

「PAJEK を活用した社会ネットワーク分析」の実装と解説（1）

安 田 雪

Intoduction to Pajek – Mannual and Tips (Part I)

Yuki YASUDA

Abstract

This paper is an introductory manual for the social network analysis software, "Pajek." Pajek is developed by Andrej. Mrvar, Vladimir. Batagelj et. al., mainly in Slovenia. Pajek is strong at analyzing large-scale networks with relatively low computing-power machine, yet it does not have user-friendly interface, especially for novice users. This paper explains details of Part I and Part II (Chapter I to Chapter V) of *Exploratory Social Network Analysis with Pajek* (Wouter de Nooy et al., 2005). I and Yasuda-seminar members have installed, run all the procedures mentioned in the book and tested if the procedures written in the book are usable with Japanese windows machines. This paper adds some new tips and techniques which are not explained by the developers of the software and authors of the above book.

Keywords: Social network analysis, Pajek, Visualization

抄 録

本稿は、大規模ネットワークの描画と探索的分析を行うソフトウェア「Pajek」の解説である。Pajek とはスロヴェニアにおいて Wouter de Nooy、Andrej. Mrvar、Vladimir Batagelj らによって開発されたソフトである。比較的、計算能力の高くないコンピュータでも高速に大規模ネットワークの分析が可能である一方、直感的に使いにくいインターフェースをもつため、初学者はとりつきがたい。本稿では、Pajek の解説書である『Pajek を活用した社会ネットワーク分析』（Wouter de Nooy et. al., (著), 2005; 安田雪（監訳）, 2009）のうち、第一部と第二部（第一章から第五章）で論じられている概念を解説するとともに、プログラムを実装し、正しい出力結果を得るために必要な操作手順について図解をする。さらに上述書では論じられていないが描画等の出力に必要な手作業について補足の説明を行う。

キーワード：社会ネットワーク分析、Pajek、可視化

解説の目的と構造

本解説の目的は、Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj (2005) *Exploratory Social Network Analysis with Pajek*, Cambridge University Press、邦訳『Pajek を活用した社会ネットワーク分析』（2009）安田雪監訳（東京電機大学出版局）の内容を章別に要約するとともに、安田研究室における Pajek の実装とプログラミングによる再現可能性

検証の結果を示すことである。原著（2005年版）は、Part IからPart Vまでの12章と付録により構成されている。同書（2011年版）ではランダムグラフモデルに関する13章が追加されている。

本解説では、ネットワークの可視化を重視し、一部はマニュアルとしても利用可能な構成を目指している。原著が全397頁の大著のため解説は分冊化し、今回は、原著のPart IとPart IIを扱う。Part Iは実装と基礎的情報にあたる、1章 探索、2章 属性、Part IIでは3章 凝集的サブグループ、4章 バランス理論、5章 所属（2部グラフ）を論じる。Part IIIでは仲介、Part IVでは序列、Part Vでは役割を扱う。

第1章 社会構造を探索する

はじめに

◆この章における目的

- ・社会ネットワークの基本要素を紹介し、社会ネットワークを構築し、描画する方法を示す。

◆ソシオメトリー

- ・個人間の関係を研究する社会科学である。
- ・社会的選択は社会関係の最も重要な表現であり、ソシオメトリックチョイスとして知られる。
- ・ソシオメトリーでは、社会的選択を社会関係で最も重要な表現と考える。

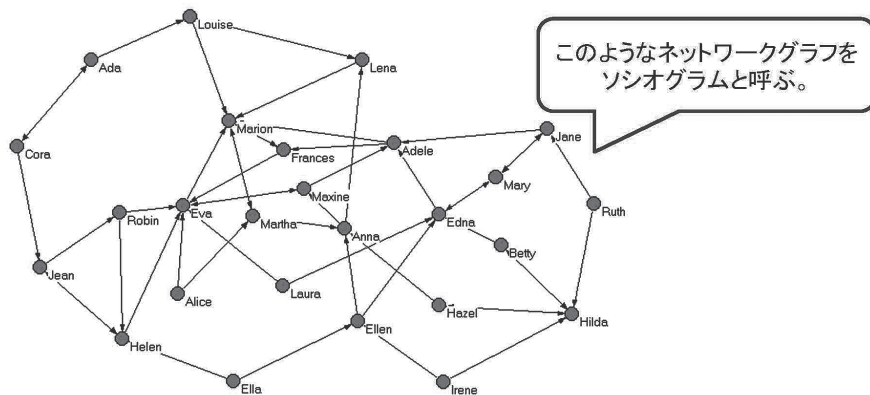
◆ソシオグラム

- ・グループ構造をグラフで表現した図である。
- ・ソシオメトリーに起源をもつ最も重要な方法で、社会ネットワークを可視化する基礎である。

第1章で例として扱うネットワーク

◆『夕食の同席者のネットワーク』

- ・ニューヨーク州立職業訓練校の寮に住む26人の少女たちの選択を示したデータである。
- ・少女たちが夕食の同席者として最も好ましい相手を選択したデータのうち、一番目と二



番目までを今回用いるデータとする。

探索的社会ネットワーク分析

◆社会ネットワーク分析

- ・ 行為者の社会的紐帯を研究する方法である。
- ・ 行為者間の社会的のパターンの解明とその解釈が目的となる。

◆探索的社会ネットワーク分析

- ・ 構造的な仮説検証は行わないが、意味のあるパターンを求めて社会的ネットワークを探索する。
- ・ ネットワークの定義、ネットワークの操作、構造属性の決定、視覚的検討の4つの部分から構成される。
- ・ ネットワークは膨大なので内外の境界を設定するが、ネットワークの境界は構造に大きな影響を及ぼす。したがって、境界の明確化は注意深く行う必要がある。

ネットワークの定義

◆点（ポイント、ノード）

- ・ ネットワークを構成する最小単位である。
- ・ 行為者（例えば、寮の少女、組織、国）を表し、通常は番号をつけて識別する。

◆線

- ・ ネットワークの2点を結ぶ紐帯であり、どのような社会関係でもありうる。
 - ・ 隣接している2つの端点を結ぶ紐帯と定義される。
- {
- ・ 弧（アーク）…方向性のある線である。順序関係のある「対の点」であり、送り手から受け手に向けられる。矢印。
 - ・ 辺（エッジ）…方向性のない線。順序関係のない対の点であり、双方向の弧でもある。
 - ・ ループ…ある点をその点自身と結びつける特殊な線である。

◆グラフ

- ・ 対になっている点を結ぶ線の集合と点の集合で、ネットワークの構造を表現するために用いる。
- {
- ・ 有向グラフ（ダイグラフ）…1つ以上の弧を含む無方向の社会関係である。
 - ・ 有向グラフ…弧を含まず、全ての線がエッジで構成される。

◆単純グラフ

- ・ 多重線を持たないグラフである。
- ・ 多重辺もループも含まないものを単純無向グラフと呼び、多重弧を含まないものを単純有向グラフと呼ぶ。

◆ネットワーク

- ・ ネットワークとは何か、という概念を定義するためにグラフ理論と言う数学の一分野を用いる。
- ・ グラフと、そのグラフの点や線に付与された追加的な情報で構成される。
- ・ ただし付与的な情報は、そのネットワークの構造と無関係である。

Pajek

◆Pajek

- ・ 社会ネットワークを分析・描写するために使用するためのコンピュータプログラムである。スロヴェニア語で蜘蛛を意味する。発音はパジェックではなく、パニエックに近い。
- ・ Pajekでは、ネットワークはグラフ理論にしたがって定義される。

- ・ HP (<http://vlado.fmf.uni-lj.si/pub/networks/Pajek/>) からソフトと、必要なデータセットのダウンロードができる。最新の URL は <http://mrvar.fdv.uni-lj.si/pajek> である。



ネットワークのデータファイルを確認

◆テキストエディタ (メモ帳など) でネットワークのデータファイルを確認 (Dining-table_ partners.net)

- ・ データファイルは点の数で識別される。
- ・ 点 (Vertices) にはネットワークを描くための3次元空間の点の位置を示す0から1の間の3つの実数が付与される。
- ・ 弧 (Arcs) は、送り手の点のシリアルナンバー、受け手の点のシリアルナンバー、弧の重み付けによって各行に区別される。

```
*Vertices 26
1 "Ada" 0.1646 0.2144 0.5000
2 "Cora" 0.0481 0.3869 0.5000
3 "Louise" 0.3472 0.1913 0.5000
4 "Jean" 0.1063 0.5935 0.5000
[...]
25 "Laura" 0.5101 0.6133 0.5000
26 "Irene" 0.7478 0.8087 0.5000

*Arcs
1 3 2
1 1 1
2 1 1
2 4 2
3 9 1
3 11 2
[...]
25 15 1
25 17 2
26 13 1
26 24 2

*Edges
```

点や弧の一部が[...]で置きかえられている。

Pajek の使い方

太字はメインスクリーンでの操作、
下線はドロースクリーンでの操作を表す。

◆ File → Network → Read

Pajek を起動したら、File メニューの Read コマンドをクリックする。(Network と書かれた文字の下フォルダアイコンをクリックしてもよい)

- ・ ネットワークファイル (Dining-table_partners.net) を選択。

◆ Draw → Network

- ・ Draw メニューの Draw コマンドをクリックすることで、夕食の同席者のネットワークを描画できる。(ドロップダウンメニューの右のペンアイコンをクリックしてもよい)



操作と計算

◆ 操作

- ・ 調べるネットワークが巨大で描写できない場合、意味のある一部分だけを抽出する。
- ・ 可視化は大規模ネットワークよりも小規模（数十の点）から中規模（数百の点）程度のネットワークで効果をより発揮する。
- ・ ネットワークが異なる種類の関係性を含む場合は1つの関係性だけに焦点をあてるべきである。

◆ 計算

- ・ 社会ネットワーク分析では多くの構造的属性が計量的に表現される。
- ・ 計算はネットワーク全体の属性の場合には単一の値を出力し、サブネットワークや点の場合には一連の値を出力する。
- ・ ネットワークの構造分析には、ネットワークを視覚的に調べることで構造的な指標の計算の双方を実施する。

可視化

◆ 可視化

- ・ 分析者はネットワーク描写をする際に、ネットワークを自動的に最適なレイアウトで作図する手順を踏むべきである。
- ・ ネットワーク描画の基本原則として最も重要なことは、点と点の間の距離を可能な限りその紐帯の強さと数で表し、描写は紐帯が交差しないよう点を配置する。
- ・ 見やすさを追求するために、グラフのビジュアル性には条件を付ける。

