

Artisoc で作成する Girvan-Newman コミュニティ 抽出シミュレーション

末 木 俊 之*

Girvan-Newman Community Extraction Simulation with Artisoc

Toshiyuki SUEKI*

キーワード：媒介中心性 モジュラリティ (modularity) ランダムネットワーク スケールフリー
ネットワーク

1. 媒介中心性指数とコミュニティ抽出

2010年度研究紀要⁽¹⁾では、artisoc textbook⁽²⁾を使用し、ネットワークでつながったノードを円内に表示する簡単なシミュレーションを検討した。(1)重要度の高いノードが、より円の中心に来るように配置すること、(2)より密なリンクで互いに結合し合っているノードを集めて表示すること、以上の2点がポイントであった。そして、各ノードの重要度を測る指標として媒介中心性指数を計算し表示するシミュレーションを artisoc で作成した。単なる計算処理をわざわざシミュレーションソフト artisoc で行わなくとも良いが、artisoc ではネットワークでつながったものを処理すること、計算処理の経過・結果をグラフ表示することが容易で便利である。

昨年の段階で各ノードの媒介中心性指数を計算できる環境が整ったので、ノードの重要度(媒介中心性)に基づいて表示位置を制御することが可能になった。今回はさらに媒介中心性指数計算を応用し、切断法によるコミュニティ抽出⁽³⁾を行う artisoc シミュレーションを作成した。

グラフ上のいくつかのリンクを仮切断した後、リンクをたどってつながるノードグループに同じコミュニティ番号を付与することによりコミュニティ抽出(コミュニティ分割)が実現できる。切断するリンクを選択するのに辺の媒介中心性指数が使われる。

切断法については、『ネットワーク科学の道具箱 つながりに隠れた現象をひもとく』⁽⁴⁾などを参考にできる。同書 p. 50にて Girvan-Newman コミュニティ抽出(GN法)として紹介されている。(図1.)は疑似プログラム風に表現された Girvan-Newman コミュニティ抽出アルゴリズムである。

$G(V, E)$ は、頂点 V (Vertex) と辺 E (edge) の集合であるグラフを意味する。アルゴリズムを文章で表現すると「グラフの全ての辺の媒介中心性指数(C_e)を計算し、最大の C_e 値を持つ辺をグラフから除去する。それをグラフから全ての辺が無くなるまで繰り返し、最後に Q 値の(局所的)最大値とそれが得られた時点のグラフ(頂点と辺情報)を報告する。」となる。

*人間健康学部 健康栄養学科

```

1  do
2    calculate betweenness  $C_e$  for all edge  $E$  in  $G(V, E)$ 
3    delete edge  $e$  with maximum value of  $C_e$ 
4     $G \leftarrow G(V, E - \{e\})$ 
5  while  $G$  is not empty
6    report  $G$  with (local) maximum value of  $Q$ 

```

(図 1.) Girvan-Newman コミュニティ抽出アルゴリズム⁽⁵⁾

Q 値とはモジュラリティ (modularity) と呼ばれる数値である。この数値が大きいほど同じ属性を持つグループ(コミュニティ)に属するノード同士をつなぐリンクの割合が、異なる属性を持つノード同士をつなぐリンクの割合よりも多いことを示す。リンクをたどりお互いにつながる全ノードには同じグループ番号を付与する。そして同グループ番号を持つノードは同属性を持つと考え、大元のグラフ(リンクが削除されていないグラフ)の Q 値を計算する。その値が最大となった時点が同じグループ(コミュニティ)に属するノード間をつなぐリンクの割合が最大であると判定できる。

Q 値は、(図 2.) の式で定義されている。

$$Q = \sum_p e_{pp} - \sum_p a_p b_p$$

(図 2.) Q 値モジュラリティ (modularity)⁽⁶⁾

e_{ij} は、(図 3.) の式で定義される。グラフが有効グラフである場合に、リンクの始点ノードがグループ i に属し、リンクの終点ノードがグループ j に属している属性ペア (i, j) を持つリンクの全リンクに占める割合である。 n_{ij} は属性ペア (i, j) を持つリンクの数、 M はグラフのリンク総数である。

$$e_{ij} \stackrel{\text{def}}{=} n_{ij}/M$$

(図 3.) e_{ij} ⁽⁷⁾

従って(図 2.) 右辺の 1 つ目の Σ 式は、始点と終点と同じグループに属するリンクの割合を意味する。

a_i, b_j は、(図 4.) で定義される。 a_i は、始点がグループ i に属するリンク割合(全リンクからランダムに選択したリンクの始点がグループ i に属する確率)を、 b_j は終点がグループ j に属するリンクの割合(全リンクからランダムに選択したリンクの終点がグループ j に属する確率)である。

$$a_i \stackrel{\text{def}}{=} \sum_j e_{ij}$$

$$b_j \stackrel{\text{def}}{=} \sum_i e_{ij}$$

(図 4.) a_j, b_j の定義⁽⁸⁾

従って(図 2.) 右辺の 2 つ目の Σ 式は、は、ランダムネットワーク(ノードを連結するリンクを完全ランダムに付与したと仮定したネットワーク)において始点と終点と同じグループに属するリンクの割合(存在確率)である。従っ

て Q 値 = 0 であれば、そのネットワークは完全ランダムネットワークである。 Q 値が大きくなるほど「ランダムネットワークの性質が低い」、「始点と終点と同じ属性を持つノードになるリンクの割合が多い」、「同じグループに属するノード同士が密につながる性質を持つネットワークである」と判定できる。

具体的には、以下の(a)～(d)の処理を1ステップとしてグラフから全てのリンクが消えるまでステップを繰り返すことになる。大元のグラフをA、辺を仮削除したグラフをBと呼ぶことにする。最初は、グラフA=グラフBである。

