De-anonymization

De-anonymization

A. Narayanan, V. Shmatikov, *De-anonymizing Social Networks.* International Symposium on Security and Privacy, 2009

Goal

Demonstrate the feasibility of large-scale, passive de-anonymization of real-world social networks

Anonymization and Privacy

- Suppression (anomymization) often misinterpreted as removal of "personally identifiable information" (PII)

- the EU privacy directive defines "personal data" as:
as "any information relating to an identified or identifiable natural person [. . .]; an *identifiable person* is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity"

- After a New York court ordering Google to hand over viewing data of over 100 million YouTube users to Viacom, revised agreement to anonymize

- The CEO of NebuAd (a U.S. company that offers targeted advertising based on browsing histories gathered from ISPs): "We don't have any raw data on the identifiable individual. Everything is anonymous".

- Phorm, a similar U.K. company collect the data on Web-surfing habits of 70% of British broadband users; only privacy protection user identities are mapped to random identifiers

Roadmap

1. survey the current state of data sharing in social networks,

2. develop a generic re-identification algorithm for anonymized social networks,

3. apply the de-anonymization algorithm to Flickr and Twitter and show that:

   "a third of the users who are verifiable members of both Flickr and Twitter can be recognized in the completely anonymous Twitter graph with only 12% error rate, even though the overlap in the relationships for these members is less than 15%"

Survey: Applications I

Academic and government data-mining

Phone-networks

- (in published papers) mobile-phone call graphs of, 7, 3, and 2.5 million customers, the land-line phone graph of 2.1 million Hungarian users

- Corporations like AT&T, database of 1.9 trillion phone calls decades back
3,000 wireless companies in the U.S.

- commonly used to detect illicit activity (e.g., calling fraud) and for national security (e.g., identifying the command-and-control structures of terrorist cells by sub-network topologies)

Sociologists, epidemiologists, and health-care professionals:

data about geographic, friendship, family, and sexual networks to study disease propagation and risk.
E.g, AddHealth dataset: the sexual-relationship network of almost 1,000 students of anonymous Midwestern high school (published in an anonymized form)

For online social networks,

- collected by crawling either via an API, or "screen-scraping"
- anonymized graphs available by request only

In some online networks, (e.g., LiveJournal, the Experience Project), user profiles and relationship data are public, but many users maintain pseudonymous profiles

Survey: Applications II

Advertising

- concrete evidence that social-network data makes commerce much more profitable

- network operators increasingly sharing their graphs with advertising partners to enable better social targeting of advertisements.

  For example, Facebook explicitly says that users' profiles may be shared for the purpose of personalizing advertisements and promotions, as long as the individual is not explicitly identified

- Both Facebook and MySpace allow advertisers to use friends' profile data for ad targeting

- Social-network-driven advertising pursued by many startups, even Google, typically relying on anonymity to prevent privacy breaches

Survey: Applications III

Third-party applications.

on Facebook alone in the tens of thousands and rapidly growing

data from multiple applications can be aggregated and used for targeted advertising (e.g., as done by SocialMedia)

on the Ning platform (a platform for cerating social websites), over 275,000 networks, each can be considered a third-party application.

data given to third-party applications usually not anonymized,
poor privacy preservetion

- a security hole in a Facebook application developed by Slide, Inc. “exposed the birthdays, gender, and relationship status of strangers, including Facebook executives, [and] the wife of Google co-founder Larry Page”

- WidgetLaboratory, one of the most popular developers for the Ning platform, was banned permanently after “gathering credentials from users and otherwise creating havoc on Ning networks”

Survey: Applications IV

Aggregation from many social networks

facilitated by projects such as OpenID, DataPortability, the “social graph” project, various microformats

Existing aggregators include FriendFeed, MyBlogLog, Jaiku (recently acquired by Google), and Plaxo; the latter even provides an open-source “social graph crawler”

an excellent source of auxiliary information for attacks.

Other data-release scenarios.

WellNet

In “friend-to-friend networking,” a peer-to-peer file-sharing network is overlaid on social links to defeat censor nodes such as the RIAA.

photographs published online - accuracy of face recognition improved by the fact that users who appear together in photographs are likely to be neighbors in the social network

State-of-the-Art

all k-isomorphic neighborhoods have the same value of some sensitive attribute

active attacks:

(1) restricted to online social networks

(2) little control over the incoming edges to the nodes it creates

a subgraph with no incoming edge will stand out

5.3 million nodes + 77 million edges

7-node subgraphs containing Hamiltonian paths

(3) many OSNs require a link to be mutual before information is available in any form