自動描画

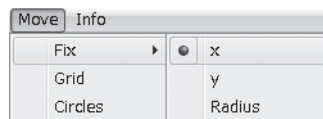
◆自動描画

- ・ 出てくる結果の図が調査者の先入観や誤解に依存しにくく、手動描画よりも基本的なレイアウトを素早く作成出来る。
- ・ Pajek の自動的レイアウトにおいて 2 つのコマンドはエナジーコマンドと呼ばれ、Layout → Energy メニューから選択可能である。
- ・ 点が均衡状態に達するまで点をより良い位置に引っ張ることから、ばねモデルとして知られる。
- ・ 1 回のエナジーコマンドで描画すると、作図結果が点の初期状態の位置に依存したり、改善幅が比較的小さくなったり、再配置が停止するという限界が存在する。

手動描画

◆手動描画

- ・ 点をクリックし、そのままドラッグすることで点の移動ができる。
- ・ Move → Fix → x, y, Radius で点の移動を制限できる。
- ・ マウスの右ボタンを押しながら、拡大したいエリアに長方形にドラッグすることで、図をズームできる。（全体のネットワークに戻りたいときは Redraw コマンドを使用）
- ・ Graph Only メニューは点と線だけを表示させるため、大規模ネットワークを描写するときには便利である。



図の保存

◆描画のレイアウトの保存

- ・ ネットワークそのものをメインスクリーンの Save コマンドで保存する。

◆ネットワークの図を Pajek からエクスポートする方法

- ・ ドロースクリーンの Export メニュー ESP/P, SVG, Bitmap は 2 次元の作図をする。
VRML, MDL, MOLfile, Kinemages は 3 次元の作図をする。

社会ネットワークデータ

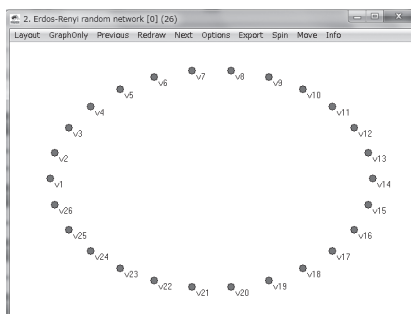
◆社会ネットワークデータの収集

- ・ ネットワーク分析を実施するには社会関係を測定しコード化することが必要である。
- ・ ソシオメトリストが注目する方法→グループ内の社会構造の選択。
- ・ 例：教室の子供たちに隣に座りたい子の名を述べるように依頼する。回答者は自ら選んだ子の名を書き出すか、リストの名にチェックする。前者の方法をフリーコール、後者をラスターと呼ぶ。
- ・ ソシオメトリーでは選択数を3に限定するのが一般的である。
- ・ 条件付きの選択、自由な選択、ランキング、二者間比較は質問によって社会関係データを抽出する技法だが、今や電子データの蓄積が急増しており大規模なネットワークのデータを収集する新しい機会が生まれている。

ネットワークの作成

◆データ収集後の2通りのネットワーク作成法

- ①テキストエディタを用いてネットワークをテキストファイル形式で保存する。(ファイルの拡張子は「.net」)
- ②Pajekでネットワークファイルを作り出す。
 - ・ Network → Create Random Network → Total No. of Arcs → 必要な点の数を入力 → 弧の本数を指定
 - ・ File → Network → View/Edit
 エディットネットワークスクリーンでNewlineをダブルクリックして弧を付与する。



Editing Network: 2. Erdos-Renyi random network [0] (26). Vertex:1				
File				
1:	4.1	val=1.00000	/	v4.v1
1:	1.3	val=1.00000	/	v1.v3
1:	1-2	val=1.00000	/	v1-v2
Newline				

第1章のまとめ

- ◆ ネットワークは点のセットと点の間を結ぶ線のセット、さらに点と線に付加された情報を含めて定義される。

- ・そのために、一部分の構造から全体の構造に至るまで、多種多様な実際の現象をネットワークとしてモデル化できる。

◆社会ネットワーク分析では、人々の関係や人々のグループを代表する社会主体の関係に注目する。

◆分析手順の結果は分析するネットワークに依存するので、ネットワークの種類を知ることが重要である。

◆実際にやってみることが、社会ネットワーク分析を学ぶ最良の方法であるため Pajek の訓練が必要である。

第2章 属性と関係

はじめに

◆この章の目的

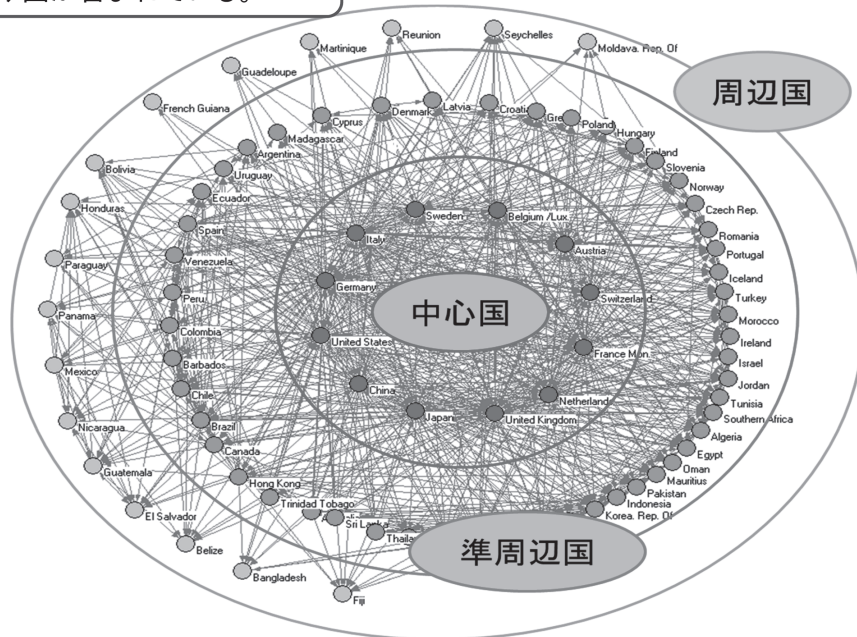
- ・「関係についてのネットワークデータ」と「ネットワークの点を持つ非関係的な属性（心理的・社会的・経済的・地理的属性）」とを結合させる技法を紹介する。
- ・統計的分析において、ネットワークにおける点の位置をどのように扱うかを議論する。
- ・ネットワーク分析で使用する基本的なデータを理解し、関係についてのデータと非関係的なデータを結合出来るようになる。

パーティション

◆ネットワークのパーティション

- ・「ネットワーク内の点をクラス分け／クラスタ分けすること」で、ネットワークの大きさや複雑さを削減するために有効である。
- ・各点は、必ず1つのクラス（クラスタ）に振り分けられ、負でない整数のリストになる。
- ・パーティションのクラスナンバーの順序は任意だが、明らかな順位を反映しないものは好ましくない。（例えば、準周辺国1，中心国2，周辺国3など）
- ・クラスナンバーが実際の数（点に接続する線の数など）を表現することもある。

1994年の世界の貿易統計
80ヶ国が含まれている。



ネットワークの縮約

◆パーティションの分割方法

- ・ ネットワークを削減するためのパーティションを3つの方法で使用する。

◆局所図

- ・ ネットワークの一部を抽出する。

◆全体図

- ・ 各クラスに属する複数の点を新しい点に縮約する。

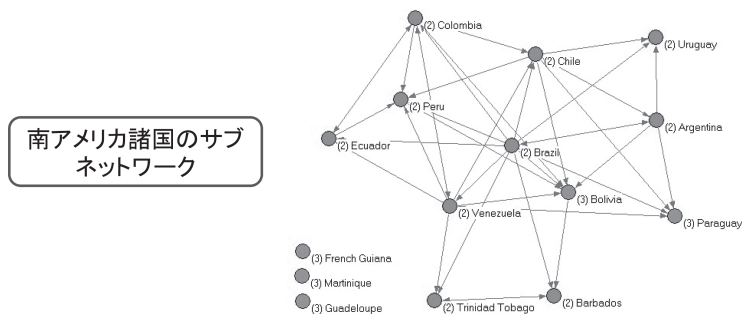
◆概況図

- ・ 特定のクラスを選択し、そのクラスの内部構造とネットワーク全体におけるそのクラスの位置付けに注目するために、隣のクラスを縮約させる。

局所図

◆局所図

- ・「1つのクラスに属する点だけを抽出」することで作成出来る。
- ・「点の集合」と「その点だけに接続するすべての線」を選択することで、局所の構造を示すサブネットワークが取り出せる。



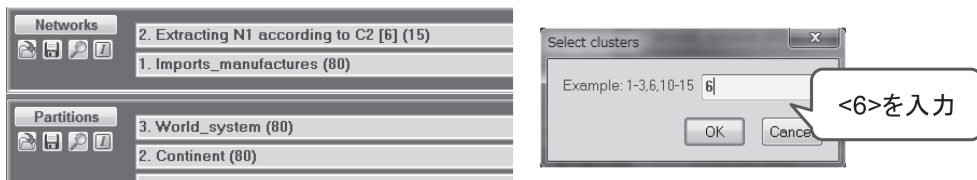
局所図の描き方 (Pajek の操作)

◆局所図の描画

- ・ Operations → Network + Partition → Extract SubNetwork → 数値の選択
- ・ Draw → Network + First Partition



- ・ Partitions → Extract SubPartition (Second from First) → 数値の選択
- ・ Draw → Network + First Partition



全体図

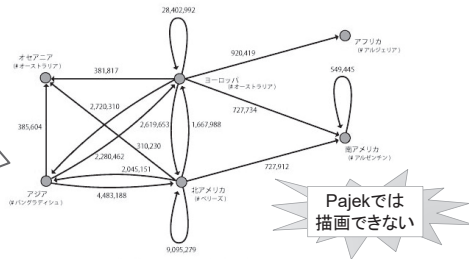
◆全体図とは

・「あるクラスに属するすべての点を新しい1つの点に置き換えること」で作成出来る。

・点同士の関係ではな

く、クラス間の関係を見ることが出来る。

大陸内のすべての点を大陸全体を表す新しい1つの点に縮約する。



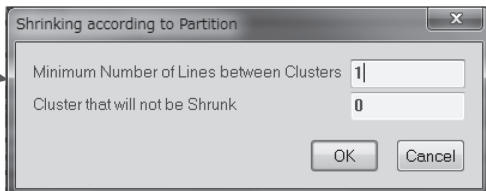
全体図の描き方 (Pajek の操作)

◆全体図の描画

・ Operations → Network + Partition → Shrink Network → 数値の入力

・ 新しいパーティション (Shrinking) が作成される。

上段に初期値である<1>を入力
下段に初期値である<0>を入力

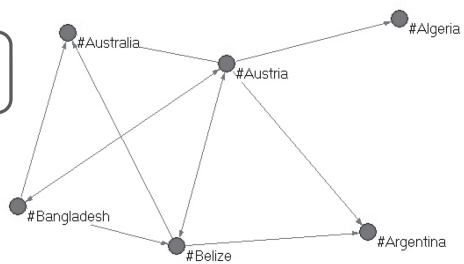
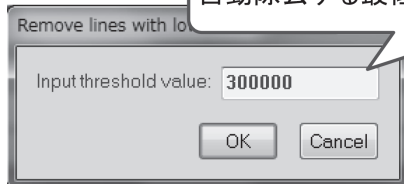


