

# Graphs I: Fundamental Algorithms

R6-1

## Answering Graph Pattern Queries Using Views

Author: Wenfei Fan(Univ. of Edinburgh), Xin Wang, Yinghui Wu

Presenter: 熊本 和正(筑波大)

ICDE勉強会 (2014/5/31)

# 概要

- 問題意識

- グラフが巨大なときの問い合わせはコストがかかる  $O(|Q_s|^2 + |Q_s||G| + |G|^2)$
- キャッシュされた複数のグラフビューをうまく利用して検索効率を上げたい

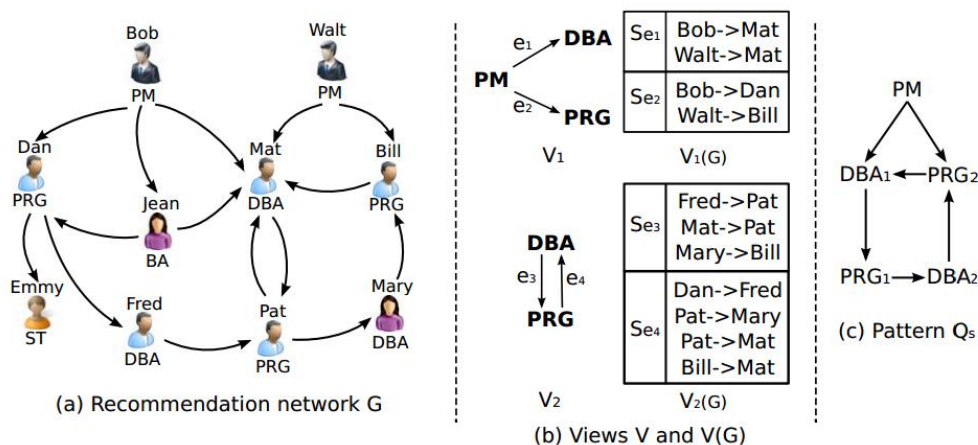


Fig. 1: Data graph, views and pattern queries

- この論文でやること

- ビュー集合からの検索結果の検出の効率化
- ビュー集合中から検索結果を検出するためのビュー選択の効率化

# 提案手法

- ビュー集合からの検索結果の検出の効率化
  - ビュー集合をエッジの組Viに変換して、その組み合わせにQsがマッチするかを見る
- ビュー集合中から検索結果を検出するためのビュー選択の効率化
  - 変換されたエッジの組を組み合わせてビューVから新しいビューV'を作る

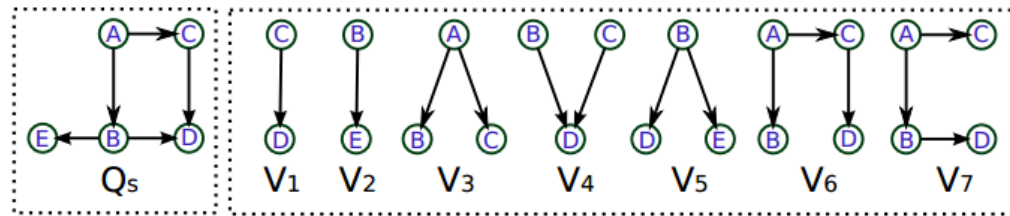


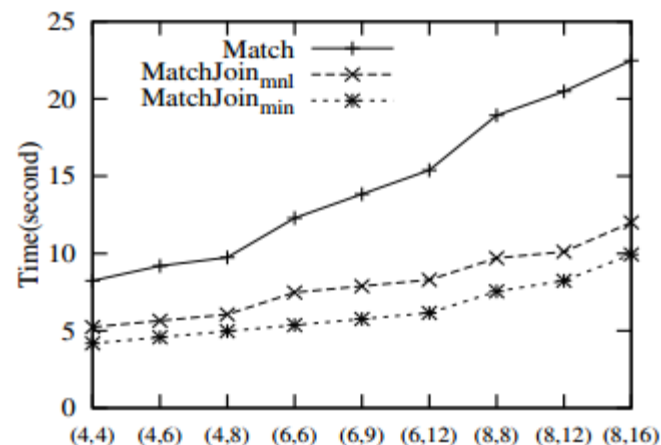
Fig. 4: Containment for pattern queries

$V_i$	$M_{V_i}^{Q_s}$	$V_i$	$M_{V_i}^{Q_s}$
$V_1$	$\{(C, D)\}$	$V_2$	$\{(B, E)\}$
$V_3$	$\{(A, B), (A, C)\}$	$V_4$	$\{(B, D), (C, D)\}$
$V_5$	$\{(B, D), (B, E)\}$	$V_6$	$\{(A, B), (A, C), (C, D)\}$
$V_7$	$\{(A, B), (A, C), (B, D)\}$		

# 実験

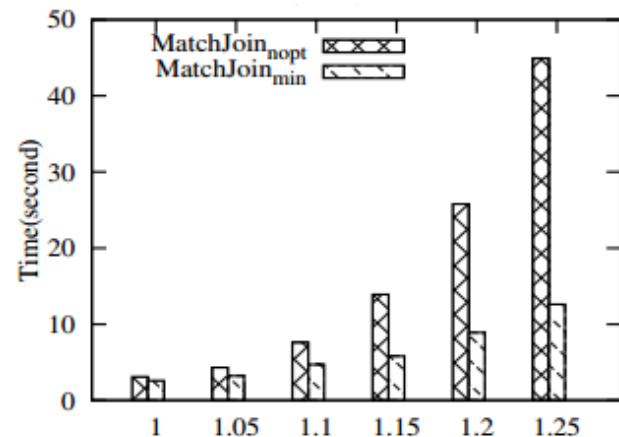
## ● 実世界データ

- ノード数548K エッジ数1,780K
- 平均ノード数・エッジ数5Kのビュー12個
- 横軸は(ノード数,エッジ数)
- ビューを使わないものと提案手法を比較
- 平均実行時間47%(minimal) 57%(minimum)に



## ● 合成データ

- ノード数 200K エッジ数  $|E| = |V|^\alpha$
- ビューの選択を効率化するかどうかで比較
- 平均実行時間54%に
- グラフが密であるほど有効



(f) Varying  $\alpha$  (synthetic)