

ICDE2015勉強会 : R17-3

Scalable SPARQL Querying using Path Partitioning

Buwen Wu ^{†#}, Yongluan Zhou ^{#*}, Pingpeng Yuan ^{†*}, Ling Liu [§], Hai Jin [†]

[†]SCTS/CGCL, Huazhong University of Science and Technology, Wuhan, China

[#]University of Southern Denmark, Denmark

[§]Georgia Institute of Technology, Atlanta, USA

Email: [†]{wubuwen, ppyuan, hjin}@hust.edu.cn, [#]zhou@imada.sdu.dk, [§]lingliu@cc.gatech.edu

担当 : 駒水@筑波大



概要

- 目的：SPARQLクエリ処理高速化のためのRDFグラフ分割
- 問題意識：
 - ~~分割をまたぐようなクエリの処理にオーバーヘッド~~
 - 分割したクエリの結果をマージする際にオーバーヘッド
 - 複雑なSPARQLクエリにありがち
- 解決策：良い感じの分割 **Balanced path partitioning (BPP)**
- 課題：BPP は難しい (NP-Hard かつ APX-Hard)
- 貢献：
 - BPP の困難性 (NP-Hard, APX-Hard) の証明
 - BPP の近似計算の提案
 - 分割に合わせたクエリの分割手法の提案
 - ベンチマーク (LUBM[11], SP²Bench[25]) と実データ (UniProt[6]) を用いた有効性検証

Path partitioning

- k -way path partitioning

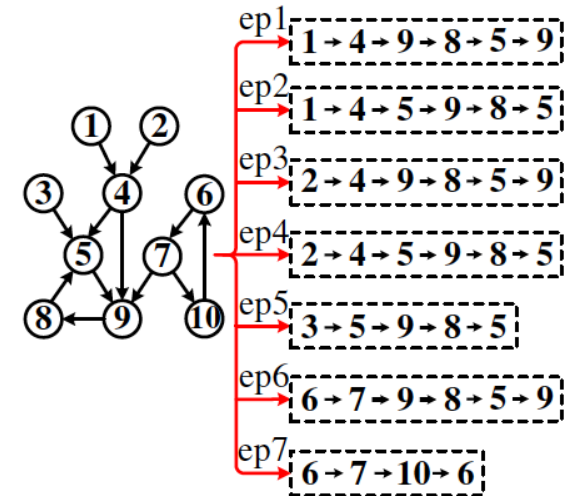
- 入力 : グラフ $G(V, E)$, パラメータ k
- 出力 : end-to-end paths 集合 $\{P_1, \dots, P_k\}$
 - end-to-end path: 源点から沈点までのパス
(注) 厳密な定義は論文を参照
 - メリット : S-O 結合を P_i 内で可能

- 指標

- バランス : パーティション毎のパス数
- ~~データ重複 : エッジの重複率~~
- クエリ効率 : merged vertex の数
 - merged: 異なるパス中の同一ノードが同一のパーティションに存在

- アルゴリズム

- クエリ効率を最大化 s.t. パーティション毎のパス数は **たかだか** $\text{ceil}(n/k)$
- ➔ NP-Hard (reduced from the balanced hypergraph partitioning)

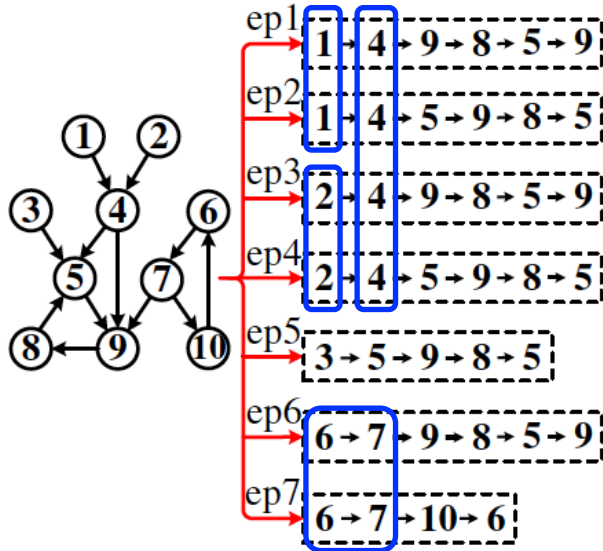


(a) End-to-End Paths

近似アルゴリズム

- 条件の緩和:

