An Automated Social Graph De-anonymization Technique

Kumar Sharad [1] George Danezis [2]

[1]

[2]

November 3, 2014

Workshop on Privacy in the Electronic Society, Scottsdale, Arizona, USA

This Talk

1. The Art of Data Anonymization
2. The D4D Challenge
3. An Ad-hoc Attack
4. Learning De-anonymization
5. Results

The Art of Data Anonymization

Releasing Anonymized Data

- **Motivation**: Process data without jeopardizing privacy.
- **Popular**: Randomize identifiers and/or perturb data.
- **Pros**: Cheap, preserves utility, provides legal immunity.
- **Cons**: Practiced as an **art form**.

The Data for Development (D4D) Challenge

The D4D Challenge[1]

- Introduced by a large Telco for research related to social development in Ivory Coast.
- Four datasets of anonymized call patterns released.
- Datasets include: Antenna-to-antenna calls, individual trajectories of varying spatial resolution and call graphs.
- Ivory Coast facts:
  - Population – 22.4 million.
  - Mobile phone users – 17.3 million.
  - Telco subscribers – 5 million.
  - A country fraught with civil war.

[1] http://www.d4d.orange.com/

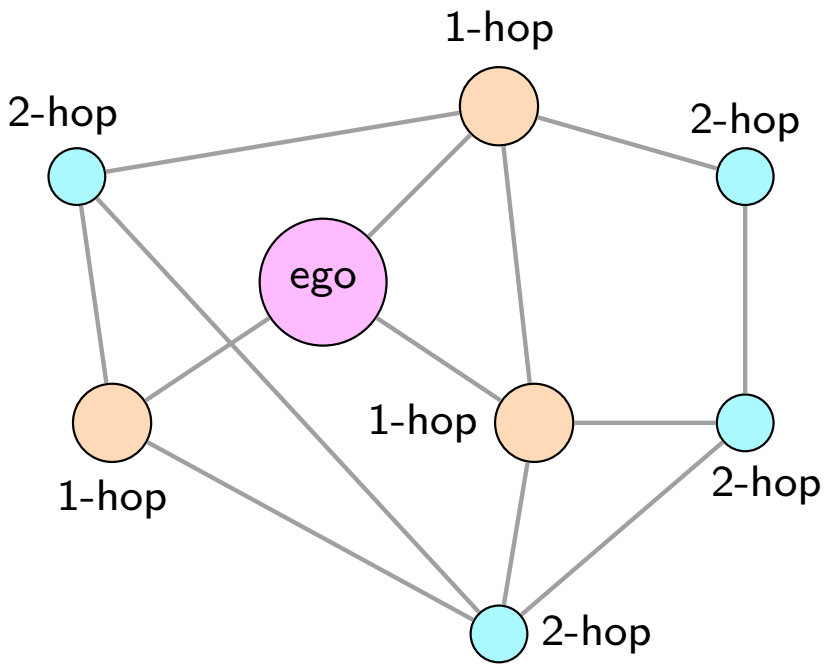
Timeline

- **July 2012**: A preliminary version of the datasets made available to us for evaluation.
- **September 2012**: We provide feedback depicting weaknesses of the scheme, specifically the anonymized call graphs.
- **Late 2012**: The challenge goes live after strengthening the anonymization. Released under strict NDA.

The Dataset 4: Anonymized Call Graphs

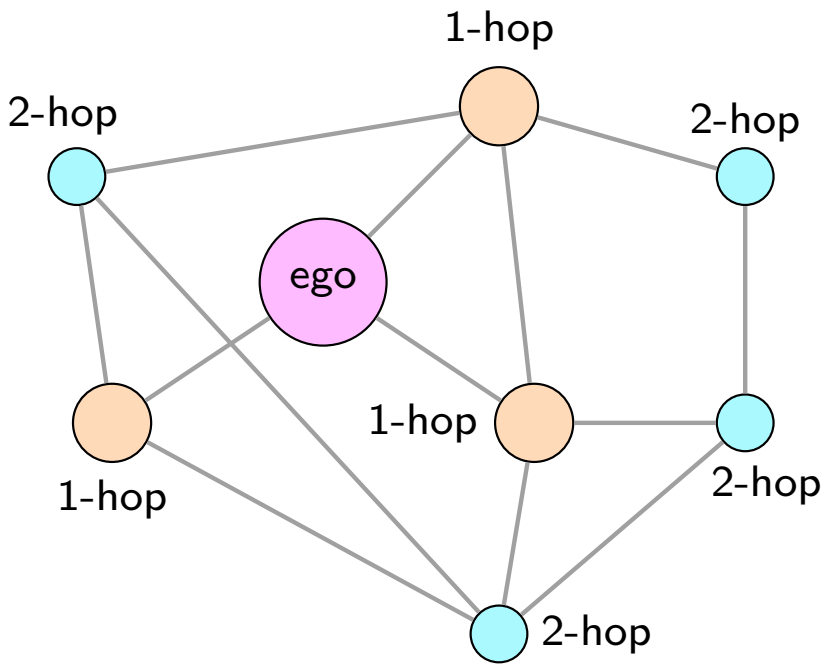
- 2-hop communication network (egonet) of an individual.
- Vertices represent users and edges their interactions.
- **Scheme 1** (pre-review):
  - 8300 egonets.
  - Edge attributes: call volume, duration and directionality.
- **Scheme 2** (post-review):
  - 5000 egonets.
  - All edges between 2-hop nodes are **removed**.
  - Edge attributes: **redacted**.

