

コミュニティ抽出と Pajek の自動レイアウトについて

末 木 俊 之*

Community Extraction and Automatic layout of Pajek

Toshiyuki SUEKI*

Abstract

In complex network science, community extraction has become a hot topic, but it is not a topic in social network analysis. In social network analysis, tools such as Pajek are used to visualize social networks. After community extraction with artisoc, I have set the network data to Pajek and ran Fruchterman Reingold energy command for automatic layout of Pajek. Draw speed is fast and the number of crossing lines are minimized. It is possible to extract community roughly from the display of automatic drawing. However, to obtain the best community extraction is difficult.

1. コミュニティ抽出と Pajek の自動レイアウト表示

artisoc textbook⁽¹⁾を使用して、コミュニティ抽出とネットワーク表示を行うシミュレーションを作成したが、Pajek等のネットワークを扱うツールの存在も気になっていた。コミュニティ抽出は、複雑ネットワークの科学分野には登場するが、社会ネットワーク分析分野では話題になっていない。ネットワークを構成する線には方向性の有るものと無いものがある。方向性の有る線を孤 (Arc)、方向性の無い線を辺あるいはエッジ (Edge) と呼ぶならば、コミュニティ抽出を行うアルゴリズムではネットワークを構成する線は、方向性の無い辺であるか、または点同士を双方向につなぐ2つの孤の重なりとして扱うのが一般的だと思われる。ここ数年作成したコミュニティ抽出シミュレーション

でも任意の2つの点同士をつなぐ線は、双方向の孤として扱うモデルを使っている。「点Aから出て点Bに入る孤」がある場合には、その逆の「点Bから出て点Aに入る孤」も必ず存在するネットワークモデルである。社会ネットワーク分析で扱われるネットワークはソシオグラムと呼ばれている。ソシオグラムでは、片方向のみの孤、多重線、ループ (自分から出て自分に入る線) なども登場する。その違いで、社会ネットワーク分析ではコミュニティ抽出が話題になりにくいと言えるのかもしれない。

社会ネットワーク分析では、ソシオグラムの点と線を可視化して自動的に表示する Pajek などのツールを使い分析作業がおこなわれている。Pajek はネットワーク (ソシオグラム) を最適なレイアウトで自動的に表示する機能を持ち、またマウスのドラッグ操作により手動で任意の

*人間健康学部 健康栄養学科

点を好みの位置に移動することもできる。

Pajek はリュブリアナ大学（スロベニア）のホームページ⁽²⁾ から展開した Pajek wiki ダウンロードサイト⁽³⁾ より「Pajek32.exe」をダウンロードしインストールすることにより容易に使用可能となった。

Pajek では、スペース区切りのテキストファイルを用意し、それを読み込ませることによりソシオグラムを表示させることができる。テキストファイルは、点（Vertice）と、弧（Arc）に関するデータで構成される。X 軸・Y 軸・Z 軸の3D 表示が可能である。各点（Vertice）のデータは、シリアルナンバー（1 からはじまる点を識別する整数値）、点の名称、表示位置 X 座標値、表示位置 Y 座標値、表示位置 Z 座標値で構成される。必要であれば、各点の近くに点の名称を表示することができる。各弧（Arc）のデータは、始点のシリアルナンバー、終点のシリアルナンバー、弧の重み付け値で構成される。

ここ数年 artisoc textbook⁽¹⁾ で作成したシミュレーションでは、グラフを構成する点（ノード）は node エージェント、辺（リンク）は link エージェントで表現している。コミュニティ抽出計算後に、node エージェント、link エージェントデータをカンマ区切りの csv 形式テキストファイルとして書き出す。それをエクセルシート上に読み込み、Pajek 向けのテキストファイル仕様に合わせるために手入力を少し加える。そしてスペース区切りのテキストファイルとして書き出せば、Pajek で表示させるソシオグラムデータが容易に作成できた。ただし、点（Vertice）の表示座標値（X・Y 座標値）は、Pajek の表示スクリーン座標に合わせて変換しつつ artisoc から書き出す必要がある。Pajek の表示スクリーンの X 座標（水平軸）は、スクリーンの左端が 0 で、スクリーンの右端が 1

である。Y 座標（垂直軸）は、スクリーンの上端が 0 で、スクリーンの下端が 1 となっている。

Pajek で自動レイアウトコマンドを動作させれば、美しいネットワーク表示が簡単に実現される。その場合コミュニティが綺麗に分割され表示されるのか興味があった。Pajek には必要であれば、各点の近くに点の名称を表示する機能がある。今回は点の名称の代わりに各点が属するコミュニティ ID を入れたテキストファイルを用意して Pajek に読み込ませた。そうすれば自動ノードレイアウトコマンドの実行により、見た目でコミュニティ分割可能な表示が生成されるのか確認することができる。

2. Pajek 表示用ソシオグラムデータ構築

（図 1.）は Pajek に読み込ませるソシオグラムデータとなるテキストファイルをエクセルシート上で作成している様子である。2013年度研究紀要⁽⁴⁾にてコミュニティ抽出計算したノード数1000、リンク数1998、コミュニティ数36の『Barabasi-Albert モデル』ネットワーク（ $m=0, m=1$ BA モデル）のノードとリンクデータを基に作成している。そのデータの先頭数行の様子である。1 行目の「*Vertices」「1000」はそれぞれセル A1、B1に手入力でデータを追加入力している。点（Vertice）の総数が1000個という宣言である。2 行目以降の各行が1つの点を表現するデータである。B 列は、点のシリアルナンバー（1 から始まる点を識別する整数値）、D 列は点の名称（ただし今回は名称の代わりに各点の属するコミュニティの ID 番号である）、F 列は表示位置 X 座標値、H 列は表示位置 Y 座標値、J 列は表示位置 Z 座標値（今回は Z 軸データは無く、全点 0.5 にしてある）である。