◆綺麗な図を作成するためネットワークを縮約する

・ Network → Create New Network → Transform → Remove → lines with value → lower than → 数値の入力

・ Draw → Network + First Partition

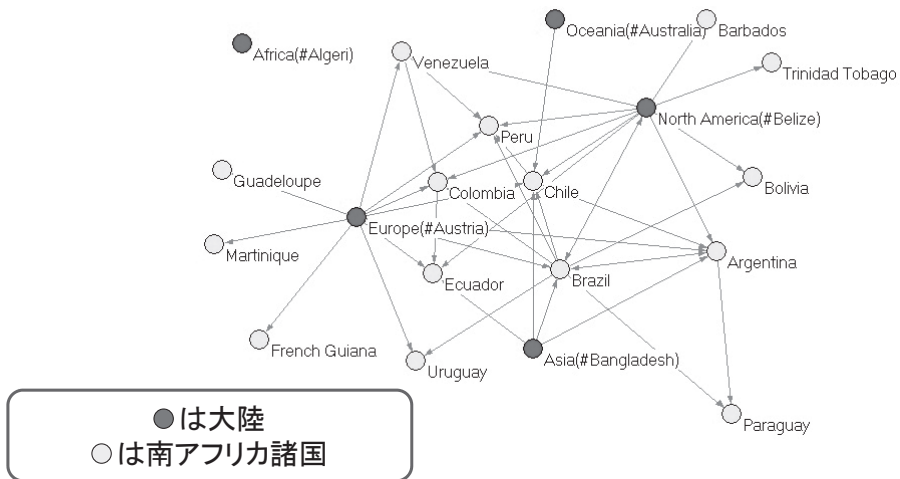
自動除去する最低値を入力



概況図

◆概況図とは

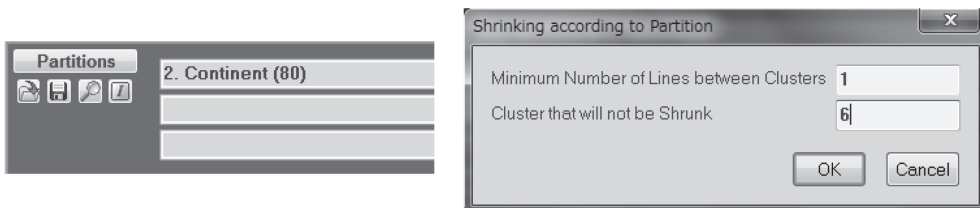
- ・「特に興味があるクラス以外のすべてのクラスを縮約する」ことで作成出来る。
- ・あるクラスの点と、それ以外のクラスとの関係を詳しく見る事が出来る。



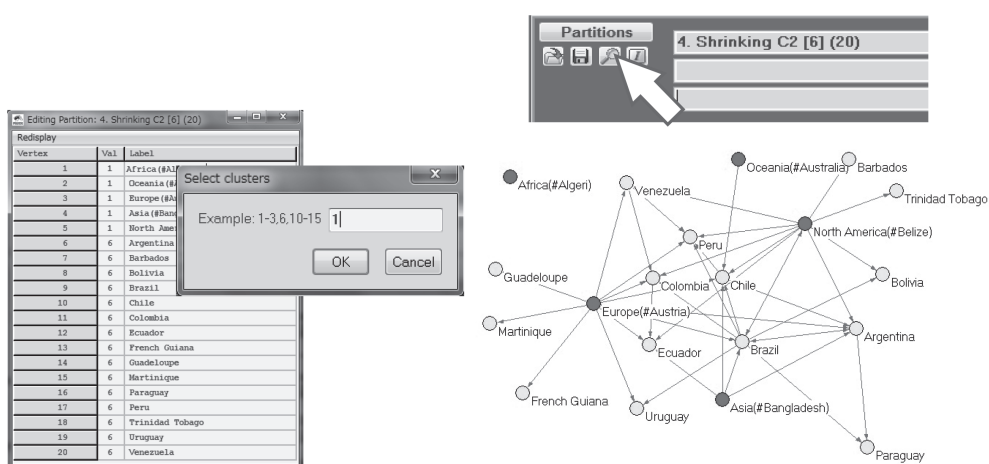
概況図（Pajek の操作）

◆概況図の描き方

- ・ Operations → Network + Partition → Shrink Network
- ・ 第一のダイアログボックスは初期値である <1> を入力。
- ・ 第二のダイアログボックスは縮約させないクラス、今回は南アメリカのクラス <6> を入力する。
- ・ Network → Create New Network → Transform → Remove → lines with value → lower than
- ・ 今回は総額が1千万米ドル未満の線を自動除去する。



- ・ 縮約された5大陸にクラス1を与えるため、[Edit] 欄を書き換えるクラス内の線を削除する。
- ・ Operations → Network + Partition → Transform → Remove Line → Inside Clusters → ダイアログボックスには<1>を入力する。
- ・ Draw → Network + First Partition



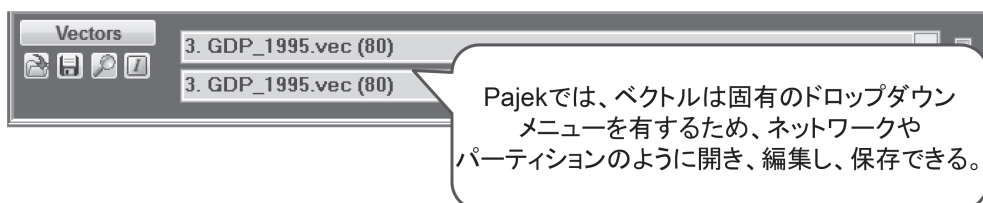
ベクトルと座標

◆ベクトルとは

- ・ 連続的な値をネットワークの各点に付与するものである。(例えば、国の面積や1人あたりGDPを点の大きさなどで表現する等)。

◆ベクトルの特徴

- ・ 実数（整数でなくても良い）を含み、負の数でも良い。
- ・ 連続的な属性は点をクラスにグループ分けするためのものではないので、ネットワークを縮約することはできない。

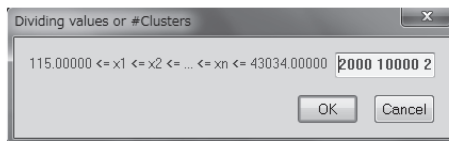


ベクトルと座標

◆基本的な情報の集め方

- ・ Info ボタンをクリックする。

第一のダイアログボックスは初期値<0>のまま入力
第二のダイアログボックスでは、頻度分布で報告
される価値のクラスの数や範囲を指定できる。
(今回は2000 10000 20000と入力)



Vector Values	Frequency	Freq%	CumFreq	CumFreq%
(2000.0000 ... 2000.0000]	22	27.5000	22	27.5000
(2000.0000 ... 10000.0000]	27	33.7500	49	61.2500
(10000.0000 ... 20000.0000]	15	18.7500	64	80.0000
(20000.0000 ... 43034.0000]	16	20.0000	80	100.0000
Total	80	100.0000		

ベクトルと座標

◆パーティションとベクトル

- ・ パーティション…ネットワークから点のサブセットを選択する役目を持つ。
- ・ ベクトル…計算で使用する数的な特徴を特定する。
- ・ 以上2つは互いに変換が可能である。

◆パーティションからベクトルへ

- ・ Partition → Copy to Vector

◆ベクトルからパーティションへ

- ・ Vector → Make Partition → Copy to Partition by Truncating(Abs)
- ・ Vector → Make Partition → by Intervals → First Threshold and Step

ベクトルと座標

◆ベクトルの描画

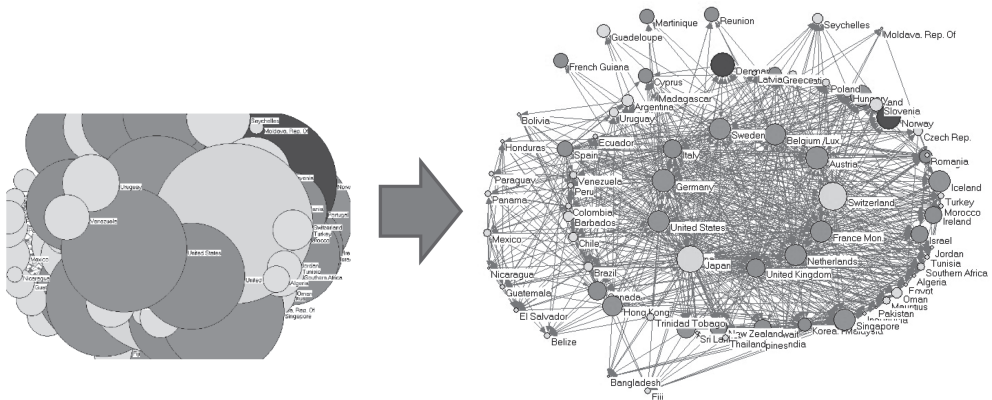
- ・ Draw → Network + First Vector

もしくは

・ Draw → Network + First Partition + First Vector

◆点が大きくなりすぎるので、点のサイズを縮小する

・ Option → Size → of Vertices

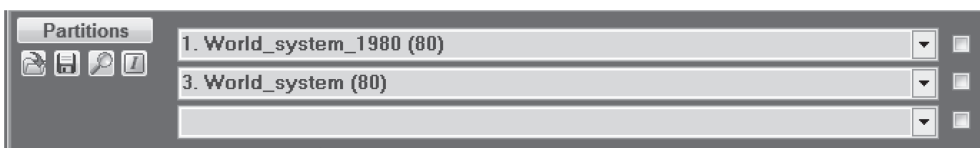


ネットワーク分析と統計

◆ Pajek のもつ基本的な統計技法

- ・ 二つのパーティションのクロス集計表を作成する。
- ・ 二つのパーティションによるクラス間の関連性を測定する。

◆ Partitions → Info → Cramer' s V, Rajski, Adjusted Rand Index



第一のパーティションに1980年の世界システム、
第二のパーティションに1994年の世界システムを選択

クロス集計表の作成例 1

◆1980年の世界システムの位置 (行) × 1994年の世界システムの位置 (列)

Rows: 1. World_system_1980 (80)					
Columns: 3. World_system (80)					
Crosstabs					
	1	2	3	Total	
1	10	1	0	11	
2	1	16	0	17	
3	0	15	0	15	
4	0	4	5	9	
9999998	1	15	12	28	
Total	12	51	17	80	

1980年の世界システムに占める各国の位置に従い4つのクラスに分割

- 1 中心国
- 2 強い準周辺国
- 3 弱い準周辺国
- 4 周辺国

クロス集計表の作成例 2

◆1980年の世界システムの位置 (行) × 1994年の世界システムの位置 (列)

Rows: 3. World_system (80)						
Columns: 1. World_system_1980 (80)						
Crosstabs						
	1	2	3	4	9999998	Total
1	10	1	0	0	1	12
2	1	16	15	4	15	51
3	0	0	0	5	12	17
Total	11	17	15	9	28	80