New type of Attack

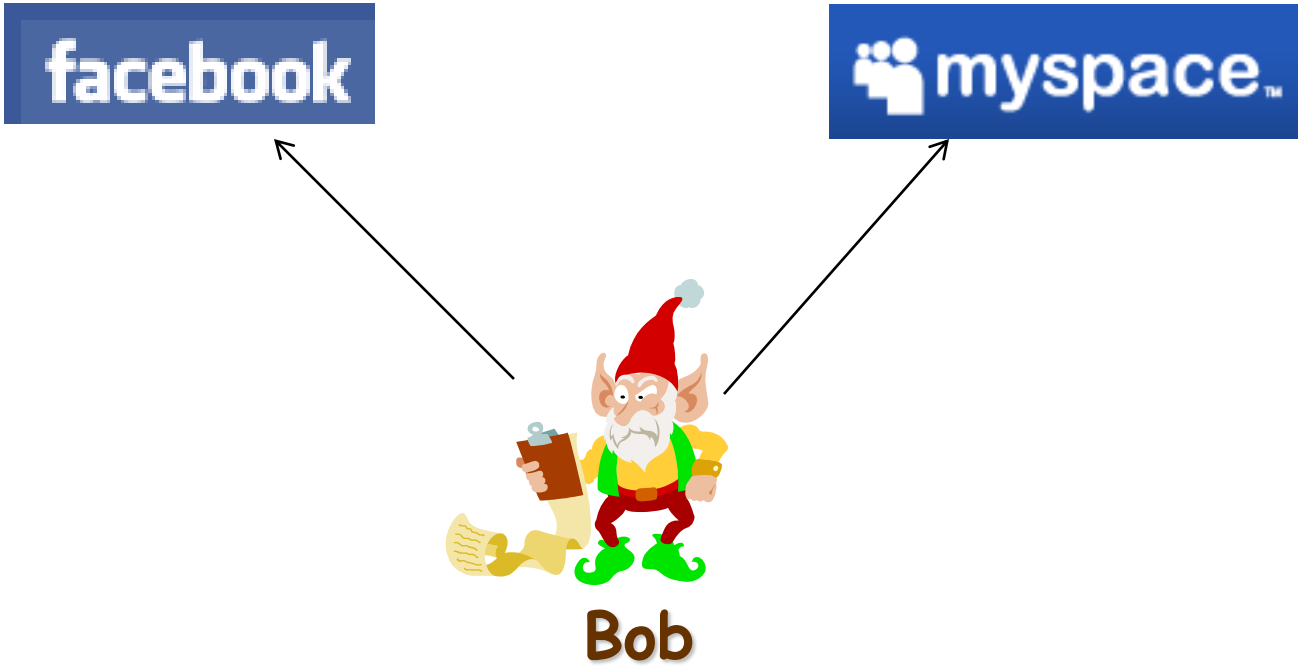
So far background information local (restricted to the neighborhood a a node)

Global information:

Background knowledge:

another social network with partially overlapping membership

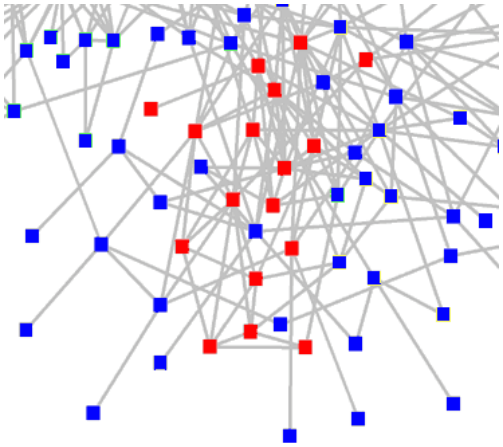
New type of Attack



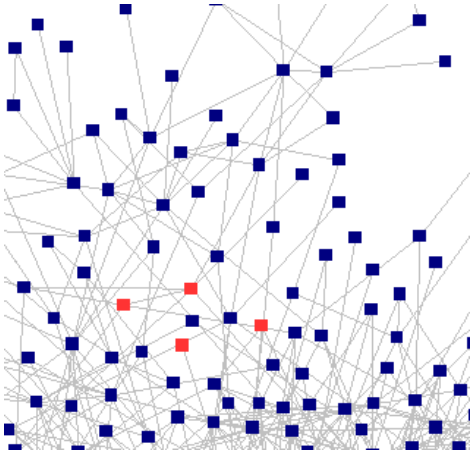
Bob has accounts in both facebook and myspace