Scheme 1 vs. Scheme 2: Illustrated

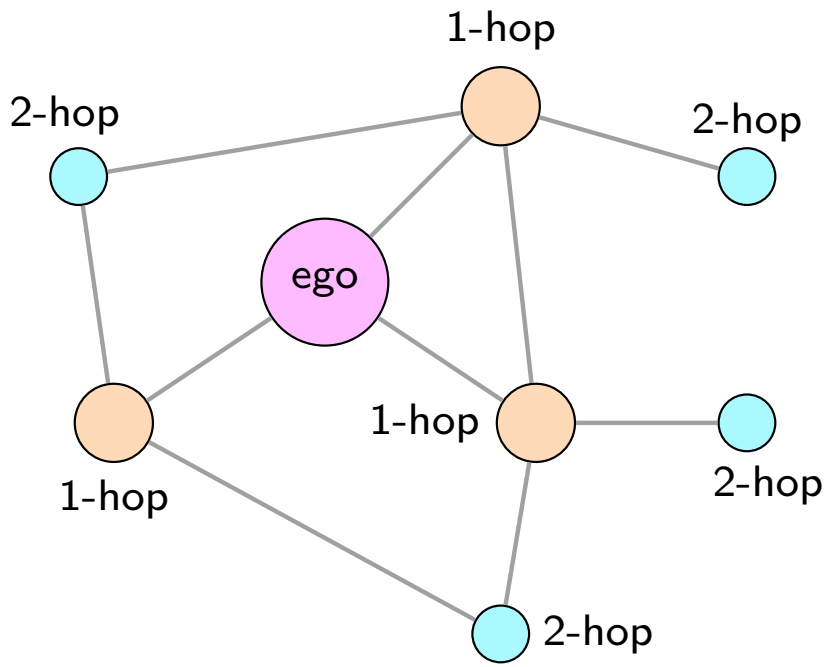


Scheme 1: Pre-review

Scheme 1 vs. Scheme 2: Illustrated

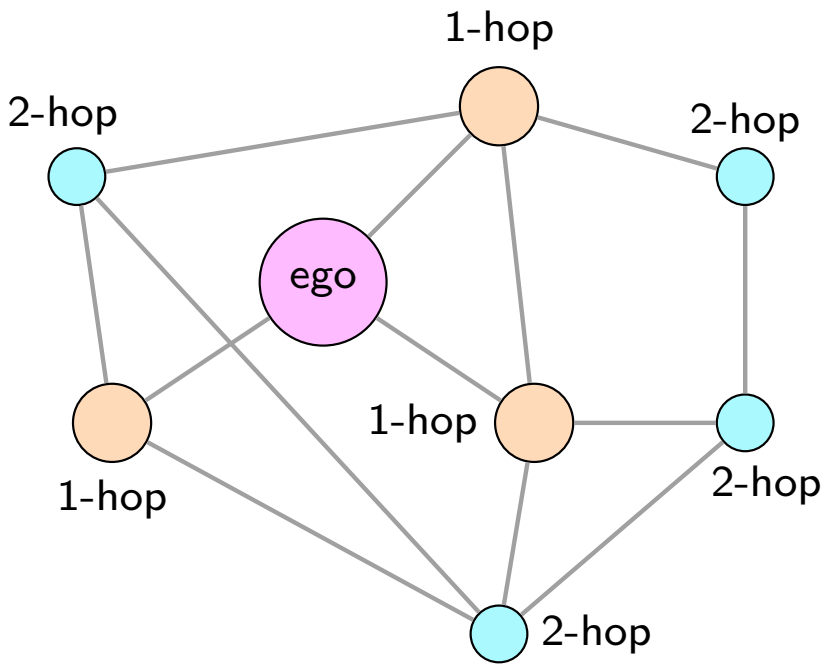


Scheme 1: Pre-review

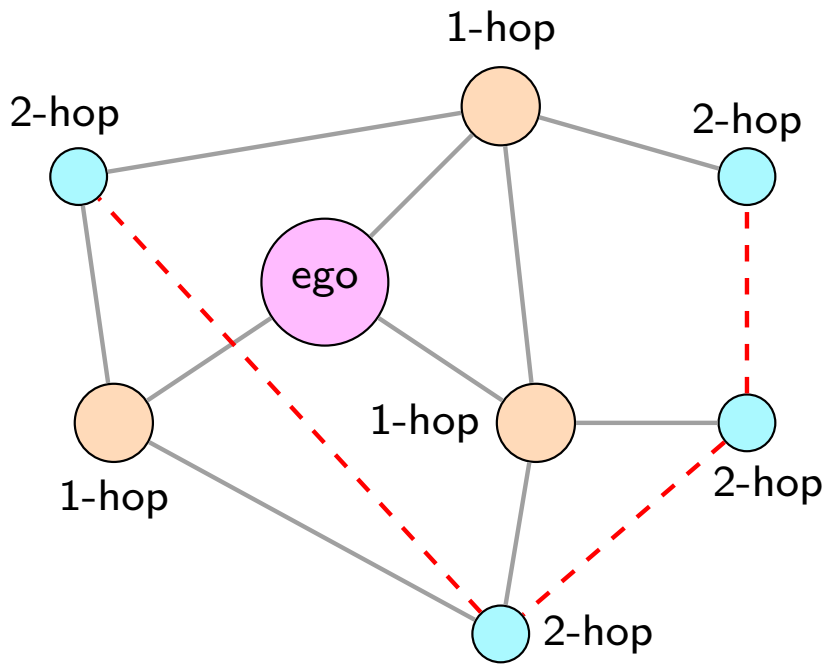


Scheme 2: Post-review

Scheme 1 vs. Scheme 2: Illustrated



Scheme 1: Pre-review



Scheme 2: Post-review

Anonymization Strategy

- Individuals picked at random.
- Identifiers randomized in each egonet.
- Tries to conceal the larger graph.
- Hope: Facilitate analysis while preserving privacy.
- Anonymity strengthened by redacting information.

How to Evaluate Anonymization Schemes?

- **Option 1**: We believe the scheme is secure.
  - Hard to merge the egonets.
  - Difficulty of linking egonets should be quantifiable.

How to Evaluate Anonymization Schemes?

- **Option 1**: We believe the scheme is secure.
  - Hard to merge the egonets.
  - Difficulty of linking egonets should be quantifiable.
- **Option 2**: We believe the scheme is insecure.
  - Show that a significant fraction of egonets can be re-linked.
  - Discern real world identities.
  - Recover full communication graph.

How to Evaluate Anonymization Schemes?

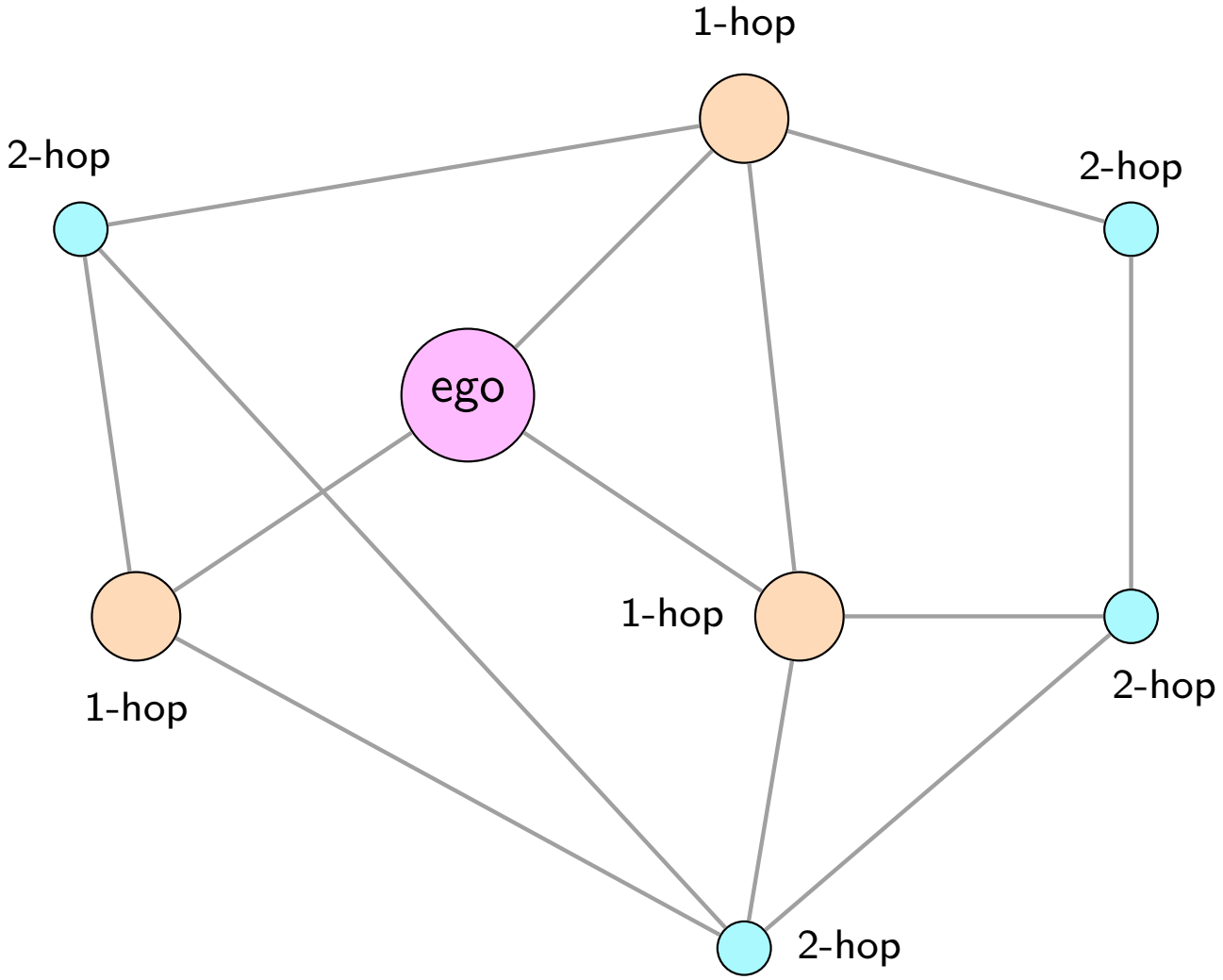
- **Option 1**: We believe the scheme is secure.
  - Hard to merge the egonets.
  - Difficulty of linking egonets should be quantifiable.
- **Option 2**: We believe the scheme is insecure.
  - Show that a significant fraction of egonets can be re-linked.
  - Discern real world identities.
  - Recover full communication graph.
- **Gap**: Lack of an attack does not imply security.

