

【VLDB 2013勉強会】

Session 6:
Web Data and Information Dissemination

担当：杉浦 健人(名大)

Some figures are copied from VLDB 2013 proceedings.

DisC Diversity: Result Diversification based on Dissimilarity and Coverage

Marina Drosou, Evaggelia Pitoura (U. Ioannina)

問題設定

入力：オブジェクトの集合, 半径 r

目的：多様な情報を持つ部分集合の選択

- 最も良いオブジェクトを1つ選択するよりも, 様々なオブジェクトの集合の方が有益な場合もある

例：性能の良いカメラが知りたい

- 高価格 高性能なカメラ1つ
- 価格帯に応じた性能の良いカメラの集合

情報として有益

DisC diversity

- 新しい diversity の考え方
- この尺度に基づいて部分集合を選択

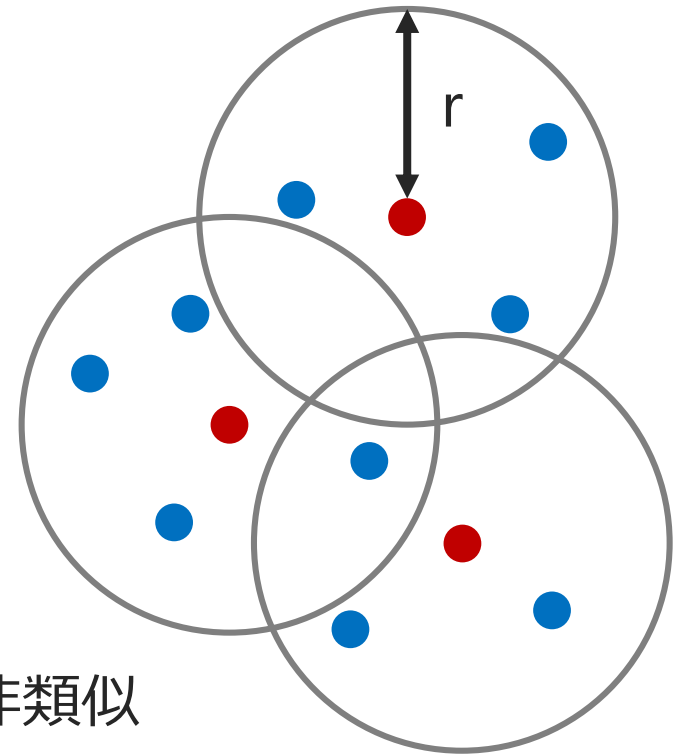
r-DisC diverse subset

Dissimilarity

- 部分集合内のオブジェクトは互いに非類似

Coverage

- 部分集合に含まれないオブジェクトは、部分集合のいずれかのオブジェクトと類似

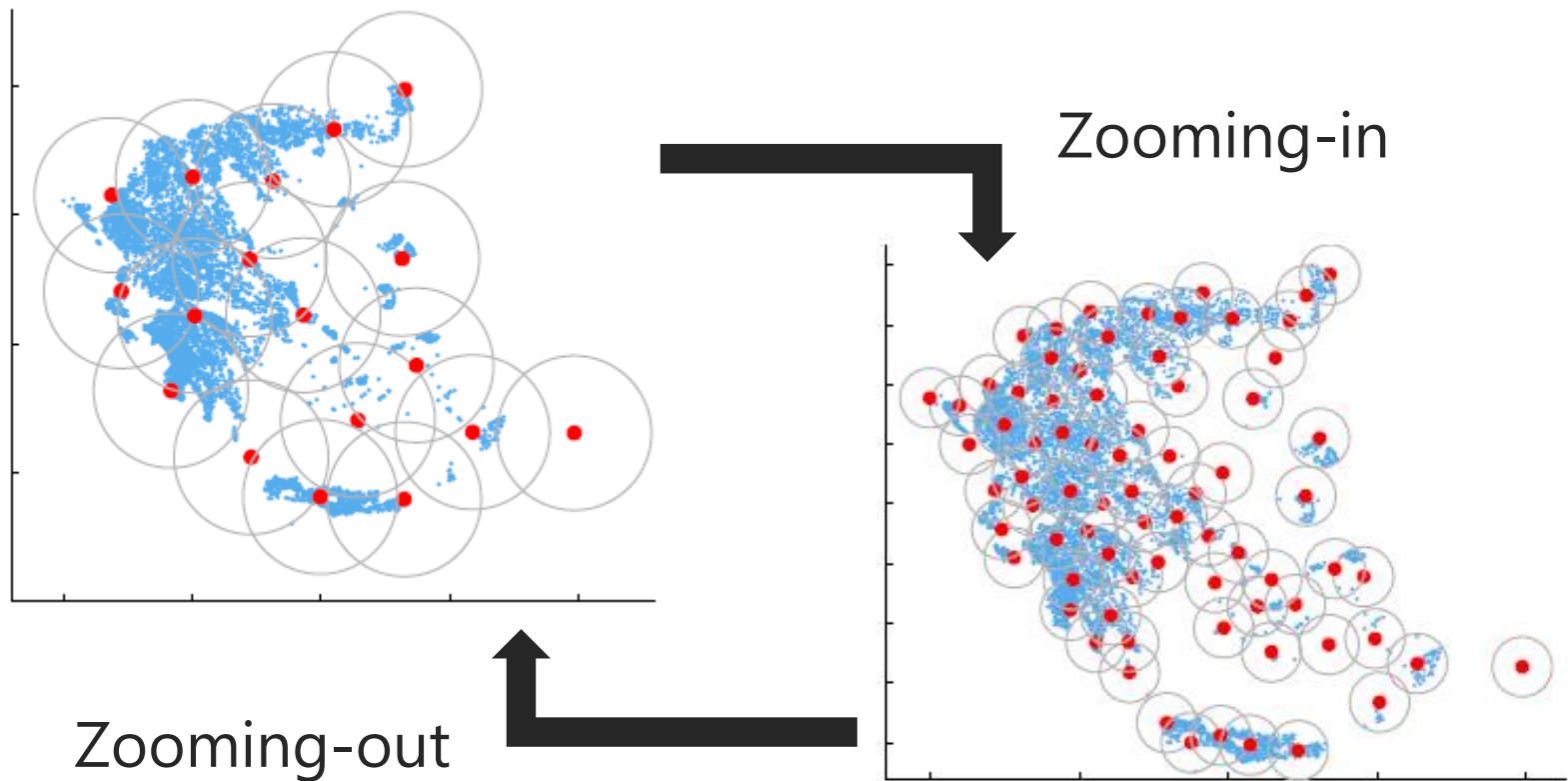


※ オブジェクト間の距離 = 類似度

※ 距離 $< r \rightarrow$ 類似

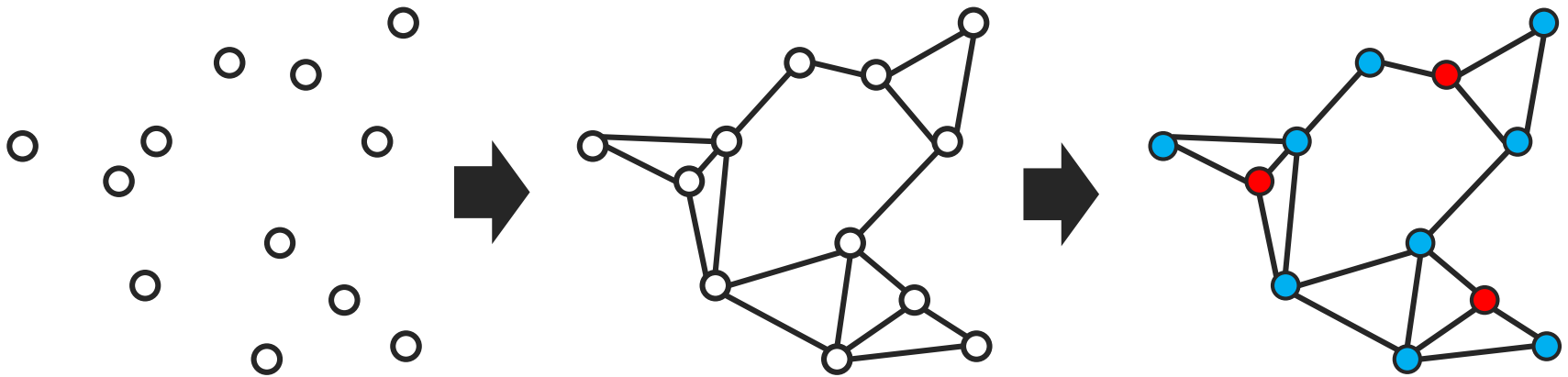
DisC diverse subset の利用方法

- ユーザは半径 r をチューニングパラメータとして、自分の知りたい粒度の情報を探す



DisC diverse subset の生成方法

- オブジェクト集合を**グラフに変換して処理**
 - 半径 r 以内にあるオブジェクトを辺でつなぐ
 - 部分集合の選択を**頂点の色付け問題**として処理



- 論文中では効率のよい色の塗り方や
Zooming-in / out の実行方法についても紹介