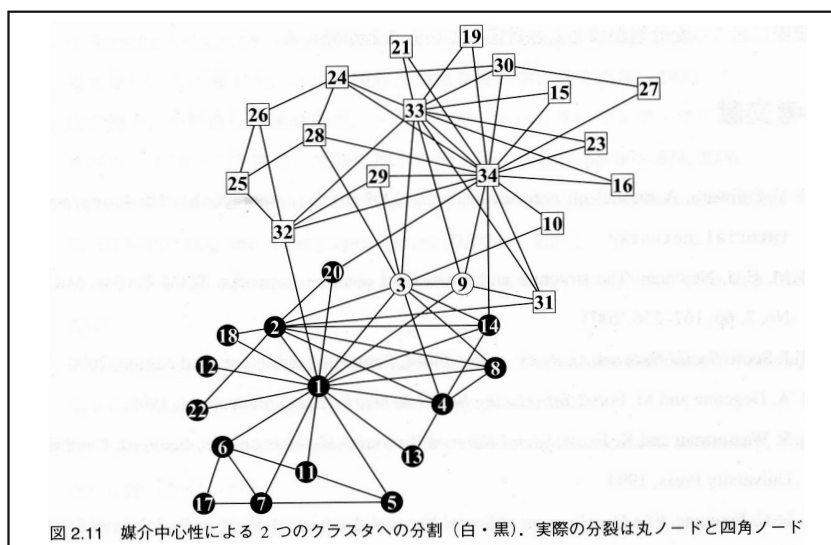
- (a) グラフB全辺について辺の媒介中心性指数(C_e)を計算する。
- (b) 最大の C_e 値を持つ辺をグラフBから仮除去する。大元のグラフAには手を加えない。
- (c) 全ノードにグループ番号(コミュニティ番号)を付与する。グラフBのリンクをたどってお互いにつながる全ノードに同じグループ番号(コミュニティ番号)を付与する。グラフAのノードにも同じグループ番号

を付与する必要がある。

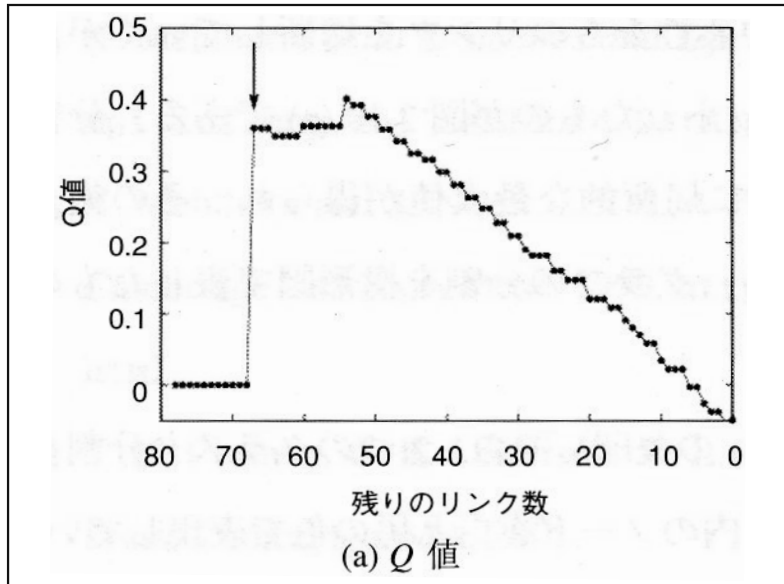
- (d) グラフA上で Q 値を計算し報告する。
(ノードの属性としてグループ番号を使い計算する)

最大 Q 値が得られた時点(ステップ)で各ノードに付与されたグループ番号(コミュニティ番号)を用いてノードのグループ分けを行うと、同じグループに属する(同じ属性を持つ)ノード同士のリンクの割合が最大となるグループ分けが実現されていることになる。

『ネットワーク科学の道具箱 つなぐに隠れた現象をひもとく』⁽⁴⁾ではGirvan-Newmanコミュニティ抽出アルゴリズムでコミュニティ抽出を行った実例が紹介されている。(図5.)はそこに登場する米国のある大学の空手部部員の友人関係を表現したグラフ図である。1から34までの番号がつけられている丸ノード、四角ノードが人を表し、リンクは部活動外での友人関係を表している。番号1のノードが師範、番号34のノードが部長である。この空手部は結果的に(図5.)に示される丸ノードと四角ノードの2つの派閥(グループ)に分裂したそうで



(図5.) 米国のある大学の空手部部員の友人関係を表現したグラフ図⁽⁹⁾



(図6.) Q 値推移図⁽⁹⁾

ある。

(図6.) はリンクを削除していくに従って変化するグラフの Q 値を折れ線グラフで表示したものである。最初 0 であった Q 値が急激に増加し、その直後に局所的な最大値が得られたステップが矢印で示されている。その後、Q 値は最大値を経て減少に向かっている。(図6.) の矢印で示された Q 値の局所的最大値に対応するステップ時点でのグループ分割 (コミュニティ抽出) が、(図5.) に示された実際に分裂した空手部の 2 つの派閥とほぼ対応する結果となっている。

今回作成した artisoc シミュレーションでは、この空手部の友人関係情報 (ノードとリンク) をエクセルで csv 形式のテキストファイルとして作成したものを読み込ませた後シミュレーションを動作させ、(図5.) (図6.) で示した Q 値計算とグループ分割 (コミュニティ抽出) を検証することを最初の目標とした。