2 行目以下は、artisoc textbook シミュレーションから書き出した csv 形式テキストファイ

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	*Vertices	1000											
2		1		1		0.519244		0.147442		0.5			
3		2		2		0.56725		0.528569		0.5			
4		3		3		0.666667		0.341198		0.5			
5		4		4		0.893105		0.512662		0.5			
6		5		5		0.285654		0.491593		0.5			
7		6		1		0.839532		0.568713		0.5			
8		7		3		0.323287		0.761817		0.5			
9		8		1		0.426389		0.580351		0.5			
10		9		6		0.228727		0.964814		0.5			
11		10		7		0.362502		0.4091		0.5			
12		11		3		0.265351		0.236187		0.5			
13		12		5		0.498331		0.654013		0.5			
14		13		2		1		0.252937		0.5			
15		14		8		0.751967		0.563426		0.5			
16		15		9		0.498277		0.848419		0.5			
17		16		10		0.195396		0.321338		0.5			

図1. ソシオグラムデータ上部の様子

	A	B	C	D	E	F	G	H	I	J	K	L	M
996		995		16		0.207265		0.786086		0.5			
997		996		13		0.455328		0.634102		0.5			
998		997		2		0.185441		0.778142		0.5			
999		998		10		0.578829		0.275436		0.5			
1000		999		1		0.214593		0.397602		0.5			
1001		1000		14		0.401408		0.171514		0.5			
1002	*Arcs												
1003	1		2										
1004	1		6										
1005	1		8										
1006	1		19										
1007	1		22										
1008	1		71										
1009	1		106										
1010	1		191										
1011	1		316										
1012	1		386										

図2. ソシオグラムデータ中間部の様子

ルをエクセルシート上に読み込ませただけで、追加手入力はない。

(図2.) はテキストファイルの中間付近の様子である。1001行目は、最後の1000個目の点のデータ行である。1002行目の「*Arcs」は、セル A1002に手入力している。孤 (Arc) に関するデータが次行 (1003行目) から始まる。1003行目以降は各孤 (Arc) に関するデータである。A 列は孤の始点のシリアルナンバー、C 列は終点のシリアルナンバーである。今回は孤の重み付け値は省略してある。

3. ソシオグラム表示結果

3-1. ソシオグラム表示結果 (初期 X・Y 座標値通りの表示)

Pajek にソシオグラムデータテキストファイ

ルを読み込ませた後、「Draw → Network」コマンド操作で表示した結果が (図3.) である。点 (ノード) 数1000、線 (リンク) 数1998の『Barabasi-Albert モデル』のネットワーク ($m_0 = 0$, $m = 1$ BA モデル) の様子である。(1) 全点が円形内に表示されていること。(2) 重要度の高い (より多くの点とリンクしている) 点 がより円の中心に配置されている。(3) 同一コミュニティに属する点は扇型内に集めて表示されている。図上の数字は、各点の属するコミュニティ番号になっている。artisoc から書き出した点の座標位置を基にデータを作成している ので、以上 (1) ~ (3) の表示上の性質は、artisoc でのネットワーク表示と同じである。

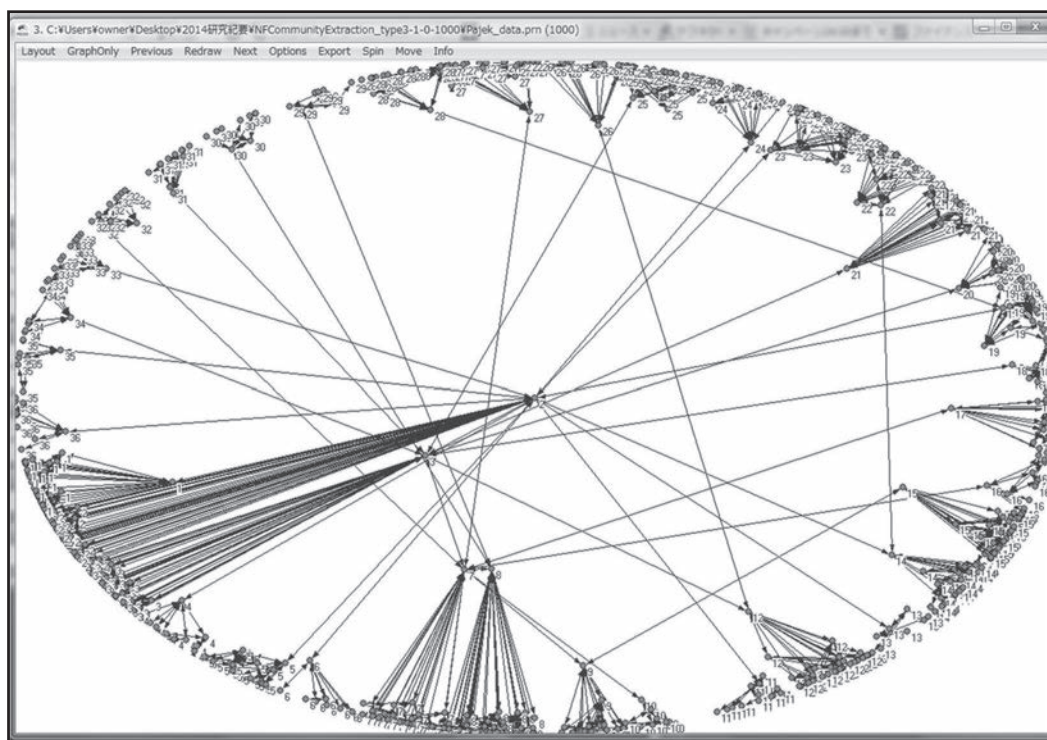


図3. ノード数1000、リンク数1998 BAモデルネットワーク表示

3-2. ソシオグラム表示結果 (Fruchterman Reingold 自動レイアウト表示)

