

第4回社会ネットワーク輪読会

第6章 Components, Cores and Cliques

東京大学 医学部附属病院
臨床医学オントロジー研究ユニット

今井 健

-1. 自己紹介

■ 今井 “wanderer” 健

学部: 数学科 (Riemannian Geometry)

修士: 情報学環・伊東研 (Citation Network)

博士: 情報学環・小野木研
(自然言語処理を用いた診療文書からの
情報抽出)

現在: 東大病院・臨床医学オントロジー研究ユニット

0. 今日の内容

- (1) Introduction
この章の導入 - 'cliques' の定式化に向けて-
- (2) Components, Cycles and Knots
いくつかの“まとまり”の概念
- (3) The Contours of Components
まとまったら、それらの階層関係を書いてみたい。
- (4) Cliques and their Intersections
'cliques' の再定義とそれらの交わり、'circle' へ
- (5) Components and Citation Circles
社会学において、実際に適用された研究-引用解析

1. Introduction

- Social Network Analysis をする人の興味はネットワークを分割する「派閥」や密集した「sub-group」を発見することである。
cf) Hawthorne, Yankee City Studies
- 様々なsub-groupの理論的なモデルを考えることができる。
Cliques, Clusters, Components, Cores, Circles etc.
- グループ構造解析の first step は Sub-group の特定だが、どのように特定するか。
random, gender ...
- この章ではsub-groupの特定に使用する特徴について、その理論的な基盤を述べる。

2 . Components, Cycles and Knots

- 最も簡単な Sub-group の概念は Component (極大連結 sub-graph のこと)
- Network の構造記述の第一段階は component の数やサイズを調べることである。

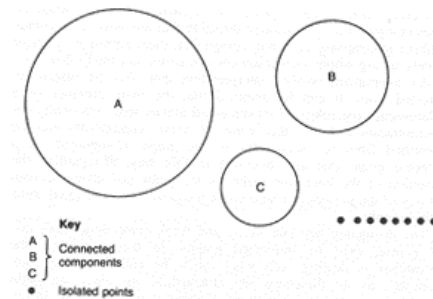


Figure 6.1 Components in a network

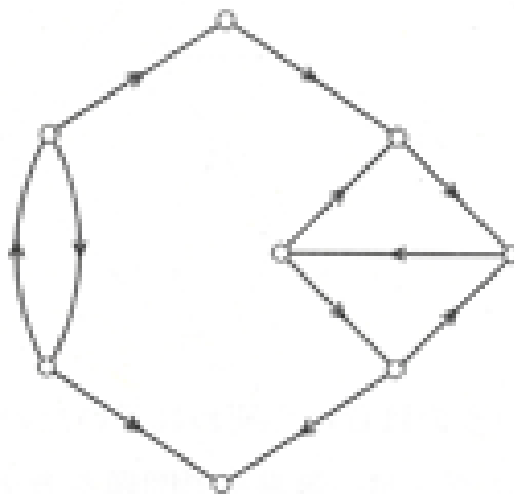
2 . Components, Cycles and Knots

- Component を特定するための単純なアルゴリズムは全ての可能な最短パスを検索すること
- 非常に効率が悪く、計算量が多いため、多くの解析パッケージでは、back-tracking を用いて全域木を構築することで component を発見している。
- Component は directed, undirected 双方で見つけることができるが、重要な違いがある。

{Strong, Weak, Simple} component

- Directed の場合
 - ・ strong component (強連結成分)
 - => パス中の向きが一定
 - ・ weak component (弱連結成分)
 - => 向きを無視した無向グラフが連結。
- Undirected の場合
 - ・ 当然これらの区分はなし。
 - ・ このような場合は Simple component という。

演習：強連結成分を探せ



Cyclic components

- コンポーネントの内部構造を調べるためには、Strong, Weak 以外の概念が必要。
- Everett の “block” を元にした提案 “Cyclic Components”
- “k-cycle”, “Maximum cycle length”, “bridge”, “k-bridge”

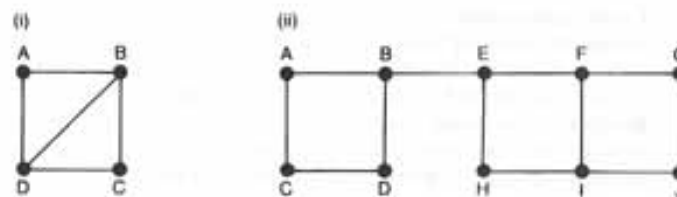


Figure 6.2 Cyclic components

Cyclic components 解析における 連結成分の分類 by Everett

- (1) Cyclic Components
 - (2) Hangers
 - (3) Bridgers
 - (4) Isolated Trees
 - (5) Isolates
- 発見された Cyclic Components に社会学的な解釈を与えられるか、というのは難しい問題。
(3 or 4 の cyclic length で考えるのが現実的 by Everett)