2. Girvan-Newman コミュニティ抽出 artisoc シミュレーションの概要

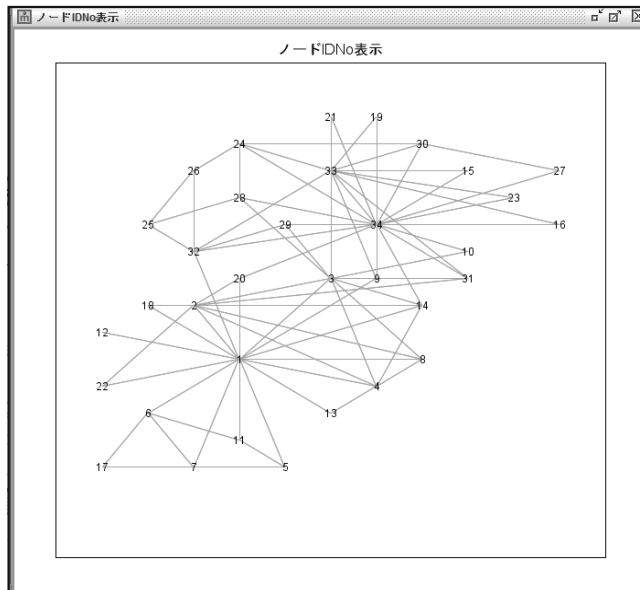
2.1. 図・グラフ表示の概要

2010年度研究紀要で作成した媒介中心性指数計算シミュレーション⁽¹⁾ と同じく、「ノード IDNo 表示図」(図7.)、「辺の媒介中心性指数 (C_e) 表示図」などを用意した。

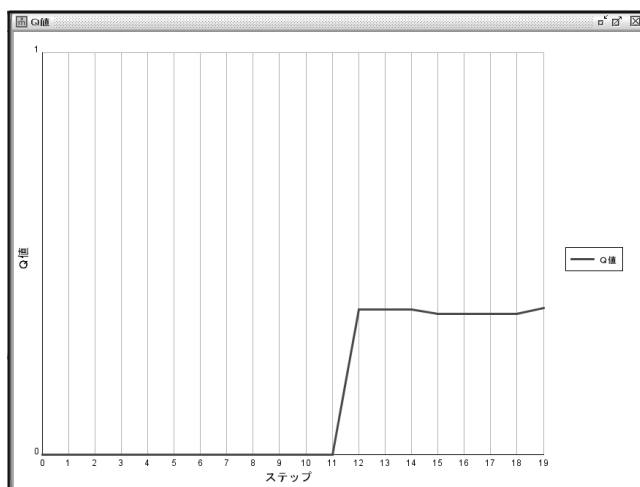
(図7.) では、(図5.) の空手部の友人関係図の人の配置・リンクが再現されている様子が窺える。

(図8.) は、シミュレーションステップ19終了時点の「Q 値表示図」である。「Q 値表示図」を観察すると、ステップ12の時点で Q 値の局所的最大値が現れていることが分かる。シミュレーションのステップごとの処理については後述するが、ステップ11で辺をグラフから仮削除した効果 (Q 値、ノードのコミュニティ分割) は 1 ステップ後のステップ12の図に現れてくる。ここが少々少しわかりにくい点になっている。

(図9.) はステップ12終了後の「コミュニティ ID 図」である。この図からステップ11での辺



(図7.) 「ノード IDNo 表示図」



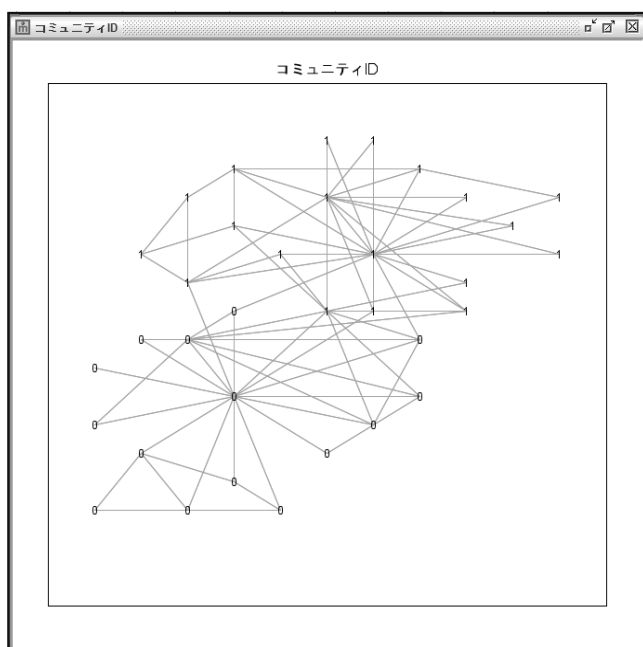
(図8.) ステップ19終了後の「Q 値表示図」

の仮削除によりコミュニティ ID = 0、1 の 2 つのコミュニティができている様子が窺える。

2.2. 計算処理の概要

Girvan-Newman コミュニティ抽出アルゴリズムに基づいてコミュニティ抽出を行なうために、以下(1)、(2)の2つの機能を別々に実行できるようにした。

(1)各処理ステップで全辺の媒介中心性指数(C_e)を計算し、全ノードにグループ番号(コミュニティ番号)を付与する。そしてQ 値を計算する。さらに最大の媒介中心性指数 C_e 値を持つ辺をグラフから仮削除し、削除したリンク情報等を csv 形式のテキストファイル「output.txt」に書き出して報告する。ステップ処理終了後グラフ表示(「Q 値表示図」など)が更新



(図9.) ステップ12終了後の「コミュニティ ID 図」

される。全てのリンクが削除されるまでこのステップ処理を続け停止する。(図10.)は「output.txt」ファイルをエクセルで開いた様子である。

「FromNode」変数と「ToNode」変数のペアが各処理ステップにおいて削除されたリンクを記録したものである。

例えば図の2行目の記録からは、ステップ1にてIDNo=32のノードとIDNo=1のノードをつなぐリンクが削除されたことが確認できる。

(1) 処理終了後、「Q 値表示図」または「output.txt」テキストファイルによって Q 値の局所的最大値が出現するステップを探す。そしてその1つ前のステップ終了時点で得られるグループ分割に基づき付与された全ノードのグループ番号 (コミュニティ番号) 情報を csv 形式のテキストファイル等で取得する必要がある。これを実現するために2つ目の機能(2)を設けた。