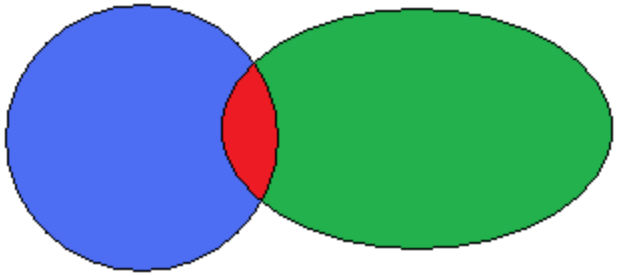
overlap

Facebook graph



Myspace graph





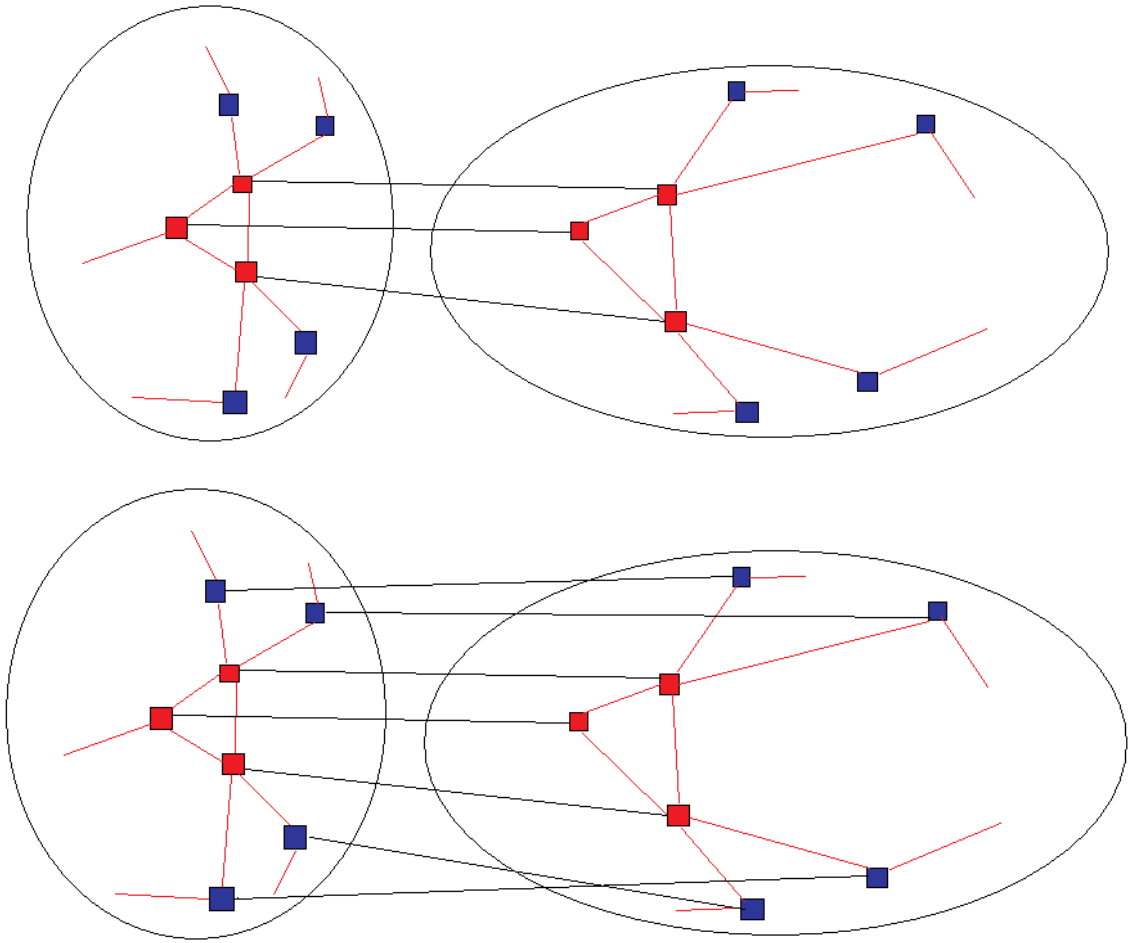
Goal

Given the auxiliary information, (nodes, edges) about one social network (auxiliary network) and a small number of members of target network, the attacker wants to learn sensitive information about other members of target network.

partial mapping -> learn the rest

Facebook graph

Myspace graph



Model and Definitions

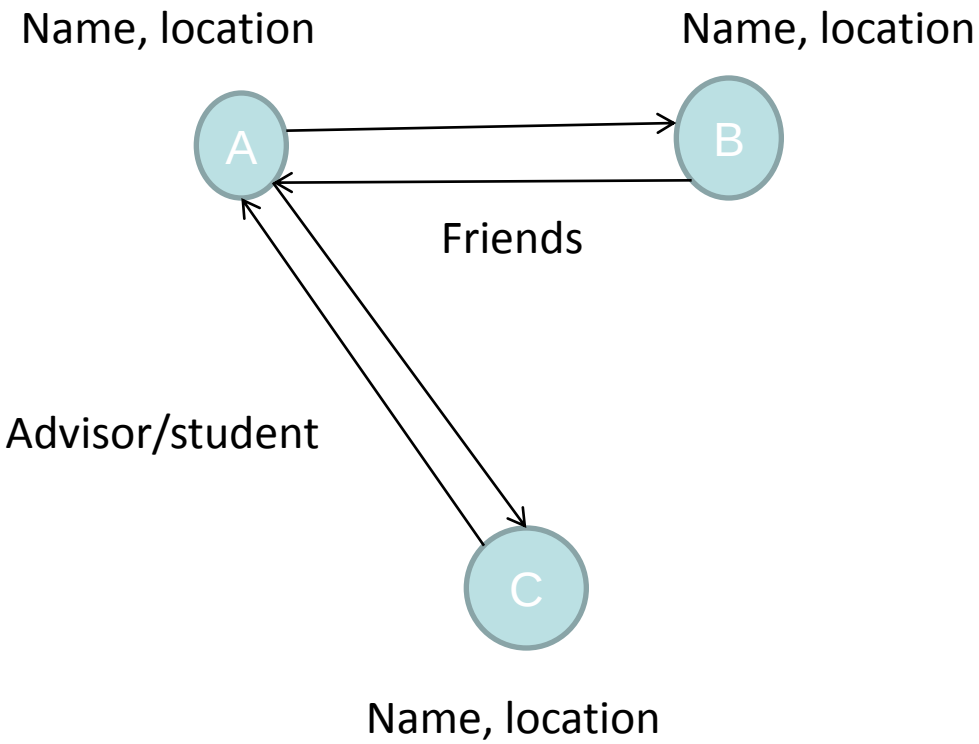
Model: Social Network

A social network S consists of:
(1) a directed graph G = (V; E), and
(2) a set of attributes (i.e., labels) *X* for each node in V and a set of attributes *Y* for each edge in E

- Attributes take atomic values from a discrete domain (real-valued attributes must be discretized)
  - edges as attributes in Y with values in {0, 1}

- Implicit attributes, i.e., properties of a node or an edge based purely on the graph structure (e.g., node degree)

- Implicit attributes may be leaked without disclosing any explicit attributes


Name, location
Name, location
A
B
Friends
Advisor/student
C
Name, location

Data Released

What type of data are released?

