

Social Network Analysis

第3章

Handling Relational Data

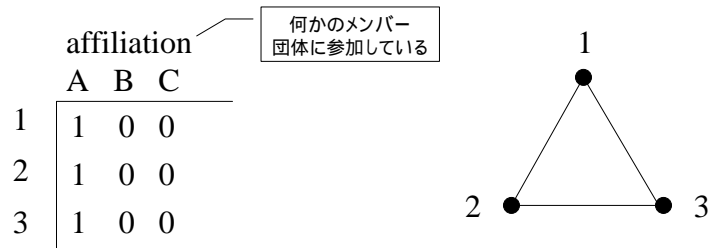
経済学研究科博士課程
桑島由芙

この章の目的

- 社会ネットワーク分析は社会構造を分析するツールであり、構造の関係について明らかにするためのものである。
- relational dataは属性データよりもavailabilityが重要であるので、この章ではどうやってrelational dataを集めて、蓄積し、社会ネットワーク分析に対して用意をするのかについて説明する。
- relational dataに特有でない部分に関しては他の本を見るべし。ただし、他の本で十分に説明し切れてないことに関しては扱う。

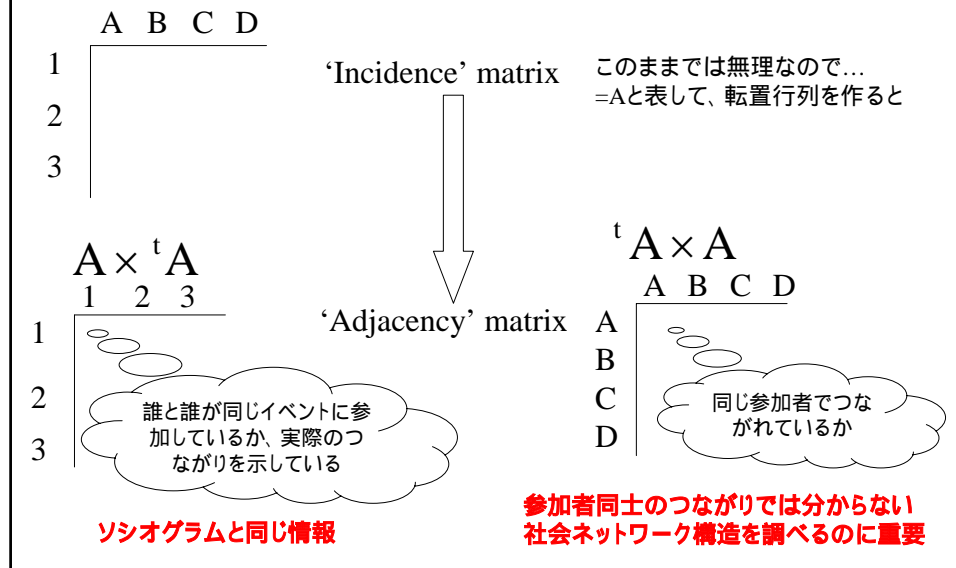
The Organization of Relational Data

Data matrix と sociogram



1,2,3は何かのイ
ベントで会う

Two-mode dataからone-mode dataへ



調査上の注意点(1)

- まずは‘incidence’ matrixを書いてから、‘adjacency’ matrixに変換するのが良い。
- どの変数をcaseにしてどの変数をaffiliationにするのかは調査の目的によって考える。
- ただしどちらにしても分析上の手続きに置いては変わりがなく、慣習的なものである。
- 会社と取締役の例 (figure3.5)

調査上の注意点(2)

- 変数をどう測定するかという問題 (binary or valued) (undirected or directed) の2軸で分けられ、4種類の方法がある。 <figure3.6>
- 変数を valued から binary に変えるのは簡単 (例えば、ある値以上なら1、未満なら0という slicing という方法がある)。 <5章で詳しく扱う>
- Directed data の場合、(3,6) と書く場合は3から6への影響、(6,3) と書く場合は6から3への影響を表すのでアシンメトリーなデータになる。
- すべてのデータは binary で undirected のデータに変換できるのでこのタイプが最も多用な分析ができる。