	A	B	C	D	E	F	G	H	I	J	K
1	ID	X	Y	Layer	Direction	IDNo	処理前Q値	最大媒介中心性指数	コミュニティ総数	FromNode	ToNode
2	0	0	0	0	0	0	0	71.39285714	1	32	1
3	0	0	0	0	0	0	0	66.89517705	1	1	3
4	0	0	0	0	0	0	0	77.31739927	1	1	9
5	0	0	0	0	0	0	0	82.00290592	1	14	34
6	0	0	0	0	0	0	0	123.2329171	1	20	34
7	0	0	0	0	0	0	0	100.2055556	1	3	33
8	0	0	0	0	0	0	0	143.6269841	1	2	31
9	0	0	0	0	0	0	0	109.25	1	2	3
10	0	0	0	0	0	0	0	107.6666667	1	3	4
11	0	0	0	0	0	0	0	142.75	1	3	8
12	0	0	0	0	0	0	0	285	1	14	3
13	0	0	0	0	0	0	0.359960552	16.83333333	2	10	34
14	0	0	0	0	0	0	0.359960552	18.18333333	2	34	28
15	0	0	0	0	0	0	0.359960552	18	2	3	10

(図10.) エクセルで開いた「output.txt」テキストファイル

(1)処理終了後に csv 形式テキストファイルとして書き出された「output.txt」テキストファイル（各処理ステップで削除したリンク情報）情報を読み込むためのエージェント集合型 Universe 変数（グローバル変数）「deleteLink」を新たに用意した。「deleteLink」変数に情報をセットしてから処理(2)を実行する。この操作は少々面倒であるが、(1)の処理結果がメモリに残らないので必要な処理である。

(2)指定されたステップまで、記録された削除リンクを仮削除する。「deleteLink」変数に読み込まれているリンクの削除記録通りに、指定されたステップまで単純にリンクを仮削除する。そしてリンクが仮削除されたグラフで全ノードにグループ番号（コミュニティ番号）を付与する。また全ノードの媒介中心性指数も計算する。そして全ノード情報を csv 形式のテキストファイル（「input.txt」ファイル）に書き出してシミュレーションを終了する。

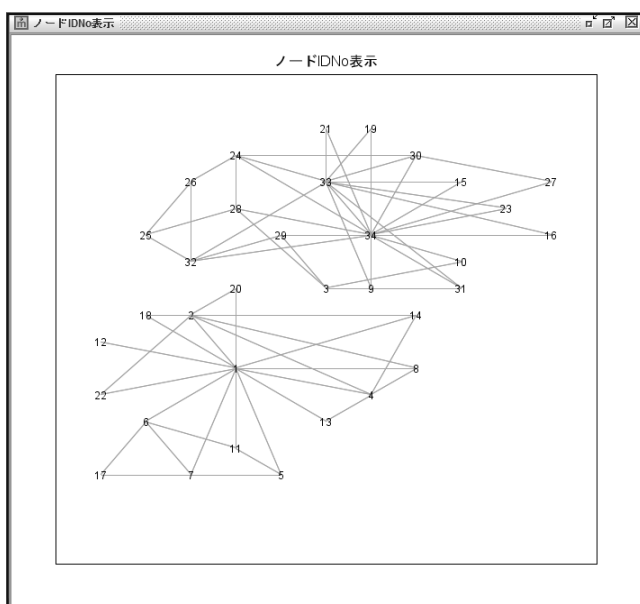
1つのシミュレーションで(1)、(2)処理を切り替えて実行させるために、今回もコント

ロールパネルを使用した。コントロールパネルのスライダーを0にセットしてシミュレーションを起動すると(1)処理を実行し、スライダーが0以外であれば、(2)処理を実行するように作った。

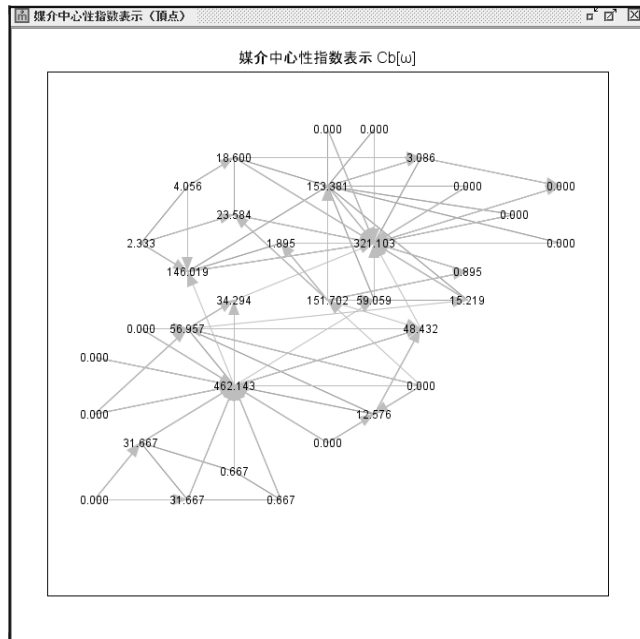
(2)処理における媒介中心性指数計算は、(1)処理と同じものは使えない。(1)処理ではリンクが仮削除されたグラフで計算するが、(2)処理では大元のグラフで計算しなければならない。結局2種類の媒介中心性指数計算サブルーチンを用意した。

(図11.)は(2)処理終了後のノード IDNo 表示図である。11個のリンクが仮削除された結果、2つのグループに分裂している状態で停止している様子が窺える。この「ノード IDNo 表示図」では仮削除されたリンクを表示しないのでグループ分割が分かりやすい。

(図12.)も(2)処理終了後のノードの媒介中心性指数表示図であるが、リンクが仮削除されたグラフではなく大元のグラフに基づいて媒介中心性指数計算されている様子が窺える。



