

## 第8章 (pp.146-174) Dimensions and Displays

2005-06-29

杉山達彦

1

## 概要

- 標準 (hub-and-spoke) 型のソシオグラムは大規模ネットワークの可視化に適さない
  - 点の数が10~20以上になると、比較的低密度のネットワークでも関係 (線) の交差数が増えて、理解しづらい
- Grieco (1987)、Scott and Hughes (1980)
  - 円周状に配置すると線が交差しづらくなり見やすい (図8.1、circle diagram)
  - 点の配置は、線の交差数が減るように、人間が、美的感覚に基づいて、trial-and-errorで最適化する
- 関係 (線) の見やすさは向上したが、  
点の位置、線の長さには何の意味もない

2

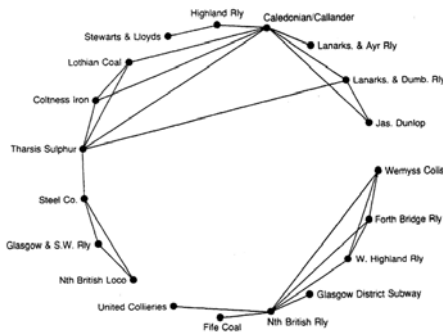


Figure 8.1 Scottish companies: a circle diagram

3

- McGraph et al. (1997)
  - データに対する感覚的な理解と視覚的な構造の関連性を実験により示した
    - 被験者は、同じネットワークでも、空間的配置が異なる場合には、異なる数のサブグループを識別
  - 読者に計量社会的特性 (sociometric property) を推理させたいなら、2点間の物理的な距離を、(グラフ)理論的な距離と可能な限り一致させるべき、と結論付けた
- 本章の目的
  - 多次元尺度構成法 (multidimensional scaling, MDS) を用いたネットワークの可視化方法
    - 地図を見て新しい国を発見するように、ネットワークの空間的な構成から何かを発見できるように

4

## 本章の構成

- 概要
- Distance, Space and Metrics
- Principal Components and Factors
- Non-metric Methods
- Advances in Network Visualization
- Elites, Communities and Influence

5

## Distance, Space and Metrics 距離、空間、メトリクス (計量指標)

- MDS (多次元尺度構成法)
  - 空間内の距離によって変数間の強さを表す (安田先生)
  - 関係データを空間 (位置、距離) にマップする方法
    - メトリクス (計量指標) に基づいてマップする
    - 距離 (distances) や方向 (directions) が意味づけされる
  - 例 ユークリッドメトリック (Euclidean metric)
    - 2点間の距離は最短距離 (図8.2, 2cm) as the crow flies
  - 対 グラフ理論
    - 2点間の距離は間に存在する線の数 (図8.2, 3) path distance
  - ユークリッドメトリックの算出例
    - グラフ理論の距離を物理的な距離に変換することで、作成できる
      - path distanceが3のとき2cmとすれば、図8.2のように配置される

6

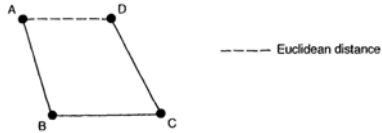


Figure 8.2 Euclidean distance

- 2点 (A, B) 間の距離
- ・グラフ理論の距離は3
- ・ユークリッド距離は2 cm
- 直感的に2点 (A, B) 間の関係強度を把握できる

7

- グラフ理論の点の近さ (closeness) を、MDSは近接データ (proximity data) として表現
- 算出方法
  - graph theoretical measures (ソシオグラム) から、case-by-case proximity matrixを算出

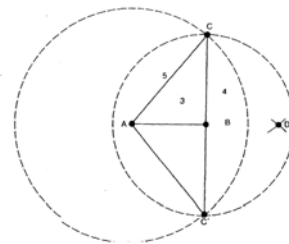
- proximity data
  - 人間間の対話頻度、企業間の株式保有量、コミュニティ間の重複メンバー数に基づく相関係数
  - 人間間、企業間、コミュニティ間の相関係数
- similarityを表現したいなら、相関係数に比例した値をセルに設定
- dissimilarityを表現したいなら、反比例した値をセルに設定

