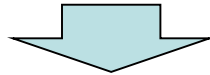


ビッグデータにおけるグラフの位置づけ

- 構造化データにおけるデータ間の関係性を捉え、また非構造化データを処理するのにグラフが適する

構造化データ

RDBに格納可能で、従来の技術は個々のデータを解析することを目的



・非構造化データに対する関心が高まる以前は非常に研究されていた

非構造化データ

従来のDBモデルに適合しない、構造定義の持たないデータ



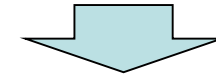
SNS



インターネット



Eメール



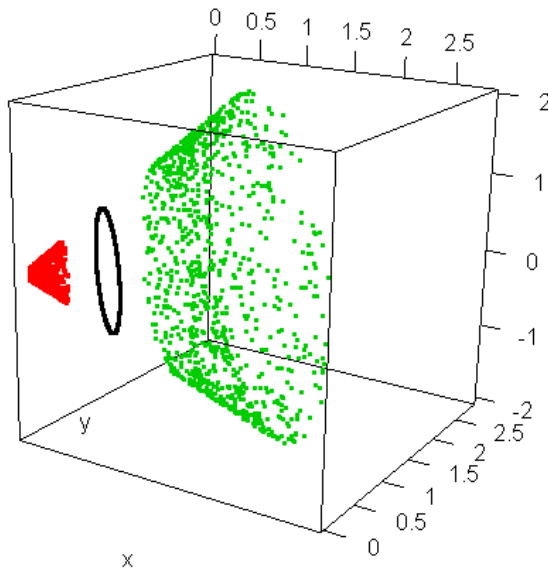
・データは大規模であるがデータ間の関係性は非常にスパースで、すべてのデータ間には関係性はない

実用上における課題

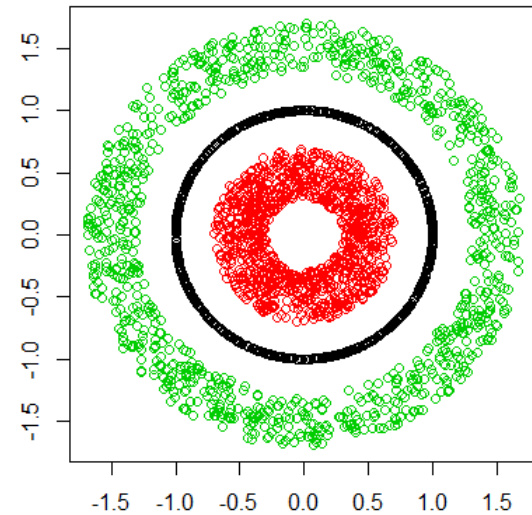
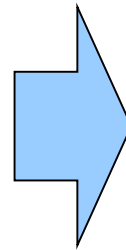
- 可視化
 - グラフを視覚的に理解できるため、技術に明るくない方に対してインパクトがある
- 多次元化
 - グラフデータに対して直接にSVMなどの機械学習手法を適用することは困難
 - 機械学習手法を適用するためにグラフデータを多次元化することが必要

Embedding

- Embedding とは多次元データを小さい次元に埋め込む技術
 - もともとは主に可視化のために利用
 - 埋め込みによりグラフを多次元化し, 従来研究されてきた機械学習手法 (SVM など) に適用することが可能になる



Original dataset (3-D)



Embedding (2-D)