Pajek で使える自動レイアウト機能のうちの1つ、「Fruchterman Reingold」コマンドによる自動レイアウト表示を試みた。点（ノード）の初期位置により自動レイアウト表示結果が影響を受けるので、まず点の初期位置を Random（任意の位置）にするコマンドを実行してから

表示させた。（図4.）のように、まず「Layout → Energy → Starting position → Random」の操作を行う。

そして「Layout → Energy → Fruchterman Reingold → 2D」コマンドを実行するが、実行前にサブメニュー「Factor」で点と点の最適な距離を指定する。参考にした書籍によると『このコマンドは、正の数字を入れるダイアログ

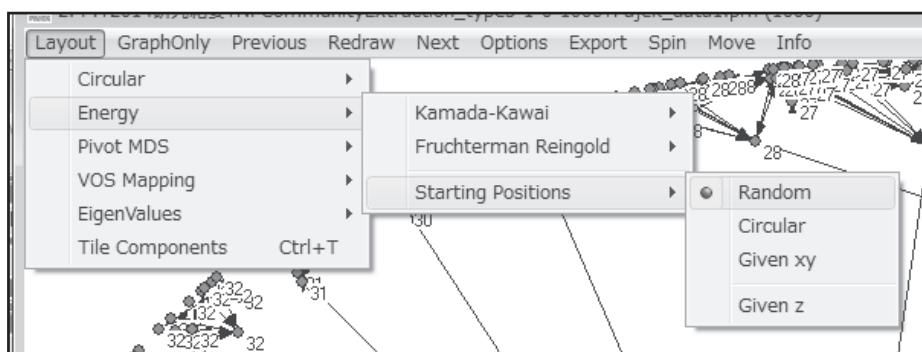


図4. Pajek Layout コマンド操作の様子

ボックスを表示する。小さい数字は点と点の間の距離を小さくするので、多数の点がスクリーンの中央に配置される。大きい数字は点を中央から周囲の位置に押しやる。初めの点の位置としては、最適距離を1にするのが良い。』⁽⁵⁾と記述されている。表示が落ち着くまで2分ほど時間がかかった。

(図5.) が Factor = 1.0での Fruchterman

Reingold 自動レイアウト表示の様子である。

artisoc での表示との類似点が3点ほどある。

- (1) 全点が円形内に表示されていること。(2) 重要度の高い(より多くの点とリンクしている)点がより円の中心に来るように配置されている。(3) 同一コミュニティに属する点が扇型に近い枠内に集めて表示されている。