(図11.) (2)処理終了後のノード IDNo 表示図



(図12.) (2)処理終了後の媒介中心性指数表示図

米国のある大学の空手部が丸ノードと四角ノードで表わされた2つのグループ（コミュニティ）に分裂した実際の状況図（図5.）と、今回 artisoc で作成した Girvan-Newman コミュニティ抽出シミュレーションでコミュニティ分割した（図11.）を比較すると、ほぼグループ（コミュニティ）分割が一致していることが分かる。

ノード番号3、9の2名の所属するコミュニティが食い違っているだけであった。

3. Universe 変数（グローバルに使う変数）

今回のシミュレーションは2010年度研究紀要⁽¹⁾ とほぼ同じであるが、Q 値、コミュニティ情報、仮削除されたリンク情報を保持する変数など追加した。新たに追加した変数のみ表に掲載する。

4. シミュレーションのエージェント

今回作成したシミュレーションは2010年度研

変数名	型	用途
コミュニティ	二次元配列 長整数型	コミュニティに関する情報を保存する2次元整数型変数。 一次元目の配列数は、現状1000（コミュニティ ID=0～999に対応）。 二次元目の配列数は、2。 コミュニティ (i, 0)……コミュニティ ID=iであるノード同士をつなぐリンク総数。 コミュニティ (i, 1)……コミュニティ ID=iであるノードの持つリンクの総数。
コミュニティ数	長整数型	グラフの持つコミュニティ総数（グループ分割数）。
Q 値	実数型	Q 値、モジュラリティ（modularity）値。
実行ステップ数	整数型	コントロールパネルのスライダーと連動する変数。

究紀要で作成したシミュレーション⁽¹⁾を元に作成した。node エージェント、link エージェントについては説明を省略する。

各処理ステップで仮削除されたリンク（辺）等の情報を cvs テキストファイルから読み込み保持するために deleteLink エージェントを新たに追加した。

（１）deleteLink エージェント

各処理ステップで仮削除されたリンク（辺）等の情報を csv テキストファイルから読み込んで初期化して使う。このエージェントに保持された情報通りにリンクの仮削除を実行する。重要な変数のみ表に掲載する。

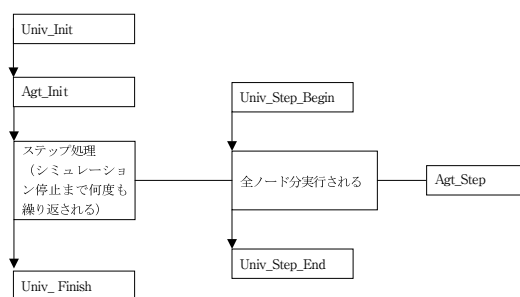
変数名	型	用途
処理前 Q 値	実数型	前ステップ処理終了後のグラフの Q 値、モジュラリティ（modularity）値。
最大媒介中心性指数	実数型	辺の媒介中心性指数（ C_e ）の最大値。
コミュニティ総数	整数型	前ステップ処理終了後のグラフのコミュニティ総数。
FromNode	整数型	仮削除したリンクの連結元 node エージェントの IDNo を保持する。
ToNode	整数型	仮削除したリンクの連結先 node エージェントの IDNo を保持する。

5. シミュレーションのスクリプト

5.1 シミュレーションスクリプトの構成

artisoc のスクリプトは、独立したセクションに分かれている。

（図13.）がスクリプトの構成概略図である。



（図13.）スクリプトの構成外略図

今回のシミュレーションでは、エージェントが実行するスクリプトは無い。従って Agt_Init、Agt_Step 部分にスクリプトは無い。

5.2 Univ_Init セクション

ここでの処理は2010年度研究紀要で作成した

シミュレーション⁽¹⁾のものを元に作成した。

今回はリンクの仮削除が行われるので、node エージェントの「link」変数（ノードが連結する全てのノードを保持するエージェント集合型変数）を別変数「全 Link」にコピーする初期化処理を追加した。

コントロールパネルのスライドバーと連動するグローバル変数「実行ステップ数」の値により処理が変化する。「実行ステップ数」が0以外の場合には、このセクションにて「実行ステップ数」値のステップ数分リンクの仮削除を行う。

5.3 Univ_Step_Begin セクション

コントロールパネルのスライドバーと連動するグローバル変数「実行ステップ数」値により動作が異なる。

実行ステップ数 = 0 の場合は、媒介中心性指数計算を行い、同時にノードにコミュニティ ID を付与する。そして Q 値を計算し、最大媒介中心性指数を持つ辺を仮削除する。

「実行ステップ数」が0以外の場合は、ノー

ドにコミュニティ ID を付与する。そして、大元にて媒介中心性指数計算を行う。
元のグラフ(仮削除された辺を除かないグラフ)

1	Univ_Step_Begin {
2	
3	Dim one As Agt
4	Dim start Node As Agt
5	Dim 結果 As Integer
6	
7	if universe. 動作モード=1 then
8	
8	// コミュニティ情報の初期化
10	ClearCommunity ()
11	
12	if Universe. 実行ステップ数=0 then
13	// 媒介中心性指数計算を行い、同時にノードにコミュニティ ID を付与する
14	// そして Q 値を計算し、最大媒介中心性指数を持つ辺を仮削除する
15	// 仮削除された辺を除いたグラフにて媒介中心性指数を計算する
16	
17	// 全ノードを始点にして、各ノードの媒介中心性指数計算を行う
18	For each one in Universe. 全ノード
19	CalcDependency (one) // 媒介中心性指数計算
20	next one
21	
22	CalcModularity () // Q 値 (モジュラリティ) 計算
23	
24	// 最大媒介中心性指数を持つ辺を仮削除する
25	結果 = DeleteEdge (-1, -1)
26	if 結果=1 then
27	universe. 動作モード=2 // シミュレーション動作モードアップ
28	end if
29	else
30	// 実行ステップ数が 0 以外の場合には
31	// ノードにコミュニティ ID を付与する
32	// そして、大元のグラフ (仮削除された辺を除かないグラフ) にて
33	// 各ノードの媒介中心性指数計算を行う
34	
35	For each one in Universe. 全ノード
36	if one. コミュニティ ID=-1 then
37	SetCommunityID (one) // ノードにコミュニティ ID を付与する
38	end if
39	CalcDependency2 (one) // 媒介中心性指数計算
40	next one
41	
42	universe. 動作モード=2 // シミュレーション動作モードアップ
43	end if
44	
45	else
46	universe. 動作モード = universe. 動作モード+1 // シミュレーション動作モードアップ
47	end if
48	
49	}