An Ad-hoc Attack

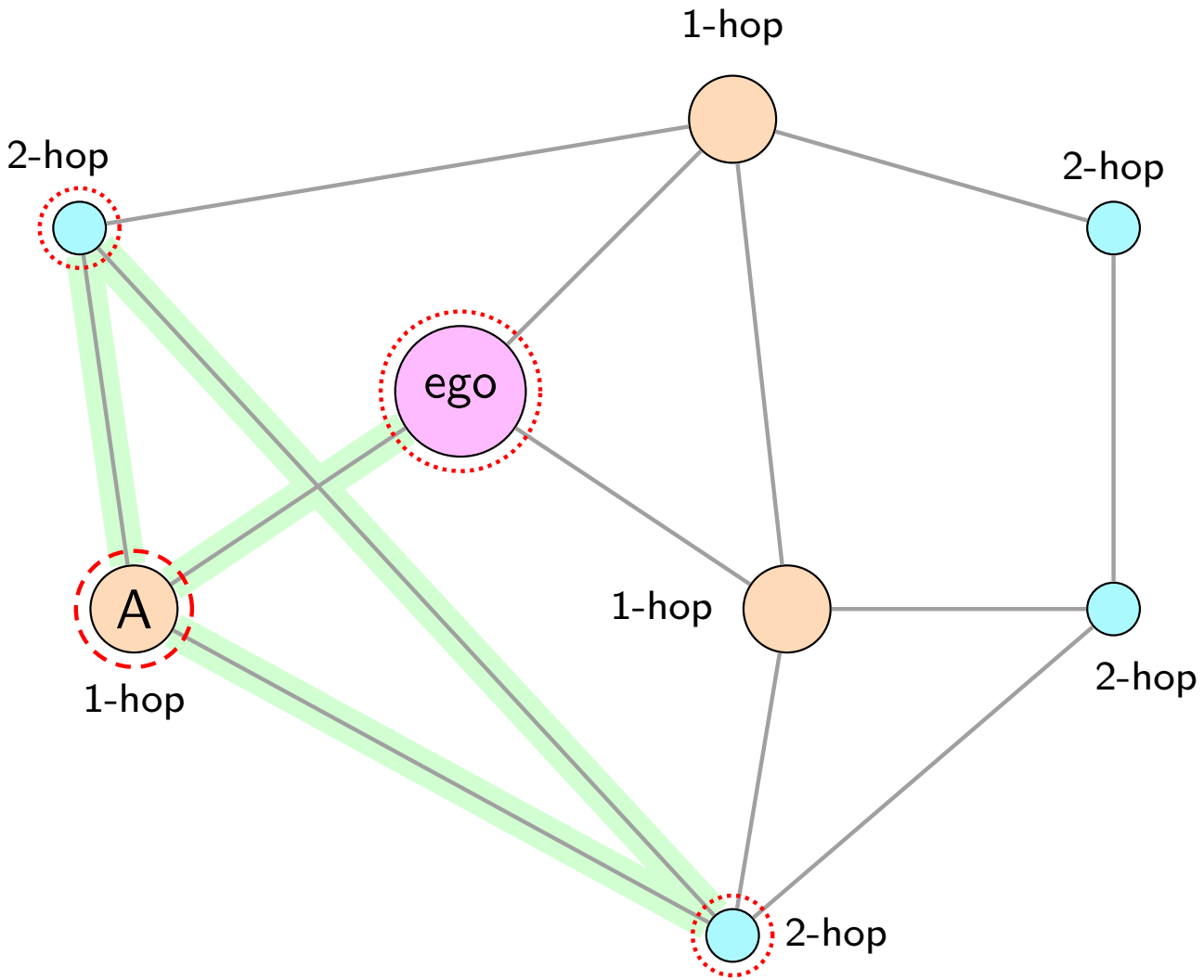
Ad-hoc Attack on Scheme 1

- Transformation into egonets preserves an important variant.
- The degree of egos and 1-hop nodes is preserved.
- Degrees of the 1-hop sub-graph of 1-hop nodes is preserved.
- Can be used as a stable signature.