- Advertisers often given access to the entire graph in a (presumably) anonymized form and a limited number of relevant attributes for each node.
- Application developers get access to a subgraph via user opt-in and most or all of the attributes within this subgraph (including the identifying attributes)
- Researchers may receive the entire graph or a subgraph (up to the discretion of the network owner) and a limited set of non-identifying attributes.

"Anonymization" is modeled by publishing only **a subset of attributes**

No distinction between identifying and non-identifying attributes

Suppressed attributes not limited to the demographic quasi-identifiers a priori; simply assume that the published attributes by themselves are insufficient for re-identification.

The released graph: Ssan

Data release may involve perturbation or sanitization that changes the graph structure in some way to make re-identification attacks harder

Model as follows:

1. Select a subset of nodes, Vsan $\subset$ V , and subsets Xsan $\subseteq$ X; Ysan $\subseteq$ Y of node and edge attributes to be released.

2. Compute the induced subgraph on Vsan.

   For simplicity, no complex criteria for releasing edge, e.g., based on edge attributes.

3. Remove some edges and add fake edges.

4. Release Ssan = (Vsan;Esan); {X(v)$\forall$v $\in$ Vsan;X $\in$ Xsan}; {Y (e)$\forall$e $\in$ Esan; Y $\in$ Ysan})

The attackers

Threat Model I

Attack scenarios.

Attackers into different categories depending on capabilities and goals.

The strongest adversary - a government-level agency interested in global surveillance.

- Assumed to have access to a large auxiliary network Saux
- Objective large-scale collection of detailed information about as many individuals as possible
- Aggregating the anonymous network Ssan with Saux by recognizing nodes that correspond to the same individuals.

Abusive marketing.

- Obtain an anonymized social-network graph from the network operator for advertising purposes.
- Using publicly available data, engage in abusive marketing aimed at specific individuals.
- Phishing and spamming (craft a highly individualized, believable message)

Threat Model II

targeted deanonymization of specific individuals by stalkers, investigators, nosy colleagues, employers, or neighbors.

- attacker has detailed contextual information about a single individual, e.g., some of her attributes, a few of her social relationships, membership in other networks, and so on.

- objective is to use this information to recognize the victim's node in the anonymized network and to learn sensitive information including social relationships in that network.

The Attacker

In addition to the anonymized, sanitized target network Ssan,

the attacker has access to a different network Saux whose membership partially overlaps with S.

How?

- possible to extract Saux directly from original S:
  - e.g.,, parts of some online networks can be automatically crawled, or a malicious third-party application can provide information about the subgraph of users who installed it.

- the attacker may collude with an operator of a different network whose membership overlaps with S.

- the attacker may take advantage of several ongoing aggregation projects

- government-level aggregators, such as intelligence and law enforcement agencies, can collect data via surveillance and court-authorized searches.

- the nodes of Saux may be a subset, a superset, or overlap with those of the target network.

NOTE

consider edge relationship to be a binary attribute in Y and all edge attributes $Y \in \mathcal{Y}$ to be defined over V^2 instead of E.

If $(u; v) = \notin E$, then $Y [u; v] = \perp \forall Y \in \mathcal{Y}$.

The Attacker

Saux: a graph Gaux = {Vaux;Eaux} and a set of probability distributions AuxX and AuxY, one for each attribute of every node in Vaux and each attribute of every edge in Eaux.

- Represent the adversary's (imperfect) knowledge of the corresponding attribute value.
  - For example, the adversary may be 80% certain that an edge between two nodes is a "friendship" and 20% that it is a mere "contact."

- Since edges as attributes, also captures the attacker's uncertain knowledge about the existence of individual edges.

- Does not capture some types of auxiliary information, such as "node v1 is connected to either node v2, or node v3."

For an attribute X of a node v (respectively, attribute Y of an edge e), we represent by Aux[X; v] (resp., Aux[Y; e]) the attacker's prior probability distribution (i.e., distribution given by his auxiliary information) of the attribute's value. The set AuxX (resp., AuxY) can be thought of as a union of Aux[X; v] (resp., Aux[Y; e]) over all attributes and nodes (resp., edges).

The Attacker

Detailed information about a very small number of members of the target network S (seeds).

can determine if these members are also present in Saux (e.g., by matching usernames and other contextual information).

how to collect such data:

- If the attacker is already a user of S, knows details about his own node and its neighbors
- Some networks permit manual access to profiles even if large-scale crawling is restricted (e.g., Facebook allows viewing of information about “friends” of any member by default.)
- Some users may make their details public even in networks that keep them private by default.
- The attacker may even pay a handful of users for information about themselves and their friends, or learn it from compromised computers or stolen mobile phones.
- With an active attack, the attacker may create fake nodes and edges in S with features easy to recognize in the anonymized version of S, such as a clique or an almost-clique.

Breaching Privacy

Privacy Policy: labeling that specifies for every node attribute, edge, and edge attribute whether it should be public or private.

$$PP: \mathcal{X} \cup \mathcal{Y} \times E \rightarrow \{pub, priv\}$$

Focus on node re-identification.

Ground truth: a mapping μG between the nodes of Vaux and the nodes of Vsan

a pair of nodes are mapped to each other if they belong to the same “entity”.

- 1-1 mapping
- If μG(v) = ⊥, no mapping for node v (e.g., if v was not released as part of Vsan).
- μG need not map every node in Vsan.

Node re-identification or re-labeling: finding a mapping μG between a node in Vaux and a node in Vsan.