ただし、(図5.) におけるコミュニティ ID

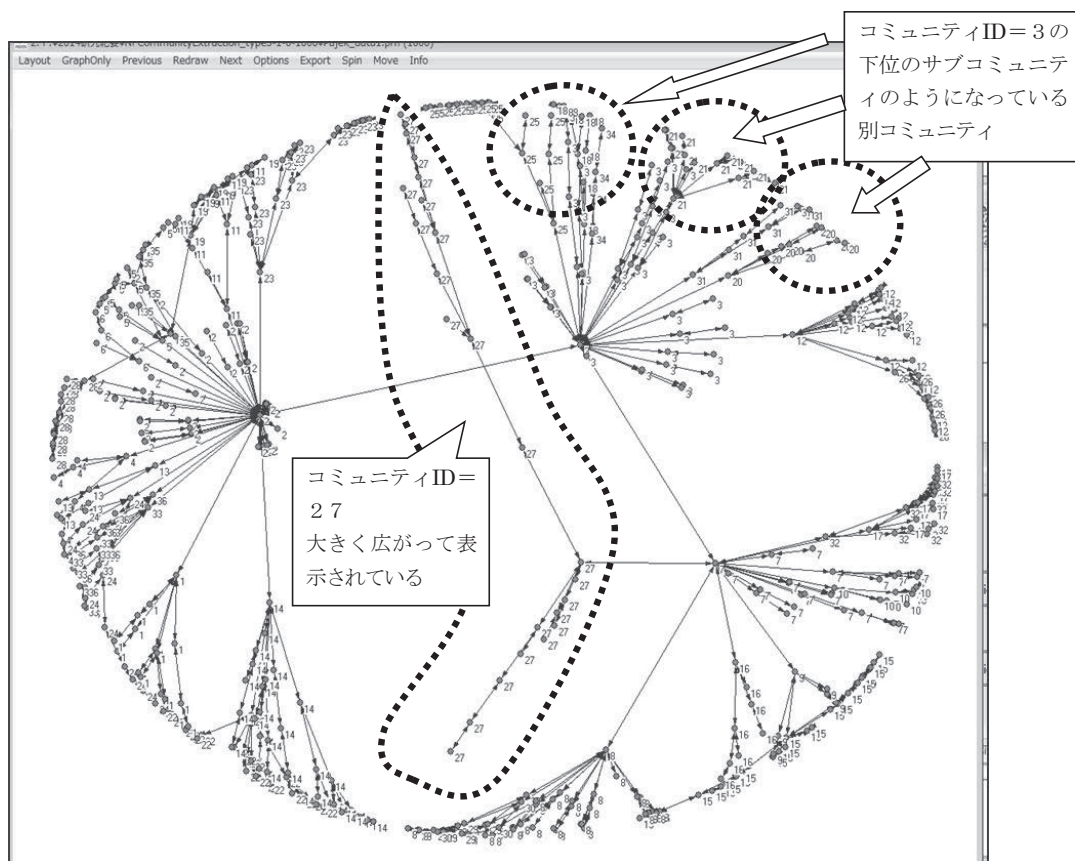


図5. Factor = 1.0 Fruchterman Reingold 自動レイアウト表示

=27の例のように、同一コミュニティに属する点が円内に大きく広がって表示される箇所や、右上のコミュニティ ID = 3の例のように、別

コミュニティが、あたかも同じコミュニティに所属する下位のサブコミュニティの様相で表示される箇所がある。

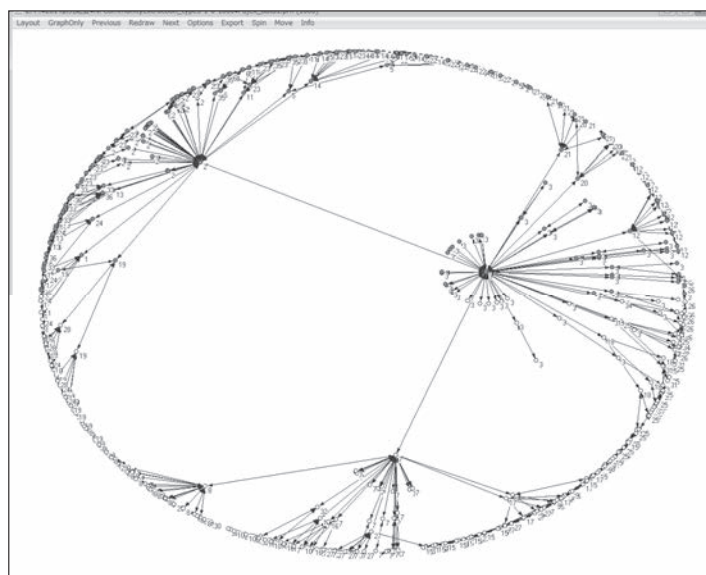


図 6. Factor = 1.6 Fruchterman Reingold 自動レイアウト表示

(図 6.) が Factor = 1.6での Fruchterman Reingold 自動レイアウト表示の様子である。

ここでも artisoc での表示との類似点がある。
 (1) さらに全点が円形内に表示される傾向が強くなっている。(2) 重要度の高い(より多くの点とリンクしている)点は円の中心で目立っている。(1)、(2)の性質は artisoc での表示と類似している。

しかし、重要度の小さい点は円の周辺に押しやられて表示されている。Factor 値が大きくなるに従い、円の周辺部に押しやられる点が増え、異なるコミュニティに属する点同士が混在して表示される傾向が強くなっている。