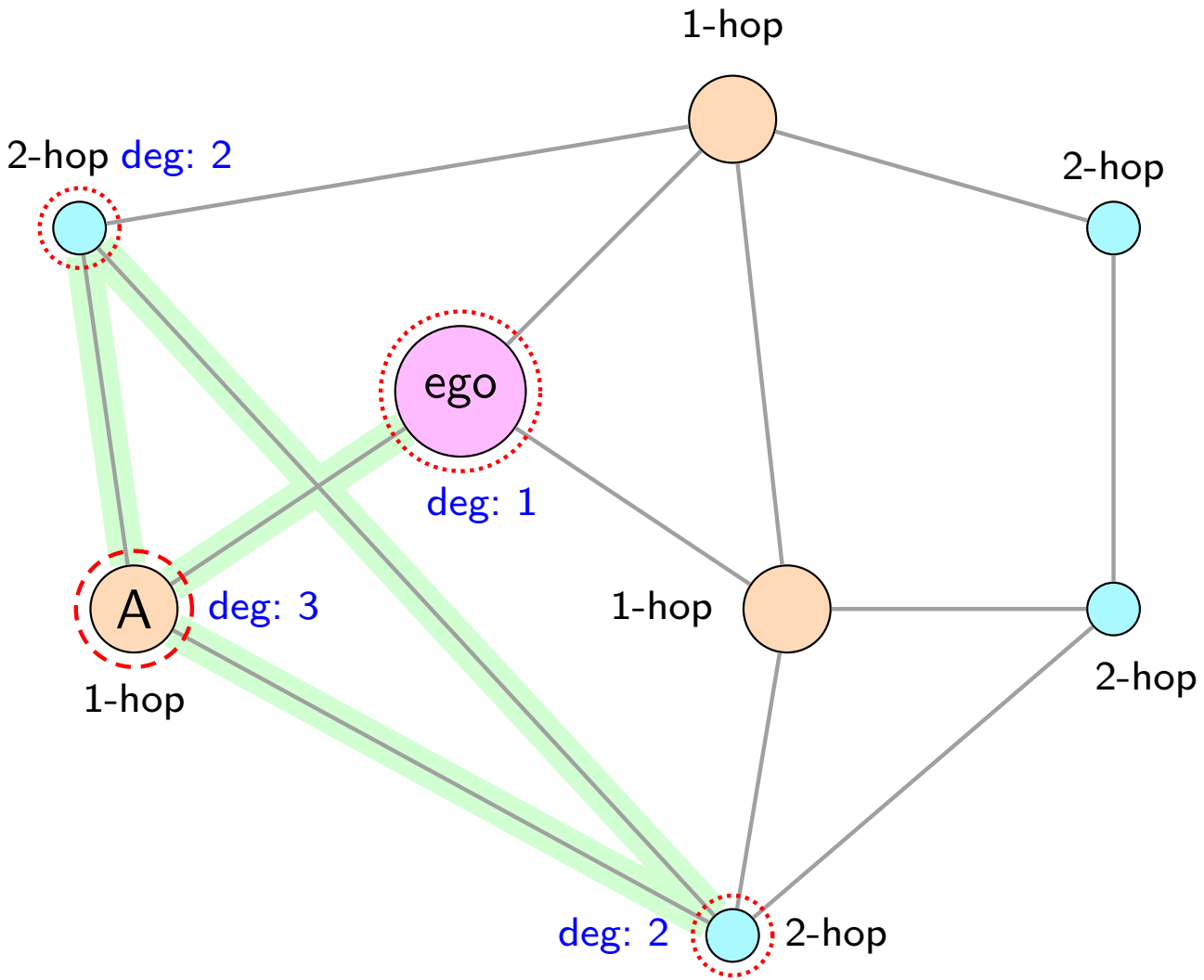
Ad-hoc Attack on Scheme 1: Illustrated



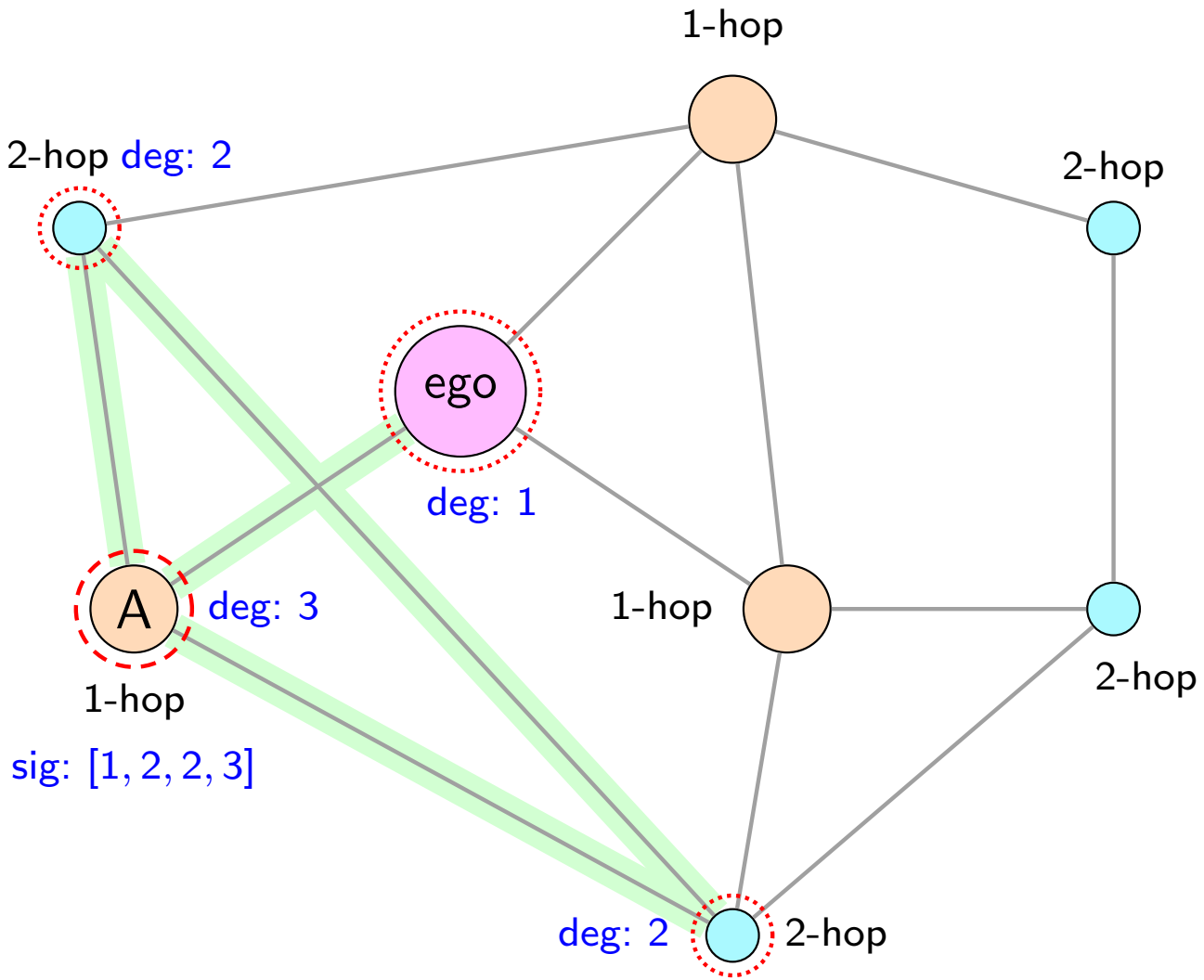
Ad-hoc Attack on Scheme 1: Illustrated



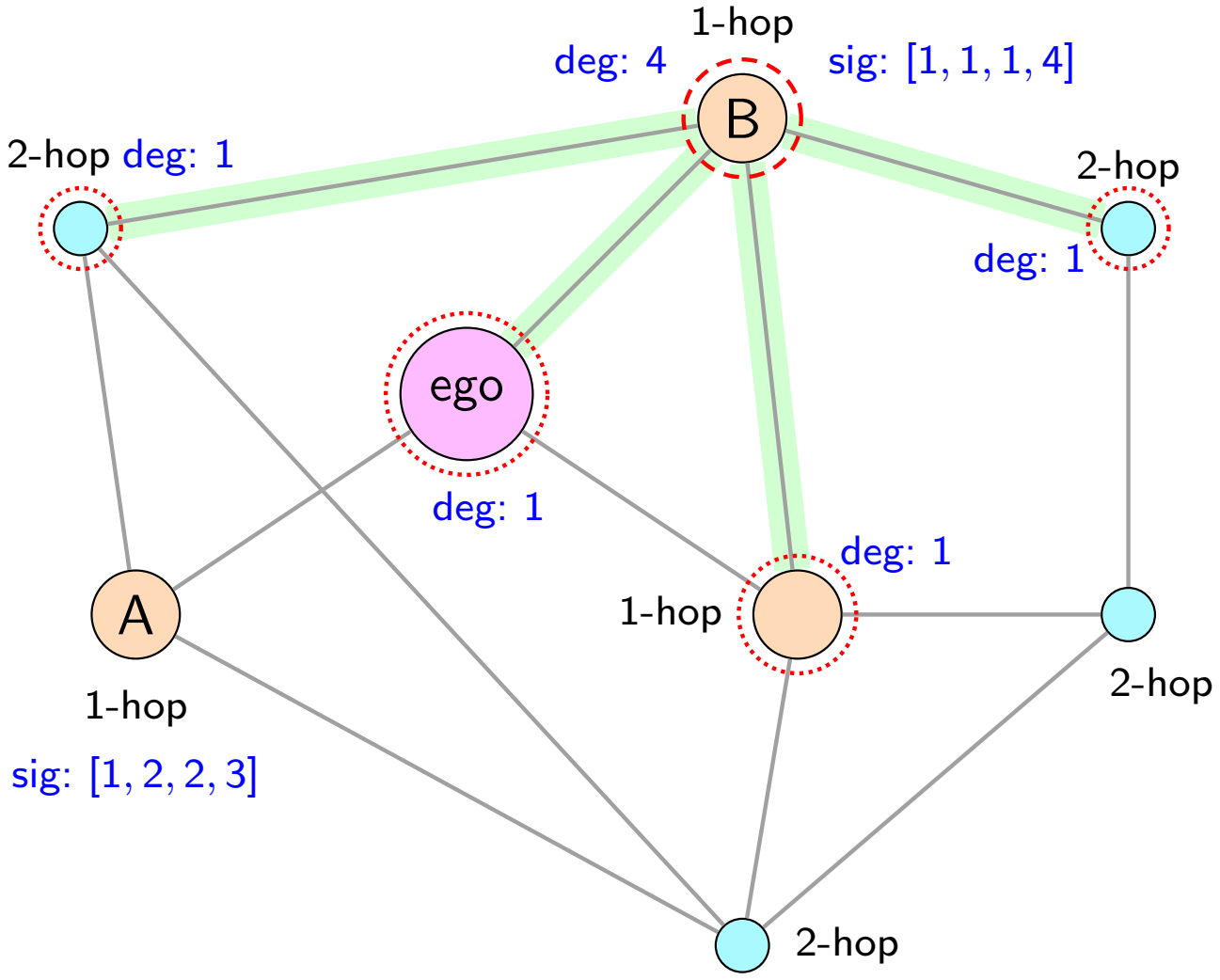
Ad-hoc Attack on Scheme 1: Illustrated



Ad-hoc Attack on Scheme 1: Illustrated



Ad-hoc Attack on Scheme 1: Illustrated



Success Rate: Scheme 1

- **100%** match for identical node pairs (theoretical).
- Over **99.9%** mismatch for non-identical node pairs.

Learning De-anonymization

Security Economics: Attacking a Class of Schemes

- Scheme 2 defeats the ad-hoc attack.
- A piecemeal approach towards de-anonymization does not scale.
- Defeating an instance of anonymization is not generalizable.
- Can we generalize attacks?

