

【VLDB2011勉強会】

Session 23: Social Networks

担当：岩井一晃 横山拓也(名古屋大学)

Structural Trend Analysis for Online Social Networks

カリフォルニア大学
Ceren Budak,
Divyakant Agrawal,
Amr El Abbadi

Social NetworkにおけるTrend

- 社会的な関心を理解するのに重要

従来 of Trend研究

- Trendを時間的, 空間的に分析
- Trendを頻出項目, 重要な人物から特定

問題点

- Spammerに弱い/違うコミュニティでの活動が不明

構造に着目したTrend解析の提案

Coordinated Trend

Uncoordinated Trend

Structural Trend Analysis for Online Social Networks

カリフォルニア大学
Ceren Budak,
Divyakant Agrawal,
Amr El Abbadi

Coordinated Trend

- ▶ 利用者同士が繋がりを持っているTrend
- ▶ 式(2)によってScoreを算出
 - ▶ あるTopicにおいて、利用者とその友人が言及した数を利用

$$g(T_x) = \sum_{n_i \in N} (C_{i,x} \sum_{n_k \in N_i} C_{k,x}) \quad (2)$$

- ▶ Incrementalな処理に変換
 - ▶ 数が非常に多いため再計算は非現実的

$$g'(T_x) = g(T_x) + \sum_{n_i \in N_i'} C_{i,x} + \sum_{n_i \in N_i} C_{i,x} \quad (6)$$

- ▶ さらに最適化を行う
 - ▶ 図4の三角形の数は式(2)の値と等しい値である
- ▶ 更新処理が上記の式で算出するよりも早く求められる

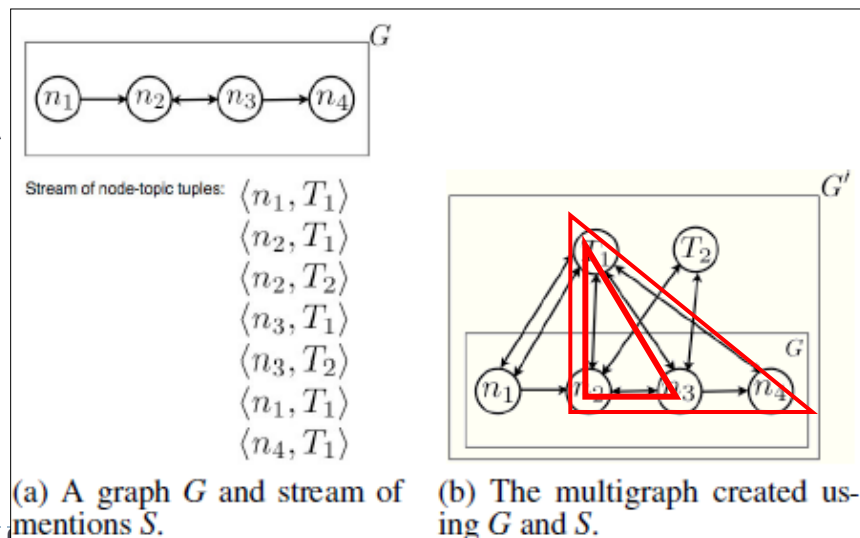


Figure 4: Reduction to counting local triangles.

Structural Trend Analysis for Online Social Networks

カリフォルニア大学
Ceren Budak,
Divyakant Agrawal,
Amr El Abbadi

実験

- ▶ データセット: 2000万人のユーザ, 4億6700万ポスト
- ▶ Topic: Hashtagにより決定 [2960495 Topic]
- ▶ User : 270万人の利用者, 2億3000万の友人関係
- ▶ p はサンプリングに関するパラメータ

- ▶ p が低くとも高精度
- ▶ Top-29604の精度が低いのは言及数が少ないTopicが多いため

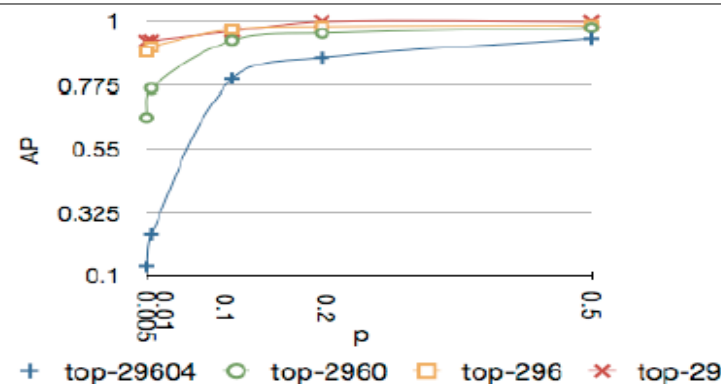


Figure 5: Average Precision of sampling for coordinated trends.