(図 7.) が Factor = 3.0での Fruchterman Reingold 自動レイアウト表示の様子である。

全点が円形内に表示されるのは artisoc での

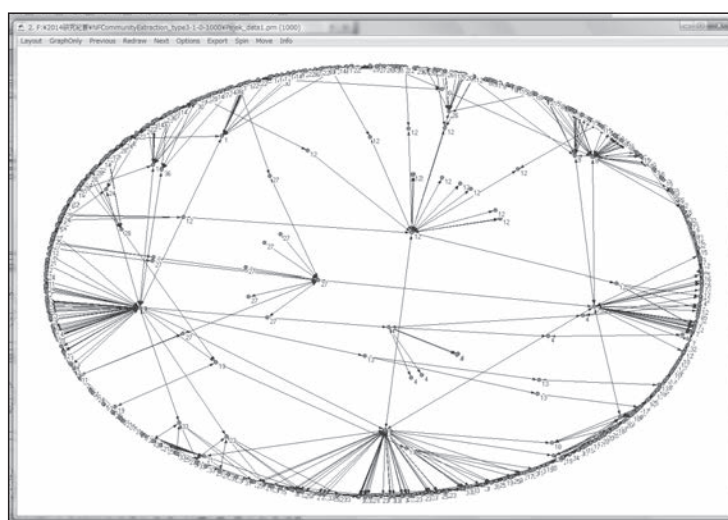


図 7. Factor = 3.0 Fruchterman Reingold 自動レイアウト表示

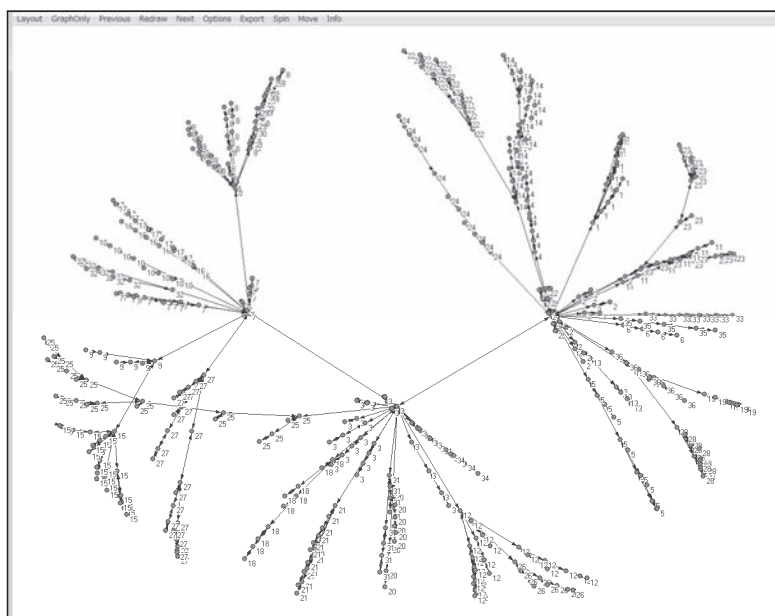


図8. Factor =0.4 Fruchterman Reingold 自動レイアウト表示

表示と同じだが、コミュニティ分割は分かりにくくなっている。なぜか、円の中心に配置されるノードが増えている。

(図3.)は、同一コミュニティに属する点は扇型内に集めて表示されているが、線の交錯が多い。コミュニティの配置(円内の扇型部分の並び順)を最適化すれば、より線の交錯が少ない表示が可能になるはずである。(図5.)(図6.)の表示では自然にコミュニティの配置の最適化が実現されており、(図3.)表示と比較すると線の交錯が少ない表示となっていることが指摘できる。

しかし、(図7.)では、線の交錯が増えてい

る。このネットワークモデル(ソシオグラム)のケースでは、PajekのFruchterman Reingold自動レイアウト表示におけるFactor =1.0~1.6の表示が線の交錯が少ない。しかし、それ以上のFactor値では線の交錯が増加する印象である。Factor値が大きすぎると有用な表示とはなり得ないと言える。

(図8.)がFactor =0.4でのFruchterman Reingold自動レイアウト表示の様子である。

円形表示と言うより、樹状表示(幹と枝)に変化している。重要度の高い(より多くの点とリンクしている)点がより中央に表示されている様子は窺える。

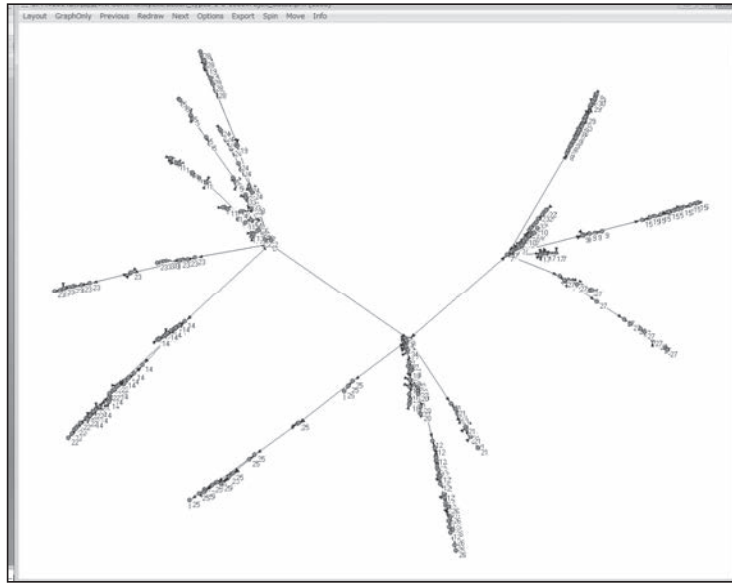


図9. Factor =0.2 Fruchterman Reingold 自動レイアウト表示

(図9.) は Factor =0.2での Fruchterman Reingold 自動レイアウト表示の様子である。

Factor 値が小さくなるに従って、さらに樹状表示(幹と枝)に変化している。重要度の高い点同士をつなぐ線は距離が長く、幹のように目立って表示されている。重要度の低い点同士をつなぐ線は短く、点が重なり合って表示され、その点の属するコミュニティはより小さな枝として表示され目立たなくなっている。

線の交錯に着目すると、Factor 値がより小さい Fruchterman Reingold 自動レイアウト表示ほど、線の交錯が少ない表示が実現されている。

今回使用したネットワークモデル(ソシオグラム)のケースでは、artisoc での表示と同様な円形表示で、コミュニティ分割が分かりやすく、かつ線の交錯が少ない表示を実現するには、Factor 値が0.6～1.2の Fruchterman Reingold 自動レイアウト表示が良い。

3-3. ソシオグラム表示結果 (Kamada-Kawai 自動レイアウト表示)

もう1つ Pajek で使える自動レイアウト機能「Kamada-Kawai」コマンドによる自動レイアウト表示を試みた。「Layout → Energy → Kamada-Kawai → Free」コマンド操作で表示させたのが(図10.)の図である。

artisoc で行った円形表示とは全く異なる表示である。同一コミュニティに属する点は、ある程度は近くにまとまって表示されている。重要性の低い点は周囲に表示されているが、重要な点と同様に目立っている様子が窺える。コミュニティの分割は見た目ではわかりにくい。

3-4. BAモデル($m_0=0$, $m=1$ 、ノード数2000、リンク数3998)表示結果 Fruchterman Reingold 自動レイアウト表示)

比較のために点(ノード)数が多いネットワークモデルで Pajek Fruchterman Reingold 自動レイアウト表示を行ってみた。2013年度研究紀要⁽⁴⁾にてコミュニティ抽出計算したノード数2000、リンク数3998、コミュニティ数53の

