



Title	コミュニティ分割されたコミュニティの特性解析
Author(s)	三好, 栄次; 鈴木, 育男; 山本, 雅人 他
Citation	情報処理北海道シンポジウム2008講演論文集, B-04
Issue Date	2008-09-19
Doc URL	https://hdl.handle.net/2115/51359
Rights	ここに掲載した著作物の利用に関する注意 本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うをお願いいたします。
Type	journal article
File Information	2008B04.pdf



コミュニティ分割されたコミュニティの特性解析

三好栄次* 鈴木育男 山本雅人 古川正志
(北大工)[†]

1 はじめに

近年, web技術の進歩, SNSなどによりweb上のネットワークが飛躍的に拡大されてきた. またそれに伴いネットワークからコミュニティを抽出する研究も盛んに行われている. 大規模なネットワークにおけるコミュニティ抽出の手法についてはGirvan, Newman, Clauset^[1]らによって提案されている.

本研究の目的は, コミュニティ分割されたコミュニティに対し, (1)平均クラスタリング係数, (2)各コミュニティ内における平均パス長を求め, 分割されたコミュニティ内でのコミュニティの特性解析を行った結果を報告する.

2 コミュニティ分割

2.1 コミュニティ分割とは

コミュニティ分割とは, ネットワーク上のノードをある指標に基づき, 特定の集団として抽出する操作であり, 本研究では指標にモジュラリティ (Modularity) を用いる. モジュラリティはコミュニティ内にエッジが多く, コミュニティ間にエッジが少ないノード集合が良く分割されたコミュニティである前提で, コミュニティの分割の評価を数値化したものである.

2.2 分割手法

Clauset^[1]の提案したコミュニティ分割手法は, コミュニティ同士を結合した際のモジュラリティの変化が最も大きい結合を選びながらコミュニティを抽出する手法である. 今, 対象となるネットワーク N を $N=(V, E)$, $V=\{v_1, v_2, \dots, v_i, \dots, v_n\}$, $E=\{e_{ij}=(v_i, v_j) | v_i, v_j \in V\}$ とし, m を総エッジ数, k_i をノード v_i の次数, Q_{ij} をモジュラリティ, 及び ΔQ をモジュラリティの変化分とする. このとき分割アルゴリズムは以下となる.

1. 全てのノードをコミュニティ内ノード数1のコミュニティとする.
2. ノード間エッジ $e_{ij}=(v_i, v_j)$ において, 全てのエッジ集合を E とし, ΔQ_{ij} の初期値をノードの次数 k_i, k_j を用い以下の条件で求める.

$$\Delta Q_{ij} = \begin{cases} 1/m - k_i k_j / 2m^2; & e_{ij} \in E \\ 0; & e_{ij} \notin E \end{cases} \quad (2)$$

3. それぞれのノードに関して,

$$a_i = \frac{k_i}{2m} \quad (3)$$

を求める.

4. 最も値の大きい ΔQ_{ij} を選び, i と j に相当するコミュニティを統合する.
5. コミュニティ間のエッジ $e_{ij}=(c_i, c_j)$ に関して, i, j に関連するエッジの集合を E_{ij} とする. ΔQ の値を以下に従い更新する.

$$\Delta Q'_{jk} = \begin{cases} \Delta Q_{ik} + \Delta Q_{jk}; & e_{ik}, e_{jk} \in E_{ij} \\ \Delta Q_{ik} - 2a_j a_k; & e_{ik} \in E_{ij}, e_{jk} \notin E_{ij} \\ \Delta Q_{jk} - 2a_i a_k; & e_{jk} \in E_{ij}, e_{ik} \notin E_{ij} \end{cases} \quad (4)$$

6. a_j を更新する.
7. ΔQ が正である, またはコミュニティが一つになるまでステップ3からステップ6までを繰り返す.

3 数値計算実験

3.1 実験条件

本研究ではAmazon.comの米国政治に関する本のオンライン販売のネットワーク^[2]をコミュニティ分割の対象ネットワークとする. このネットワークでは本をノードとし, ある本を購入した人が別の本を購入したとき, ノード間にエッジが張られる.

実験に用いられるデータは $|V|=105$, $|E|=441$ である.

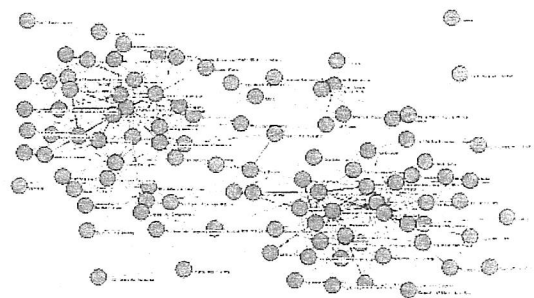


Fig.1 オンライン販売のネットワーク. 可視化ツールとしてPajek^[3]を使用

*miyoshi@complex.eng.hokudai.ac.jp

[†]札幌市北区北14条西9丁目北海道大学工学部

使用したネットワークを図 1 に示す. なお今回使用した計算機の性能は, Pentium4 2.60GHz メインメモリ 1.0GB である.

