

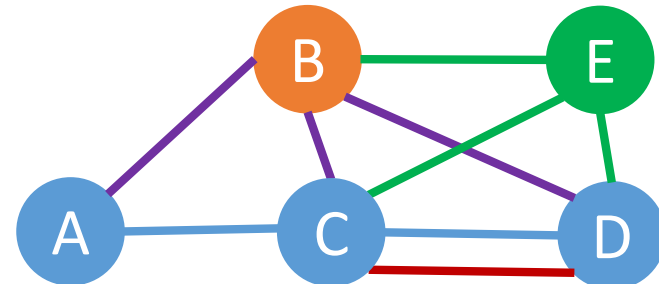
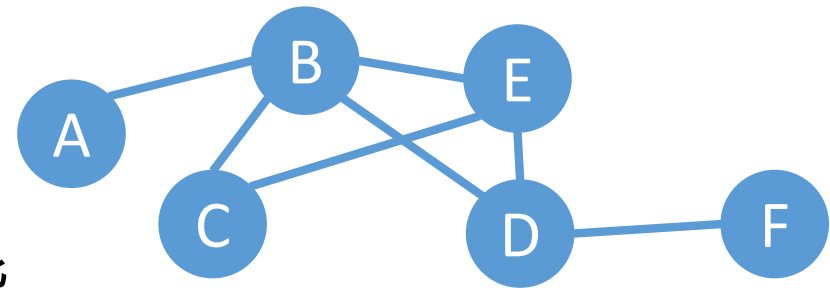
グラフデータ分析のための 特徴分析に関する研究開発

名古屋大学情報基盤センター

駒水 孝裕

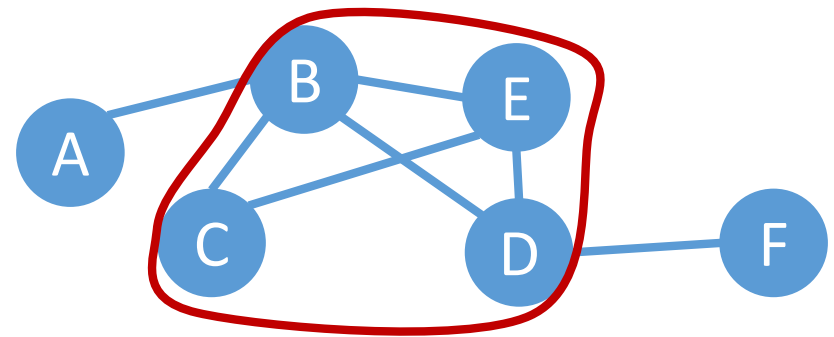
グラフ (Graph)

- データをノード同士をリンクで結ぶ表現方法
 - SNS, 知識データ, 科学データ, など
 - 例
 - ノード : 人
 - リンク : 友人関係
- 複雑なデータを記述可能
 - 例
 - ノード : 人, 会社, 場所
 - リンク : 友人関係, 同郷, 雇用関係, 所在地, etc.

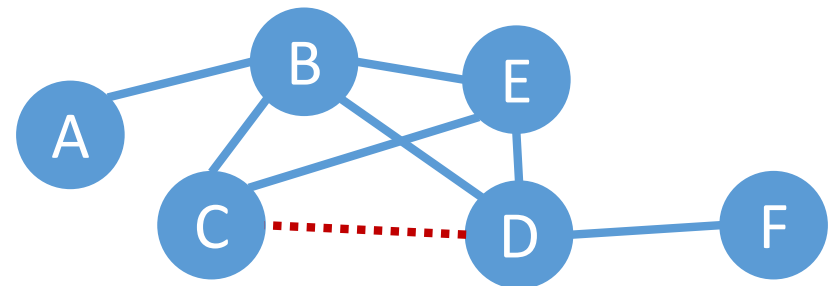


グラフデータ分析

- コミュニティ分析
 - 似た興味を持つユーザグループの発見
 - 近い研究を行っている研究者グループの検出



- リンク予測
 - 友達になりそうな人の発見
 - 将来の共同研究者の発見



研究目標

複雑なグラフデータの分析

- 対象：複雑なグラフデータ
 - ノードやリンクの種類が多様
- 本研究の目的：特徴設計
 - 分析をするためにグラフデータをどう特徴づけるか
 - 様々な従来手法
 - トポロジーに基づく設計
 - 近接ノード
 - ノード間のパス（リンクを数珠つなぎにしたもの）
 - ノードやリンクの属性に基づく設計
 - 埋め込み (embedding) に基づく手法

注目手法：meta-path [Sun et al. ASONAM 2011]

- ノードの種類に基づく（抽象）パス

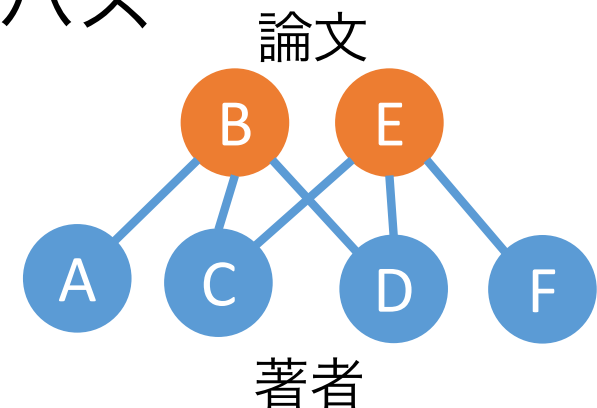
- 例：文献データ

- 著者→論文→著者

- 共著

- 著者→論文→会議→論文→著者

- 同じ会議に投稿した論文の著者



- パスの設計は手動

- データの専門家が設計

- グラフデータのノードやリンクの種類数が多くなると設計が大変

本研究の取組みと期待される成果

グラフデータ分析における特徴分析

- meta-path 設計の自動化
 - 今後の取組：
 - 類似度のサンプルから meta-path を推定
 - 期待される成果：
 - meta-path の手動設計の手間の軽減
 - meta-path を利用した手法への応用
 - metapath2vec [Dong et al. KDD 2017]
 - PathSim [Sun et al. PVLDB 2011], etc.
- 他手法による特徴分析の高度化

- Dong et al. metapath2vec: Scalable Representation Learning for Heterogeneous Networks, KDD, 2017
- Sun et al., PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks, PVLDB, 2011