A Machine Learning Approach

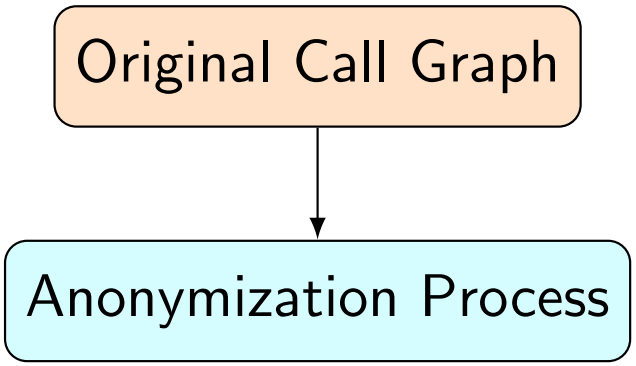
- Traditional approach:
  1. An anonymization strategy is designed.
  2. Manually construct an attack.
  3. Strategy is tweaked.
  4. GO TO 2.

- Machine learning approach:
  1. An anonymization strategy is designed.
  2. Generate training and test data based on the algorithm.
  3. Extract features.
  4. Train the model.
  5. Evaluate the performance

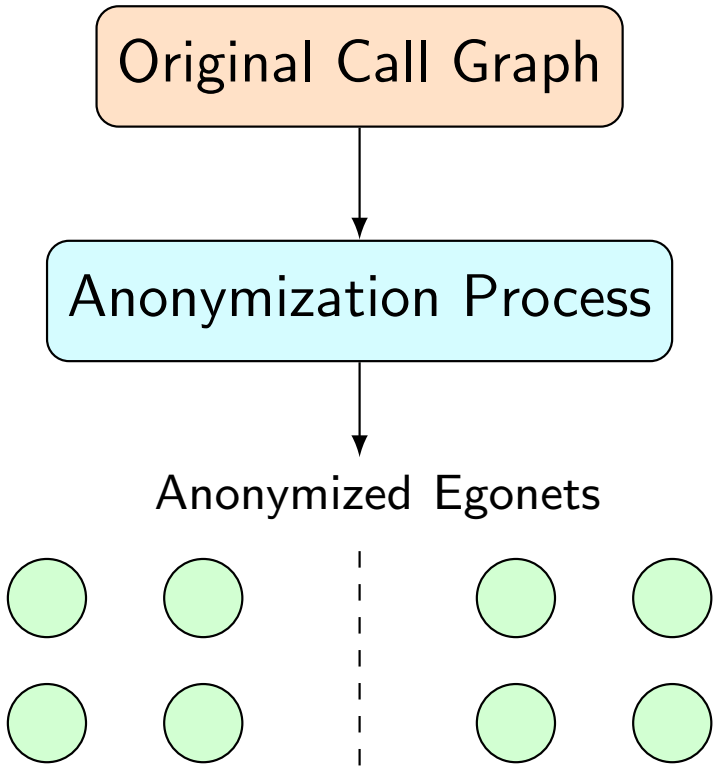
The Model for D4D Learning Task

Original Call Graph

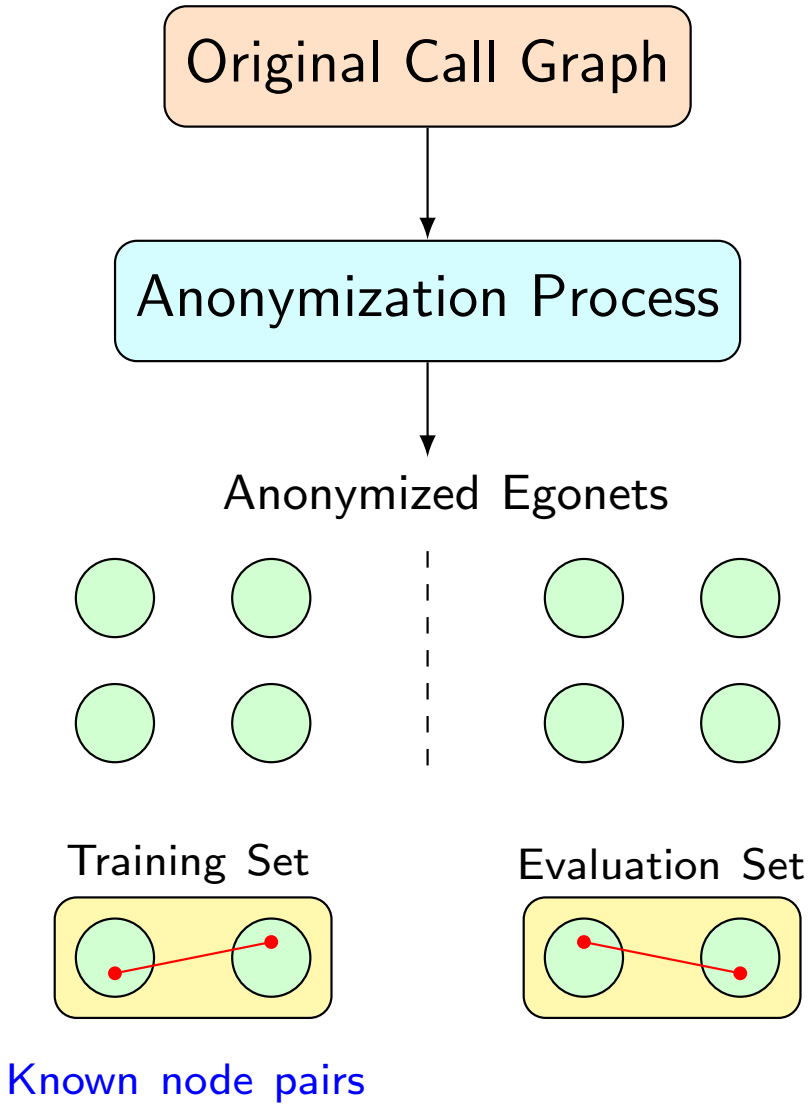
The Model for D4D Learning Task



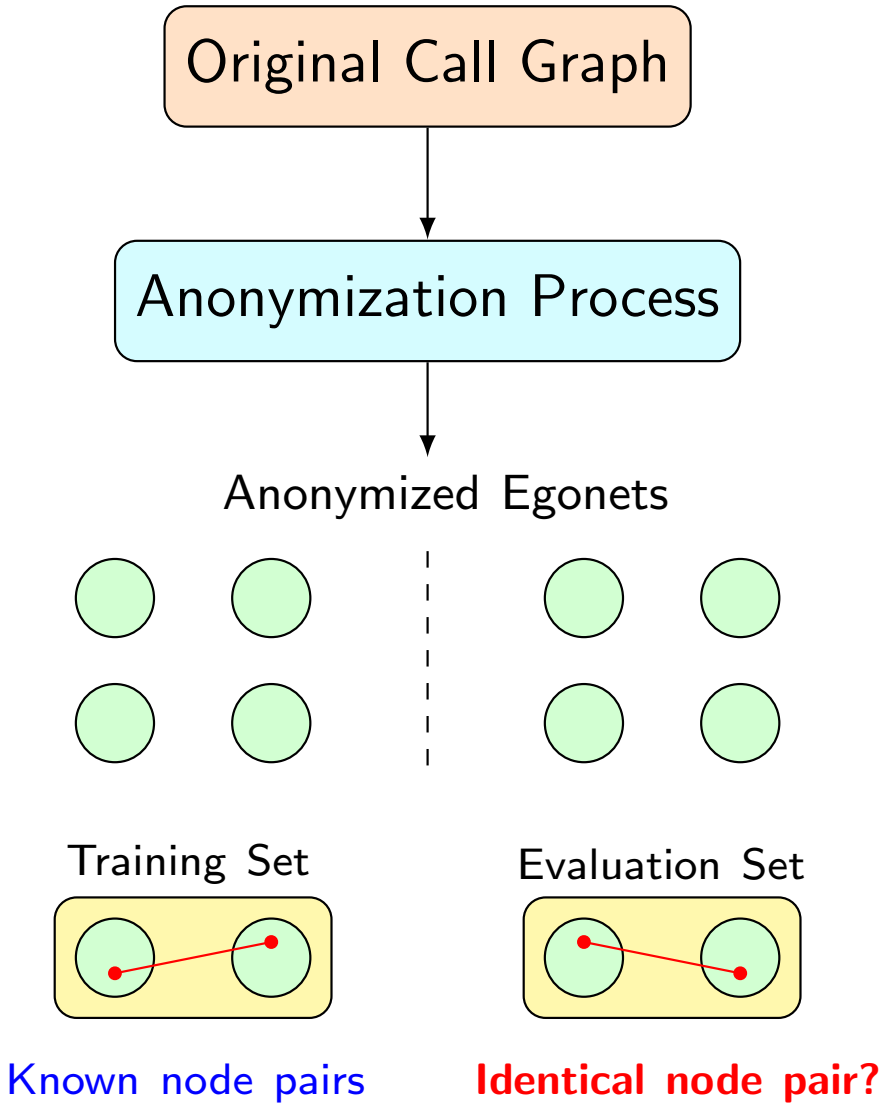
The Model for D4D Learning Task



The Model for D4D Learning Task



The Model for D4D Learning Task



Node Features

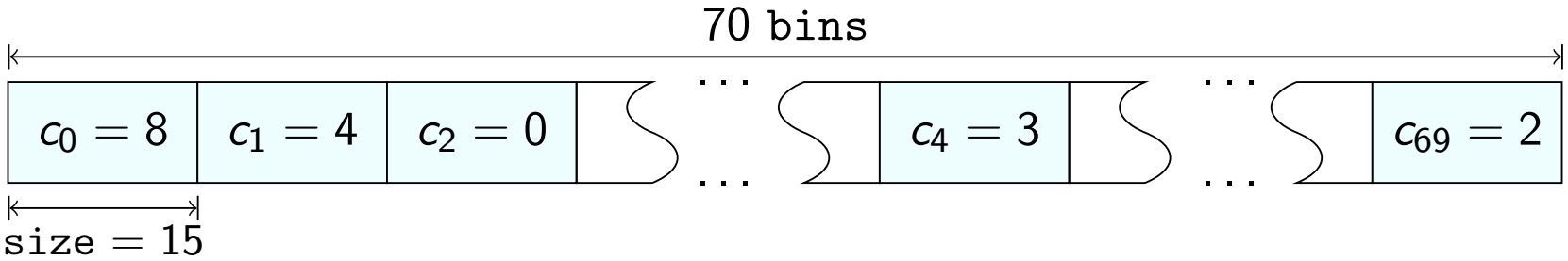
- Must distinguish identical and non-identical node pairs.