1994年の世界システムに占める各国の位置に従い3つのクラスに分割

- 1 中心国
- 2 準周辺国
- 3 周辺国

第2章のまとめ

◆パーティションとベクトルはどちらも数のリストであり、各点はひとつの値を持つ。

- ・パーティション：離散的なクラスを表し、クラスは負の数を取らない整数である。
- ・ベクトル：点の連続的な特徴を表し、実数を含み、負の数でも良い。

◆社会ネットワークは大きく複雑であるので簡略化すると良い。

- ・抽出：点の一部のセットを紐帯と共に選び出す。(局所図)
- ・縮約：点の一部のセットと紐帯をまとめて1つとして扱う。(全体図、概況図)

◆パーティションとベクトルで表現された点の特徴や特質を属性と呼び、どちらも相互関係データには使えない統計分析に活用出来る。

第3章 凝集的サブグループ

はじめに

◆この章における目的

- ・凝集的サブグループを見つける技法を紹介する。
- ・「ネットワークの構造から見つかったサブグループ」と「社会的な特徴から分類されたグループ」とを比較し、異なるものかどうかを検証する。社会的な特徴とは規範、行動、アイデンティティなどを指す。

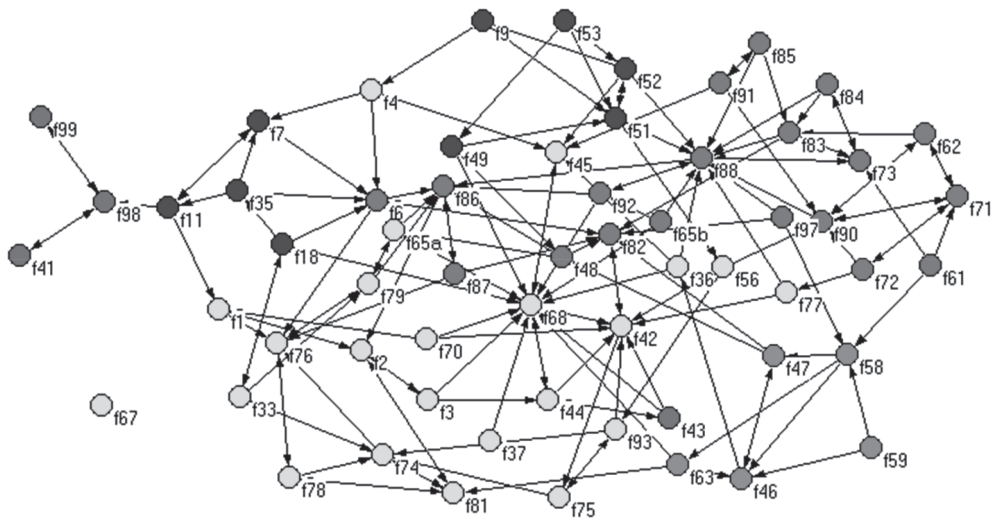
◆凝集的サブグループ

- ・社会ネットワークにおける、「互いに密接」な人々の集合体である。
- ・必ずしも社会的なグループである必要はない。

第3章で例として扱うネットワーク

◆『Attiro の訪問関係』

- ・コスタリカ共和国にある Attiro と呼ばれる地域の大農場に住む家族の訪問関係のネットワークである。
- ・各弧が1つの家庭から他の家族への「頻繁な訪問」を示す。
- ・民俗学上の特徴により、友情関係を六つにグループ(色)分けをしている。



密度

◆密度

点の数が多いと存在可能な線の数も増加するが、
それぞれの点を持てる紐帯の数は限られている。

||
大きなネットワークでは密度が低くなる。

- ネットワークに現在存在している線の数
・ 密度 = $\frac{\text{ネットワークに現在存在している線の数}}{\text{すべての存在しうる線の最大数}}$
- ・ 多重線と線の値は無視し、いずれも 1 つの線として扱う。
 - ・ 密度 (最大値は 1) が高いほど「互いに密接」な関係にある。
 - ・ ネットワークの大きさに左右されるため、異なる大きさのネットワークの凝集性を比較する指標としては適さない。
 - ・ 完全な単純ネットワークにおいて密度は最大になる。= 完全ネットワーク。

密度 (Pajek の操作)

◆密度の計算

・ Network → Info → General

```
Density1 [loops allowed] = 0.0447222
Density2 [no loops allowed] = 0.0454802
Average Degree = 5.3666667
```

× ループありの場合の密度

○ ループなしの場合の密度

このネットワークにループは無意味。
(自分自身を訪問することは不可能)



次数

◆次数

- ・ 点に接続している紐帯の数である。
- ・ 有向グラフでは点が受け取っている弧の数（入次数）と、発している弧の数（出次数）を区別しなければならない。
- ・ すべての点の平均の次数が多いほど、「互いに密接」な関係にある。
- ・ 異なる大きさのネットワークの凝集性を比較するのに適している。
- ・ 無向の単純ネットワークが最も簡単に解釈できるため、有向ネットワークを対称化して考えることもある。
- ・ 対称化……有向ネットワークの一方向の弧と双方向の弧を辺に置換する。

次数 (Pajek の操作)

◆ネットワークを対称化する

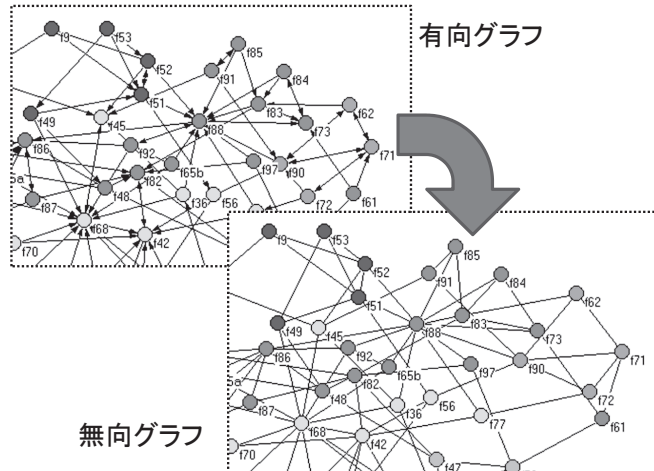
- Network → Create New Network → Transform → Arcs → Edges
→ All → 条件を選択する。

1-Sum
新しい線にまとめる複数の線の値を合計する。

2-Number
新しい線にまとめる複数の線の数を数える。

3-Min
まとめた複数の線の最小値をとる。

4-Max
まとめた複数の線の最大値をとる。



Input…入次線を数える。

Output…出次線の数进行数える。

All…入次線と出次線の双方を数える。

◆次数のパーティションを作成

- Network → Create Partition → Degree → 条件を選択する。

◆次数の度数分布表を作成

- Partition → Info

◆次数パーティションをベクトルに変換

- Partition → Copy to Vector

Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
0	1	1.6667	1	1.6667	#67
1	3	5.0000	4	6.6667	#37
2	1	1.6667	5	8.3333	#59
3	19	31.6667	24	40.0000	#3
4	20	33.3333	44	73.3333	#1
5	4	6.6667	48	80.0000	#45
6	6	10.0000	54	90.0000	#51
7	2	3.3333	56	93.3333	#6
8	1	1.6667	57	95.0000	#66
9	1	1.6667	58	96.6667	#42
13	1	1.6667	59	98.3333	#8
14	1	1.6667	60	100.0000	#68
Sum	60	100.0000			

◆次数の平均値を計算

- Vector → Info

Arithmetic mean: 4.2666667
Median: 4.0000000
Standard deviation: 2.3442601

Attiroの家族が平均4つ以上の家族を定期的に訪問することを示している。

— 次数の平均値

コンポーネント

◆グラフ理論の概念

- ・ウォーク (例 $A \rightarrow B \rightarrow D \rightarrow C \rightarrow B \rightarrow D \rightarrow C$)

その内部の線のいずれの終点もが弧の出発点になっていないという条件を満たすセミウォーク。

- ・セミウォーク (例 $A \rightarrow B \rightarrow C \rightarrow D \rightarrow B \rightarrow C$)

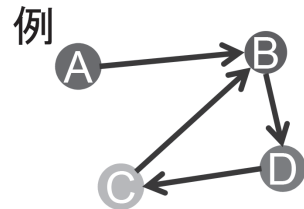
ある点から始まり他のある点で終わる、弧の方向を考慮に入れない一続きの線。(同じ点を何度通ってもよい)

- ・パス (例 $A \rightarrow B \rightarrow C$)

1つの点を複数回通らないウォーク。

- ・セミパス (例 $A \rightarrow B \rightarrow D \rightarrow C$)

1つの点を複数回通らないセミウォーク。



「最大の」

- ・ そのサブネットワークの特徴(ここでは連結性)を壊すような点を含まない。
- ・ 特徴を満たしている点をすべて含んでいる。

コンポーネント

◆ (弱) 連結のネットワーク

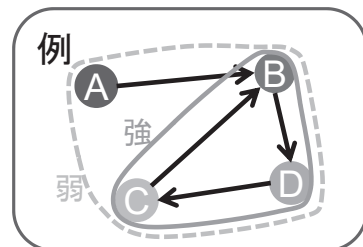
- ・ 各点のペアがそれぞれセミパスによって連結しているネットワークである。

◆強連結のネットワーク (B, C, D)

- ・ 各点のペアがそれぞれパスによって連結しているネットワークである。
- ・ パスに沿って各点からどの点にも到達できる。

◆ (弱連結) コンポーネント (A, B, C, D)

- ・ 最大の (弱連結した) サブネットワークである。
- ・ 他のコンポーネントとの間に線 (弧) はない。



- ・最大の強連結したサブネットワークである。

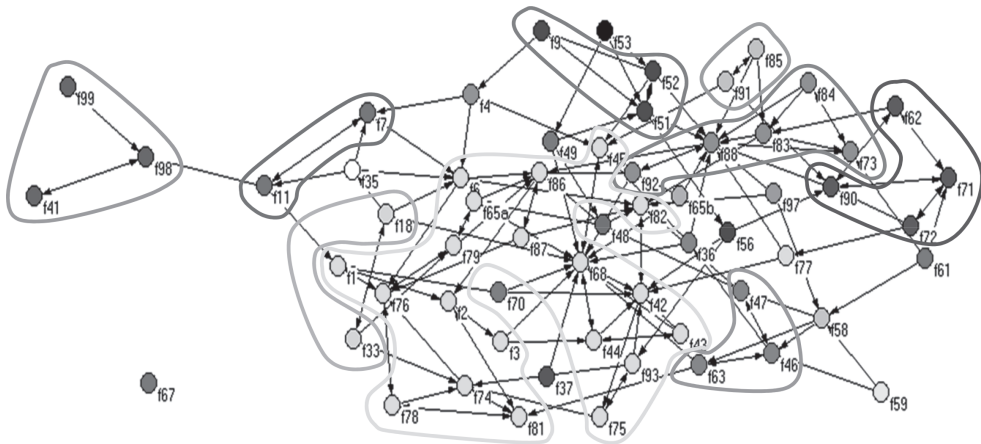
コンポーネント (Pajek の操作)