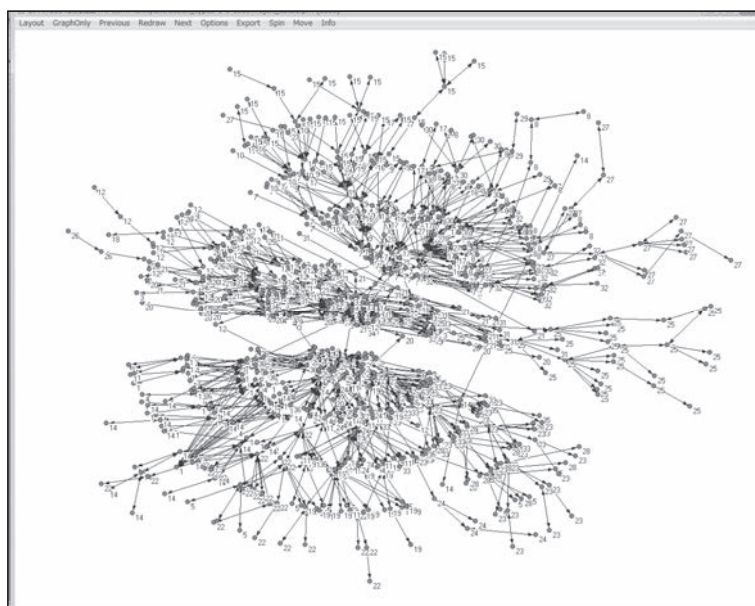


図10. Kamada-Kawai Free 自動レイアウト表示

『Barabasi-Albert モデル』($m_0=0$, $m=1$) のネットワークを使った。

ノード数が2000に増えると表示が落ち着くまで5分ほど時間がかかった。

(図11.) が Factor = 0.8 での Fruchterman Reingold 自動レイアウト表示の様子である。

artisoc での表示と同様、円形内に全点が表示され、扇型に近い様相でコミュニティ部分が表示される傾向もある。ただし、あたかも同じ

コミュニティに所属する下位のサブコミュニティの様相で混在して表示される別コミュニティ部分もある。また異なるコミュニティに属する点同士が混在している部分も多い。

3-5. BA モデル ($m_0=0$, $m=3$, ノード数1000、リンク数5988) 表示結果 Fruchterman Reingold 自動レイアウト表示)

比較のために線(リンク)数が多いネットワークモデルで Pajek Fruchterman Reingold 自動

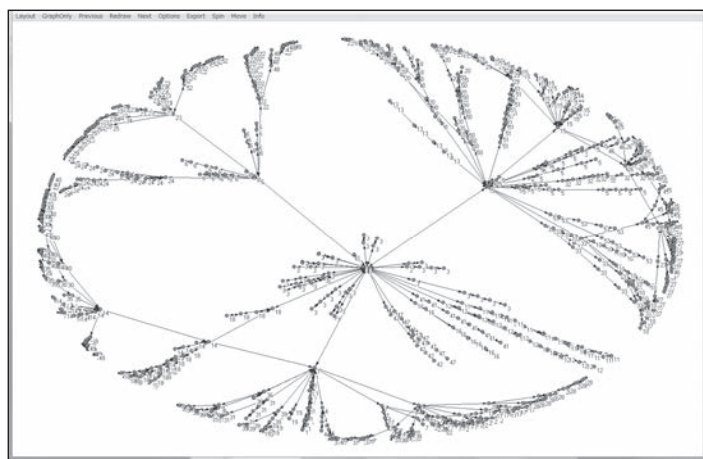


図11. Factor = 0.8 Fruchterman Reingold 自動レイアウト表示

レイアウト表示を行ってみた。2013年度研究紀要⁽⁴⁾にてコミュニティ抽出計算を行ったノード数1000、リンク数5988、コミュニティ数5の『Barabasi-Albert モデル』($m_0=0$, $m=3$)のネットワークを使った。

表示が落ち着くまで5分ほど時間がかかった。

(図12.) が Factor = 1.0での Fruchterman Reingold 自動レイアウト表示の様子である。

リンク数が多いネットワークは円形状に表示されにくい傾向がある。また同一コミュニティ

に属する点がきれいに近くに集まって表示されていない。

(図13.) が Factor = 30での Fruchterman Reingold 自動レイアウト表示の様子である。

ほとんどの点は円の周辺部に押しやられている。異なるコミュニティに属する点が混在している部分も見られるが、一応コミュニティごとに分かれている。また artisoc での表示よりも円内の線の交錯が少なくすっきりとした表示に見える。