- Feature vector purely based on topology (no edge weights or directionality).
- Too generic: high false positives.
- Too specific: low true positives.
- Extend the signature by quantizing it.

Internals: Feature Vector

Feature vector of a node with neighbors of degrees –
$[1, 1, 3, 3, 5, 6, 7, 13, 16, 20, 21, 30, 65, 69, 72, 1030, 1100]$.

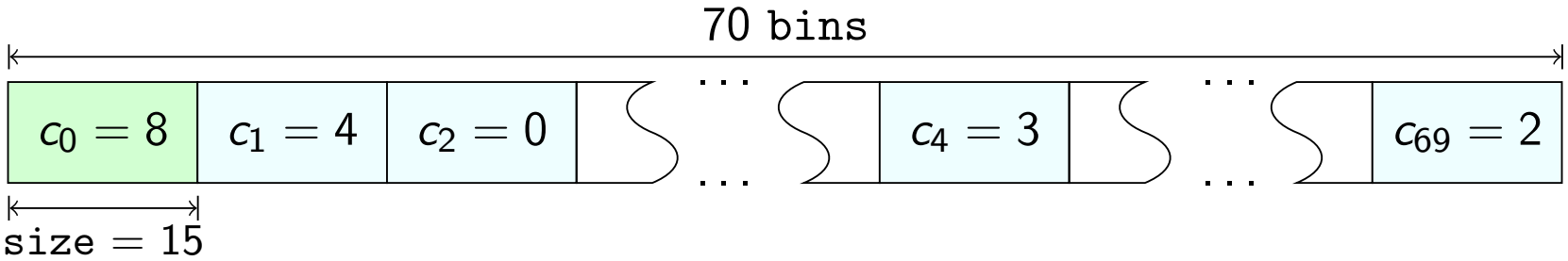
Internals: Feature Vector

Feature vector of a node with neighbors of degrees –
$[1, 1, 3, 3, 5, 6, 7, 13, 16, 20, 21, 30, 65, 69, 72, 1030, 1100]$.



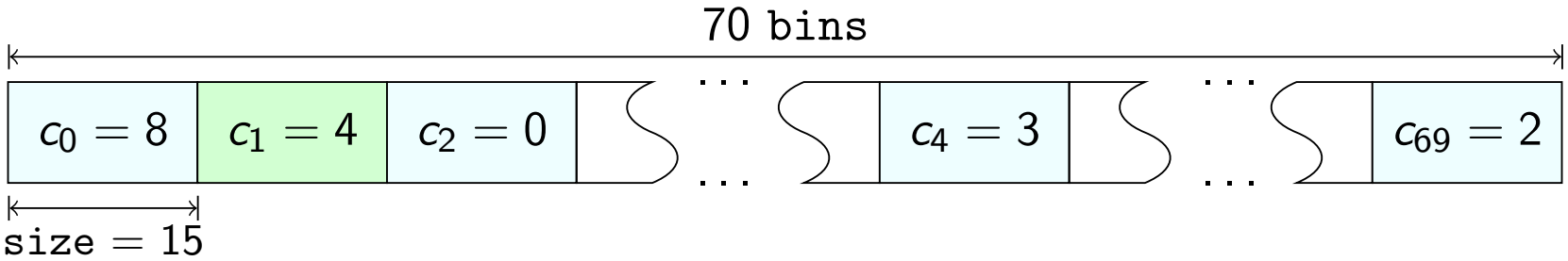
Internals: Feature Vector

Feature vector of a node with neighbors of degrees –
$[1, 1, 3, 3, 5, 6, 7, 13, 16, 20, 21, 30, 65, 69, 72, 1030, 1100]$.



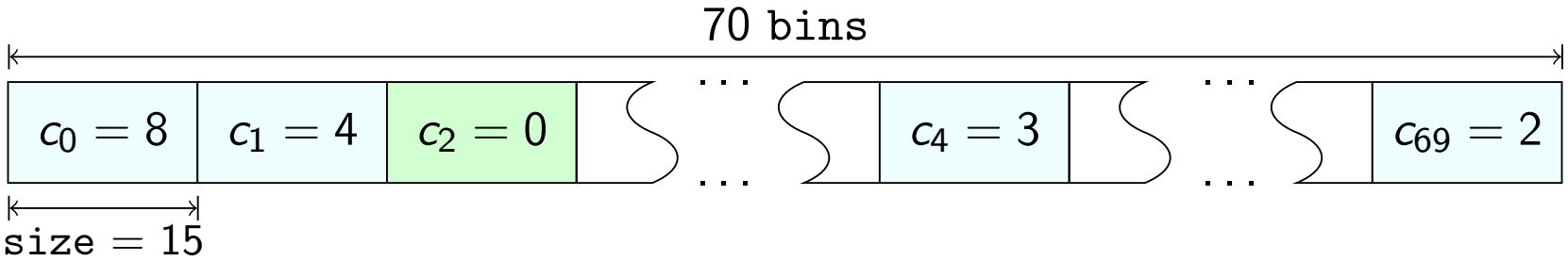
Internals: Feature Vector

Feature vector of a node with neighbors of degrees –
$[1, 1, 3, 3, 5, 6, 7, 13, 16, 20, 21, 30, 65, 69, 72, 1030, 1100]$.



