Adaptive Anonymity via b-Matching

Sharing data with variable privacy levels for each user

Krzysztof Choromanski Tony Jebara Kui Tang
Columbia University, KC → Google Research

Liberal  ...

Conservative  ...

Paranoid 

Privacy expectations are not uniform!

- To share, merge and model data, we must maintain privacy
- Current privacy-preserving approaches (k-anonymity, l-diversity, differential privacy) assume privacy-level is uniform
- But individuals/organizations have different privacy requirements (from liberal to conservative to paranoid)
- Traditional k-anonymity deletes until $\exists$ k copies of each user ... one paranoid user requiring $k = n$ ruins the whole data-set!

Definition (Adaptive anonymity)

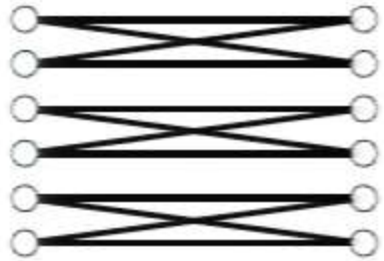
For users $i = 1, \ldots, n$ with inputs $\mathbf{x}_i \in \mathbb{Z}^d$ and anonymity levels δ_i, output $\mathbf{y}_i \in \{\mathbb{Z}, *\}^d$ with minimal suppressions $*$ where user i has de-anonymization probability at most $1/\delta_i$

Adapatative anonymity generalizes k-anonymity

We relax k-anonymity or k-cliques to a b-matching:
$\mathbf{y}_i$ compatible with δ_i inputs $\mathbf{x}_j$ & $\mathbf{x}_i$ compatible with δ_i outputs $\mathbf{y}_j$

k-anonymity

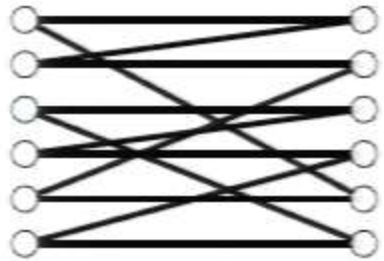
username				
alice	1	0	0	0
bob	0	0	0	0
carol	0	0	1	1
dave	1	0	1	1
eve	1	1	0	0
fred	0	1	1	1



				key
*	0	0	0	ggacta
*	0	0	0	tacaga
*	0	1	1	ctagag
*	0	1	1	tatgaa
*	1	*	*	caacgc
*	1	*	*	tgttga

b-matching

username				
alice	1	0	0	0
bob	0	0	0	0
carol	0	0	1	1
dave	1	0	1	1
eve	1	1	0	0
fred	0	1	1	1



				key
*	0	0	0	ggacta
*	*	0	0	tacaga
*	0	1	1	ctagag
*	*	1	1	tatgaa
1	*	0	0	caacgc
0	*	1	1	tgttga

We maximize utility via a polynomial time approximation scheme
Our algorithm uses variational b-matching to minimize *'s
It gives more utility than k-anonymity... even if δ_i are all equal!

As strong privacy theorems, improved utility experiments

Theorem

Given a b-matching returned by our algorithm, assume the adversary knows c edges of the matching and selects uniformly at random a vertex to attack. Then, he succeeds with probability $1/(\delta_i - c)$ *where* δ_i *is the degree (desired privacy level) of victim* i*.*

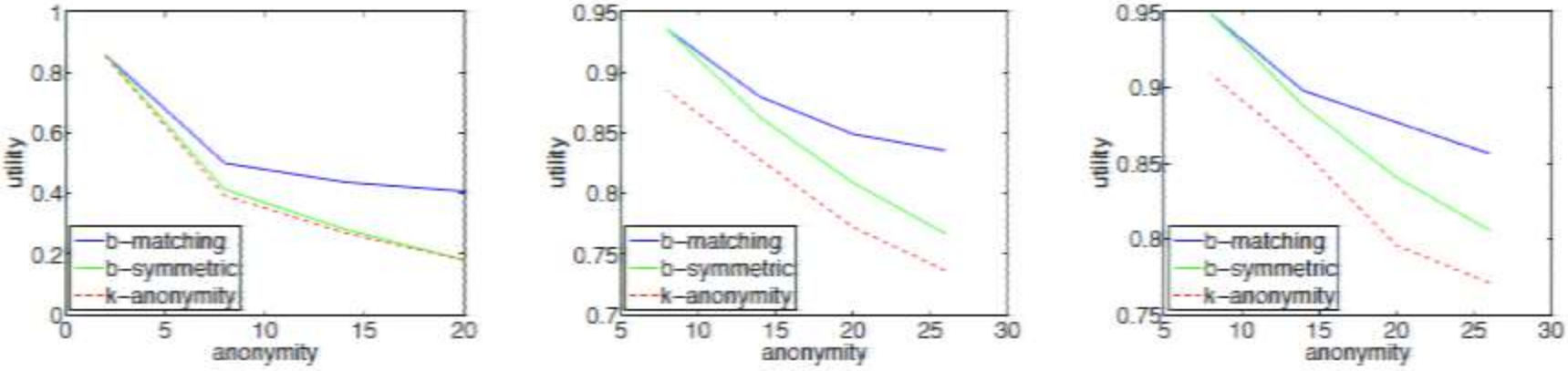


Figure: Utility $(1 - \frac{\#(*)}{nd})$ vs. Average Anonymity on Hepatitis, CalTech University Facebook, and Reed University Facebook data-sets