5.4 サブルーチン

2010年度研究紀要で作成した媒介中心性指数計算シミュレーション⁽¹⁾と同じ（ほぼ同じ）ものは説明を省略する。

(1) サブルーチン DeleteEdge

リンクを検索してグラフから仮削除する。第1引数によって動作が異なる。第1引数が-1

以外の場合には、第1引数と第2引数ペアと一致するリンクを削除する。第1引数=-1の場合には、最大媒介中心性指数を持つリンクを削除する。

(2) サブルーチン CalcModularity

グラフのQ値(modularity)を計算する。

```

1 // Q値(モジュラリティ)を計算する
2 Sub CalcModularity () {
3
4   Dim node As Agt
5   Dim link As Agt
6   Dim コミュニティ ID As Integer
7   Dim 総リンク数 As Integer
8   Dim i As Integer
9   Dim コミュニティリンク率 As Double
10
11  // コミュニティ内外のリンク数のカウント等
12  For each node in Universe. 全ノード
13
14     コミュニティ ID=node. コミュニティ ID
15
16     For each link in node. 全 link
17         Universe. コミュニティ (コミュニティ ID, 1)=Universe. コミュニティ (コミュニティ ID, 1)+1
18         if link. コミュニティ ID=コミュニティ ID then
19             Universe. コミュニティ (コミュニティ ID, 0)=Universe. コミュニティ (コミュニティ ID, 0)+1
20         end if
21     next link
22
23 next node
24
25 //////////////////////////////////////
26 // Q値(モジュラリティ)の計算
27
28
29 Universe.Q値=0
30 総リンク数=0
31
32 For i=0 to (Universe. コミュニティ数-1)
33     総リンク数=総リンク数+Universe. コミュニティ (i, 1)
34 next i
35
36 if 総リンク数=0 then
37     Universe.Q値=0
38 else
39     For i=0 to (Universe. コミュニティ数-1)
40         コミュニティリンク率=(Universe. コミュニティ (i, 1)/総リンク数)
41         Universe.Q値=Universe.Q値+(Universe. コミュニティ (i, 0)/総リンク数)-(コミュニティリンク率*コミュ
42 ニティリンク率)
43     next i
44 end if
45
46 }
```

(3) サブルーチン CalcDependency

リンクの仮削除されたグラフにて、1つの頂点(ノード)を始点にする他の全ノードの情報媒介度(媒介中心性指数)を計算する。

また同時にノードにコミュニティ ID を付与する。リンクをたどってつながる全てのノードには同じコミュニティ ID を付与する。それ以外は、2010年度研究紀要で作成した媒介中心性指数計算シミュレーション⁽¹⁾とほぼ同じである。

(4) サブルーチン CalcDependency2

リンクの仮削除がされていない大元のグラフにて、1つの頂点(ノード)を始点にする他の全ノードの情報媒介度(媒介中心性指数)を計算する。

2010年度研究紀要で作成した媒介中心性指数計算シミュレーション⁽¹⁾とほぼ同じである。

(5) サブルーチン SetCommunityID

1つの頂点(ノード)を始点にリンクをたどってつながる全ノードに同じコミュニティ ID を付与する。

6. 結び

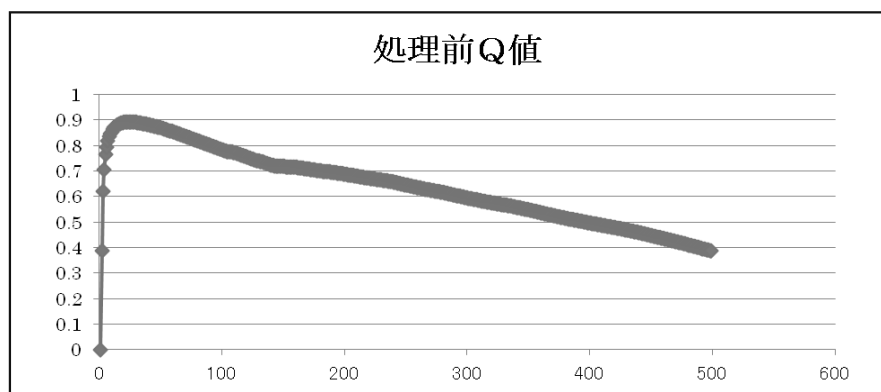
今回作成したコミュニティ抽出シミュレーションでは、まず参考にした書籍⁽⁴⁾で解説されていた米国のある空手部の友人関係の実例と同じノードとリンク情報をエクセルで作成しシ

ミュレーションに読み込ませコミュニティ抽出を実行し、同じコミュニティ分割結果が得られることが検証できた。

次に2010年度研究紀要⁽¹⁾で媒介中心性指数計算を行ったより大規模なネットワークでもコミュニティ抽出を行ってみた。MAS コミュニティ (<http://mas.kke.co.jp/index.php>)⁽¹⁰⁾で紹介されている『Barabasi-Albert モデル』artisoc シミュレーションをダウンロードして使い、人工的に生成させたノード数500個のスケールフリー・ネットワークである。

このネットワークのノードとリンク情報を csv テキストファイル形式で書き出し、シミュレーションに読み込ませ、処理を実行した。処理 1 ステップごとに最大の媒介中心性指数 C_v 値を持つ辺をグラフから除去し、Q 値(modularity)を計算した結果を表示すると(図14.)の結果が得られた。横軸は処理ステップ数である。リンク総数は998本であるが、有向グラフを想定したシミュレーションなので、1回の処理ステップで双方向の2つのリンクが消える。998/2=499ステップで処理が終了する。