◆強連結コンポーネントを見つける

- ・ **Network** → Create Partition → Components → Strong → Minimum size 入力（コンポーネントの最小の大きさ）

◆弱連結コンポーネントを見つける

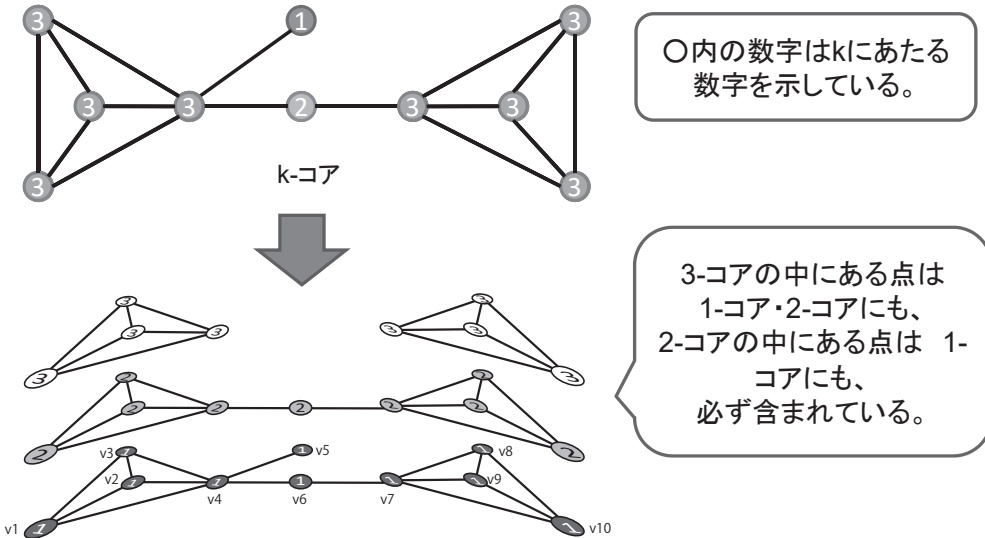
- ・ Network → Create Partition → Component → Weak → Minimum size 入力



k- コア

◆ k- コア

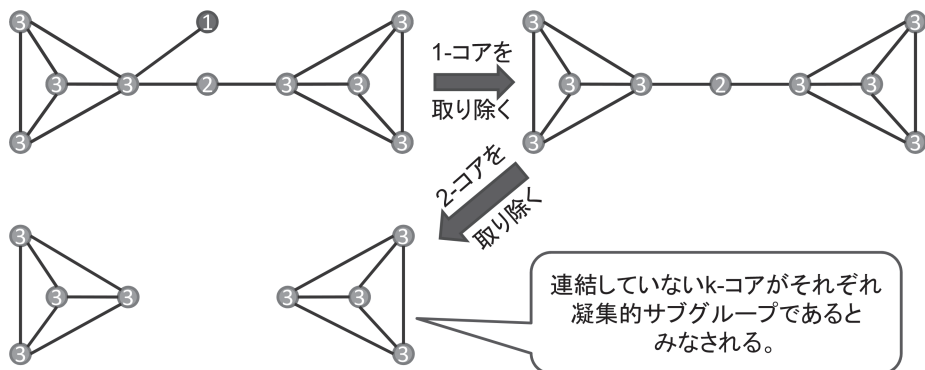
- ・ コア内でそれぞれの点が少なくとも k 個の点と繋がる最大のサブネットワークである。



k- コア

◆ k- コアを使って凝集的サブグループを見つける

- ・ ネットワークが密度の高いコンポーネント群に分解されるまで、低い k -コアから順に取り除いていく。
- ・ 無向グラフにのみ適用できる。

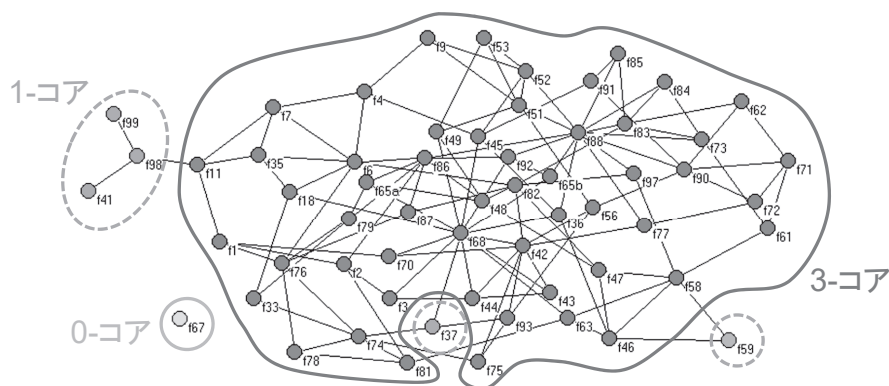


k- コア (Pajek の操作)

◆ k- コアを見つける

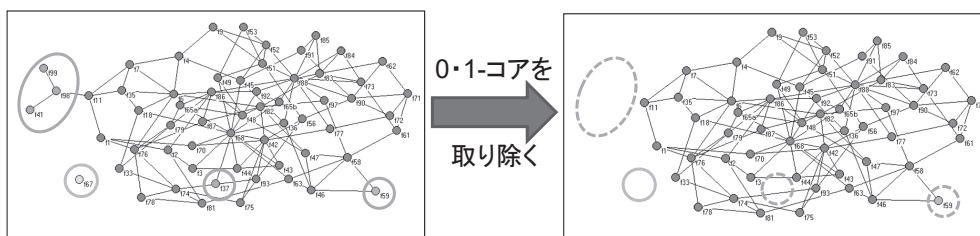
- ・ Network → Create Partitions → k-Core →
 - Input (入次数のコア)
 - Output (出次数のコア)
 - All (紐帯の方向を無視したコア)

ここでは対称化されたネットワークを使用するため、Allを選択



◆低いk- コアを取り除く (k- コア以上を抜き出す)

- ・ Operations → Network + Partition → Extract SubNetwork
→ Select clusters を入力
(抜き出したい 1 番低い k- コア - 抜き出したい 1 番高い k- コア)



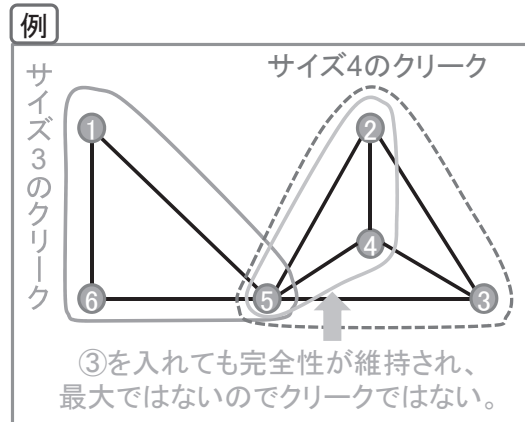
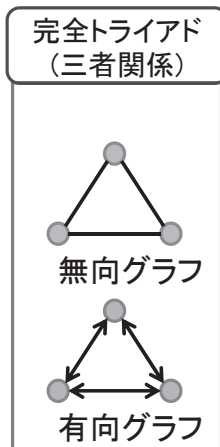
◆ 2 つ以上のコンポーネントに分断されるか調べる

- ・ Network → Create Partition → Components → Strong

クリーク

◆クリーク

- ・ 3つ以上の点を含む、密度が最大の完全サブネットワークである。
- ・ クリークと完全サブグラフは重なり合える。
- ・ 重なり合うクリークの構造が凝集的サブグループとみなされる。



クリーク (Pajek の操作)

◆クリークを見つける

すべての完全なトライアドを見つける。

- ・ 1つ目のネットワークに部分ネットワーク（完全トライアドのネットワーク）を選択する。
- ・ 2つ目のネットワークに対称化されたネットワークを選択する。



・ Networks → Fragment (First in Second) → Extract subnetworkのみチェック → Find

```
100 fragments found.
200 fragments found.
216 fragments found. ——— 完全なトライアドの数
Time spent: 0:00:00
```

◆完全なトライアドの度数分布の確認

- ・ Partition → Info

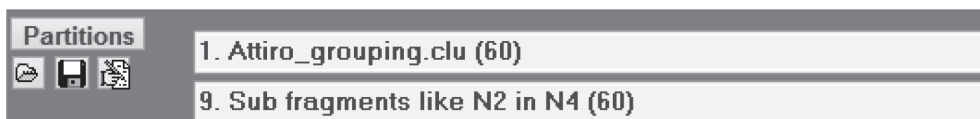
Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
0	13	21.6667	13	21.6667	1
1	19	31.6667	32	53.3333	3
2	9	15.0000	41	68.3333	20
3	13	21.6667	54	90.0000	5
4	3	5.0000	57	95.0000	17
5	1	1.6667	58	96.6667	15
6	1	1.6667	59	98.3333	15
7	2	3.3333	60	100.0000	35
Sum	60	100.0000			

ヒエラルキー

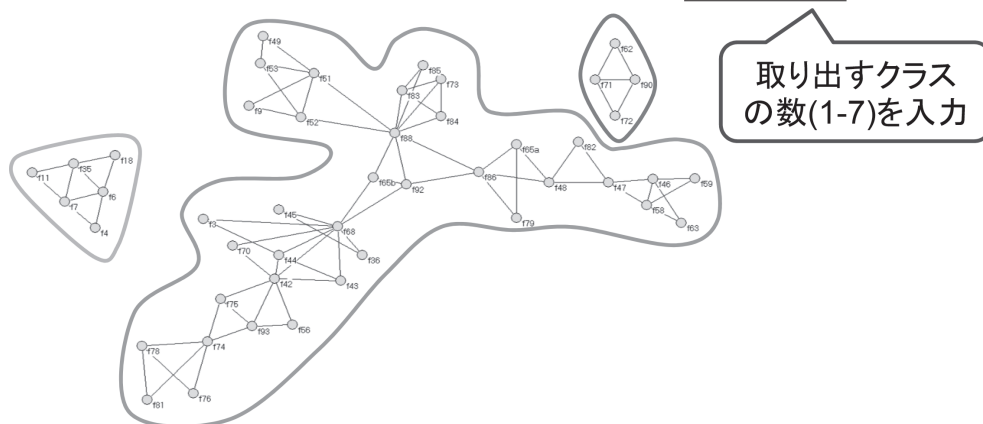
File→Hierarchy→Editで、どの点がどのトライアドに属しているかを見ることができる。