{Strong, Weak} cyclic components

- directed cycle と semi cycle
- Strong ... directed, acceptable (semi)
Weak ... unacceptable (semi)

- Strong の中にさらに
hangers-on
hangers-off
の区別を考えられる



Figure 6.3 Cycles and semi-cycles

Cutpoints & Knots

- コンポーネントの内部構造を解析するためのもう1つのアプローチ Hage, Harary(1983)
「複数のcomponentsを結ぶ、要となる点を考慮する手法」
- Cutpoint (“component” をくっつける要)
Knot (連結されている個々の “component” (のようなもの))

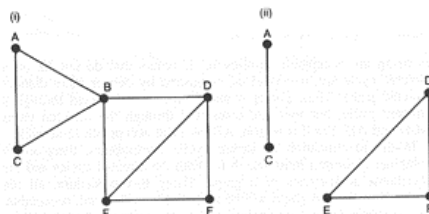


Figure 6.4 Knots and cut-points

3.The Contours of Coponents

- この節では、第3章でも少し述べた nesting という手法を用いて、component の内部構造を systematic に追っていく。
- A・・・最も tight に結合。Component の核。
B・・・より弱い結合レベル。
C・・・全ての連結成分を含む。

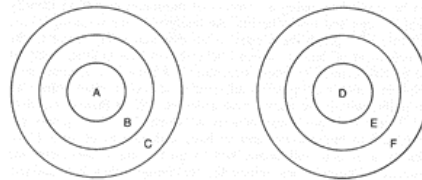


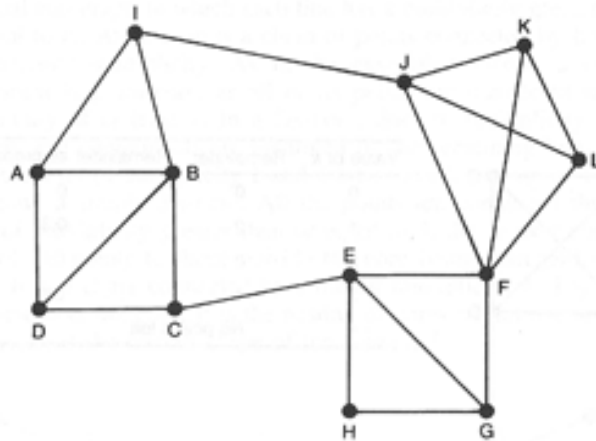
Figure 6.5 Nested components

Nesting を行う2つの方法

- 密集や結合の度合いを測る尺度として・・・
 - (1) Degree-based => 'k-core'
「点の次数を用いる」
 - (2) Multiplicity-based => 'm-core'
「辺の多重性を用いる」
- 'k-core'
各点が少なくともk以上の次数を持つ極大subgraph
- 'm-core'
各辺が多重度m以上を持つ極大subgraph

(1) 'k-core' と 'core collapse sequence'

■ 3k-core



(1) 'k-core' と 'core collapse sequence'

- Seidman はまた 'core collapse sequence' を見ることで、ネットワークの断片化状態を評価する方法を示した。

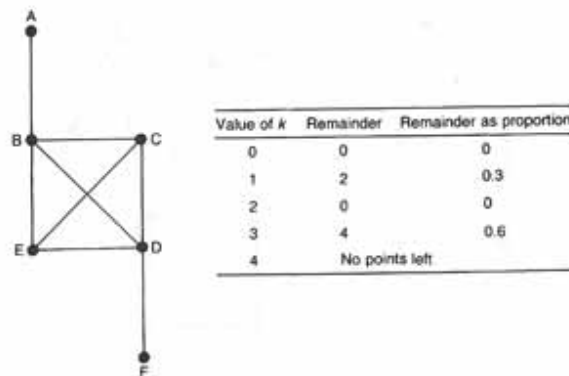


Figure 6.7 Collapse of a k -core

(2) 'm-core' と 'core collapse sequence'

■ 3m-core

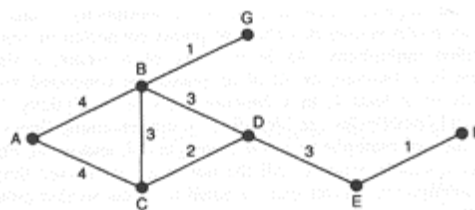
(B,C は多重度4の辺で結ばれており、4m-coreを形成している。)



