

【VLDB2013勉強会】

Session 12: Spatial and Text

担当: 趙(名古屋大学)

Session 12: Spatial and Text 担当: 趙(名大)

A General Framework for GeoSocial Query Processing

▶ Nikos Armenatzoglou , Stavros Papadopoulos, Dimitris Papadias

▶ Hong Kong University of Science and Technology

▶ 対象

▶ **Geo-social networks(GeoSNs)**

▶ **Social network** + **Location-based service**

▶ 例: Foursquare(一日に30M以上のユーザ,何百万の“check-ins“)

▶ **Geo-SN queries**

▶ **ソーシャル関係**と**現在の位置**を含む役立つ情報を抽出

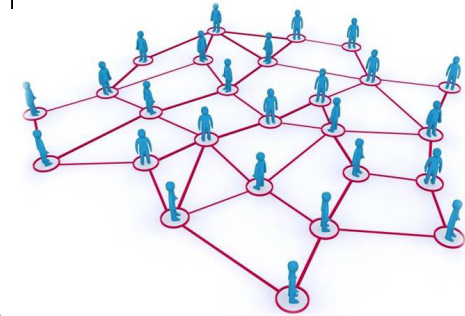
▶ 動機

▶ 既存研究はアルゴリズムのみに着目し,
重要な**データマネジメント問題**を見落としている

▶ 提案手法

▶ GeoSN問合せ処理に対する**汎用フレームワーク**

▶ **柔軟なデータマネジメントとアルゴリズムデザイン**



Query processing

- ▶ GeoSN アーキテクチャ(Figure 1)
 - ▶ SM: ソーシャルデータのみを保存 (e.g., 友達関係)
 - ▶ GM: 地理的情報のみを保存 (e.g., check-ins)
 - ▶ QM: ユーザからのクエリを受け取り, 実行し, 結果を返す
- ▶ 基本問合せ
 - ▶ ソーシャル基本問合せ
 - ▶ $GetFriends(u)$, $AreFriends(u_i, u_j)$
 - ▶ 地理的基本問合せ
 - ▶ $GetUserLocation(u)$, $Range\ Users(q,r)$, $NearestUsers(q,k)$
- ▶ GeoSN問合せ
 - ▶ $Range\ Friends(RF)$, $Nearest\ Friends(NF)$
 - ▶ **Nearest Star Group(NSG): $NSG(q,m,k)$**
 - ▶ 与えられたポイント q に最近傍の k 個のグループ (m 個のユーザを含む)を返す
 - ▶ アルゴリズム: $NSG\ eager$, $NSG\ lazy$, $NSG^*\ eager$

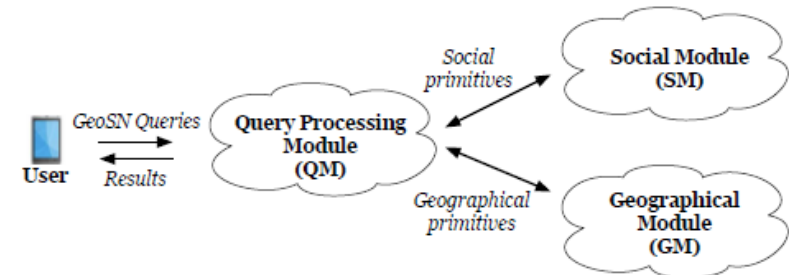
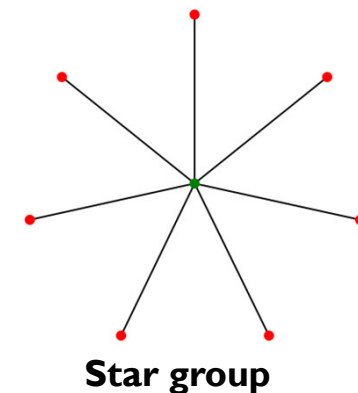


Figure 1: Proposed GeoSN architecture



実験

▶ Datasets

- ▶ 実データ(Foursquare, Twitter: 1日に12,652 users' check-ins)
- ▶ エンジェデータ(五つ)

▶ Storage approach

- ▶ Disk-based: centralized, distributed architecture
- ▶ Memory-based: centralized, distributed architecture

▶ Parameter(table 1) : r, s, k, m, N

▶ クエリ時間

- ▶ 各クエリの100回の平均時間

▶ 実データでNSGアルゴリズムの性能

- ▶ パラメーター: $m=5, k(1\sim6)$
- ▶ NSG* eagerがベストアルゴリズム
- ▶ k が高いほど、各メソッドのコストが高くなる

Table 1: System parameters and their ranges

Parameter	r	s	k	m	N
Range	0.5-5 km	1K-5K	1-10	2-7	1M-5M

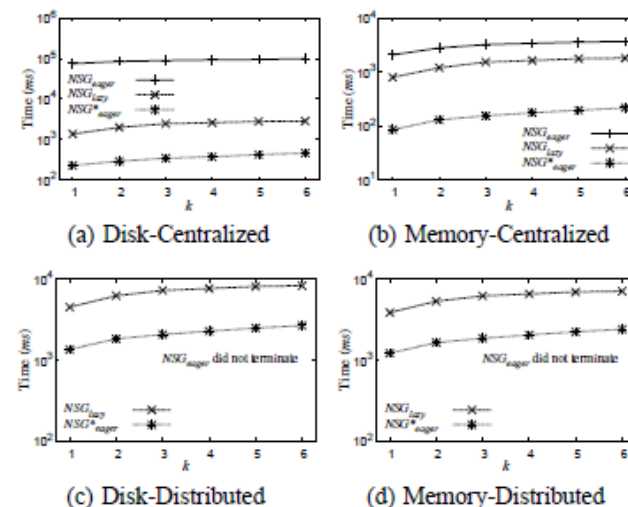


Figure 18: Query time for NSG vs. k (Real dataset)