Definition 1 (Re-identification algorithm): A node re-identification algorithm takes as input S_{san} and S_{aux} and produces a probabilistic mapping $\tilde{\mu}: V_{\text{san}} \times (V_{\text{aux}} \cup \{\perp\}) \rightarrow [0, 1]$, where $\tilde{\mu}(v_{\text{aux}}, v_{\text{san}})$ is the probability that v_{aux} maps to v_{san}.

use this mapping to derive a distribution of the attributes (labels) of the nodes and edges

Definition 2 (Mapping adversary): A mapping adversary corresponding to a probabilistic mapping $\tilde{\mu}$ outputs a probability distribution calculated as follows:

$$\mathsf{Adv}[X, v_{\mathsf{aux}}, x] = \frac{\sum_{v \in V_{\mathsf{san}}, X[v]=x} \mu(v_{\mathsf{aux}}, v)}{\sum_{v \in V_{\mathsf{san}}, X[v] \neq \perp} \mu(v_{\mathsf{aux}}, v)}$$

$$\mathsf{Adv}[Y, u_{\mathsf{aux}}, v_{\mathsf{aux}}, y] = \frac{\sum_{u,v \in V_{\mathsf{san}}, Y[u,v]=y} \tilde{\mu}(u_{\mathsf{aux}}, u)\tilde{\mu}(v_{\mathsf{aux}}, v)}{\sum_{u,v \in V_{\mathsf{san}}, Y[u,v] \neq \perp} \tilde{\mu}(u_{\mathsf{aux}}, u)\tilde{\mu}(v_{\mathsf{aux}}, v)}$$

Adv is defined if there is a non-zero number of nodes v ∈ Vsan such that μ((vaux, v) > 0 and X[v] ≠⊥.

can be computed in other ways, e.g., by looking only at the most likely mapping

- We say that privacy of vsan is compromised if, for some attribute X which takes value x in Ssan and is designated as "private" by the privacy policy, the adversary's belief that X[vaux] = x increases by more than δ, which is a pre-specified privacy parameter.

- For simplicity, we assume that the privacy policy PP is global, i.e., the attribute is either public, or private for all nodes (respectively, edges).

Definition 3 (Privacy breach): For nodes $u_{\mathsf{aux}}, v_{\mathsf{aux}} \in V_{\mathsf{aux}}$, let $\mu_G(u_{\mathsf{aux}}) = u_{\mathsf{san}}$ and $\mu_G(v_{\mathsf{aux}}) = v_{\mathsf{san}}$. We say that the privacy of v_{san} is breached w.r.t. adversary Adv and privacy parameter δ if

(a) for some attribute X such that $\mathsf{PP}[X] = \mathsf{priv}$, $\mathsf{Adv}[X, v_{\mathsf{aux}}, x] - \mathsf{Aux}[X, v_{\mathsf{aux}}, x] > \delta$ where $x = X[v_{\mathsf{aux}}]$, or
(b) for some attribute Y such that $\mathsf{PP}[Y] = \mathsf{priv}$, $\mathsf{Adv}[Y, u_{\mathsf{aux}}, v_{\mathsf{aux}}, y] - \mathsf{Aux}[Y, u_{\mathsf{aux}}, v_{\mathsf{aux}}, y] > \delta$ where $y = Y[u_{\mathsf{aux}}, v_{\mathsf{aux}}]$.

What proportion of entities that are active in a social network and for which non-trivial auxiliary information is available can be re-identified?

degree centrality: each node is weighted in proportion to its degree

Definition 4 (Success of de-anonymization): Let $V_{\mathsf{mapped}} = \{v \in V_{\mathsf{aux}} : \mu_G(v) \neq \perp\}$. The *success rate* of a de-anonymization algorithm outputting a probabilistic mapping $\tilde{\mu}$, w.r.t. a centrality measure ν, is the probability that μ sampled from $\tilde{\mu}$ maps a node v to $\mu_G(v)$ if v is selected according to ν:

$$\frac{\sum_{v \in V_{\mathsf{mapped}}} \mathsf{PR}[\mu(v) = \mu_G(v)]\nu(v)}{\sum_{v \in V_{\mathsf{mapped}}} \nu(v)}$$

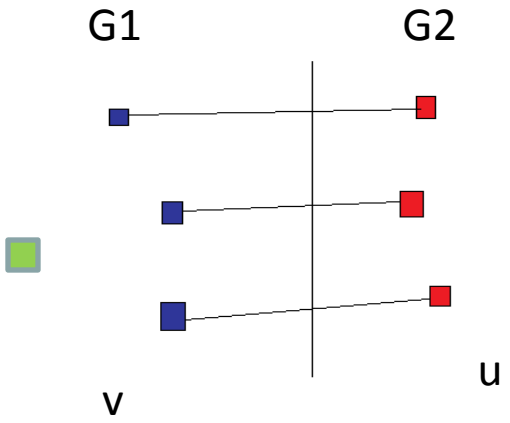
The *error rate* is the probability that μ maps a node v to any node other than $\mu_G(v)$:

$$\frac{\sum_{v \in V_{\mathsf{mapped}}} \mathsf{PR}[\mu(v) \neq \perp \wedge \mu(v) \neq \mu_G(v)]\nu(v)}{\sum_{v \in V_{\mathsf{mapped}}} \nu(v)}$$

- Accuracy(acc) on mappings

$$\text{acc} = \frac{\sum_{v \in V_{\text{mapped}}} \text{PR}[\mu(v) = \mu_G(v)]\nu(v)}{\sum_{v \in V_{\text{mapped}}} \nu(v)}$$

$$\text{error} = \frac{\sum_{v \in V_{\text{mapped}}} \text{PR}[\mu(v) \neq \perp \wedge \mu(v) \neq \mu_G(v)]\nu(v)}{\sum_{v \in V_{\text{mapped}}} \nu(v)}$$