8

### 最も簡単なMDSの例

- 町のネットワークを、地図のように2次元座標上に視覚化すること
  - East/West, North/South
- 2町間の道路の距離行列は、近接データであり、地図上のcm距離に容易に変換
- 地図は使いやすいが、不都合な点も
  - 2町間の移動時に常時同じ道路を通るとは限らない
    - 2町間の距離は一定ではない
  - 高さの次元が捨象

9



	A	B	C	D
A	0			
B	3	0		
C	5	4	0	
D	6	3	5	0

距離行列から、コンパスを使って点の位置を決定可能

Figure 8.3 Constructing a metric model

10

### rotation (回転)

- 点の位置関係を解釈しやすい軸を与えること  
discovery meaningful dimensions
  - 位置関係を変えずに、軸は自由に回転
  - 原理は下記ページに詳しい
    - <http://kogolab.jp/elearn/icecream/chap8/sec4.html>
  - 例えば、距離行列の場合、南北軸、東西軸を、北を上、東を右に配置すれば、解釈しやすい
  - Torgerson(1952)が最初に取り扱った
- 幾何学的原理に基づくMDSアルゴリズムはデータの近接性の扱いに適す

11

### Principal Components and Factors 主成分、因子

- 主成分分析  
principal components analysis PCA
- 因子分析 factor analysis
  - 古典的には異なるが、今日は同一視
  - SPSSというソフトウェアが有名

12

- 属性データのcase-by-variable matrix
  - <http://kogolab.jp/elearn/icecream/chap7/sec1.html>
- 相関係数に基づくvariable-by-variable matrix
  - 変数間の相関に基づいて、変換する
  - <http://kogolab.jp/elearn/icecream/chap8/sec1.html>
- 因子分析に基づくvariable-by-component matrix
  - 元データを2次元にマップするために、
  - 第1主成分、第2主成分(共通因子)等を固有値に基づいて算出し、caseを表現するための変数(観測変数)を減らしていく
    - <http://kogolab.jp/elearn/icecream/chap8/sec4.html>
- 因子分析の原理は下記ページに詳しい
  - <http://kogolab.jp/elearn/icecream/chap8/sec0.html>
  - <http://gucchi24.hp.infoseek.co.jp/SHUSEI.htm>

13

## factor loading 因子負荷(量)

- variable-component matrixの各セルの値
  - 観測変数(variable)に対して共通因子(component)がどれくらいの強さで影響を与えているか
- 原理は下記ページが詳しい
  - <http://kogolab.jp/elearn/icecream/chap8/sec3.html>

14

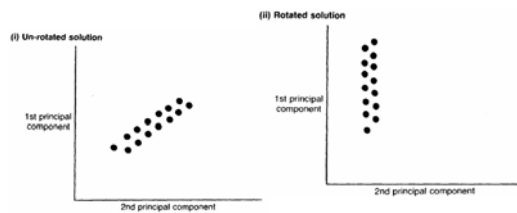


Figure 8.4 Rotation

15

- 別の計算方法(Qモード主成分分析)
  - case-by-variable matrixの転置行列を生成
  - case-by-case correlation(相関) matrixの算出
  - ... 因子分析 ...
  - case-by-component matrix
- 前者はRモード主成分分析
  - variable-variable matrixを利用
    - オリジナルデータのカラム(項目名)を利用

16

- 関係データを調べる場合
  - persons-by-organizations matrix
  - ... Rモード ...
  - organizations-by-components matrix
  - persons-by-organizations matrix
  - ... Qモード ...
  - persons-by-components matrix
- CONCOR
  - ネットワークを特定のブロックに分割するアルゴリズムの一種
  - 相関係数を距離もしくは近接データと捉え、近接性に基づいて分割

17

## Non-metric Methods 測定基準(距離関数)のない方法

- MDSはマトリクス(計量指標)ありきの手法
  - 因子分析では、最初に相関係数(近接性)を算出
- 関係データにおいては、相関係数はどのくらい近いか、を示しづらい
  - 4人の重役が共通な2企業
  - 2人の重役が共通な2企業
  - 前者は後者の2倍近いといえるか?