◆元のパーティションを新しいパーティションに適応

- ・ 1つ目のPartition元のパーティションを選択する。
- ・ 2つ目Partitionに前述の工程で作った新しいパーティションを選択する。



- ・ Partitions → Extract SubPartition (Second from First) → Select clusters



第3章のまとめ

◆凝集的サブグループ

- ・社会ネットワークにおける、「互いに密接」な人々の集合体である。
- ・必ずしも社会的なグループである必要はない。

◆ネットワークの密度と次数

- ・ネットワークの凝集性を測る指標として、密度と次数を紹介した。
- ・ネットワークの大きさが同程度なら密度、異なる場合は次数で比較する。

◆ネットワークの抽出方法

- ・ネットワークを抽出して分析する場合に用いる大まかな3つの分類として、k-コア、クリーク、コンポーネントを紹介した。

◆密度の高いネットワークを抽出するには、クリークを用いる。

◆密度の低いネットワークを抽出するには、コンポーネント、k-コアを用いる。

◆探索的に調べるには

- ・無向ネットワークの場合、コンポーネント → k-コア → クリークの順に分類すると良い。
- ・有向ネットワークの場合、強連結コンポーネント → 弱連結コンポーネント → クリーク → 対称化 → k-コア → クリークの順に分類すると良い。

第4章 バランス理論

はじめに

◆この章における目的

- ・前章で学んだ考え方を、友情と敵意、好きと嫌いのように肯定もしくは否定的な感情に基づく関係にまで広げていく。
- ・バランス理論を紹介し、これを主観的・心理的な社会関係に応用する。

◆社会心理学

- ・「人々が集団を形成する過程」と、「それが個人の行動や感情に与える影響」を研究する。

◆バランス理論

- ・社会心理学から派生した関係のパターンに関する仮説と理論である。

バランス理論

◆フリッツ・ハイダー

- ・1940年代にバランス理論の元になる、「人は友人が自分の考えを否定したときに不快になる」という原則を定式化した。
- ・人はバランスの取れていない状況にストレスを感じ、バランスのとれた状況へ変わりやすくなると予測した。

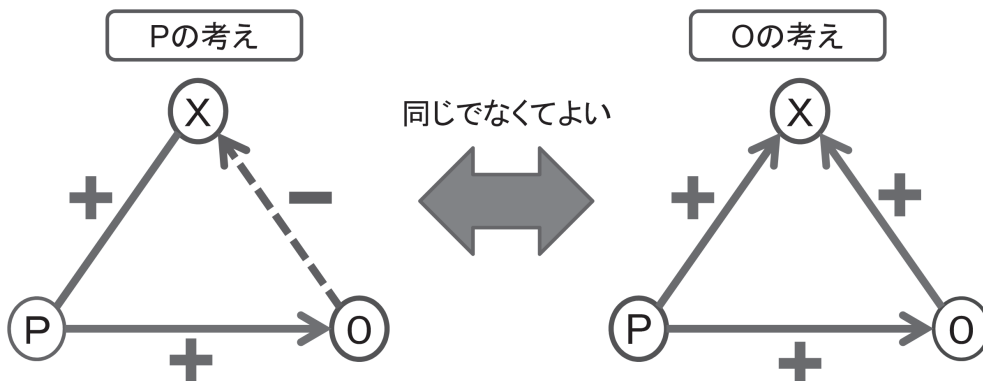
◆ドーウィン・カートライトとフランク・ハラリー

- ・ハイダーの考えをネットワーク分析に移行した。
- ・感情の紐帯構造を表すネットワークである、符号付きグラフを定義付けした。
- ・符号付きグラフがバランスするための正確な状態を定式化した。

フリッツ・ハイダー

◆帰属（アトリビューション）

- ・OからXへの矢印をPの知覚として捉える。
- ・OからXへの紐帯は、「OはXについてそう考えている」というPの考えであるので、Oの考えと同一である必要はない。



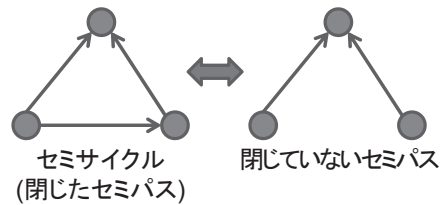
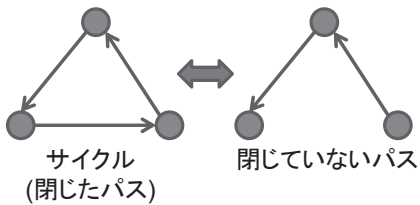
符号付きグラフ

◆符号付きグラフ

- ・それぞれの線が正か負の符号を持つグラフで、負の線は点線で表される。(正の線：肯定的な感情、負の線：否定的な感情を持つ)。
- ・偶数の負の線があるか、もしくは負の線が1つもないサイクルは、バランスしたサイクルとされる。

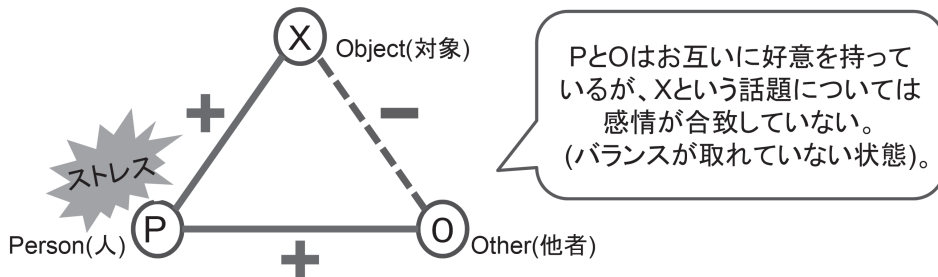
◆サイクル

- ・サイクルとは閉じたパスのことである。
- ・セミサイクルとは閉じたセミパスである。
- ・負の線を奇数本含んでいなければ、(セミ)サイクルはバランスしている。



ハイダーの原則

- ・ 1940年代にバランス理論の元になる、「人は友人が自分の考えを否定したときに不快になる」という原則である。
- ・ ハイダーの「人－他者－対象」という三角形（下の図）は、最初と最後の点が一致するサイクルで表される。

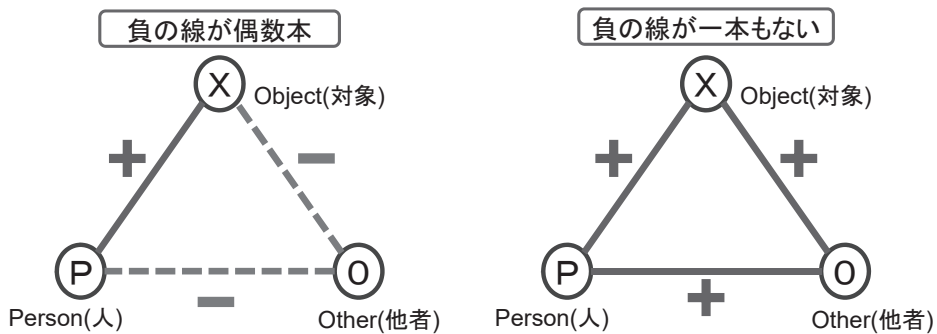


ハイダーは、この時Pが以下いずれかの行動を取るだろうと予測した。

- ・ Xに対するOの意見を変える。
- ・ Oに対する感情を変える。
- ・ Oは本当はXを嫌いではないと思い込む。

符号付きグラフ（バランスがとれているグラフ）

◆ バランスがとれているグラフの例

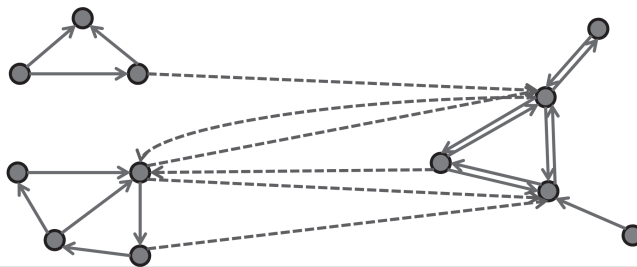


- ・ 上の2つのグラフは負の線が偶数、もしくは負の線がないのでバランスがとれている。
- ・ バランスしているので、Pはストレスを感じていない。

符号付きグラフ（バランスがとれているグラフ）

◆符号付きグラフがバランスするための条件（カートライトとハラリー）

- ・すべての正の弧が含まれている2つのクラスタに分けることができ、すべての負の弧は2つのクラスタの間に位置する。
- ・すべての（セミ）サイクルがバランスしている。
- ・すべての正の紐帯がクラスタ内に含まれ、すべての負の紐帯がクラスタ間に位置している。



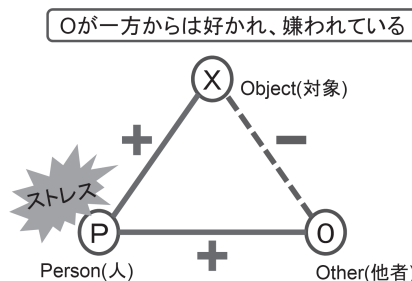
クラスタ化

- ・（セミ）サイクルは1つも負の弧を含んでいなければクラスタ化できる。
- ・符号付きグラフは、すべての正の弧がクラスタ内に含まれ、すべての負の弧がクラスタ間に位置するならば、クラスタ化できる。

符号付きグラフ（バランスしていないグラフ）

◆バランスしていないグラフの例

- ・奇数の負の線があるため、このグラフはバランスがとれていない。
- ・バランスがとれていないので、Pはストレスを感じている。



第4章で例として扱うネットワーク

◆『修道院のネットワークとその変化』

- ・ サミュエル・F・サンプソンのニューイングランドの修道院におけるコミュニティの構造のエスノグラフィーを使用する。修道士の間の感情的な関係データである。
- ・ 彼らが誰を一番好きか、誰を一番好きではないかのデータを収集し、以下のように値をつける（複数の時点で測定）。

好きな人として選ばれた人

一番目 → + 3

二番目 → + 2

三番目 → + 1

好きではない人に選ばれた人

一番目 → - 3

二番目 → - 2

三番目 → - 1

- ・ ここでは4時点目 (T4) ...