正確な最大 Q 値を示すステップは、結果を書き出した「output.txt」ファイル(図15.)を調べれば分かる。処理ステップ26(27行目)の「処理前 Q 値」が0.892606857で局所的最大値で



(図14.) スケールフリー・ネットワーク(ノード数500、リンク数998のネットワーク)のQ値表示図

	A	B	C	D	E	F	G	H	I	J	K	
1	ID	X	Y	Layer	Direction	IDNo	処理前Q値	最大媒介中心性指数	コミュニティ総数	FromNode	ToNode	
2	0	0	0	0	0	0	0	48576	1	2	4	
3	0	0	0	0	0	0	0.386157084	29500	2	2	8	
4	0	0	0	0	0	0	0.620095903	10864	3	2	3	
5	0	0	0	0	0	0	0.704796366	7645	4	1	2	
6	0	0	0	0	0	0	0.764077253	4067	5	4	35	
7	0	0	0	0	0	0	0.794406448	3225	6	9	12	
8	0	0	0	0	0	0	0.818004747	2280	7	2	29	
9	0	0	0	0	0	0	0.834054883	1664	8	2	5	
10	0	0	0	0	0	0	0.845247609	1350	9	8	9	
11	0	0	0	0	0	0	0.854086931	1342	10	4	27	
12	0	0	0	0	0	0	0.862703363	1335	11	2	22	
13	0	0	0	0	0	0	0.871355938	775	12	6	16	
14	0	0	0	0	0	0	0.875476404	744	13	1	44	
15	0	0	0	0	0	0	0.879323778	648	14	2	57	
16	0	0	0	0	0	0	0.88237999	518	15	2	45	
17	0	0	0	0	0	0	0.884422151	510	16	4	20	
18	0	0	0	0	0	0	0.886385998	490	17	35	66	
19	0	0	0	0	0	0	0.888177156	460	18	82	97	
20	0	0	0	0	0	0	0.889775543	408	19	2	24	
21	0	0	0	0	0	0	0.890952245	372	20	2	170	
22	0	0	0	0	0	0	0.891885976	308	21	4	13	
23	0	0	0	0	0	0	0.892257461	285	22	2	43	
24	0	0	0	0	0	0	0.892494408	266	23	9	139	
25	0	0	0	0	0	0	0.892500432	260	24	2	50	
26	0	0	0	0	0	0	0.892574729	259	25	4	15	
27	0	0	0	0	0	0	0.892606857	235	26	2	172	
28	0	0	0	0	0	0	0.892518504	220	27	16	49	
29	0	0	0	0	0	0	0.892201236	176	28	8	38	
30	0	0	0	0	0	0	0.891552644	172	29	2	131	
31	0	0	0	0	0	0	0.890940197	160	30	4	41	
32	0	0	0	0	0	0	0.890183172	150	31	1	36	

(図15.) 書き出された「output.txt」ファイルをエクセル上で開いたもの（一部表示）

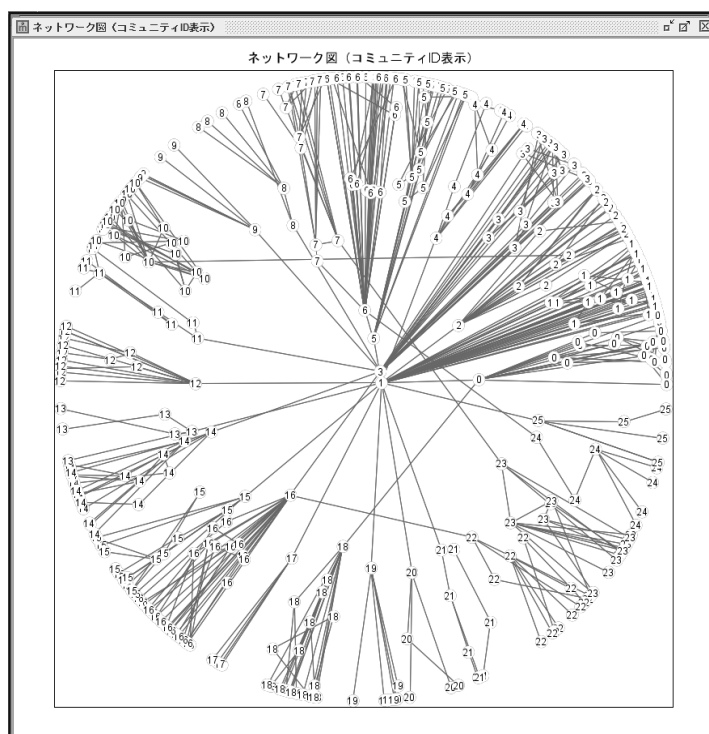
ある。また「コミュニティ総数」は26であった。「処理前 Q 値」、「コミュニティ総数」は実際には1ステップ前の処理終了直後の値を示しているので、これより処理ステップ=25の段階で得られるコミュニティ分割が適切であるという結論が得られた。

この「output.txt」テキストファイルを元にした情報を deleteLink 変数に読み込ませコントロールパネルを25に設定してシミュレーションを再起動した。そして「input.txt」ファイルに書き出された node エージェントデータを2010年度研究紀要⁽¹⁾で作成した全ノードを円内に配置するシミュレーションに読み込んで表示させた結果が(図16.)と(図17.)である。