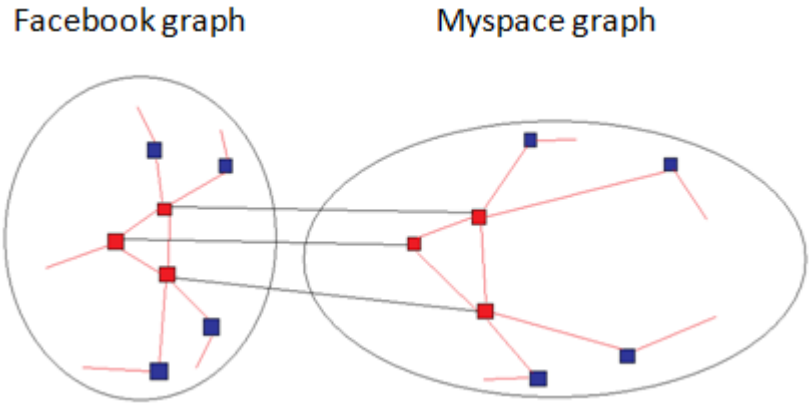
Algorithm

2 stages

1. Seed identification
2. Propagation

Seed identification

- Identify a small number of seed nodes in both target graph and auxiliary graph, and map them to each other
- assumes a clique of k nodes in both graphs
- suffices to know the degree of each node and the number of common neighbors for each pair



Seed identification

Input:

1. the target graph
2. k seed nodes in the auxiliary graph
3. k nodes' information, such as, degree values and pairs of common-neighbor counts
4. Error parameter ε

Method

Search the target graph for a unique k-clique with matching (within a factor of $1 \pm \varepsilon$) nodes degrees and common-neighbor counts

Output

If found, it maps the nodes in the clique to the corresponding nodes in the auxiliary graph

Seed identification

- does not guarantee a unique k-clique in target graph.
- the running time is exponential in k.
  - Once we find a matched clique in target graph, stop searching

Propagation

Input

1. G1(V1, E1)
2. G2(V2, E2)
3. A partial “seed” mapping between the two.

No distinction which is the auxiliary graph or the target graph

Output

mapping μ, focus on deterministic

Propagation

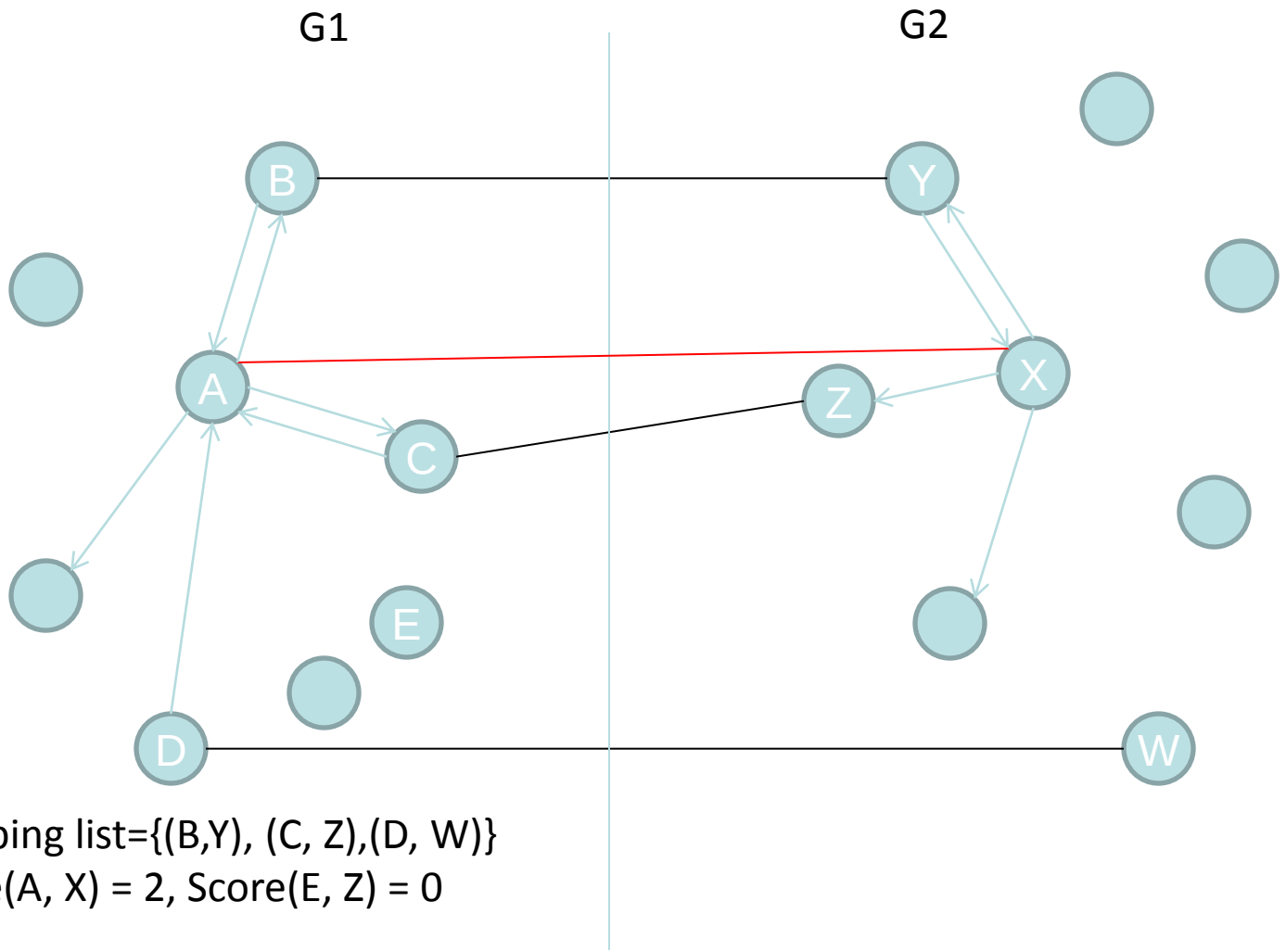
The algorithm finds new mappings using the topological structure of the network and the feedback from previously constructed mappings.