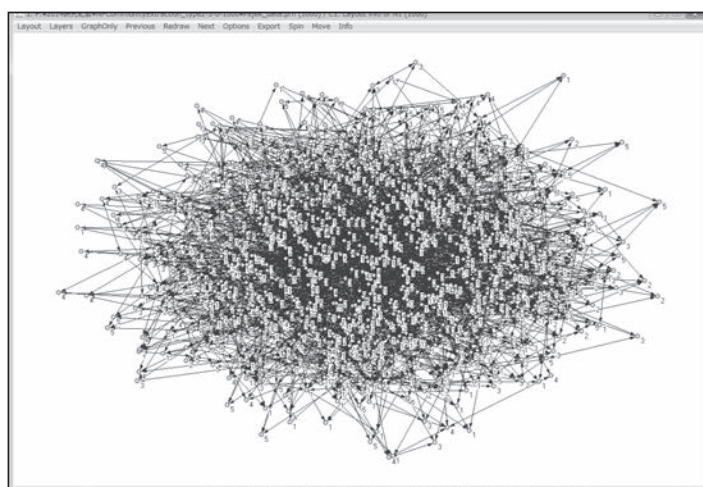


図12. Factor = 1.0 Fruchterman Reingold 自動レイアウト表示

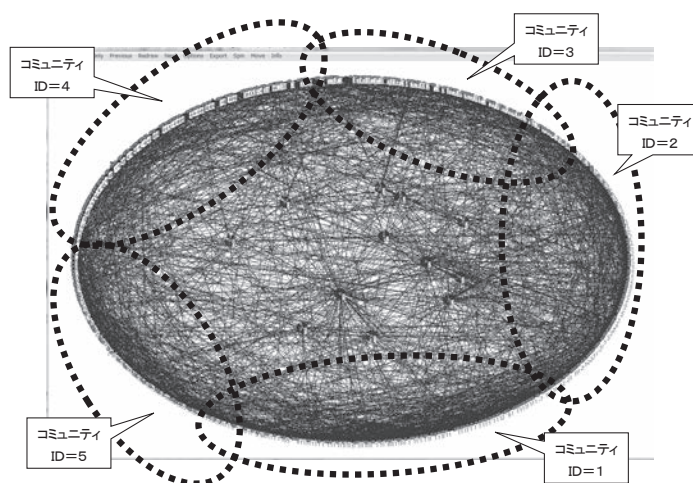


図13. Factor = 30 Fruchterman Reingold 自動レイアウト表示

3-6. 米国のある大学の空手部部員の友人関係を表したグラフ図表示結果 (Fruchterman Reingold自動レイアウト表示)

比較のために、参考にした書籍で扱われていた点（ノード）の少ないネットワーク⁽⁶⁾についても Pajek で自動レイアウト表示をさせてみた。ノード数34、コミュニティ数2のネットワーク例である。(図14.) は書籍に記述されているものと同じノード配置の初期表示である。

(図15.) が Factor = 3.0での Fruchterman Reingold 自動レイアウト表示の様子である。

artisoc での表示と同様、ほぼ円形内に全点が表示されている。2つのコミュニティは綺麗に分離して表示されている。Factor = 2.0でも2つのコミュニティは分離されているが、少し樹状の配置になっている。Factor = 2.0未満だと樹状表示で、重要度の低い点同士の距離が近く、小さな枝状の表示となり点が目立たない。Factor = 3.0より大きいと、コミュニティの分離状況が悪くなる。

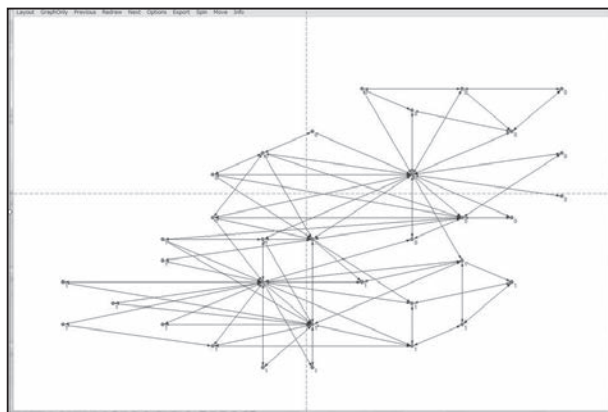


図14. 米国のある大学の空手部部員の友人関係を表したグラフ図⁽⁶⁾

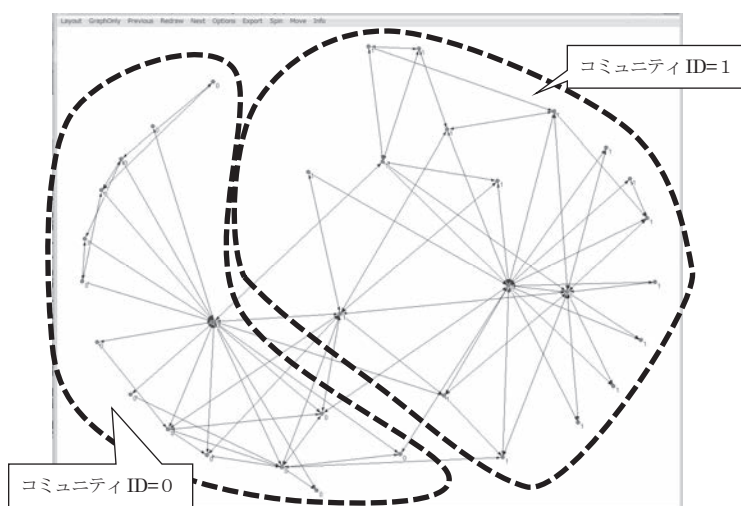


図15. 米国のある大学の空手部部員の友人関係を表したグラフ図 Pajek Factor = 3.0 Fruchterman Reingold 自動レイアウト表示図

4. 結び

Pajek は商用でなければ、誰でも無料で使えるツールである。自動的にネットワークを美しくレイアウトして表示する機能がある。ここ数年 artisoc 使用し、コミュニティ抽出計算させた結果に基づいてネットワーク表示をさせていたが、Pajek で同じネットワークを自動レイアウト表示すると、どのような表示となるか興味が湧いてきた。Pajek は主に社会ネットワーク分析分野の研究者が使用しているツールであるが、自分も実際に使用する良いきっかけが得られた。インストールは簡単であった。

artisoc で使用したネットワークのデータ（点、辺）をカンマ区切りの csv 形式のテキストファイルに書き出し、エクセルシート上に読み込み、Pajek の仕様に合わせて少々手直しし、スペース区切りのテキストファイルで書き出せば容易に Pajek 用のデータが準備できた。それを Pajek に読み込み、Fruchterman Reingold 自動レイアウトコマンドを実行すると、artisoc でネットワークを円形状に表示させたものと似た表示がされた。

