

P033 : ネットワーク科学の方法を用いたWebページ ネットワークの構造による分類

中川 帝人、鈴木 泰博(名古屋大学情報科学研究科) nakagawa.teito@b.mbox.nagyoa-u.ac.jp

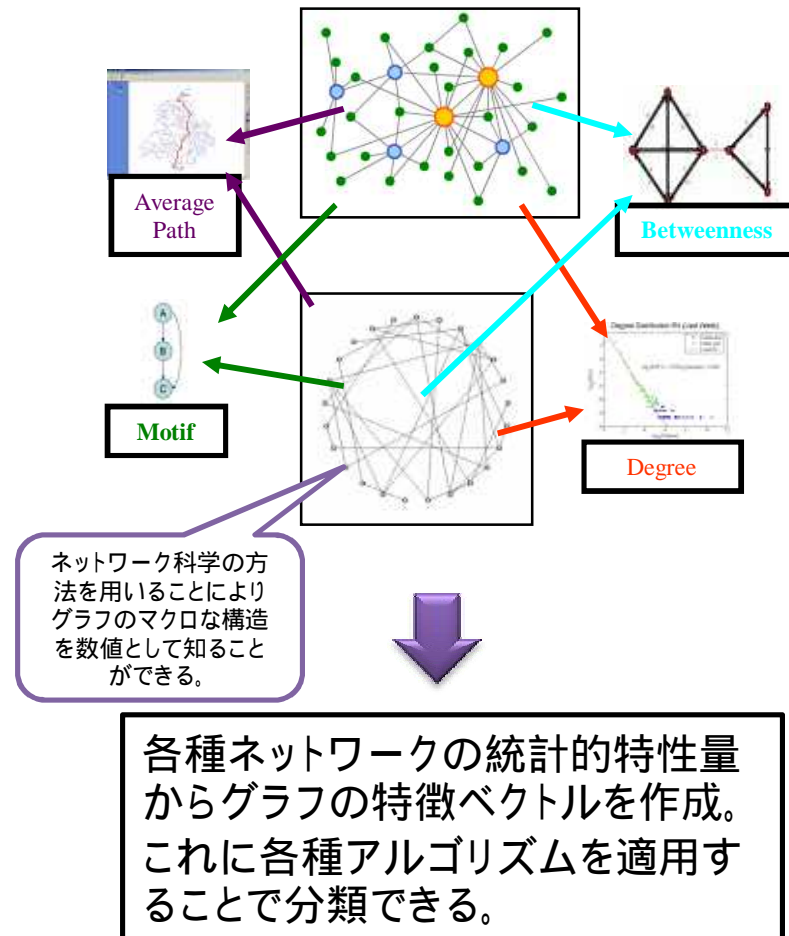
問題: グラフ分類問題

目的: 複数のグラフをマクロな構造から分類したい。

アイデア: ネットワーク科学の統計的特性量をグラフの特徴とする。

長所

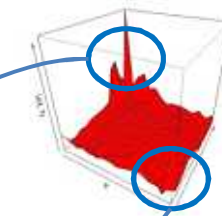
1. **グラフのマクロな特徴量を用いることにより、グラフの構造を理解できる。**
2. **計算量が少ない。**



実験と結果

- 対象: Webページのグラフ (N=1090)
- 特徴ベクトル:
 1. 平均次数 $\langle k \rangle$
 2. クラスタリング係数 $\langle C \rangle$
 3. 平均最短経路長 $\langle d \rangle$
 4. コミュニティ数 S
 5. 次数相関 r
 6. 次数1のノードの比率 I
- アルゴリズム: 自己組織化マップ (15 × 15)
- 結論: Webページのグラフにおいては完全グラフ型とツリー型の2つの頻出する構造が見られる。

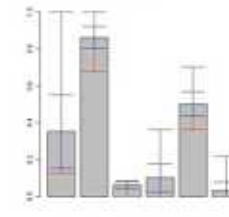
あるドメイン内のWebページを幅優先探索によって取得、グラフ化したもの。



自己組織化マップの結果

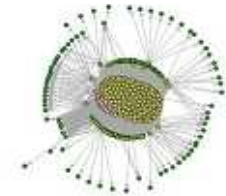
SOMの結果、2つのユニットを中心としたピークが見られた。
Webグラフは構造的に2つのパターンに類別できる。

1. 完全グラフ型 (1-15, n=68) ... 次数とクラスタリング係数が高い。

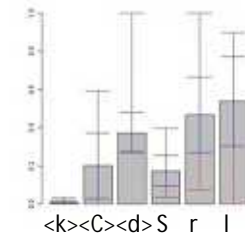


$\langle k \rangle \langle C \rangle \langle d \rangle S r I$

<http://ecs.co.uk>

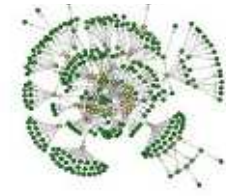


2. 木型 (15-1, n=98) ... 平均最短経路長が高く、次数が1のノードが多い。



$\langle k \rangle \langle C \rangle \langle d \rangle S r I$

<http://taxonomy.zoology.gla.ac.uk>



次数分布はべき分布に従うものが多数見られた。
スケールフリーな構造