

# 新しいグラフクラスタリング手法の開発

## Development of a novel method for graph clustering

### 1. 序論

グラフクラスタリングとは、点（ノード）とそれらをつなぐ線（リンク）からなる抽象構造としてのグラフを、リンクが密なサブグラフに分断することである。適切な分け方の全探索は計算時間の面で不可能なため、何らかのアルゴリズムが必要である。

本研究の目的は、精度と計算量のバランスが良い DA 法<sup>1-2)</sup> に改良を加えることで、計算精度を保ちながら計算量を減らすことである。

### 2. アルゴリズム

#### 2.1 DA 法

階層的クラスタリング手法の一つ。まずノードをランダムに 2 分する。次に各ノードに次数（リンク数）に応じた重みを持たせる(1)。

$$w(i) = \frac{\kappa_{r(i)}}{k_i} - a_{r(i)}, \quad (1)$$

ここで  $\kappa_{r(i)}$  は、ノード  $i$  が、自身が属するクラスター  $r$  のノードに向けられた次数、 $k_i$  はノード  $i$  の次数、 $a_{r(i)}$  はクラスター  $r$  が持つ総次数を全体の総次数で割った値。

この重みの小さい順にノードを並べ、次の確率分布に従ってノードを 1 つ選択し、他方のクラスターに移す。

$$P(q) \propto q^{-\left(1 + \frac{1}{\ln(n)}\right)}. \quad (2)$$

$q$  はノードの重みの順位、 $n$  はノードの総数。これを、 $\alpha n$  回繰り返す。 $\alpha$  は計算回数を決める任意のパラメータである。

この手順による分割を、適切なクラスター分割が行われるまで繰り返す。適切に分割が行われたかは、下記の通り定義される Modularity が極大値を取るかどうかで判断する。

$$Q = \sum_r (e_{rr} - a_r^2). \quad (3)$$

$e_{rr}$  は、クラスター  $r$  内のノード同士を繋ぐリンクの数を、全体のリンクの総数で割った値。

#### 2.2 新手法

DA 法により  $\alpha n$  計算を繰り返した後、確率的手法をやめ、重みの最も小さなノードを他のクラスターに移動させる決定論的方法に切り替え

る。これは、最初の確率的手法により局所最適解を抜け出したと考え、決定論的手法に切り替えても最適解にたどり着けると考えたためである。この切り替えによって計算時間を大幅に短縮することができる。

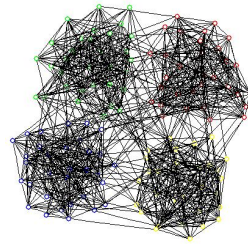


図1 実験で用いたグラフ

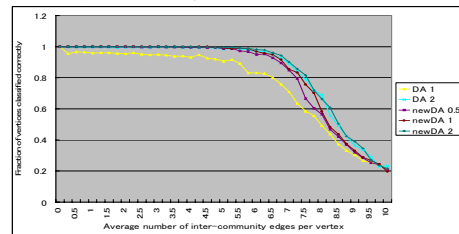


図2 DA 法と新手法の精度

### 3. 実験と考察

32 ノード×4 クラスターの、計 128 ノードからなるグラフを作成 (図 1)。各ノードが持つリンクを 16 本とし、他クラスターに接続するリンクの割合を変えながら、精度について各々 100 回のアンサンブル平均を取った。 $\alpha$  を DA 法は 1,2、新手法は 0.5,1,2 とした。図 2 は横軸が他クラスターに接続するリンクの割合、縦軸が精度。

計算量は、 $\alpha$  にほぼ比例する。 $\alpha = 0.5$  の新手法が、 $\alpha = 1$  の DA 法より明らかに精度が高い (図 2)。これは、最適解へ収束する速度の違いによるものと考えられる。

### 4. 結 論

DA 法の確率的試行を、途中から決定論的手法に切り替えることにより、計算速度、計算精度の両面において性能を向上させることができた。

### 参考文献

- 1) J. Duch and A. Arenas, Phys. Rev. E, 72, 027104 (2005)
- 2) L. Danon, A. Diaz-Guilera, J. Duch and A. Arenas, J. Stat. Mech., 5, P09008 (2005)