3.2 実験方法

Clauset のコミュニティ分割方法を上記に述べたネットワークに適用し, 分割を行う. その後, 分割された各コミュニティに対し, それぞれの平均クラスタリング係数 C , 平均パス長 L を計算し, コミュニティごとに数値を比較する. コミュニティごとのクラスタリング係数の計算では, コミュニティ間に張られているエッジは無いものとして, 計算する.

3.3 結果

数値計算実験で得られたコミュニティ分割数は 4 であり, 各コミュニティの L と C を表 1 に示す. 又, 得られたコミュニティネットワーク全体の分割を図 2 に, それぞれのコミュニティを図 3 に示す.

このとき得られたモジュラリティ Q は 0.501, であり, またネットワーク全体の平均パス長 L は 3.079, 平均クラスタリング係数 C は 0.488, となっている.

モジュラリティの値からある程度まとまりのあるコミュニティとして分割が得られている.

4 考察

ネットワーク全体に対し, 分割されたコミュニティでは平均クラスタリング係数は高く, 平均パスも 2 前後であり, コミュニティ内のノード間距離も短い. 更に分割されたコミュニティ内ではほぼ確実にセットで買われている本が存在している. 又, その本を中心としたハブが, コミュニティを形成しており, そこでは平均パス長が低くなっている. 平均クラスタリング係数においては, モジュラリティの定義から高くなったことが分かる. これらから, コミュニティ内ではスモール・ワールドを形成していることも観察できた.

5 おわりに

本研究では, コミュニティ分割とそのコミュニティの特性解析を行った. 実験より, 抽出されたコミュニティが高いクラスタリング係数, 短い平均パス長を持ち, スモール・ワールドを形成しているのが判明した. 今後の課題としてプログラムの高速化, 別の指標を用いた手法, 分割を評価するための新たな情報の追加, 異なるデータでの実験と結果の比較が必要となる.

Table 1 各コミュニティ情報

	$ V $	$ E $	L	C
(1)	49	201	2.11	0.529
(2)	3	3	1	1
(3)	12	25	1.8	0.636
(4)	41	176	2	0.554

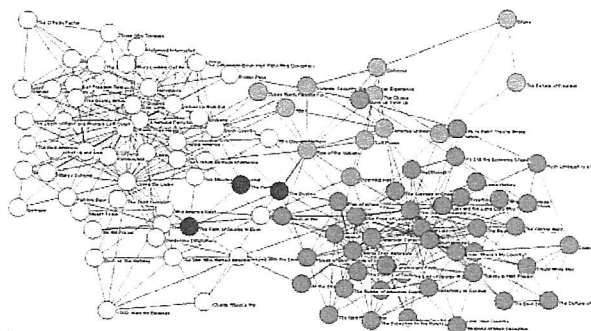


Fig.2 ネットワーク全体のコミュニティ分割

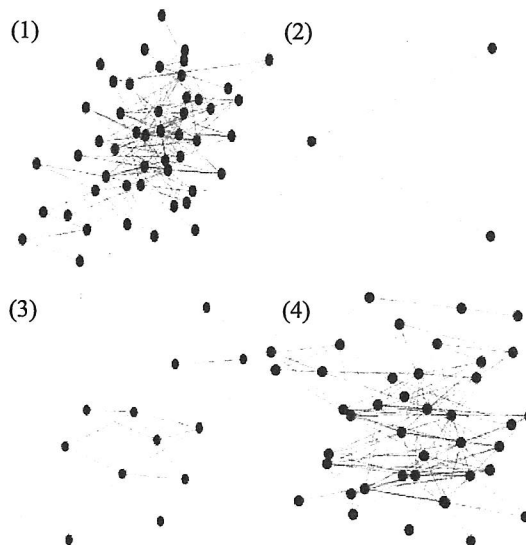


Fig.3 分割されたコミュニティをコミュニティ間のリンクを除き, コミュニティごとに表示

参考文献

- [1] Aaron Clauset, M. E. J. Newman, and Cristopher Moore, "Finding community structure in very large networks", arXiv:cond-mat/0408187v2 (2004).
- [2] Network data
<http://www-personal.umich.edu/~mejn/netdata/>
- [3] Networks / Pajek
<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>