- クエリ効率を最大化 s.t. パーティション毎のパス数はたかだか $\text{ceil}(n/k)$ $\text{ceil}(2n/k)$



(a) End-to-End Paths

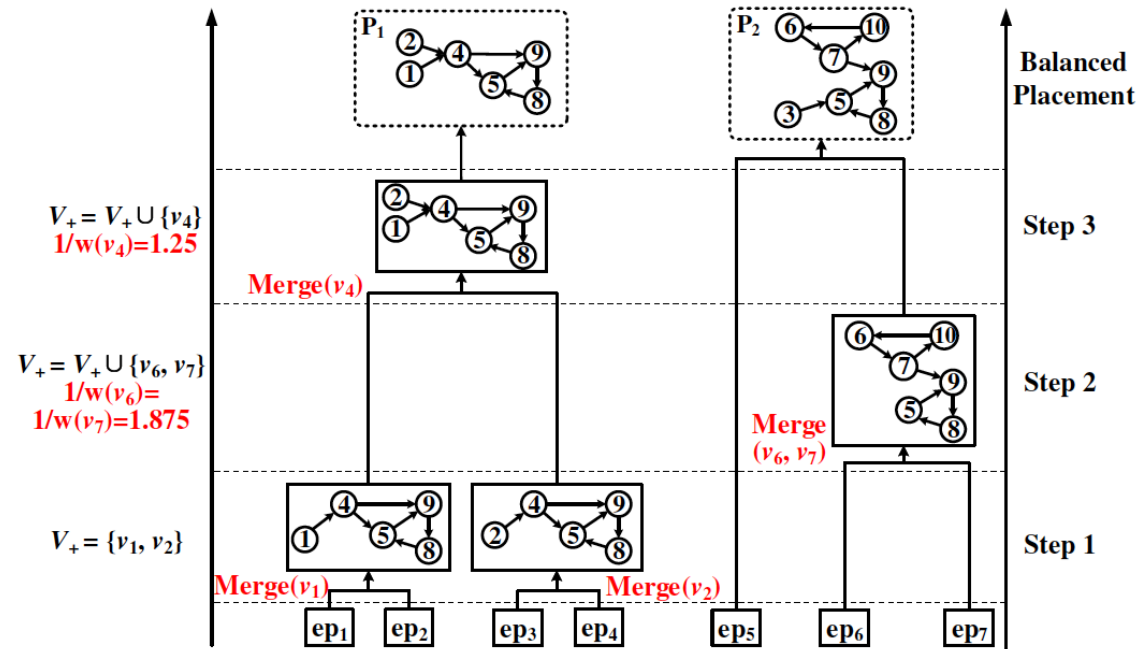


Fig. 4: Approximate Algorithm. The solid-line rectangle indicates the path group.

ICDE2015勉強会 : R17-4

Executing Queries over Schemaless RDF Databases

Güneş Aluç, M. Tamer Özsu, Khuzaima Daudjee, Olaf Hartig
Cheriton School of Computer Science, University of Waterloo

担当 : 駒水@筑波大

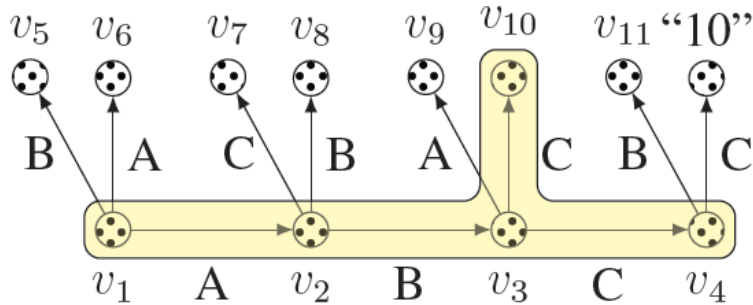


概要

- 目的：RDFデータベースに対するSPARQLクエリの高速処理
- 問題意識
 - 処理性能がデータの物理表現（列指向など）に依存
 - クエリのワークロードによって適切な物理表現が異なる.
- 課題
 - クラスタリングをどうやって決めるか？
 - 効率的な処理は？
- 解決策
 - スキーマレスな物理表現
 - ➔ **group-by-query**
 - 来そうなクエリに合わせたRDFデータのクラスタリング
- 貢献
 - **group-by-query**
 - クラスタリングの“良さ”の指標
 - クエリ評価モデル SE (Schemaless-Evaluation)
 - クエリ最適化
 - 実験

