%	65.38%	58.29%	53.80%
	0.005	49.67%	96.15%	65.50%	50.00%
	0.100	49.67%	97.44%	65.80%	50.00%
	0.200	49.68%	98.72%	66.09%	50.00%
	0.300	49.49%	99.23%	66.04%	49.62%
	0.400	49.46%	99.36%	66.04%	49.56%
	1.000	49.40%	99.62%	66.04%	49.43%
Step	3	52.16%	97.89%	68.06%	54.05%
	4	53.20%	97.44%	68.82%	55.85%
	5	53.20%	88.38%	66.41%	55.31%
	6	51.09%	67.76%	58.26%	51.45%
	7	47.55%	36.26%	41.14%	48.13%
Discount	0.4	51.07%	97.99%	67.14%	52.05%
	0.5	51.06%	92.77%	65.87%	51.93%
	0.6	47.51%	75.23%	58.24%	46.06%
	0.7	43.40%	53.93%	48.09%	41.80%
	0.8	36.88%	24.75%	29.62%	41.20%
	0.9	28.73%	5.71%	9.53%	45.77%

COMPARISONS WITH STRUCTURAL RANKING

precision 50%, half of positive predictions are incorrect.





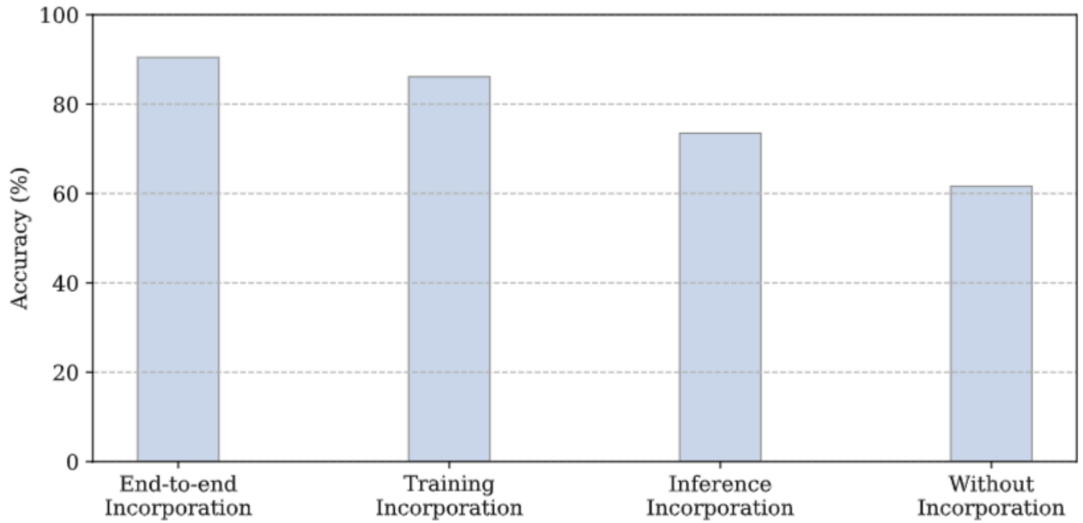
LEARNING CAPABILITY FOR SUSPICION PREDICTION

suspicion threshold as 0.5
common GNNs out of memory on Large.
various labeling rates.
Multi-hop neighbors

