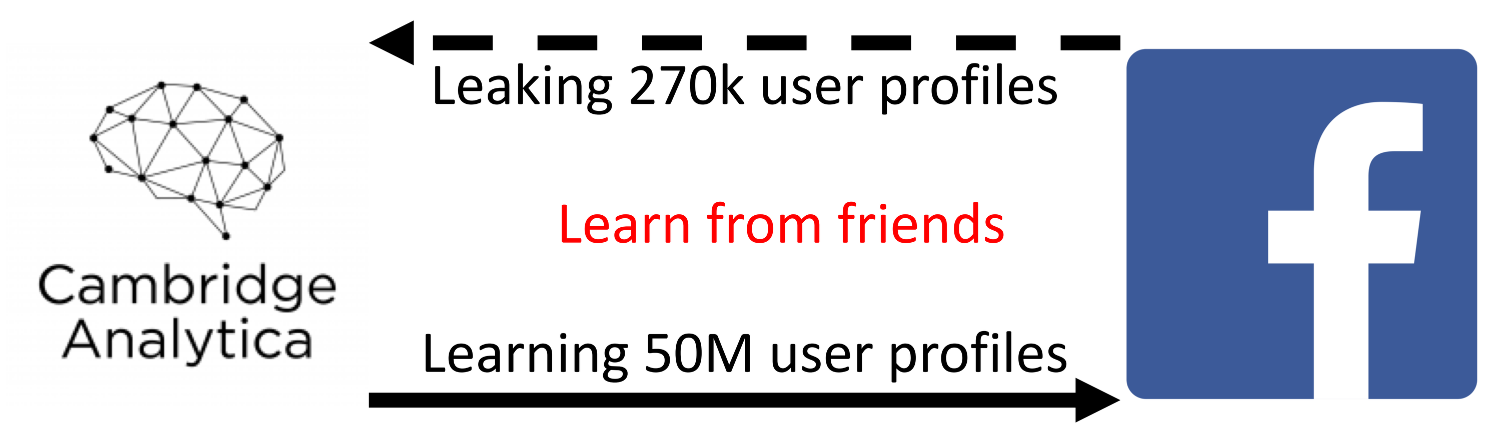
Analysis of Thompson Sampling for Graphical Bandits Without the Graphs

Fang Liu[1], Zizhan Zheng[2] and Ness Shroff[1]
[1]The Ohio State University, [2]Tulane University

The Facebook Data Leak 2018



How could Cambridge Analytica learn 50 million user profiles from only 0.5% samples of the population?

The key is to learn the information revealed by friends on online social networks.

In fact, a line of research in graphical bandits, also known as bandits with graph feedback and bandits with side observations, could have warned us about this issue.

A typical Facebook post.

Your friend accepts an event or promotion.

Your opinion is also queried and observed.

So is your other friends.

Can be modeled by graphical bandits.

Graphical Bandits Without the Graphs

Bayesian formulation of stochastic graphical bandits:

- At time t, the agent chooses an action A_t from K actions
- The agent receives an *i.i.d.* reward Y_{t,A_t}
- The agent observes the outcome $Y_{t,a}$ for each out-neighbor a in the latent graph $G_t = (\mathcal{K}, \mathcal{E}_t)$

Performance measure:

- Expected regret $\mathbb{E}[R(T)] = \mathbb{E}\left[\sum_{t=1}^{T} Y_{t,A^*} - Y_{t,A_t}\right]$

Graph numbers:

- Clique cover number: $\chi(G)$
- Maximum acyclic subgraphs: $mas(G)$
- Independence number: $\theta_0(G)$
- In general, $\theta_0(G) \leq mas(G) \leq \chi(G)$

Vanilla Thompson Sampling

TS-N Algorithm:
sampling an action according to $\pi_t = \alpha_t$

Note that α_t is the posterior distribution of A^*

Theorem 1. The regret of TS-N satisfies

$$\mathbb{E}[R(T)] \leq \sqrt{\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}[Q_t(\alpha_t)]H(\alpha_1)}$$

The graph related quantity is

$$Q_t(\pi_t) = \sum_{i\in\mathcal{K}} \frac{\pi_t(i)}{\sum_{j:j\xrightarrow{t}i}\pi_t(j)}$$

By bounding the quantity through graph numbers, we have

Corollary 1. For undirected graphs, the regret of TS-N satisfies

$$\mathbb{E}[R(T)] \leq \sqrt{\frac{1}{2}\sum_{t=1}^{T}\theta_0(G_t)H(\alpha_1)}$$

Corollary 2. For directed graphs, the regret of TS-N satisfies

$$\mathbb{E}[R(T)] \leq \sqrt{\frac{1}{2}\sum_{t=1}^{T}mas(G_t)H(\alpha_1)}$$

Remark 1. The results in Corollaries 1 & 2 hold for IDS-N [2].