18

## non-metric MDS

- smallest space analysis (最小空間分析)
  - 最初の行列のセル値を変換しない
  - セル値に関する値を算出
    - セル値に基づいてセル間の順序(ランク付け)を算出
    - セル値に基づいてセル間のユークリッド距離の順序(ランク付け)を算出
- variable-by-variable matrix
- …因子分析…
- variable-by-component (loading) matrix
  - 各データ間の距離(近接データの順序が算出された)と解釈し、ユークリッド空間にマッピング可能
    - コンパスを使うイメージ

19

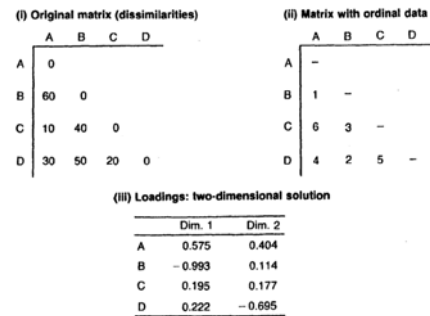


Figure 8.5 Data for non-metric multidimensional scaling

20

- ユークリッド距離の算出方法
  - trial-and-error
  - MINISSAアルゴリズムが一般的に使われる
    - 開始時の設定
- trial-and-errorの終了条件
  - シェパードダイアグラムで判定
    - ノンメトリック多次元尺度法で使う。非類似性の順位の軸と、距離の順位の軸でできた、散布図。各点の配置の適切さを評価するのに用いる。(安田先生)

21

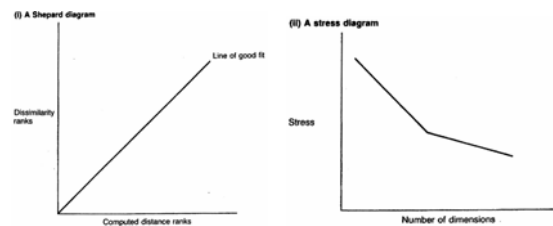


Figure 8.6 Goodness of fit in non-metric multidimensional scaling

- number of dimension (次元数) はstress値に基づいて決定
  - データとモデルの乖離量
  - 小さいほど好ましい
- 原理は下記ページに詳しい
  - <http://www.statsoft.com/textbook/stmulzca.html#computational>
  - <http://www.mindware-jp.com/xlstat/demo-mds.html>

23

- Kruskal (1964a)
  - ストレスが
    - 20%以下なら poor fit
    - 10%以下なら fair fit
    - 5%以下なら good fit
- Freeman (1983)
  - ストレスが0のときだけ good fit

24

- 2次元の散布図で多次元を表現する方法
  - Levine
    - 3次元画像の2次元への写像
      - X-Y, X-Z, Y-Z
  - Burt
    - クラスタ分析
    - 等高線で可視化
- 最後に、これら指標は人工物であることは間違いなく、改善の努力は大切、的な文章

25

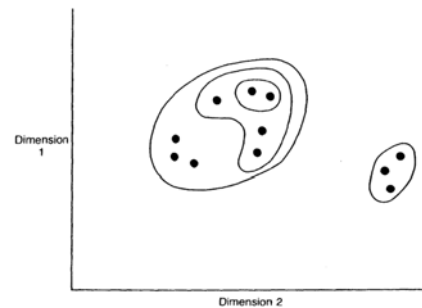


Figure 8.7 A multidimensional scaling configuration

26

## Advances in Network Visualization ネットワーク可視化の進歩

- ここまではスケールの問題は考慮していない
- Klovdahl (1981, 1986, 1989)
  - VIEW-NET
    - molecular modeling
      - 3D, ball and stick
- Krempel (1994)
  - circle shape (点の回帰曲線) に複数の点を圧縮
  - 更に、ネットワークにサブネットワークが含まれる場合
    - separate circles within a larger circle
    - macro-point
  - a hierarchy of nested circles

27

- Freeman (1996a)
  - 物理学を持ち込んだMOVIEMOL
    - molecular modelling技術
    - 3D, balls and sticks
    - computational modelというよりphysical model
  - 色
  - ネットワークの構造的な特徴に基づく配置
  - アニメーションによるネットワーク構造の時間変化の可視化
    - small networkの時間変化の分析に宗旨替え
  - 点の配置に制約(力学モデル)
- MAGE
  - 点の配置の自由度が高い
    - 強度, 多重度
  - VRML
    - 視点移動, アニメーション, 対話型