The Storage of Relational Data

分析の前の準備

- 10casesと5affiliationsを超えたらコンピュータを使う方がよい。
- データを集めたら、分析の前にいくつかやることがある。
- 変数をあとから識別しやすいようにアルファベット順等に並べる。その際に、linked listの形になっているとより良い。
- ucinetとexcelは互換性があるので、データの移動が簡単にできる。
- データがspreadsheetの形になれば、基礎的な統計的指標を計算できる。ほとんどのプログラムでは行列計算もできるのでネットワーク指標も計算できる。変数のソートもできるし、spreadsheetの一部を取り出すことも可能。有名な社会ネットワーク分析のプログラムはucinetとstructureであるが、早めにそれらのプログラムに入れてしまうのがよい。高度な統計分析をしたい時はspssなどにexpirtするのがよい。

GRADAP

- 最も強力な社会ネットワーク分析のプログラムの一つはGRADAPであるが、それで扱うデータは今までのmatrixデータと異なる。matrixの形の方が直感的に理解できる。但し、SPSSと互換性はある。
- Figure3.5のファイルと同じデータをGRADAPのファイルにしたのがfigure3.7
- GRADAPではすべてのデータが失われることなく分析することができる。
- いったん、GRADAPのファイルにしてしまえば様々分析をすることが可能であるが、グラフ理論の知識が必要である。だから初心者やたまたまネットワーク分析をする人はucinetを用いるのが良い。

The Selection of Relational Data

境界線問題(1)

- 小さなネットワークであればすべての人とすべての繋がりを調査することもできるが簡単ではない。またどこで繋がりとするかも問題である。例えば、「親友」という定義も人によって違うかもしれない。
- どのような境界も個人個人の繋がりに成り立っており、人工的なものである。
- 世の中すべてのネットワークを分析することはできず、ある一部分を取り出してその外側は無視することになるので不完全な分析であることを意識しなくてはならない。インフォーマルなネットワーク (Yablonsky, 1962)でもフォーマルなネットワーク(Kerr & Fisher, 1957)でも同じである。
- どこで境界線をひくかという問題は単に調査する上で自然に見える境界を識別するという問題ではない。

境界線問題(2)

- 関連ある境界が識別できたとしても、まずターゲット集団を定義しなくてはいけない。それには二つ方法がある。 positional approach
reputational approachである。
- positional approachでは、グループを識別し、そこから何人かランダムにサンプルを選ぶ。ただし、このアプローチではどのポジションまでが含まれるのかが問題になる(例えば、エリートの場合、トップマネジメントといった場合、どこまでか)。また、自然に感じられるサブグループ間で境界線を引いてよいかという問題もある。
- reputational approachは全体のリストがあげられない場合やポジションに関連がない場合に用いられる。ターゲット集団のことをよく知っている人を探し出し、その人に「権力を持っている人は？」などの質問をする。どの人に質問をするかということが重要な問題になる。それが難しい時にはスノーボールサンプリングをする。

サンプリング問題

- ある程度の大きさを持つ社会を調査する時などにはコンピュータでも扱えないほどの大きさになるのでサンプリングの必要性が出てくる。
- 以前は1000位のサンプリングが限界だったが、コンピュータの性能の向上によりより多くのサンプリングが可能になった。統計的にはその方が有意な結果がもたらされる。
- サンプリングすることによって失われるデータは、Burt(1983)によると、 $(100-K)\%$ である(k は全ネットワークに対するサンプルサイズ)。
- また、サンプリングデータは調査に答えていない人の存在が無視されるという問題がある。結果にバイアスがかかっている可能性がある。

サンプリング問題への対処

- 社会全体への関心をすて、個人的なエゴセントリックネットワークを調べる。
- スノーボーリングサンプリングをする。
- Burt(1983)が提案した方法(構造的に似たような位置にいる人は似たような行動をするはず)。

理想的なサンプリング