On SocialTemporal Group Query with Acquaintance Constraint

Academia Sinica **DeNian Yang**
National Taiwan Univ
YiLingChen MingSyanChen
The Penn State Univ
WangChienLee

スケジュールの共有

- ▶ ex)Google カレンダー, Facebook
- ▶ イベントの決定が共有している人の予定を見て決定可能
- ▶ 適切な参加者, 適切な時間にイベントを決定するのは困難

目的

- ▶ 自動でイベントの参加者, 時間の決定
 - ▶ 従来手法では投票方式で手動入力が必要
- ▶ 適切な参加者の決定
 - ▶ 共有情報を用いて, 出来るだけ親しい参加者を決定



On SocialTemporal Group Query with Acquaintance Constraint

Social Group Query (SGQ)

- ▶ イベント決定問題
- ▶ 参加者の最適グループが親しい範囲に収まっている
 - ▶ 検索する必要な無いほど距離がある人を枝狩り
- ▶ 参加必須の人が参加していることを満たす
 - ▶ 最適な検索範囲を利用する[最適解はNP困難]

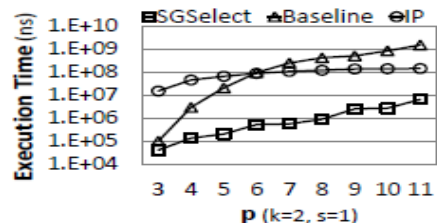
Social-Temporal Group Query (STGQ)

- ▶ SGQを期間毎に決定[最適な期間を求める]
 - ▶ 不必要な時間を検索しない[最適解はNP困難]
- ▶ 参加者について, 時間について個別に扱う必要がある

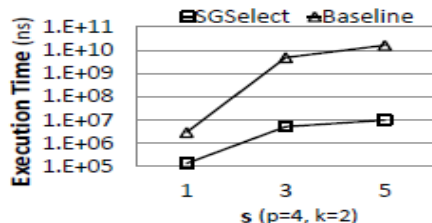
On SocialTemporal Group Query with Acquaintance Constraint

実験

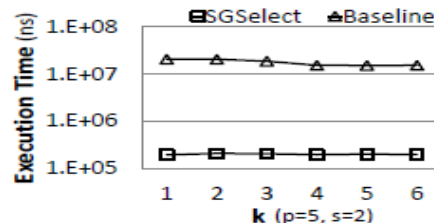
- ▶ 194人を利用 [Social Distanceは2者間の相互作用]
- ▶ PCArrangeが比較対象
 - ▶ PCArrangeには利用者制約が無いいため疑似的に与えている
- ▶ 全体的に本手法が高性能



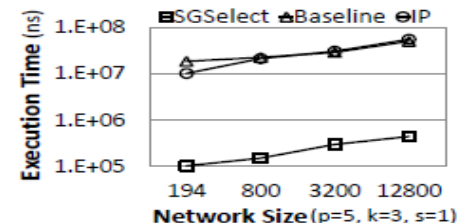
(a) Comparison of running time with different p .



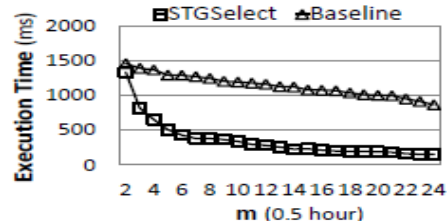
(b) Comparison of running time with different s .



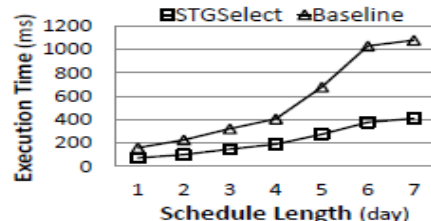
(c) Comparison of running time with different k .



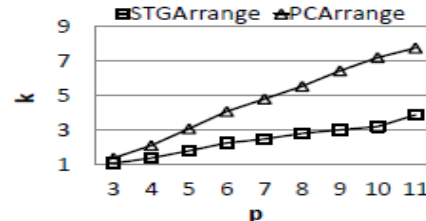
(d) Comparison of running time with different network sizes.