At each iteration,

the algorithm starts with the list of mapped pairs between V1 and V2.
Picks an arbitrary unmapped node u in V1 and computes a score for each unmapped node v in V2, equal to the number of neighbors of u that have been mapped to neighbors of v.
If the strength of the match is above a threshold,

the mapping between u and v is added to the list, and the next iteration starts.

Propagation



Mapping list={(B,Y), (C, Z),(D, W)}
Score(A, X) = 2, Score(E, Z) = 0

Propagation

We will get lots of scores such as, score(A, P), score(A, Q), score(A, R) and so on where P, Q and R are nodes in G2 and A is in G1. Which mapping should we keep?

Additional details

1. **Edge directionality**. Score = incoming edge score + outgoing edge score.

2. **Node degrees**. (to compensate for the bias towards high degree nodes) Score(u, v_i)=score(u, v_i)/√degree of v_i

3. **Eccentricity.** It measures how much an item in a set X "stands out" from the rest.

where max and max2 denote the highest and second highest values, r
espectively, and σ denotes the standard deviation.

If $\frac{\max(X) - \max_2(X)}{\sigma(X)}$ > θ, keep the mapping; otherwise, it is rejected, where θ is a parameter.

4. Does not matter whether G1 is the target and G2 is the auxiliary, each time u is mapped to v, **switch the input graph**, if v gets mapped back to u, the mapping is retained; otherwise, it is rejected.
5. **Revisiting nodes**. As the number of mapped nodes increases, we need to revisit already mapped nodes.
6. Do the iteration until convergence.

Propagation

- Complexity

  O((|E1|+|E2|)*d1*d2) where d1 is a bound on the degree of nodes in G1. D2 is a bound on the degree of nodes in G2.
- Without revisiting nodes and reverse matches

  O(|E1|*d2)
- Without reverse matches

  O(|E1|*d1*d2)

Experiment

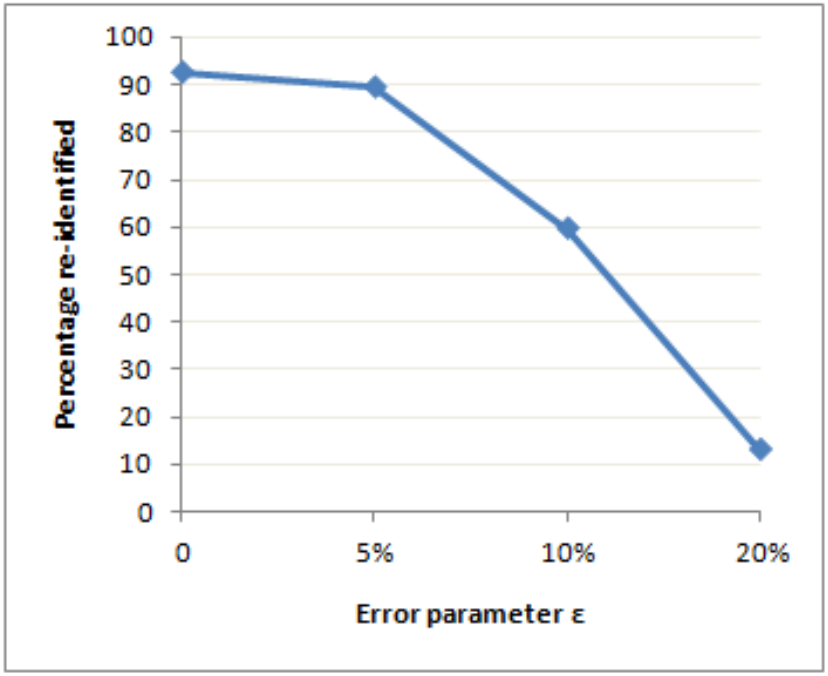
- “follow” relation on Twitter; “contact” relation on Flickr; “friend” relation on LiveJournal.