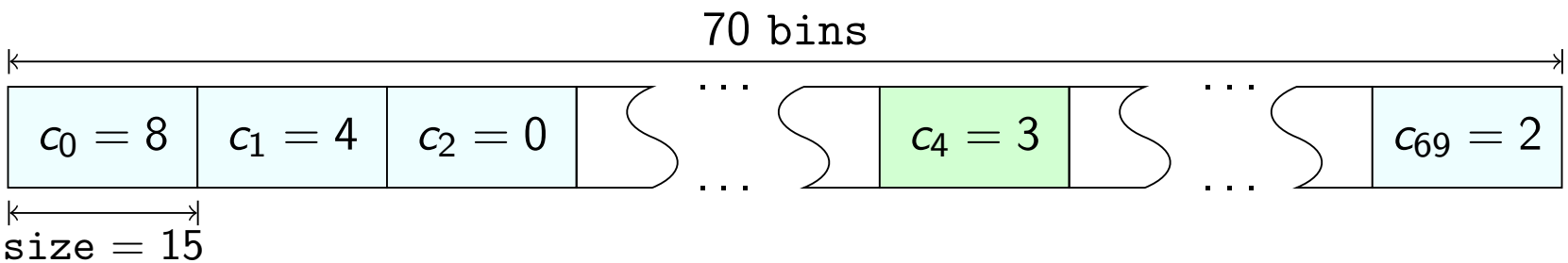
Internals: Feature Vector

Feature vector of a node with neighbors of degrees –
$[1, 1, 3, 3, 5, 6, 7, 13, 16, 20, 21, 30, 65, 69, 72, 1030, 1100]$.



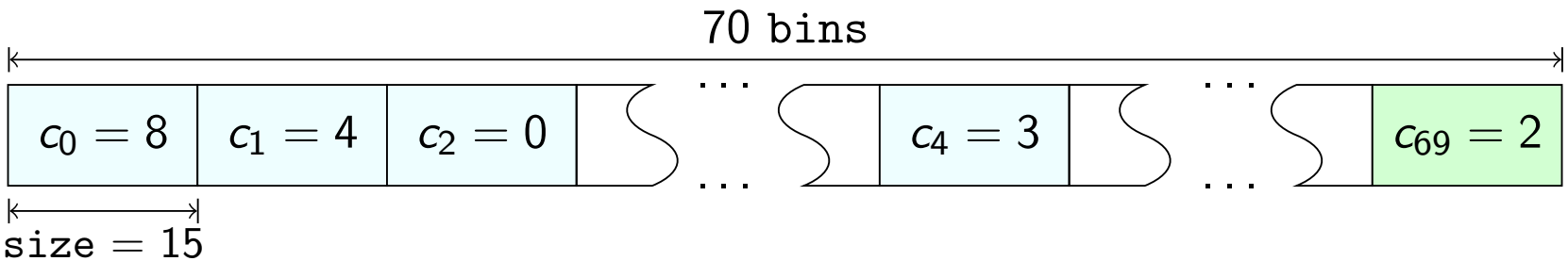
Internals: Feature Vector

Feature vector of a node with neighbors of degrees –
$[1, 1, 3, 3, 5, 6, 7, 13, 16, 20, 21, 30, 65, 69, 72, 1030, 1100]$.



Internals: Feature Vector

Feature vector of a node with neighbors of degrees –
$[1, 1, 3, 3, 5, 6, 7, 13, 16, 20, 21, 30, 65, 69, 72, 1030, 1100]$.



Internals: Random Forest

- 400 trees trained.
- Identical node pair types: 1-hop, 1,2-hop and 2-hop.
- 4 random forests trained: 1 per category + 1 generic
- Prediction: Aggregate the decision of all trees.

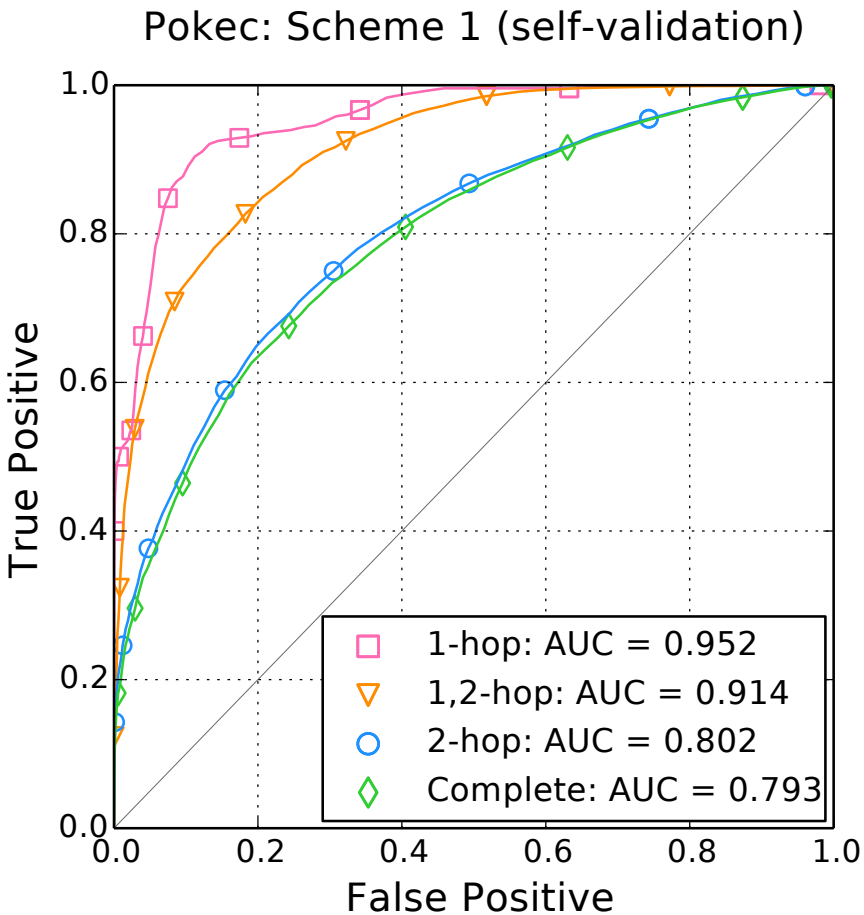
Results

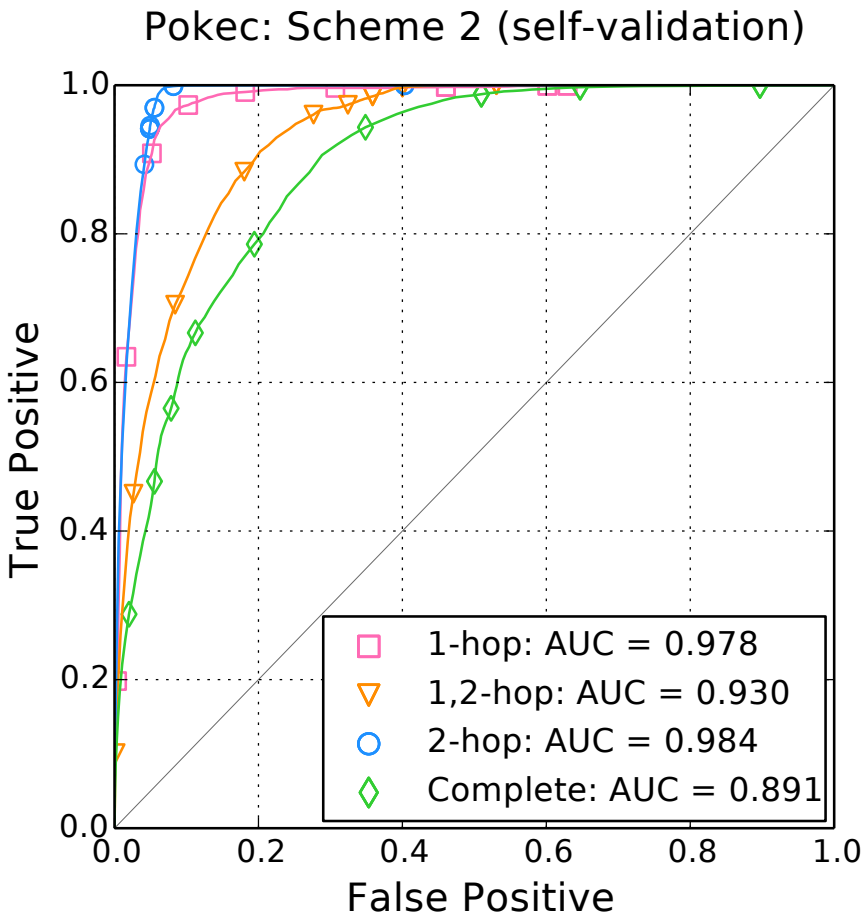
Evaluation: Datasets

- Evaluation does **NOT** use D4D datasets.
  - Ethical concerns
  - Lack of ground truth.