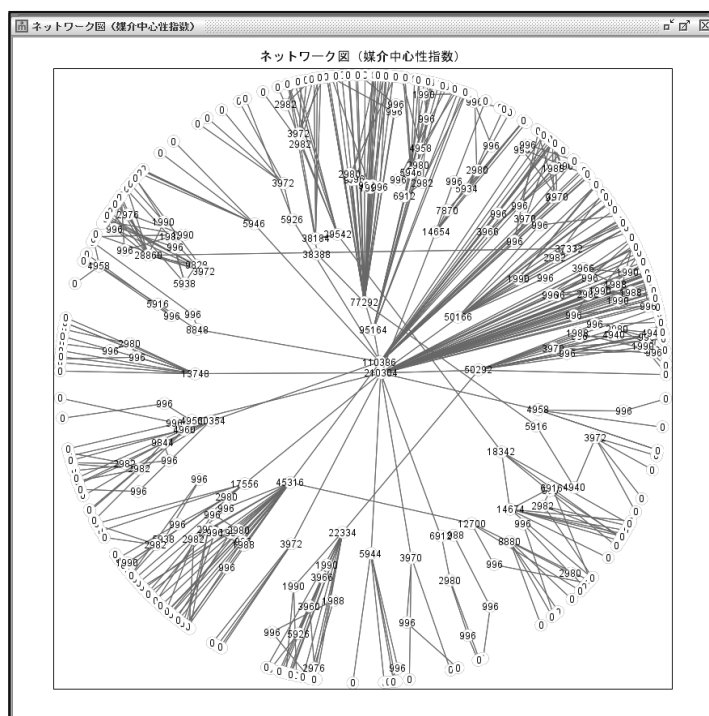
(図16.)は、ノードの「コミュニティ ID」変数情報を表示した図である。今回作成したシミュレーションにより、各ノードが0～25のコミュニティ ID を付与され(グループ分けされ)ている様子が窺える。

(図17.)は、ノードの「媒介中心性指数」変数情報を表示させた図である。今回作成したシミュレーションで計算したノードの媒介中心性指数が表示されている。

(図14.)にQ値の際立った最大値は無かったが、とりあえず最大値を示すステップで止めた時点でのコミュニティ分割を採用しただけで、(図16.) (図17.)のようなリンクの交錯の少ないすっきりとしたグラフ表示が可能となった。今回使用したグラフのケースでは、単にすっきりとしたグラフ表示が目標であるならば、特にQ値が最大値を示すステップを採用しなくても良さそうである。どのステップで止めればグラフをいくつかのコミュニティに分割できるかについての情報は、(図15.)の「output.txt」より得られる。ノード全体をいくつかのコミュニティに分割したいのか、分割数を先に決めてから、それを実現するステップでシミュレーションを停止する使い方もあり得る。



(図16.) ノードのコミュニティ ID 表示図



(図17.) ノードの媒介中心性指数表示図

今回のネットワークは各種スケールフリー・ネットワークのうちでも単純な種類のものである。もっといろいろなタイプのネットワークでこのコミュニティ抽出 artisoc シミュレーションを試してみるのも面白いであろう。

今回作成したシミュレーションでは各ステップ処理で毎回全ノードの媒介中心性指数計算を実行している。ノード数500ほどのネットワークではかなり時間がかかることを心配したが、これはさほど問題ではなかった。パソコンの性能によっても違っだろうが私が使っているパソコンでは最初のステップの計算に30秒ほどかかった。しかし、毎回最大媒介中心性指数を持つ辺が仮削除され、グラフの大きさが劇的に小さくなるので各ステップの処理時間も劇的に減少する。またQ値が局所的最大値を示すステップを探すには、全てのノードが仮削除されるまで処理を続ける必要もない。

数百のノード数レベルのネットワークでは今回作成したシミュレーションが十分使えることが確認できた。

[参考文献・出典]

- 1) 末木俊之 (2010) Artisoc による媒介中心性指数計算とネットワーク表示について、駒沢女子大学研究紀要、p. 375-394
- 2) 山影進 (2007) 人工社会構築指南 artisoc によるマルチエージェント・シミュレーション入門、有限会社書籍工房早山、東京
- 3) コミュニティ抽出とは、『相対的により密なリンクで互いに結合し合っている部分集団を塊として、複数取り出す発見的な操作を意味する。』林幸雄・大久保潤・藤原義久・上林憲行・小野直亮・湯田聴夫・相馬亘・佐藤一憲著 (2007) ネットワーク科学の道具箱 つなかりに隠れた現象をひもとく、p 50-56、近代科学社、東京
- 4) 林幸雄・大久保潤・藤原義久・上林憲行・小野直亮・湯田聴夫・相馬亘・佐藤一憲著 (2007) ネットワーク科学の道具箱 つなかりに隠れた現象をひもとく、p 51-56、近代科学社、東京
- 5) 林幸雄・大久保潤・藤原義久・上林憲行・小野直亮・湯田聴夫・相馬亘・佐藤一憲著 (2007) ネットワーク科学の道具箱 つなかりに隠れた現象をひもとく、p 51、近代科学社、東京
- 6) 林幸雄・大久保潤・藤原義久・上林憲行・小野直亮・湯田聴夫・相馬亘・佐藤一憲著 (2007) ネットワーク科学の道具箱 つなかりに隠れた現象をひもとく、p 46、近代科学社、東京
- 7) 林幸雄・大久保潤・藤原義久・上林憲行・小野直亮・湯田聴夫・相馬亘・佐藤一憲著 (2007) ネットワーク科学の道具箱 つなかりに隠れた現象をひもとく、p 44、近代科学社、東京
- 8) 林幸雄・大久保潤・藤原義久・上林憲行・小野直亮・湯田聴夫・相馬亘・佐藤一憲著 (2007) ネットワーク科学の道具箱 つなかりに隠れた現象をひもとく、p 44、近代科学社、東京
- 9) 林幸雄・大久保潤・藤原義久・上林憲行・小野直亮・湯田聴夫・相馬亘・佐藤一憲著 (2007) ネットワーク科学の道具箱 つなかりに隠れた現象をひもとく、p 55、近代科学社、東京
- 10) <http://mas.kke.co.jp/index.php>、MAS コミュニティトップホームページ