Network	Nodes	Edges	Av. Degree
Twitter	224K	8.5M	37.7
Flickr	3.3M	53M	32.2
LiveJournal	5.3M	77M	29.3

Twitter, Flickr (crawled) have APIs that expose a mandatory username and optional fields name and location

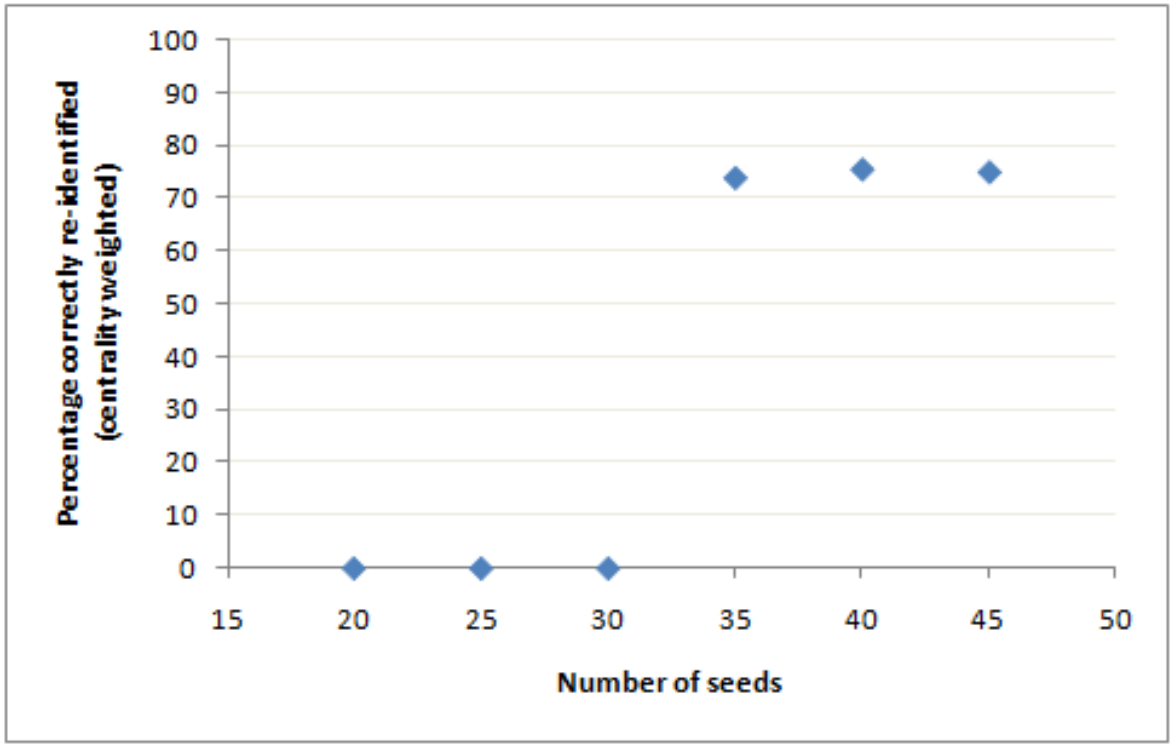
Seed indentification

Test on LiveJournal as target



Propagation

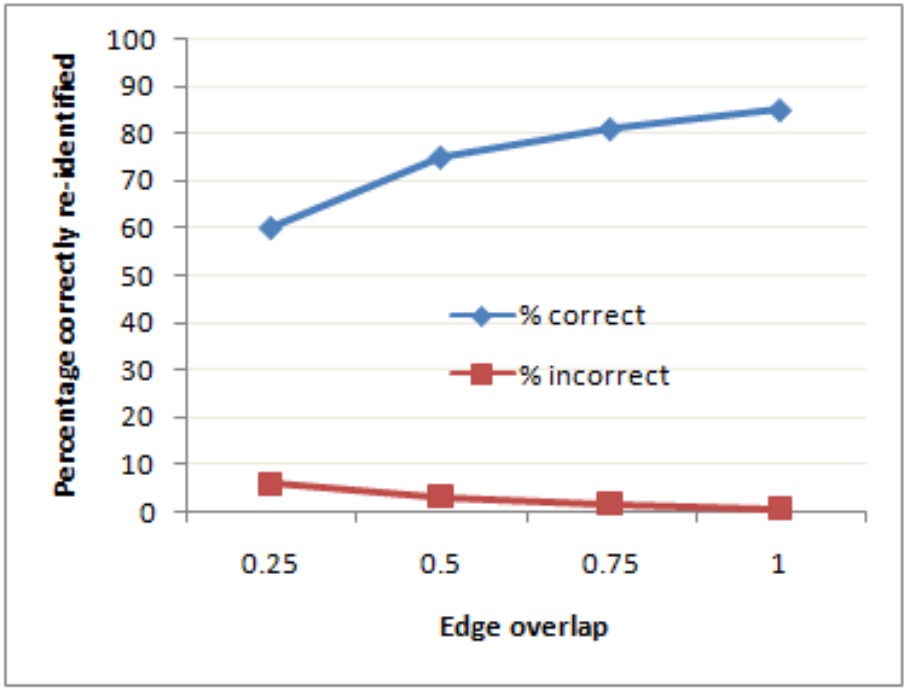
Number of seeds decides whether propagation step dies out or not. The graph is over 100,000 nodes.



Node overlap: 25% Edge overlap: 50%

Propagation

- Imprecision of auxiliary information decrease percentage of correctly re-identified rate



Node overlap: 25% Number of seeds: 50

Propagation

- Auxiliary graph: Flickr. Target graph: Twitter

ground truth (matches based on username, name, location)

27,000 mappings

- Seed mapping consists of 150 pairs of nodes with the constraints that the degree of each node in auxiliary graph is at least 80.

Result of accuracy

- 30.8% of mappings(27,000) were re-identified correctly, 12.1% were identified incorrectly, and 57% were not identified.
- 41% of the incorrectly identified mappings were mapped to nodes which are at a distance 1 from the true mapping.
- 55% of the incorrectly identified mappings were mapped to the nodes where the same location was reported.
- The above two categories overlap; only 27% of incorrect mappings are completely erroneous.