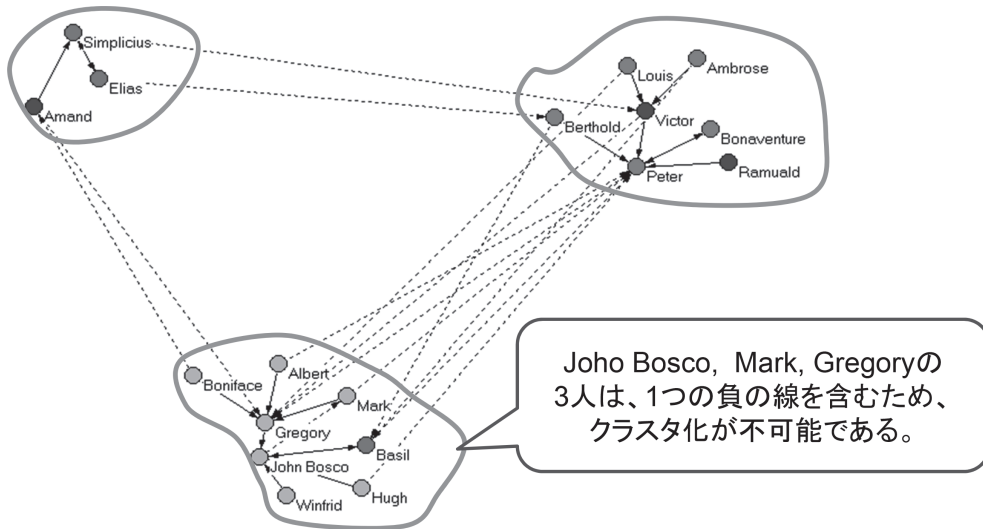
4人が修道院に選ばれる1週間前である、第一の選択だけを用いて、一番好きな人を+1、一番好きではない人を-1とした。

◆修練士のグループ分け

- ・ Young Turks (青年トルコ党) クラス1
 - …変革期に後からやってきた人たちで構成。野党のメンバーが修道院で守っている規律について疑問を投げかけた。
- ・ Loyal Opposition (野党) クラス2
 - …最初に修道院に入った人たちで構成される。
- ・ Outcasts (部外者) クラス3
 - …グループに入れてもらえない人たちである。
- ・ Interstitial (中間層) クラス4
 - …議論のどちらにも肩入れしていない人たちである。

クラスタを見つけ出す

- ・ このネットワークは3つのクラスタに分類される。
- ・ バランスしているというより、クラスタ化が可能である。



最適化手法

◆最適化するには

- ・点をクラスタの中に入れるために何度も再配置し、一番良い解を選ぶ。
- ・探索的に分析する場合は「正の線はクラスタ内に、負の線はクラスタ間にある」という法則に合致しない線が一番少なくなるようにすると良い。

◆最適化手法の3つの特徴

- ①最適化の度合いが同程度の解が出た場合、どれを選ぶかは研究者次第。
- ②例外的に、適合するクラスタが見つけれない場合がある。
- ③クラスタ数などのオプションによって異なる結果が出る可能性がある。

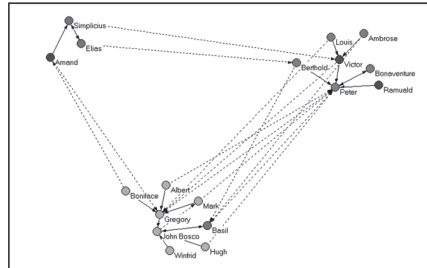


最適化手法を何度も繰り返すことが重要となる。

ソシオグラム作成法（Pajek の操作）

◆正の線の長さが最小、負の線の長さが最大のソシオグラムを作る

- ・ Draw → Network + First Partition
- ・ Options → Values of Lines → Similarities
- ・ Layout → Energy → Kamada-Kawai → Free



◆ランダムにパーティションを作る

- ・ 最初に意味のあるクラスタリングができるパーティションがない場合に使う。
- ・ Partition → Create Random Partiton → 1 -mode

→ Dimension of partition

→ Number of clusters

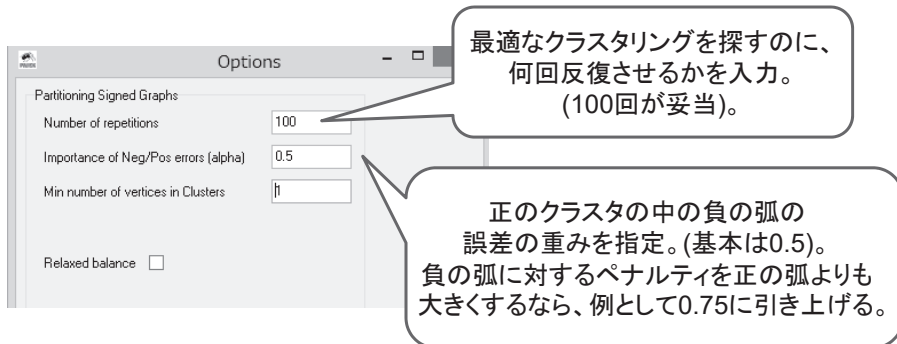
ネットワークを分割する数を
クラスタ数として入力
(この場合デフォルトの18のままでよい)

いくつのクラスタに分けたいかを入力
(ここでは3つに分けたいので3を入力)

クラスタ分けしたネットワークのバランス (Pajek の操作)

◆ Balance コマンドを使用する

- ・ 前述の工程で作成したランダムなパーティションを用いる。
- ・ Operations → Balance*



Balanceコマンドを繰り返す場合は、
常に新しいランダムなパーティションを作ってから始める。

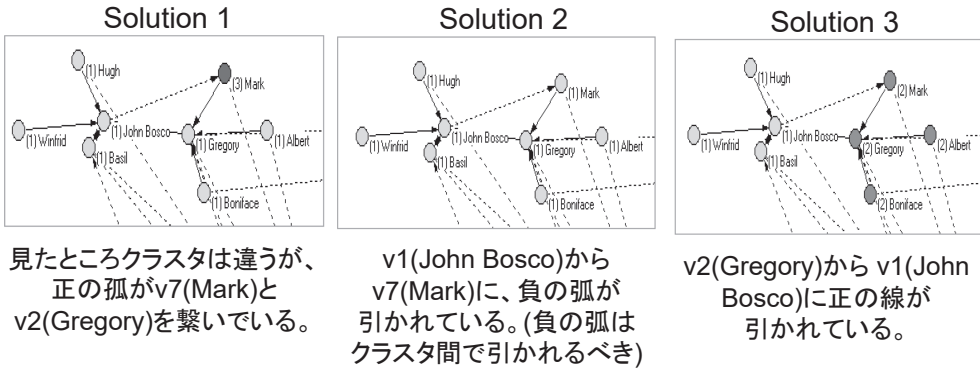
最初のパーティションが良ければ、それ以上良い解を求めずにすむ。

```
Working...
Number of clusters: 3, alpha: 0.500, min size of clusters: 1
----- Starting partition -----
Errors:      0.50      Lines
-----
1.00 :      7.2
-----
Improvements
Final partition 1-----
Errors:      0.50      Lines
-----
1.00 :      7.2
-----
Final partition 2-----
Errors:      0.50      Lines
-----
-1.00 :      1.7
-----
Final partition 3-----
Errors:      0.50      Lines
-----
1.00 :      2.1
-----
3 solutions with 0.50 inconsistencies found.
Time spent:  0:00:00
```

1つだけ間違った弧がある、3つの
解を見つける。

- ・ V7(Mark)からv2(Gregory)の正の弧
- ・ V1(John Bosco)からv7(Mark)の負の弧
- ・ V2(Gregory)からv1(John Bosco)の正の線

クラスタ分けしたネットワークのバランス（Pajek の操作）



John Bosco—Mark—Gregory の、クラスタ化できない3人が問題となる。

* 注: 安田研では現在これらは描画不可能である。

時系列の発展（Pajek の操作）

◆時系列のネットワークを作成

・ Network → Temporal Network → Generate in Time

- All（それぞれの時点でのネットワークが作る）
- Only Different（前の時点と違うネットワークだけを作る）

- ・ Select first time point（最初の時点を打ち込む）
- ・ Select last time point（最後の時点を打ち込む）
- ・ Select step（> 0）（各時点のネットワークを見たい場合は1と打ち込む）

◆時系列で見られるように設定

・ Options → Previous/Next → Apply to

・ Options → Previous/Next → Apply to

・ Operations → Balance*

時系列で変化するネットワークを自動的に
適切なネットワークへと変える。

それぞれの時点におけるバランスや
クラスタビリティを調べる。

PreviousとNextコマンドで、ネットワークの時点の切り替えが可能。

第4章のまとめ

◆バランス理論について

- ・符号付きネットワークにおける凝集的なサブグループについて論じた。
- ・関係が感情を表す場合、人々は好きな人同士で集まる傾向にあり、否定的な感情はグループをまたいで存在する傾向にある。
- ・バランスとクラスタビリティは対立するグループが分極化する中で起きる。
- ・「負の線はクラスタ内、正の線はクラスタ間」という原則に従って2つのクラスタに分けられ、符号付きネットワークのバランスがとれている。
- ・3つ以上のクラスタに分かれるならば、クラスタ化が可能なネットワークである。

◆最も適合したクラスタリングとの誤差

- ・「負の線はクラスタ内、正の線はクラスタ間」というバランス理論の原則に従わない重み付きの線は、ネットワークのバランスやクラスタビリティを表している。

第5章 所属

はじめに

◆社会的サークル

- ・ゲオルグ・ジンメル社会学で、「1つ以上の組織やイベントに集う人々のグループ」と定義される。
- ・人々はいくつもの社会的サークルに属しているので、所属のネットワークから人々の類似性を推測出来るかもしれない。
- ・これまでは「人と人」「組織と組織」の関係を扱ってきたが、第5章では人と組織の関係、その中でも「所属 (Affiliation)」と呼ばれる関係に注目する。



所属のネットワークを分析する際に、線の値をどのように扱うかを考え、社会ネットワークを三次元で表現する方法を学ぶ。

第5章で例として扱うネットワーク

◆『役員兼任と企業間ネットワーク』

- ・ 20世紀初め（1904～1905）のスコットランドにおける企業を例に、兼任役員が企業を連結させるネットワークを分析する。136の兼任取締役と108の大きな株式会社が含まれている。

（非金融企業）	64社	（銀行）	8 社
（保険会社）	14社	（投資／不動産）	22社

- ・ 2社の取締役となっている人は、兼任取締役なので2社を連結させる機能を果たす。

所属ネットワーク

◆所属ネットワーク

- ・ 所属と点の2つの集合で、通常は「イベントと行為者」と呼ばれる2つのセットで成り立つ。（例 取締役＝行為者、取締役会＝イベント）

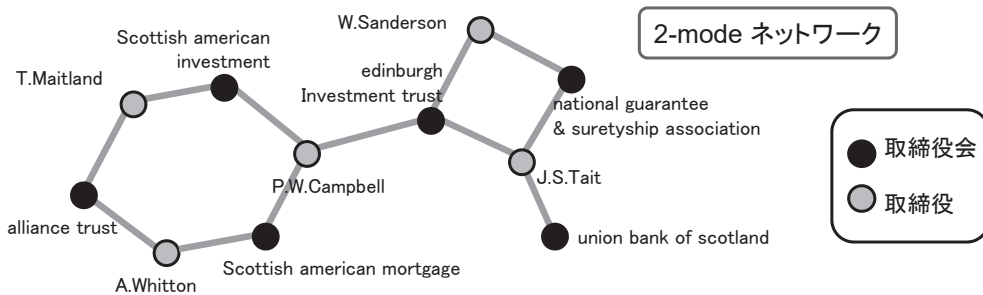


図1 スコットランドの役員兼任ネットワークの一部

* 注:この図は現在、Pajekでは出力不可能。

2-モードネットワークと1-モードネットワーク

◆1-モードネットワーク

- ・ すべての点に関係を持ちうるネットワークである（例「人－人」「組織－組織」）。

◆2-モードネットワーク（2部ネットワーク（bipartite network））

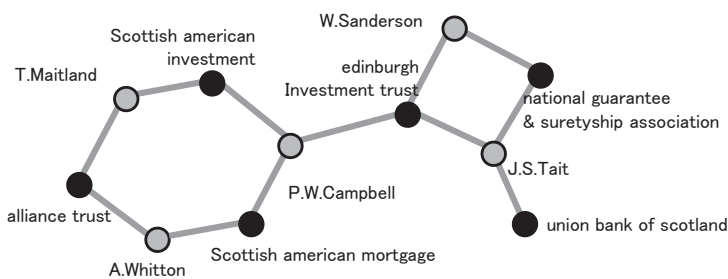
- ・ 点は2種類のセットにわかれ、各点は自分の所属しないセットの点とのみ関係を持ちう

