

WWW14勉強会 — Social系二つ —

山口祐人(筑波大)

紹介する論文

“User Profiling in an Ego Network: Co-profiling Attributes and Relationships”

Rui Li, Chi Wang, and Kevin Chen-Chuan Chang

University of Illinois

ソーシャルネットワークにおけるユーザの属性推定

“The Bursty Dynamics of the Twitter Information Network”

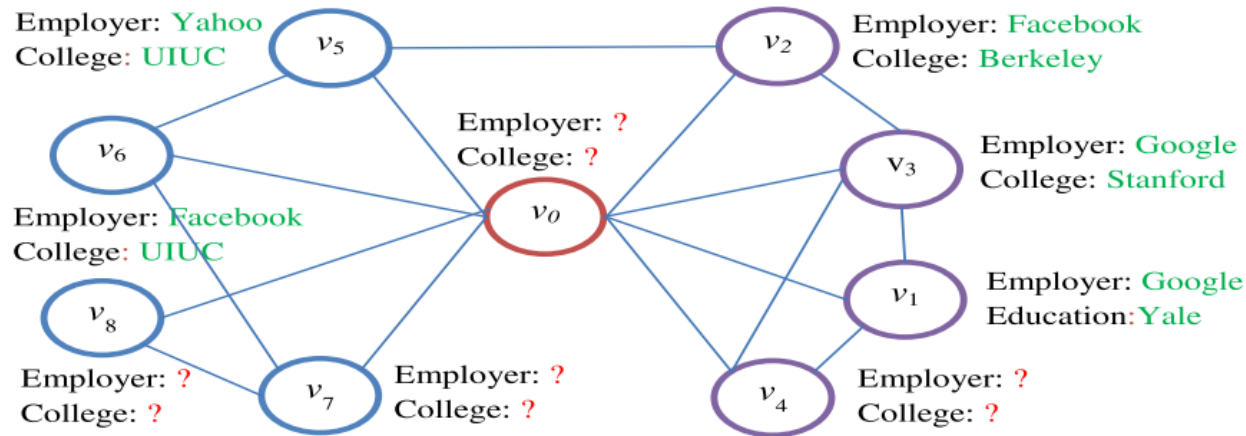
Seth Myers and Jure Leskovec

Stanford University

ソーシャルネットワーク(Twitter)の成長の分析

USER PROFILING IN AN EGO NETWORK: CO-PROFILING ATTRIBUTES AND RELATIONSHIPS

扱う問題：ソーシャルネットワーク ユーザの属性推定



論文中 Figure 1

- 部分的に属性が既知の（無向）**ego-network** が入力
 - ego-network: あるノードを中心とした1hopで到達できる部分グラフ
- 一つのノードが複数種類の属性を持つ
- 中心のノードの属性を推定するのが目的

基本的な考え方：隣り合うノードの属性は同じ可能性が高い



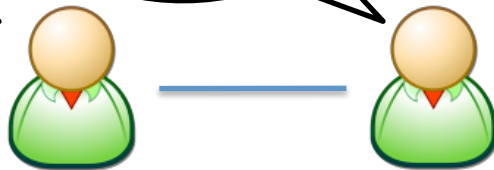
グラフを使って属性推定をする論文ではほとんどの考え方

この論文が提起する問題点

どの隣接ノードも同じ属性を持つ可能性が高いわけじゃない！

Employer
Facebook

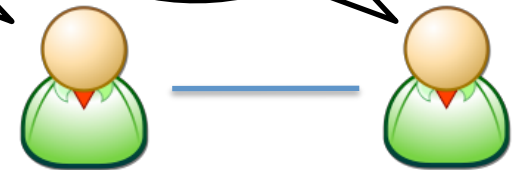
Employer
?



この二人が「テニスサークルの友達」なら同じ職場である可能性が高いとは言えない

College
Stanford

College
?