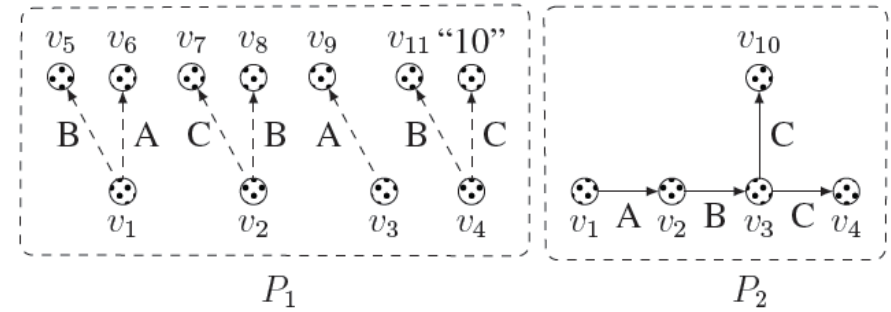
group-by-query : クラスタリング

RDFグラフデータ

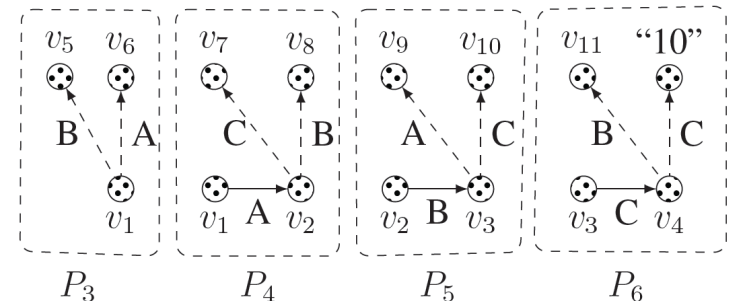


(b) Graph representation

クラスタリング例



(c) Clustering A

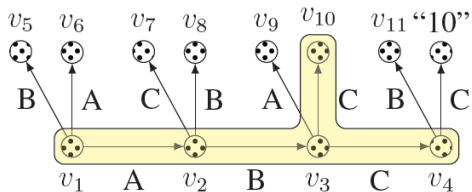


(d) Clustering B

group-by-query : クエリ処理

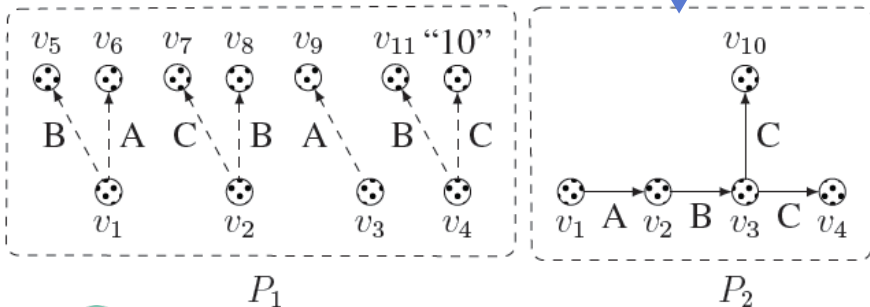
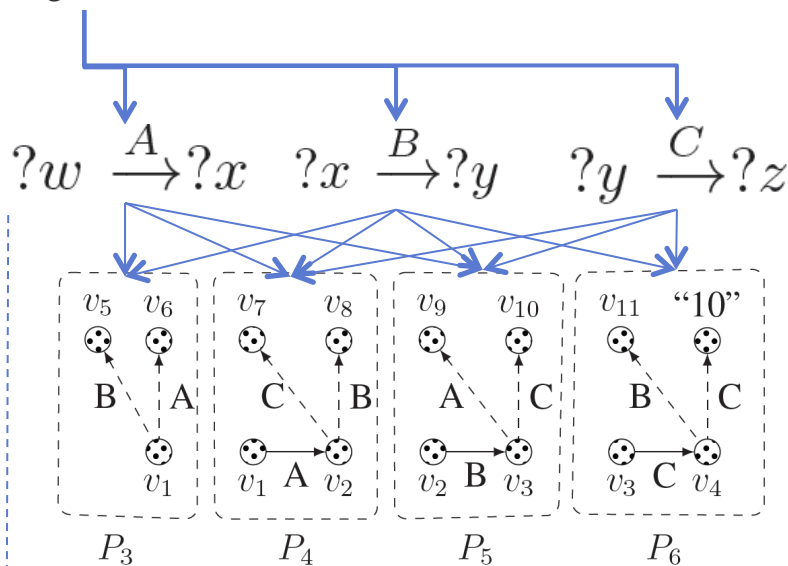
RDFグラフデータ

クエリ

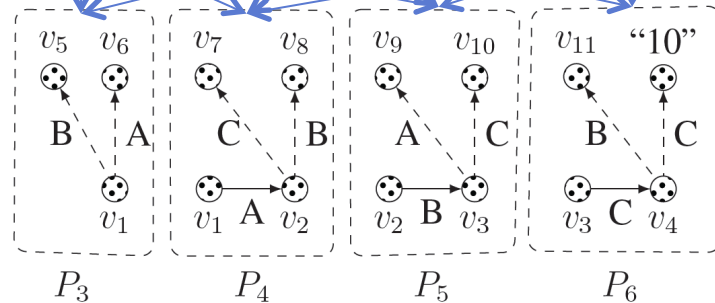


(b) Graph representation

$$?w \xrightarrow{A} ?x \xrightarrow{B} ?y \xrightarrow{C} ?z$$



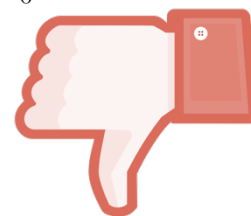
(c) Clustering A



(d) Clustering B



クラスタリング例



システム概要