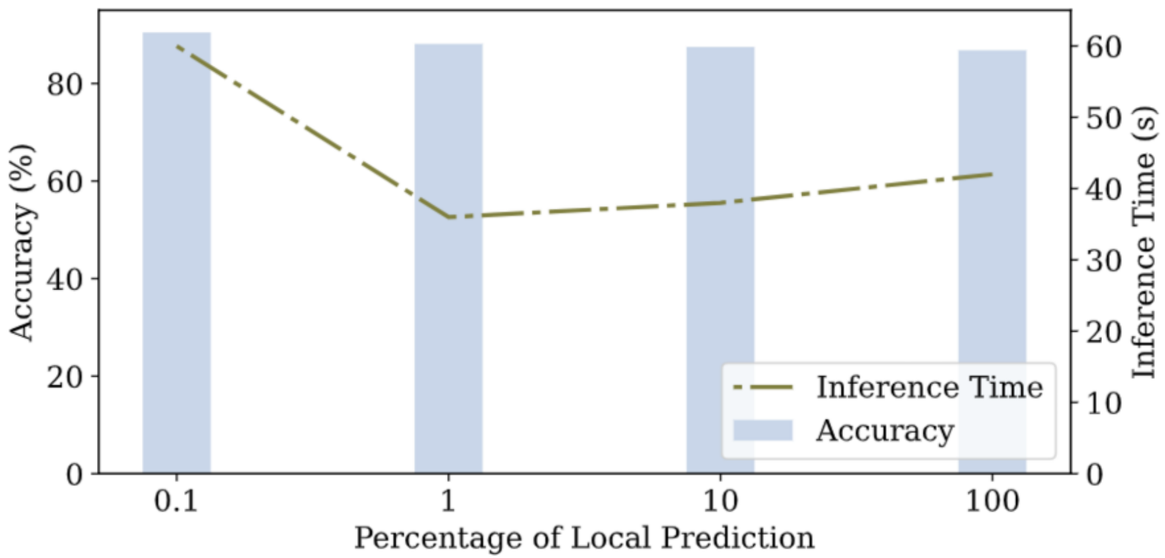
STATIC AND REAL-TIME INFERENCE EVALUATION

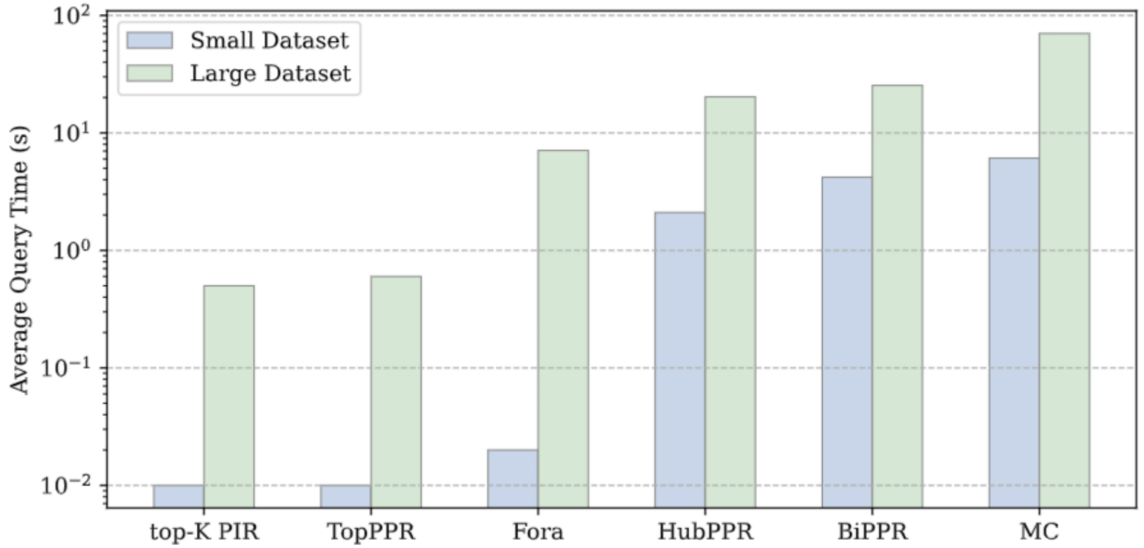
	Training		Inference	
	LP	GP	LP	GP
End-to-end Incorporation (EI)	●	●	●	●
Training Incorporation (TI)	●	●	●	○
Inference Incorporation (II)	●	○	●	●
Without Incorporation (WI)	●	○	●	○

LP: Local propagation; **GP**: Global propagation.
●: the related propagation is involved in the strategy;
○: the related propagation is not involved in the strategy.



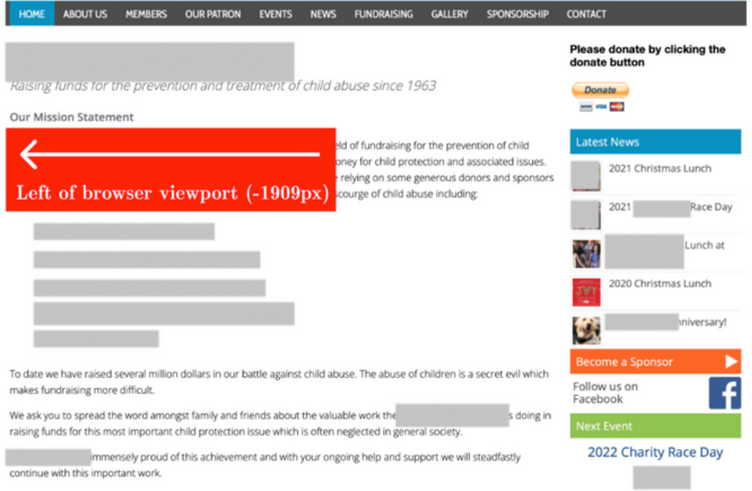
EE -WI 30%
TI 86%
II increases 12% compared to without propagation