28

## Elites, Communities and Influence えり抜き、コミュニティ、影響

- Edward Laumann (1996)
  - MDS, non-metric MDSを利用した先駆者
  - 1960コミュニティ
  - コミュニティの権力、エリート構造
  - Homans監修の博士論文
  - Parsons, Harrison Whiteの影響
    - Parsonian system theory
  - Louis Guttmanの影響

29

- Bogardus (1959) が言うところのsocial distance
  - 1920年代に盛ん
- Edward Laumann (1996)
  - social distance = 職業的地位が異なる人々間の係わり合いの差異のパターン
  - social distanceの客観的な測定、ではなく
    - どのくらい異なる地位の間での係わり合いがあったか
  - social distanceに対する主観的な感覚、を導入
    - 他人にポジティブ/ネガティブな態度を示すかどうか
    - sociological level of relational association
  - 客観的なsocial distanceを測定し、MDSを用いて社会構造の地図にマップすることが目的

30

- 白人住民、多様な職位@Boston, Massachusetts
- 友人、親族、隣人関係に関する質問を、職位別に結果をまとめた
  - top professional and business
  - semi professional and middle business
  - clerical(聖職者) and small business
  - skilled manual(体を使う労働者)
  - semi- and un-skilled manual
- position-by-attribute matrix
  - 同じ職位をどのくらい選んだか

31

- 業績は、係わり合いの差異のパターンと、生来の階層の関係を明らかにしたこと
  - 従来、職位の階層はprestige(名声)によるとされていた
  - 係わり合いのパターンが階層を形成すると考えて、研究した
  - 7つの社会関係(?)のそれぞれについて、55 x 55の行列を作成
    - 55の職位(Duncan 1961)
    - 職位間の類似性を判定
- 3次元空間(2次元+等高線)に可視化した
  - 等高線は、prestige(名声)
  - 相関係数は0.824(高い)
  - そのほかの2つの次元についてあまりうまく説明できず、目的の達成は困難だった。
- Warner, Parsons(1951)なども、職位に関する研究をしている

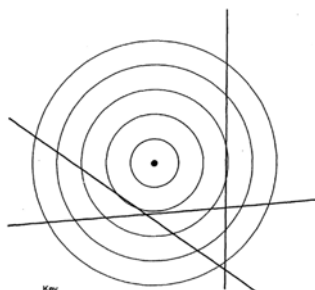
32

- Laumann@Detroit(1966, published 1973)
  - 1013白人男性、友人関係
  - 民族・宗教的な地位における友人関係
    - MDSの視点から民族と宗教を区別する必要がなく、統合して扱ようになった
    - 22民族グループ、15宗教グループ 27民族・宗教グループ
  - MDSにより3次元にマップした
    - 地位
      - プロテスタント、カソリック、ユダヤ
    - 収入の相関計数
    - 教会の訪問回数
    - 民族クラスターが識別された
  - 地位分析の結果
    - 16の地位グループにフィットする2次元のMDS指標を見出した
    - 地位・名声
    - 企業・官僚

33

- Pappi, Laumann(1973, 1976)
  - エリート層の構造(Hunter 1953, Dahl 1961)
  - 西ドイツJülich
    - 1950, 1960年代に急速に発展
    - 新住民、キリスト民主党、社会民主党の党员
    - 最も権威のある地位にフォーカス
      - Pasonian A. G. I, Lシステム
  - 46人に、この町において、他の45人で最も影響力のある人は誰かを質問
    - Herr Kは46人からの推薦を得た(自薦含)
  - 3種の社会関係を抽出
    - 仕事と専門の関係、社会的裕福さの関係、コミュニティイベントの関係
  - 関係の有無のみを調査(方向は除く)
  - ノンメトリックMDSで2次元にマップ
    - コミュニティ構造におけるintegrative centrality
    - community affairsの関係
      - コミュニティの政治的構造
      - 階層的な円状の構造(Figure 8.8)中心が影響力が強い(参加者が多い)
      - five key community issues??
      - 円と直線によりセグメント分割されている