るネットワークである（例「人－組織」）。

◆ 2つのネットワークの違い

- ・ 次数などの単純な指標が異なる意味を持つ。
- ・ 結び付ける点の数が異なるため、ネットワークに存在する最大の紐帯の数も異なる。
- ・ 最大紐帯数が異なるため、ネットワークの密度も異なる。

2-モードネットワークの分析



企業 ● の次数＝企業の兼任取締役の数	所属イベントの大きさ
取締役 ○ の次数＝取締役会に属している企業の数	行為者の参加比率

2-モードネットワークの特殊な分析手法はとても複雑である。



2-モードネットワークを1-モードネットワークに変換したうえで標準的な手法で分析する。

2-モードネットワークの分析 (Pajek の操作)

◆ 2-モードネットワークから 1-モードネットワークへ変換する。

・ Network → 2-Mode Network → 2-Mode to 1-Mode → Rows

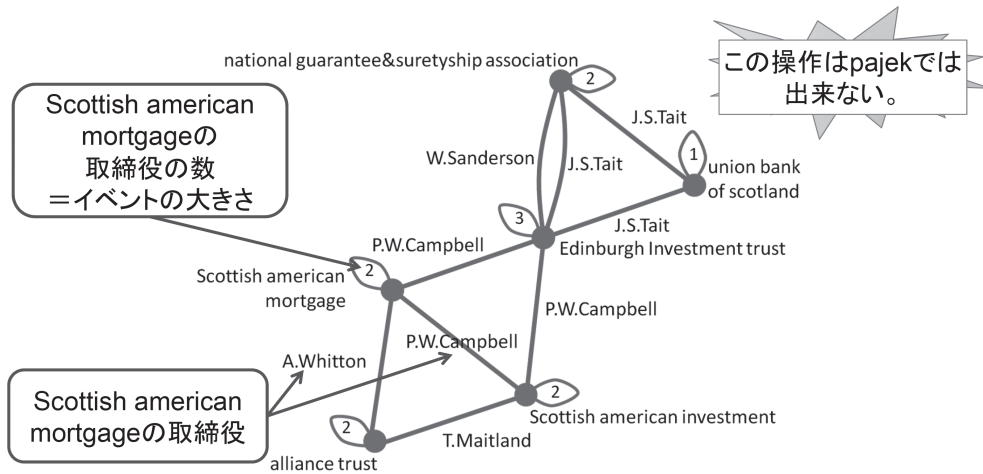


図2 図1から作成した1-モードネットワーク

値付きネットワーク

◆ 値付きネットワーク

・ 多重線を、2つの点を結びつける線の数値を付けた1本の線に置き換えて、値付きネットワークに変換出来る。

・ 線の値は線の多重性 (line multiplicity) と呼ばれる。

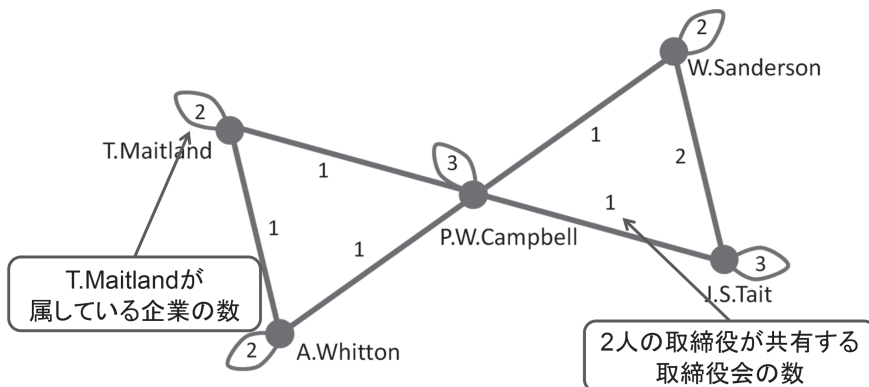


図3 図1から抽出した取締役の1-モードネットワーク

Pajekで算出しうること (Pajekの操作)

◆各点がつながっている所属先の数を調べる。

- ・ Network → Create New Network → Transform → 2-mode to 1 mode
→ IncludeLoops, Multiple lines

◆2-モードネットワークにおける企業の次数を調べる。

- ・ Network → Create Partitions → Degree → Input

Loopはpajekでは
描画できない。

◆次数パーティションから所属パーティションのクラス1を抽出する。

- ・ Partitions → Extract SubPartition (Second from First)

◆線の値の分布を調べる (線の多重性)。

- ・ Network → Info → Line Value

Line Values				Frequency	Freq%	CumFreq	CumFreq%
(---	1.0000]	231	83.6957	231	83.6957
(1.0000	---	2.0000]	28	10.1449	259	93.8406
(2.0000	---	3.0000]	7	2.5362	266	96.3768
(3.0000	---	4.0000]	3	1.0870	269	97.4638
(4.0000	---	5.0000]	1	0.3623	270	97.8261
(5.0000	---	6.0000]	0	0.0000	270	97.8261
(6.0000	---	7.0000]	0	0.0000	270	97.8261
(7.0000	---	8.0000]	6	2.1739	276	100.0000
Total				276	100.0000		

m-スライス

◆m-スライスとは

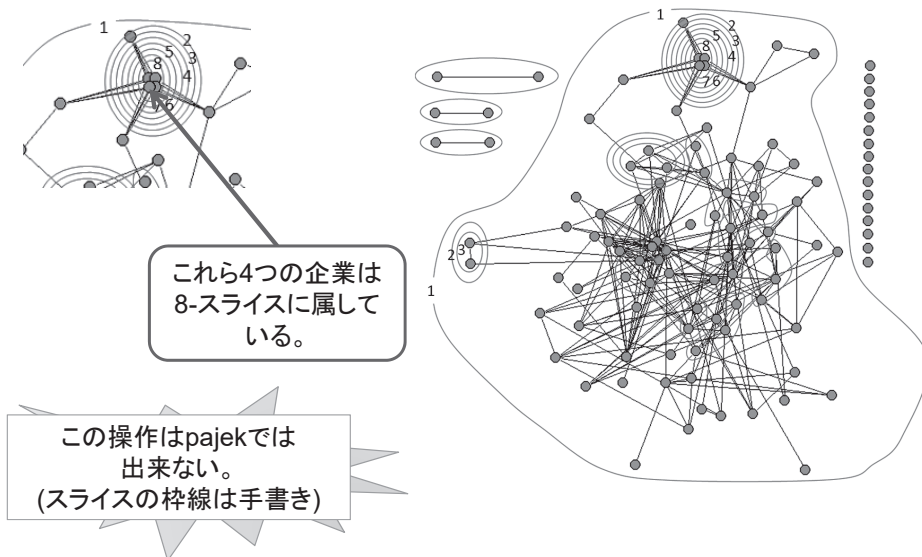
- ・ 多重性や線の値によって定義されるサブネットワークである。
- ・ m以上の多重度を持つ線と、これらの線と接続する点を含む最大のサブネットワークである (例兼任取締役が8人 = 8スライス)。
- ・ m-スライスは連結している必要はない。

◆多重線 (線の多重性)

- ・ 凝集的なサブグループの分析では、近傍の線ではなく線の多重性を察する。
- ・ 例えば、2つの企業を兼任する取締役の数が多いほうが、紐帯はより強く、より凝集的

で強い相互依存性がある。

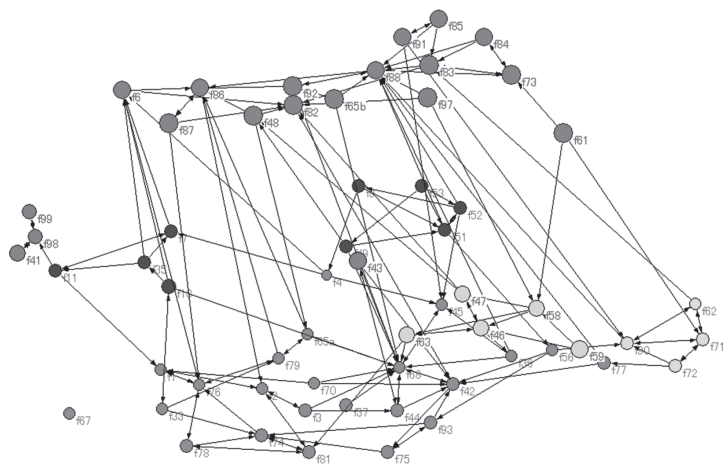
m- スライス



3次元ネットワーク (Pajek の操作)

◆ San Juan Sur の情報ネットワークを3次元ネットワークで描く

- ・ Layout → VOS Mapping → 3 D → In z direction → ScrollBar On/off
- ・ 画面上では3次元出力が可能である。



まとめ

◆所属ネットワーク

- ・ 2つの所属ネットワークの特徴と指標の違いを紹介した。
- ・ 2-モードネットワークは複雑なので、1-モードネットワークに変換した後に分析を行った。

◆値付き（あるいは重み付き）ネットワーク

- ・ 多重線を値付きネットワークに変換する方法を紹介した。
- ・ 値付きネットワークの線の値は、線の多重性（line multiplicity）と呼ばれる。
- ・ 凝集的なサブグループでは近傍の線ではなく線の多重性で考える。

◆m-スライス

- ・ 多重性や線の値によって定義されるサブネットワークである。スライスの等高線は手で描く。

分冊刊行予定

- ・ Part III 仲介
 - ・ 第6章 中心と周辺
 - ・ 第7章 仲介者とブリッジ
 - ・ 第8章 普及
- ・ Part IV 序列
 - ・ 第9章 威信
 - ・ 第10章 序列
 - ・ 第11章 引用
- ・ Part V 役割
 - ・ 第12章 ブロックモデル
 - ・ (2011年版のみ) 13章 ランダムグラフモデル
- ・ 付録

参考文献

Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj (2005) *Exploratory Social Network Analysis with Pajek*, 安田雪監訳『Pajek を活用した社会ネットワーク分析』 東京電機大学出版局.

* 安田研究室所属の学生及び大学院生の協力に謝意を表する。

—2017.12.6受稿—