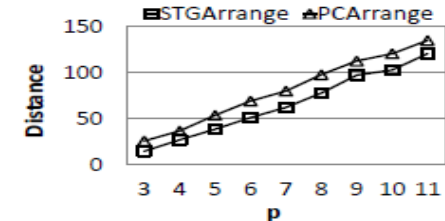
(e) Comparison of running time with different m .



(f) Comparison of running time with different schedule lengths.



(g) Relation of p and k .



(h) Relation of p and the total social distance.

Figure 1: Experimental results of SGQ and STGQ.

Social Content Matching in MapReduce

(Yahoo! Research Barcelona)

Gianmarco De Francisci Morales,
Aristides Gionis, Mauro Sozio

▶ Web上のコンテンツが多くなる中で高まる要望

▶ 受け手

- ▶ 自分の興味に合ったコンテンツを得たい

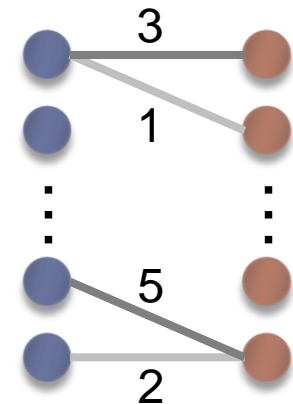
▶ 送り手

- ▶ 興味があるユーザーにコンテンツが送られ、質の高いフィードバックがほしい

▶ 2部グラフのマッチング問題として考える

- ▶ ノード間のエッジに興味に基づいて重み付け
- ▶ 各ノードのエッジ数に制限がある
 - ▶ capacity constraint

▶ 制限を満たしつつ、重みの合計が最大となるグラフをMapReduceによる処理で求める



ユーザー

アイテム

論文の貢献

- マッチングだけに注目し、
 - エッジの重みの計算
 - 候補エッジの計算
 - エッジ数の制限の計算
- はすでに行われているとする

▶ 2つのマッチングアルゴリズムを提案

StackMR

- ▶ 主双対法(primal-dual schema)がベース
- ▶ 整数線形計画問題として表す(主問題)

$$\begin{aligned} & \text{maximize } \sum_{e \in E} w(e)x_e \\ & \text{such that } \sum_{e \in E, v \in e} x_e \leq b(v) \quad \forall v \in V, \end{aligned}$$

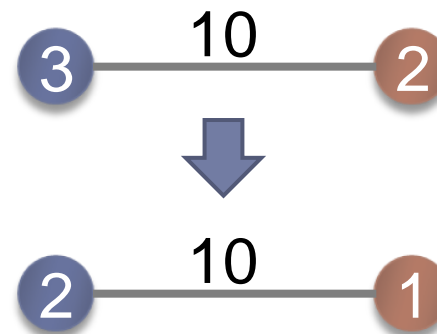
- ▶ 主問題から双対問題を導出

$$\begin{aligned} & \text{minimize } \sum_{v \in V} y_v \\ & \text{such that } \frac{y_u}{b(u)} + \frac{y_v}{b(v)} \geq w(e) \quad \forall e = (u, v) \in E, \\ & \quad y_v \geq 0 \quad \forall v \in V. \end{aligned}$$

- ▶ 双対問題の最適解が主問題の最適解
- ▶ MapReduce処理のために、エッジ数の制限を破ってもよい

GreedyMR

- ▶ 重みの大きい候補エッジから順に解として選ぶ
- ▶ エッジで繋がれたノードそれぞれのエッジ数の制限の値が0以上であるとき
- ▶ 解として選んだあと、ノードそれぞれのエッジ数の制限を1減らす



実験

データセット	アイテム	ユーザー	有効ペア
Flickr-small	2,817	526	550,667
Flickr-large	373,373	32,707	1,995,123,827
Yahoo-answers	4,852,689	1,149,714	18,847,281,236

▶ FlickrとYahoo-answersが対象

▶ 右図はYahoo-answersの結果 (Figure3)

▶ StackMR

▶ 結果の質 : Δ

▶ 計算の回数 : \circ

▶ GreedyMR

▶ 結果の質 : \circ

▶ 計算の回数 : Δ

▶ 処理を繰り返すごとに結果がよくなる回数

▶ 最初に大雑把な結果を提示

▶ その間に裏で詳細を計算

質

目的に応じて
使い分け

回数