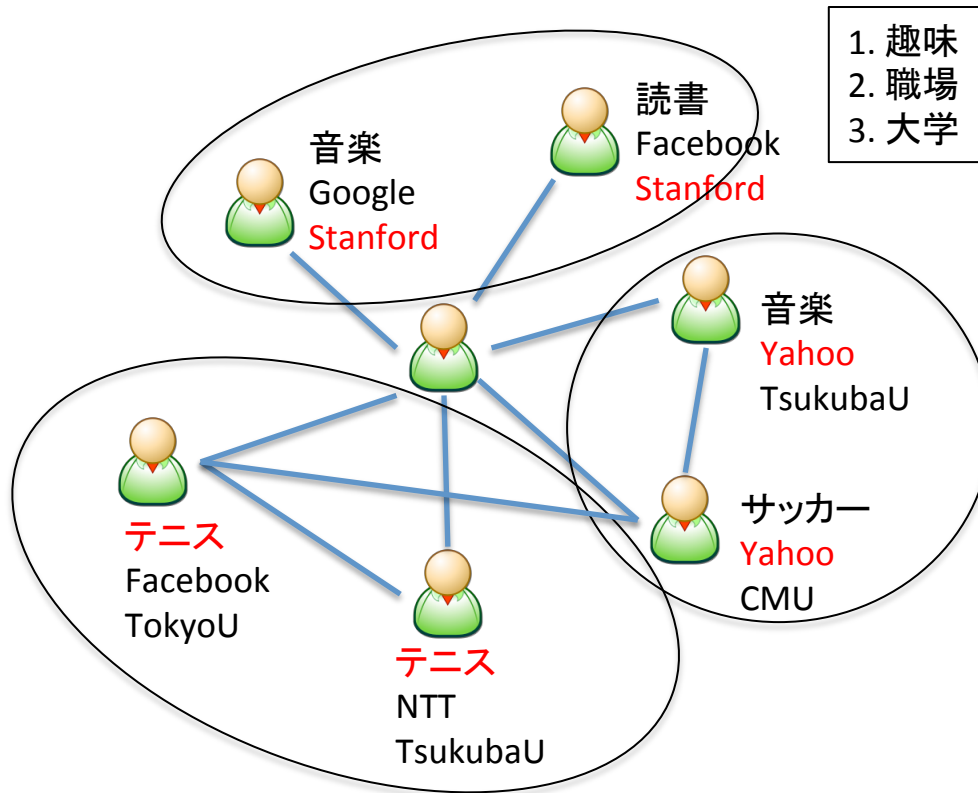


この二人が「小学校の友達」なら同じ大学である可能性が高いとは言えない

アイデア

ego-network 上で、隣接ノードを「属性を共有するグループ (Circle)」に分割する

→ Circle (隠れ変数) を推定



各Circleから対応する属性を伝播させる

→ テニス、Yahoo、Stanford

モデル

グラフ構造と既知の属性から、Circleと未知の属性を推定する

→ 最適化問題として解く

min.
属性
Circle

$$\lambda_1 \sum_{t=1}^K \left\{ \left(\sum_{e_{ij} \in E', v_i, v_j \in C_t} (w_t \cdot (f_i - f_j))^2 \right) + \sum_{v_i \in C_t} (w_t \cdot (f_0 - f_i))^2 \right\}$$

同じCircleに属すユーザ同士は
ある特定の属性が同じ

$$+ \lambda_2 \sum_{t=1}^K \sum_{v_i \in L \cap C_t} (w_t \cdot f_i - 1)^2 + \lambda_3 \sum_{e_{ij} \in E', x_i \neq x_j} 1$$

同じCircleに属す "多く" のユーザが
ある特定の属性 "のみ" を共有

別のCircleに属すユーザ間
のエッジは少ない

論文中 Equation (1)

アルゴリズム

最適化問題を
効率的に解けました!

論文を

読んで
ね!

実験

提案手法が勝ちました!
データと実装は公開しません!