(表 1.) (表 2.) にある 4 つのネットワークを Pajek で表示させてみた。Factor 値に依存し、表示の様相が変化する。Factor 値が小さいと樹状表示となり、重要度の低い点の集まりは小枝状に表示される。大まかなコミュニティ分割を

捉えるには有用かもしれない。Factor 値が大きくなるに従って円形状に変化する。Factor 値が大きいほど重要度の低い点が円の周辺（円周上）に押しやられる傾向がある。また同一コミュニティに属する点が近くに集まり表示される傾向が崩れていく。異なるコミュニティに属する点が混在して表示されるようになる。あまりに大きな Factor 値だと全ての点が円周上に押しやられ、利用価値の無い表示となる。

表中の「良いと思われる Factor 値」とは、artisoc で試みたものと同様に、以下の (1) (2) の表示条件に近いレイアウト表示がされる Factor 値のことである。

(1) 重要度の高い（より多くの点とリンクしている）点がより画面の中心に来るように配置されること。

(2) より密なリンクで互いに結合し合っている点（同一コミュニティに属する点）を近くに集めて表示すること。なるべく同一コミュニティに属する点は円の一角の扇型内に集めて表示させたい。

とは言っても、完全に (1) (2) を満たす表示にはなっていない。(図 5.) の例では、コミュニティ ID = 27 の表示のように、同一コミュニティに属する点が円内に大きく広がって表示される部分や、コミュニティ ID = 3 の表示スペース内に、あたかも同じコミュニティに所属する

表 1. BA モデル表示結果

m0	m	ノード数	リンク数	コミュニティ数	良いと思われる Factor 値	表示が落ち着くまでの時間
0	1	1000	1998	36	0.6~1.2	約 2 分
0	3	1000	5988	5	30 程度	約 2 分
0	1	2000	3998	53	0.6~0.8	約 5 分

表 2. 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表示結果

ノード数	リンク数	コミュニティ数	良いと思われる Factor 値	表示が落ち着くまでの時間
34	156	2	0.6~4	数秒

下位のサブコミュニティの様相で表示されている別コミュニティの存在も見える。今回の結果より、見た目で大まかなコミュニティ分けを把握することは可能であるが、最適なコミュニティ抽出（分割）は難しいことが分かる。

ネットワークの科学では、最適なコミュニティ分割を行う各種アルゴリズムが考案されているが、社会ネットワーク分析の分野で使用される Pajek では、ばねモデル⁽⁷⁾の自動レイアウトコマンドをまず実行し、その表示結果から分析作業を開始するのが一般的である。自動レイアウト表示では、最適なコミュニティ分割を見つけるのは難しいことがわかった。大まかなコミュニティ分割なら見た目で速やかに行えるという利点はある。社会ネットワーク分析の分野では最適なコミュニティ分割は使われないのかもしれない。

他方、複雑ネットワークの科学では、さらに高度なコミュニティ抽出アルゴリズムの開発が進められている。ばねモデル⁽⁷⁾に基づいた Fruchterman Reingold 自動レイアウトコマンドのような特殊なアルゴリズムが見つけれられるのかもしれない。見た目だけで、最適なコミュニティ分割が把握できるレイアウト表示を実現するのは、Fruchterman Reingold 自動レイアウトコマンドでは難しそうだが、表示図中心部分の線の交錯が少ない配置で点がレイアウトされることに驚かされた。これは Fruchterman Reingold 自動レイアウトコマンドの優れた点であろう。

artisoc はシミュレータなので、各ステップで Fruchterman Reingold 自動レイアウトと類似のコマンドを実行し、少しずつ理想の表示状況に近づくシミュレーションが作成できるであろう。スライドバーの操作で Factor 値を変化させながら表示の様子を観察できるシミュレーションを作成するのも面白いかもしれない。

[参考文献・出典]

- 1) 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
- 2) <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>、リュブリアナ大学(スロベニア)ホームページ
- 3) <http://pajek.imfm.si/doku.php?id=download>、Pajek wiki ダウンロードサイト
- 4) 末木俊之(2013) Artisoc による NF(Newman Fast) アルゴリズムコミュニティ抽出、駒沢女子大学研究紀要、p.181-190
- 5) ウオウター・デノイ、アンドレイ・ムルヴァル、ヴラディミール・バタゲーリ著、安田雪監訳 (2009) Pajek を活用した社会ネットワーク分析、p.24、学校法人東京電機大学東京電機大学出版局、東京
- 6) 林幸雄・大久保潤・藤原義久・上林憲行・小野直亮・湯田聡夫・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
- 7) 『線は、互いには決して近寄りすぎない性質をもった点を引っ張るバネであると想像してほしい。エナジーコマンドは、点が均衡状態に達するまで、点をより良い位置に引っ張っていく。そのため、これはばねモデル<として知られている。』と説明されている。：ウオウター・デノイ、アンドレイ・ムルヴァル、ヴラディミール・バタゲーリ著、安田雪監訳 (2009) Pajek を活用した社会ネットワーク分析、p.22、学校法人東京電機大学東京電機大学出版局、東京