34



35

- MDSの適性が証明された
  - コミュニティの権力構造
  - 地方のエリート層の識別
- Levine(1972)
  - 国家的な経済エリート層を識別するために、管理職の関係を聞いた
  - Harvardの研究者、SNAにMDSを用いた先駆者
  - アメリカの銀行業務(Patmanのデータ 1966)
    - 上位100社、14銀行、3都市
    - 70(社) x 14(銀行)行列、セル値は共通の役員数
      - preference もしくは類似性を示す
    - 3次元にマップ(Figure 8.9)
      - 地域構造(ニューヨーク、ピッツバーグ、シカゴ)、
      - 不明(うまく説明できず)
      - 企業・銀行

36

- Levineのマップは原点を中心した球のよう
  - 個々の企業・銀行の関連を考慮すると、直線状に配置できる (Figure 8.9(ii))
    - financial interest group (マルクス主義者)
    - 球に対するwedge(くさび)
      - 企業・銀行の距離は近さ
      - 中心に近いほど孤立した周辺企業
  - 原点は孤立した企業(連結なし)のため除去
    - 中心は空に
  - 影響力の球
    - 原点中心の球(地球)、銀行からみたエゴセントリックな球
    - 太陽中心の球(太陽系)、社会中心の球

37

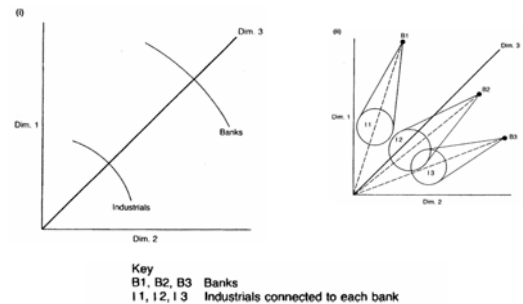


Figure 8.9 Multidimensional scaling and bank interlocks

38

- projection (射影)
  - 2次元平面に3次元データをマップしているため、分かりづらい
  - 地球の3次元構造を2次元平面にマップするための地図製作用の射影技術 (Figure 8.10)
  - parallel projection
    - 2点が2点にマップ
    - ゆがみが大きくあまり使わない
  - point projection (gnomonic projection @ cartography)
    - 2点が1点にマップ
    - 中心から投影するため、同じ角度方向にあれば同じ点にマップ
    - 銀行と、関連企業が明らかなクラスターとして投影される
      - 中心は銀行

39

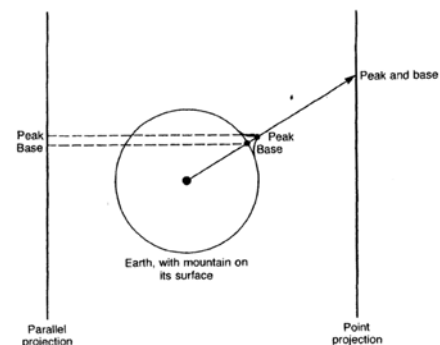


Figure 8.10 Alternative projections

40

- Levine
  - SNAにおいて重要な役割
  - ハーバードグループの第1世代を牽引
  - エゴセントリック、ソシオセントリックの両者を、多次元空間にネットワークを埋め込むことでうまくモデル化
  - 結果、グラフ理論の原理が、地域、企業の地位という視点に対して適用できるようになった
  - また、洗練された図として描けるようになった
    - 密度、中心性、社会的な円
- 現実的な応用はまだこれから
- 本書の目的は、方法論と、理論の統合を描くことで、先端の理論的アイデアを説明することだった

41

## 参考資料

- 安田雪先生の講義資料
  - [http://www.soi.wide.ad.jp/class/99001/slides/16/index\\_11.html](http://www.soi.wide.ad.jp/class/99001/slides/16/index_11.html)
- 安田雪先生の用語集
  - <http://www5.ocn.ne.jp/~yasuda/networks.htm>
- 統計基礎用語集(英和編)
  - <http://www.qmss.jp/qmss/glossary/stat-glossary-ej.htm>

42