Social Media and User Privacy

Ghazaleh Beigi, Huan Liu
Arizona State University
http://www.public.asu.edu/~gbeigi/
Contact: gbeigi@asu.edu

Motivation

- Explosive growth of the Web allows people to freely conduct activities in social media platforms.
- This user-generated data is heterogeneous and rich in content.
- This provides many opportunities for researchers to better understand users behavior and provide personalized services for them.
- Publishing user data risks users' privacy as it contains sensitive and private information.

> [redacted] · 1h
> Dr. appt Friday morning was told I need to loose 50 pounds by X-Mas , have a fatty liver, **high** cholesterol and **high blood pressure**. Today starting the #danielfast counting calories #myfitnesspal & juicing for dinner (spinach, banana, strawberries, honey, almond milk).

Privacy Risks in Heterogeneous Social Media Data

- Many anonymization techniques are introduced for social media data.
- Existing work assumes that it is sufficient to anonymize each aspect of social media data independently.
- Let us assume data consists of graph and textual information: $\mathcal{D} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{W}, \mathcal{W})$

	Case 1	Case 2	Case 3	Case 4
Structural Anonymization	✗	✗	✓	✓
Textual Anonymization	✗	✓	✗	✓

- **Question** : Is either of these cases sufficient for anonymizing social media data?

Adversarial Technique for Heterogeneous Data

Given an anonymized social media dataset D, the aim is to map each user $u \in D$ to a real identity in targeted social media T.

1. Extracting top-k posts:

- Select posts with top scores

$$s_l = \frac{\sum_{t=1, \mathbf{X}_l(t) \neq 0}^{\mathcal{W}} \mathbf{X}_l(t)}{\sum_{t=1, \mathbf{X}_l(t) \neq 0}^{\mathcal{W}} \mathbf{1}}$$

2. Finding a set of candidates from targeted social media

- Create the set of candidate users $\mathcal{C} = \{c_1, c_2, \ldots, c_{|\mathcal{C}|}\}$ by querying each $q_u^{(i)} \in \mathcal{Q}_u = \{q_u^{(1)}, q_u^{(2)}, \ldots, q_u^{(k)}\}$ in the T's search engine.

3. Matching-up candidates to target user

- Structural features
- Textual features

$$Sim(u, c_i) = \alpha Sim_{struct}(u, c_i) + (1-\alpha) Sim_{text}(u, c_i)$$

- Exploiting Homophily Theory
  - If two users match, their neighbors should also match.
  - Homophily can also help to capture hidden relations between different aspect of data.

$$Sim_{total}(u, c_i) = \beta Sim(u, c_i) + (1-\beta) Sim(\mathcal{N}(u), \mathcal{N}(c_i))$$

Evaluation

- We crawl data from Twitter and Foursquare

(a) Twitter

# of Users	# of Edges	Avg. Clustering Coefficient
6,789	244,480	0.219
Density	# of Tweets	# of Unigrams
0.005	478,129	208,483

(b) Foursquare

# of Users	# of Edges	Avg. Clustering Coefficient
22,332	229,234	0.295
Density	# of Tips	# of Unigrams
0.0005	124,744	103,264

- **Evaluation metric:** Attack success rate $= n_c/N$

(a) Twitter

	ATHD-Improved		ATHD-Simple		ADA		Narayanan et. al.	
	Naive	Diff Privacy	Naive	Diff Privacy	Naive	Diff Privacy	Naive	Diff Privacy
Naive	0.9435 (1)	0.8020 (2)	0.8200 (1)	0.6951 (2)	0.6729 (1)	0.5513 (2)	0.5073 (1)	0.4100 (2)
Sparsification	0.8087 (3)	0.6998 (4)	0.7327 (3)	0.6213 (4)	0.6099(3)	0.5114 (4)	0.4316 (3)	0.3437 (4)
k-deg(add)	0.7894 (3)	0.6814 (4)	0.6900 (3)	0.6125 (4)	0.5898 (3)	0.4982 (4)	0.3979 (3)	0.3139 (4)
k-deg(add & del)	0.7580 (3)	0.6533 (4)	0.6891 (3)	0.5821 (4)	0.5800 (3)	0.4727 (4)	0.3815 (3)	0.2997 (4)
Switching	0.6911 (3)	0.5812 (4)	0.6013 (3)	0.5186 (4)	0.4971 (3)	0.4014 (4)	0.3520 (3)	0.2618 (4)
Perturbation	0.6500 (3)	0.5685 (4)	0.5367 (3)	0.4249 (4)	0.4322 (3)	0.3618 (4)	0.2987 (3)	0.2018 (4)

- Despite anonymization of all aspects of data is essential, but it is not sufficient to anonymize each aspect independently from others.
- This is because of hidden relations between different aspects of heterogeneous social media data

Conclusion

- This work introduces new privacy risks in social media data.
- This raises the need for an anonymization approach which considers the hidden relations between different components of the data.

Sponsored by Army Research Office grant W911NF-15-1-0328 and Office of Naval Research grant N00014-17-1-2605.