

## Session 6: Top-k and Pattern Mining

担当： 王(名古屋大学)

# Finding Top-K Local Users in Geo-Tagged Social Media Data

---

- Jinling Jiang, Hua Lu, Bin Yang, Bin Cui (U. Aalborg, U. Peking)
- 問題: top-k local user search(TkLUS)
  - 条件: 位置 $q$ , 距離 $r$ , キーワードの集合 $W$
  - $q$ から $r$ 距離以内の相関ツイッターをポストしたtop-k個のユーザ
- 貢献:
  - TkLUS問題を定義
  - ユーザの評価手法を定義
  - キーワードと位置情報を認識できる雑種インデクスをデザイン
  - クエリ処理のアルゴリズムを考える
  - 評価実験を行う

# 問題の定式化

---

- A: ソシャルメディアデータモデル
  - 定義1: (Social Media Post)  $p = (uid, t, l, W)$ 
    - uid: ユーザ
    - t: タイムスタンプ
    - l: 位置情報
    - W:  $(w_1, w_2, \dots, w_n)$  キーワードの集合
  - 定義2: (Social Network)
    - $G = (U, E_{reply}, l_{reply}, E_{forward}, l_{forward})$
    - U: 頂点の集合, ユーザと対応
    - $E_{reply}$ : 返事関係の辺の集合
    - $l_{reply}(u_1, u_2) \rightarrow u_1$  が  $u_2$  に答えたポスト
    - $E_{forward}$ : リツイート関係の辺の集合
    - $l_{forward}(u_1, u_2) \rightarrow u_1$  がリツイートした  $u_2$  のポスト
- D=(P,U,G) 位置情報が付いてるポスト

# 問題の定式化

---

- B: 問題定義

- Top-k Local User Search

- 位置情報が付いてるソーシャルメディアデータ  $D=(P, U, G)$  , 位置  $l$ , 距離  $r$ , キーワード集合  $W$  に対して, Top-k local user search  $q(l, r, W)$  が条件を満たした  $k$  個のユーザを探します.

- 1)  $\forall u \in \mathcal{E}_k, \exists p \in P_u$  such that  $\|q.l, p.l\| \leq q.r$  and  $p.W \cap q.W \neq \emptyset$ .
- 2)  $\forall u \in \mathcal{E}_k$  and  $\forall u' \in \mathcal{D}.U \setminus \mathcal{E}_k$ , either  $u'$  does not satisfy condition 1 or  $score(u', q) \leq score(u, q)$ .

# ツイッターとユーザを評価する

- Popularity of tweet

- Tweet thread

*Definition 4: Popularity of Tweet*

$$\phi(p) = \begin{cases} \epsilon, & \text{if } T.h = 1; \\ \sum_{i=2}^n |T_i| \times \frac{1}{i}, & \text{otherwise.} \end{cases}$$

- ユーザを評価する

$$\text{score}(u, q) = \alpha \cdot \rho(u, q) + (1 - \alpha) \cdot \delta(u, q)$$

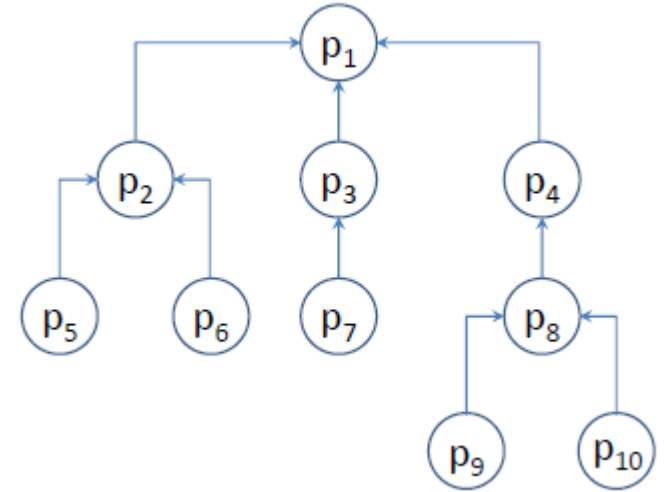
- ツイッターの評価

*Definition 5: Distance Score of Tweet*

$$\delta(p, q) = \begin{cases} \frac{r - \|q.l, p.l\|}{r}, & \text{if } \|q.l, p.l\| \leq r; \\ 0, & \text{otherwise.} \end{cases}$$

*Definition 6: Keyword Relevance Score of Tweet*

$$\rho(p, q) = \frac{|q.W \cap p.W|}{N} \cdot \phi(p)$$



# Architecture and Data organization

- System architecture & Index structure