THE BURSTY DYNAMICS OF THE TWITTER INFORMATION NETWORK

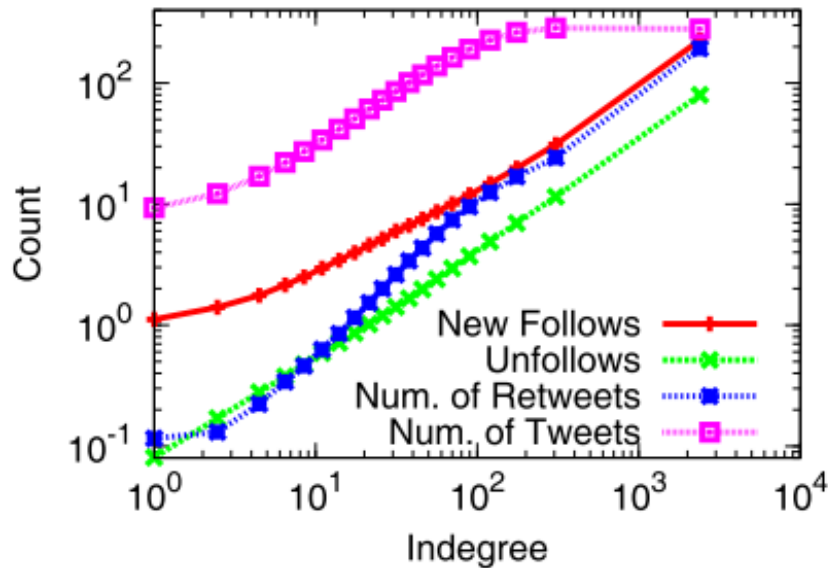
扱う問題: Twitterのソーシャルグラフ の動的な変化を分析

- 我々は全てのエッジにタイムスタンプが付いたTwitterのソーシャルグラフを持っている!!!
(◁) 欲しい!
- それを使ってグラフが動的にどう変化するかを分析する!!!
- 具体的にはあるユーザのツイートが多くのユーザにリツイートされた時(リツイートバースト)、その周辺でグラフはどう変化するか!!!?

(◁) 細かいところは置いておいて結果を見ていくよ!

Twitter graph is highly dynamic

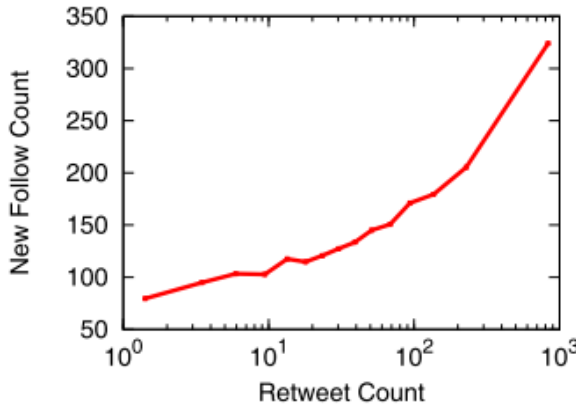
- 一ヶ月間に7%のエッジが追加され、2.3%のエッジが削除されている
→ Highly dynamic



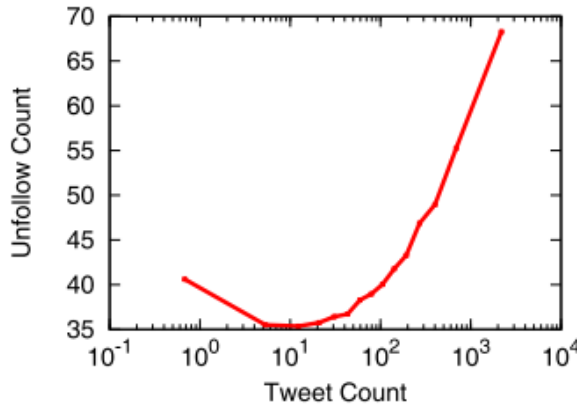
論文中 Figure 1

- 左図はある入次数を持つユーザの平均的なActivityを表す
- 例えば入次数100のユーザは一ヶ月間に10のエッジを得て、3のエッジを失う

多くのRTは多くのフォローを引き起こす



(a) New follows vs. retweets



(b) Lost follows vs. tweets

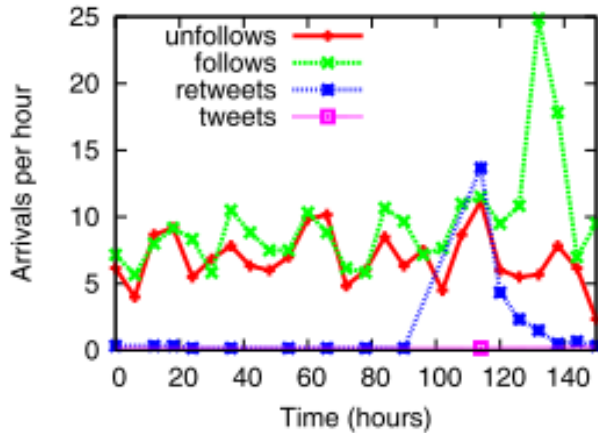
論文中 Figure 3

- 入次数1000 - 2000のユーザ
- 一ヶ月間にどれだけRTされ、tweetするか
- 一ヶ月間にどれだけフォローを得る / 失うか

- 多くのリツイートを集めるユーザは多くの新しいフォローを集める (わかる)
- ツイートしすぎる / しなさすぎるユーザは多くのフォローを失う (まあわかる)
- 一ヶ月に10回くらいの頻度でツイートすればフォローを失う確率最低 (!?)