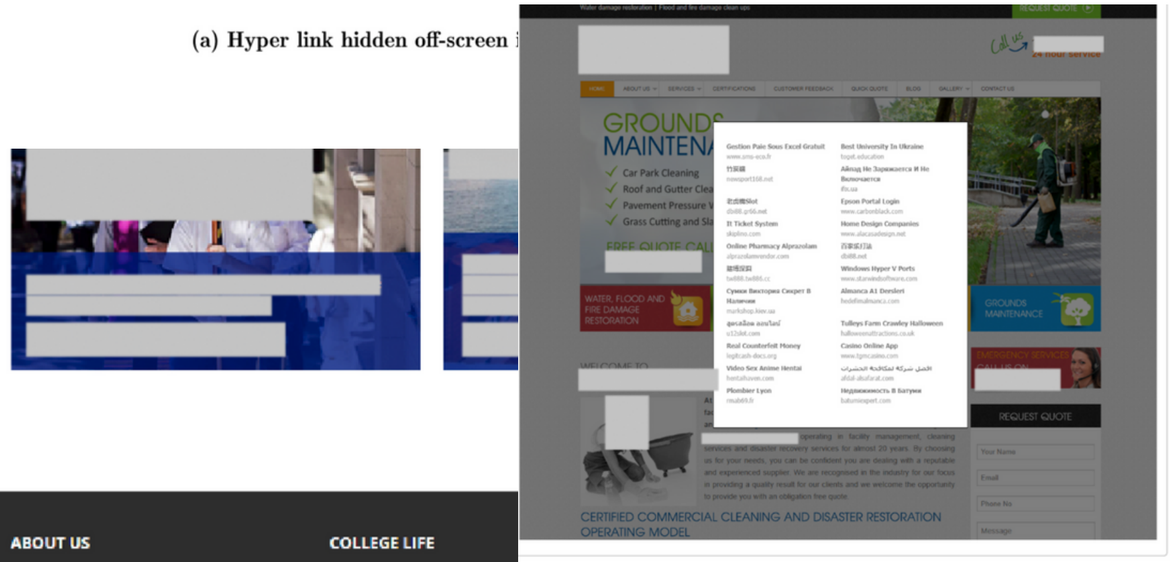


EFFICIENCY IN REAL-WORLD DEPLOYMENT

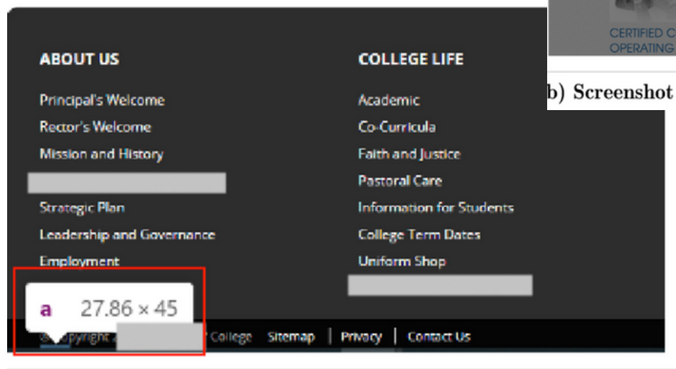
Time efficiency: training+inference < 1min
Memory efficiency: 1.5 GB and 10GB
optimization:
transferability
0.6% drop
reducing the nodes by a factor of 10.



(a) Hyper link hidden off-screen i



b) Screenshot of hidden popup (forced visible) from ca***es.com.**



(c) Hidden hyperlink in footer from s**w.edu.**

VALIDATION ON COMPROMISE

110 (10%) detected compromised nodes for manual review.
104 are true positives
Subtle-positive alarms

NDSS Symposium | 2023

NDSS Symposium | 2023

THANK YOU!

Shuo Wang
shuo.wang@csiro.au