- Publicly available datasets used
  - D4D (5M nodes) – 5000 egonets released.
  - Epinions (75K nodes) – 100 egonets extracted.
  - Pokec (1.6M nodes) – 1000 egonets extracted.

Pokec Dataset: ROC Curves





Pokec: FP vs TP (self-validation)

Scheme 1

False Positive	0.01%	0.1%	1%	10%	25%
1-hop	27.50	42.92	51.04	88.75	93.96
1,2-hop	5.25	11.58	36.16	73.24	88.68
2-hop	0.00	12.55	23.15	49.14	69.96
Complete	0.01	10.44	20.48	47.60	68.36

Scheme 2

False Positive	0.01%	0.1%	1%	10%	25%
1-hop	4.20	16.26	49.89	97.20	99.58
1,2-hop	0.79	6.41	28.32	73.88	94.66
2-hop	1.62	12.12	50.42	99.96	99.99
Complete	0.68	6.12	21.14	64.12	86.10

Pokec: FP vs TP (self-validation)

Scheme 1

False Positive	0.01%	0.1%	1%	10%	25%
1-hop	27.50	42.92	51.04	88.75	93.96
1,2-hop	5.25	11.58	36.16	73.24	88.68
2-hop	0.00	12.55	23.15	49.14	69.96
Complete	0.01	10.44	20.48	47.60	68.36

Scheme 2

False Positive	0.01%	0.1%	1%	10%	25%
1-hop	4.20	16.26	49.89	97.20	99.58
1,2-hop	0.79	6.41	28.32	73.88	94.66
2-hop	1.62	12.12	50.42	99.96	99.99
Complete	0.68	6.12	21.14	64.12	86.10

Claim of Generality

- Random forest uncovers artifacts and invariants of the anonymization algorithm not merely quirks of the input data.
- Learning de-anonymization allows it to attack previously unseen data (x-validation).
- Ideal: training and test distributions are close.
- De-anonymization is successful for a variety of schemes.

Pokec: FP vs TP (x-validation)

Scheme 1

False Positive	0.01%	0.1%	1%	10%	25%
1-hop	19.38	27.29	34.79	57.92	76.25
1,2-hop	2.98	10.10	26.52	70.37	90.72
2-hop	1.71	4.18	18.84	39.12	52.52
Complete	1.89	4.05	16.83	36.81	50.76

Scheme 2

False Positive	0.01%	0.1%	1%	10%	25%
1-hop	2.11	5.40	12.29	28.29	60.26
1,2-hop	0.18	2.08	14.34	49.25	70.76
2-hop	3.02	13.57	45.45	99.80	100.00
Complete	1.00	5.61	19.22	56.90	72.76

Pokec: FP vs TP (x-validation)

Scheme 1

False Positive	0.01%	0.1%	1%	10%	25%
1-hop	19.38	27.29	34.79	57.92	76.25
1,2-hop	2.98	10.10	26.52	70.37	90.72
2-hop	1.71	4.18	18.84	39.12	52.52
Complete	1.89	4.05	16.83	36.81	50.76

Scheme 2

False Positive	0.01%	0.1%	1%	10%	25%
1-hop	2.11	5.40	12.29	28.29	60.26
1,2-hop	0.18	2.08	14.34	49.25	70.76
2-hop	3.02	13.57	45.45	99.80	100.00
Complete	1.00	5.61	19.22	56.90	72.76

Concluding Remarks

- What TP rate is acceptable?
- What rate of de-anonymization is secure?
- Lower bound on attack performance but cheaper evaluations.
- What are the definitive set of features?

Summary

- Ad-hoc attack works but limited.
- Better: Construct attacks by using machine learning.
- Generic: Attack works even on learning from a different dataset.

Contact

An Automated Social Graph De-anonymization Technique

Kumar Sharad
University of Cambridge, UK
kumar.sharad@cl.cam.ac.uk

George Danezis
University College London, UK
g.danezis@ucl.ac.uk

ABSTRACT

We present a generic and automated approach to re-identifying nodes in anonymized social networks which enables novel anonymization techniques to be quickly evaluated. It uses machine learning (decision forests) to matching pairs of nodes in disparate anonymized sub-graphs. The technique uncovers artefacts and invariants of any black-box anonymization scheme from a small set of examples. Despite a high degree of automation, classification succeeds with significant true positive rates even when small false positive rates are sought. Our evaluation uses publicly available real world datasets to study the performance of our approach against real-world anonymization strategies, namely the schemes used to protect datasets of The Data for Development (D4D) Challenge. We show that the technique is effective even when only small numbers of samples are used for training. Further, since it detects weaknesses in the black-box anonymization scheme it can re-identify nodes in one social network when trained on another.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*; C.2.0 [**Computer-communications networks**]: General—*Security and protection*

Keywords

Privacy; de-anonymization; social networks; machine learning

1. INTRODUCTION

A number of rich datasets have recently been published for research purposes, often with only casual attempts to anonymize them. Research in de-anonymization has also seen an upswing [1, 13, 14, 19], leading to high profile data releases being followed by high profile privacy breaches.

These developments have forced organizations to make some effort to anonymize the released data. However, overly distorting data to achieve this contradicts the very purpose of a release, since it negatively impacts utility.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WPES'14, November 3, 2014, Scottsdale, Arizona, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2948-7/14/11 ...$15.00.
http://dx.doi.org/10.1145/2665943.2665960.

Social network graphs in particular are high dimensional and feature rich data sets, and it is extremely hard to preserve their anonymity. Thus, any anonymization scheme has to be evaluated in detail, including those with a sound theoretical basis [11]. Techniques have been proposed to resist de-anonymization [8, 17, 22], however, Dwork and Naor have shown [7] that preserving privacy of an individual whose data is released cannot be achieved in general.

Ad-hoc vs generic. It has been conclusively demonstrated that merely removing identifiers in social network datasets is not sufficient to guarantee privacy. Despite these results, data practitioners, continue to propose anonymization strategies in the hope that they can resist de-anonymization "in practice", such as the ones used to protect datasets from The Data for Development (D4D) challenge. This has lead to a *cat-and-mouse game*: Research thus far has focused on defeating new variants of anonymization techniques by manually devising ad-hoc de-anonymization techniques. Despite their simplicity, unraveling each anonymization technique manually requires considerable effort and time and each attack can be defeated by a small tweak to the anonymization strategy, often by destroying specific features on which the attack has been constructed. Tailoring attacks to specific scenarios [15] highlights the problem of anonymization but the expense involved in evaluating each new scheme cannot be amortized.

Better solutions are needed which attack entire classes of anonymization schemes rather than taking a piecewise approach. Such generic de-anonymization techniques will allow cheap and timely evaluation of novel anonymization schemes. In this paper, we demonstrate the efficacy of automated de-anonymization attacks on real-world anonymization schemes. They automatically uncover artefacts remaining after anonymization that allow for re-identification of nodes in social networks. The automated attacks can be used quickly and cheaply to demonstrate that a non negligible number of users would be at risk of de-anonymization for "novel" anonymization schemes. Specifically, we:

- Formulate the problem of de-anonymization in social networks as a learning task. From a set of examples of known correspondences between nodes (*training* data) we wish to learn a good de-anonymization model (Section 3.1).
- Describe a non-parametric learning algorithm tailored to the de-anonymization learning problem in social graphs. The algorithm is based on random decision forests, with custom features that match social network nodes (Sections 3.2-3.4).
- Evaluate the learning algorithm on a real-world de-anonymization task from the D4D challenge (Sections 4.1,4.2), and compare it with an ad-hoc approach (Section 4.4).
- Show that the algorithm and model learn sufficient information about the anonymization algorithm, rather than the spe-

Paper: research.ksharad.com

Authors

- Kumar Sharad
  - ✉ kumar.sharad@cl.cam.ac.uk
  - ⌂ ksharad.com
- George Danezis
  - ✉ g.danezis@ucl.ac.uk
  - ⌂ cs.ucl.ac.uk/staff/G.Danezis