

05

大きなグラフから隠れた関係を素早く見つけます

～大規模グラフ向けの高速な構造分析～

どんな研究

実世界には情報のつながりを表す膨大なグラフデータが存在します。従来の技術では、大量のデータの分析のために、膨大な時間を要していました。本研究ではグラフデータを対象にした高速な分析を行う理論的なアルゴリズムの研究を進めています。

どこが凄い

グラフデータの構造的な特徴を理論的に解析することで、**高速なグラフデータ分析アルゴリズム**を確立しました。このアルゴリズムを用いることで従来のアルゴリズムよりも**より大規模なデータを短時間で分析**することができるようになりました。

目指す未来

大規模なソーシャルネットワークやメディアのグラフを瞬時に分析し、情報の推薦・予測・理解に活用できます。本技術は、従来考えられなかった規模のデータの分析に貢献できる可能性があり、将来的に世界の至る所から得られるデータも瞬時に解析・利用できる社会を実現します。

● グラフにおけるノードの高速ラベリング技術 [1]

概要

ラベルが与えられた一部のノードからその他のノードのラベルを高速かつ高精度に推定

- 各ラベルのスコアの定義式から上限値と下限値を再帰的に推定

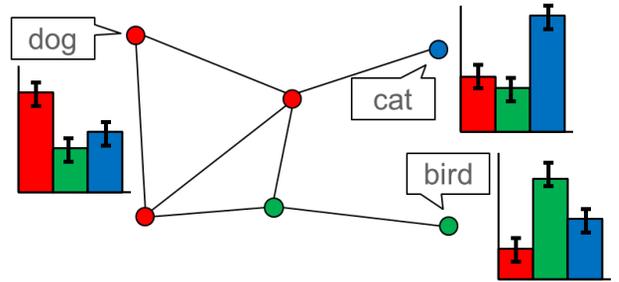
$$F = \alpha SF + (1 - \alpha)Y$$

$$= (1 - \alpha)(Y + \alpha SY + \alpha^2 S^2 Y + \dots)$$

F : ノード×スコアの行列 α : 定数

S : グラフの隣接行列 Y : 与えられるスコアの分布

- 推定した各ラベルごとのスコアの上限値と下限値からラベルを決定

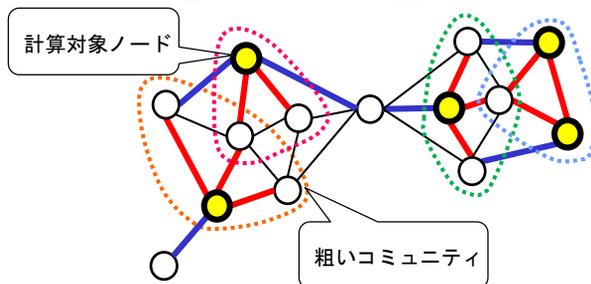


● クラスタリング高速化技術 [2]

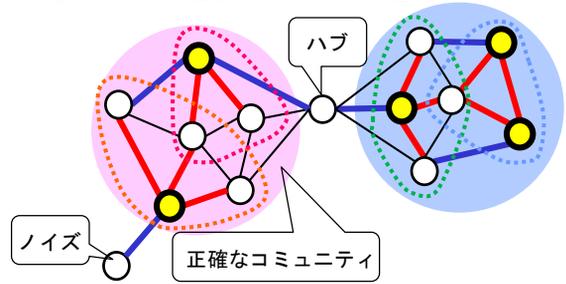
概要

人間関係や購買履歴などの大規模なグラフデータの中からコミュニティやハブ、ノイズとなるデータを高速に発見

- 最短距離が2ホップ離れたノードのみを計算し、粗いコミュニティを算出



- 複数のコミュニティに所属するノードを見つけ正確なコミュニティを修正



関連文献

- [1] Y. Fujiwara, G. Irie, "Efficient label propagation," in *Proc. the 31st International Conference on Machine Learning (ICML2014)*, 2014
 [2] 塩川浩昭, 藤原靖宏, 鬼塚真, "構造的類似度に基づくグラフクラスタリングの高速化," 第6回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2014), D6-2, 2014

連絡先

岩村相哲 (Sotetsu Iwamura) ソフトウェアイノベーションセンタ
 E-mail: iwamura.sotetsu(at)lab.ntt.co.jp