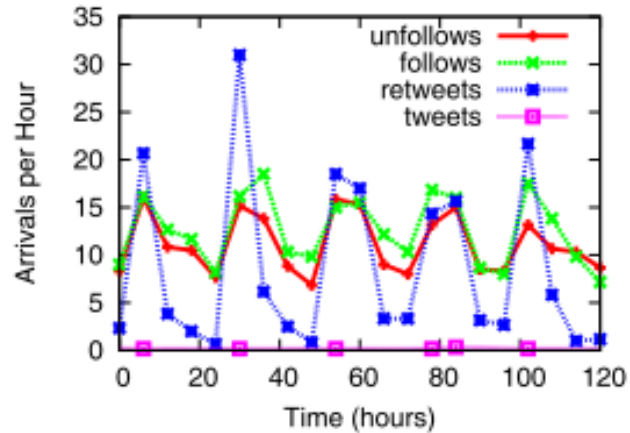
RTがフォローを引き起こすかどうかはユーザによる

論文中 Figure 4



(a) user with $d_{in} = 266,842$

RTのバーストが起こった直後に
大量の新しいフォローを得る



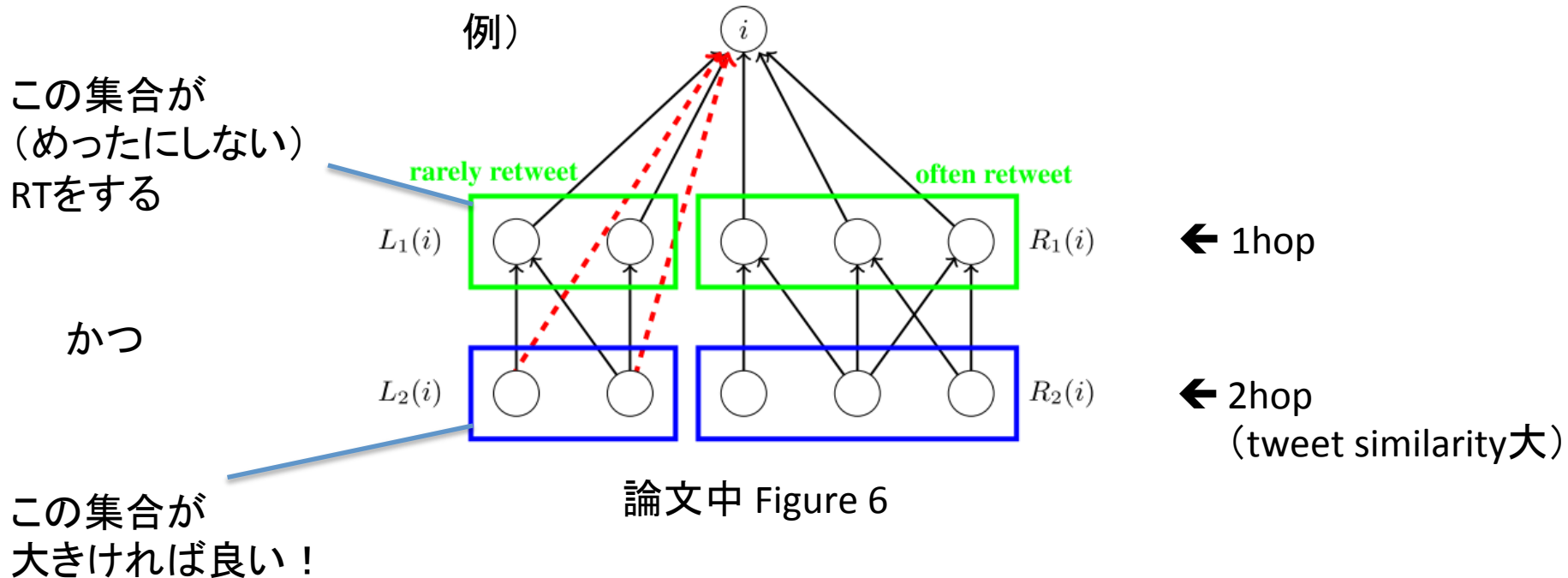
(b) user with $d_{in} = 218,045$

定期的にRTのバーストが起きているが
フォロー/アンフォローに影響はない

モデル

ある Retweetバーストが与えられた時、それが Followバーストを引き起こすかを予測

(アイデア) 対象ユーザ i とのツイートの類似度が大きく、未だに i からの RTを受け取っていないユーザが多い時、Followバーストが起きやすい



実験

提案モデルが勝ちました