Thompson Sampling with Exploration

TS-U Algorithm:
sampling an action according to $\pi_t = (1-\varepsilon)\alpha_t + \varepsilon/K$

Theorem 2. The regret of TS-U satisfies

$$\mathbb{E}[R(T)] \leq \varepsilon T + \sqrt{\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}[Q_t(\pi_t)]H(\alpha_1)}$$

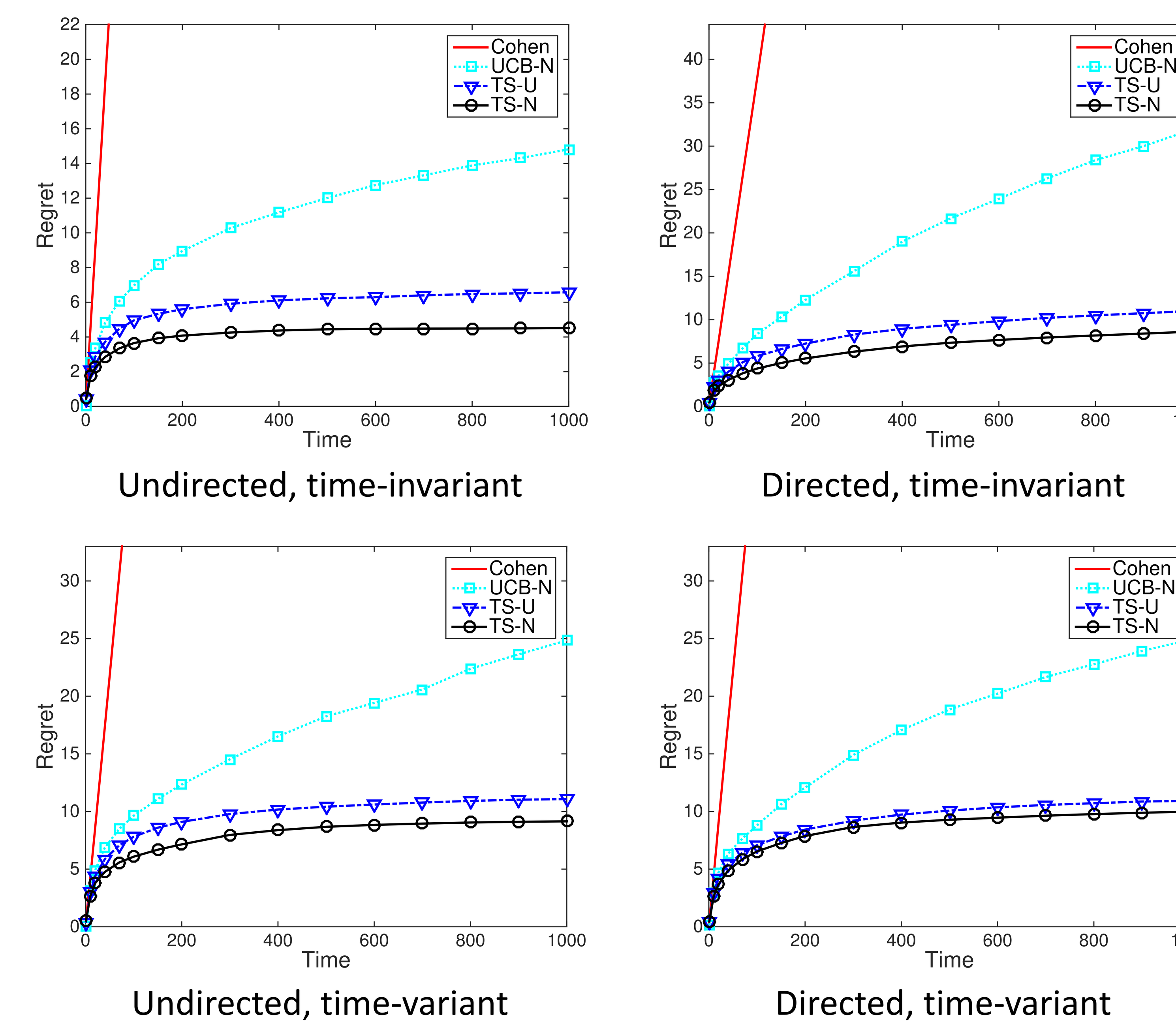
Additional exploration allows bounding Q_t by $\theta_0(G)$

Corollary 3. If $\varepsilon = 1/\sqrt{T}$, the regret of TS-U satisfies

$$\mathbb{E}[R(T)] = O\left(\sqrt{\log(KT)\sum_{t=1}^{T}\theta_0(G_t)H(\alpha_1)}\right)$$

Remark 2. The analysis in Theorem 2 holds for $\varepsilon_t = 1/t$. So we recommend using decreasing rates in practice.

Experiments



Undirected, time-invariant — Directed, time-invariant

Undirected, time-variant — Directed, time-variant

Related Work

Algorithm	Graph	Undirected	Directed
Cohen [1]	Without	$O\left(\sqrt{\theta_0(G)T\log K}\log(KT)\right)$	
IDS-N [2]	Informed	$O\left(\sqrt{\chi(G)T\log K}\right)$	
TS-N [2]	Without	$O\left(\sqrt{\chi(G)T\log K}\right)$	
TS-N	Without	$O\left(\sqrt{\theta_0(G)T\log K}\right)$	$O\left(\sqrt{mas(G)T\log K}\right)$
TS-U	Without	$O\left(\sqrt{\theta_0(G)T\log K}\log(KT)\right)$	

Reference

[1] Cohen, Alon, Tamir Hazan, and Tomer Koren. Online learning with feedback graphs without the graphs. In ICML 2016.

[2] Fang Liu, Swapna Buccapatnam, and Ness Shroff. Information directed sampling for stochastic bandits with graph feedback. In AAAI 2018.