Figure 6.8 A 3m-core

(2) 'm-core' と 'core collapse sequence'

■ Core collapse sequence の自然な拡張



Value of m	m -cores	Remainder	Remainders as proportion
0	{A, B, C, D, E, F, G}	0	0
1	{A, B, C, D, E, F, G}	2	0.28
2	{A, B, C, D, E}	0	0
3	{A, B, C, D, E}	2	0.28
4	{A, B, C}	3	0.43
5		No points left	

Figure 6.9 Collapse of an m -core

この節のまとめ

- Cyclic components についても同様に適切な閾値を考えたりして nested cyclic components を定義することもできる。
- 同様に重み付きであっても、これまでに見てきた simple components に対するアイデアを様々に拡張することで、density の測定と補完的に機能したり、network 全体の様子を解析するための様々な制限を打開していくことができる。
- Network 全体構造の比較と言ったら、
「全体の密度」、「inclusiveness」、「componentの数・サイズ」
「componentの密度」、「componentsの階層構造」、
「core collapse sequence」
などを含んでいる。

4. Cliques and their Intersections

- Hawthorne, Yankee City の話における 'Clique'
- Harary(1969), Luce & Perry (1949)
「Clique とは本質的には極大完全部分グラフである」
- Doreian(1979) 「任意の2点が 'reciprocal'なもの」
strong clique & weak clique

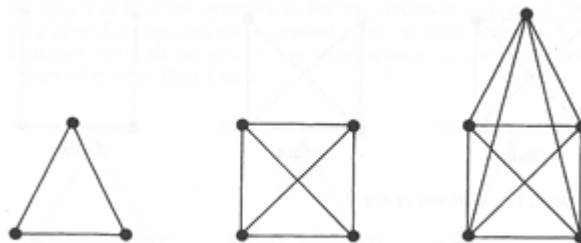


Figure 6.10 Cliques of varying sizes

Clique の定義の2つの拡張

- 極大完全部分グラフという考え方は社会的にはかなり制限がきつい。
=> clique の定義をゆるめよう。
- **(1) n-clique**
「任意の2点間の距離がn以下である部分グラフ」

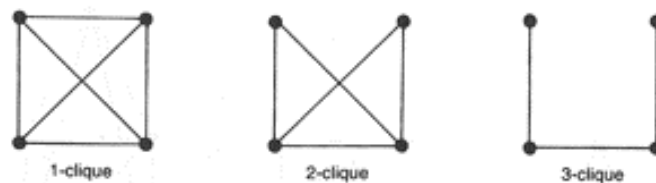


Figure 6.11 *n-cliques of size 4*

Clique の定義の2つの拡張

- n-clique の考え方には2つの限界がある。
n が2より大きくなると、社会的な意味づけが困難

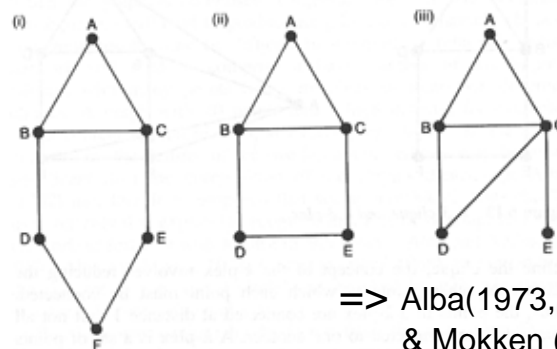


Figure 6.12 *Sub-graphs and 2-cliques*

=> Alba(1973, 1982)
& Mokken (1974) の修正
"n-clan"

Clique の定義の2つの拡張

- **(2) k-plex** Seidman & Foster (1978)
「個々のメンバーが少なくとも $n-k$ 個の他のメンバーと結合」

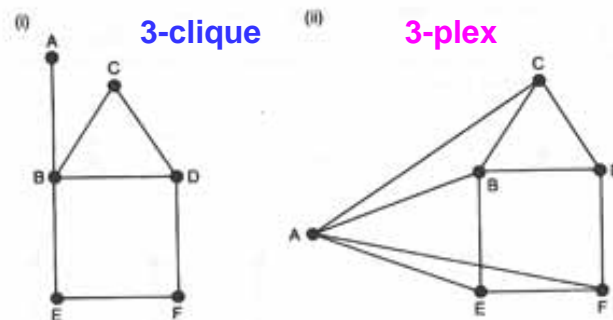


Figure 6.13 A 3-clique and a 3-plex.

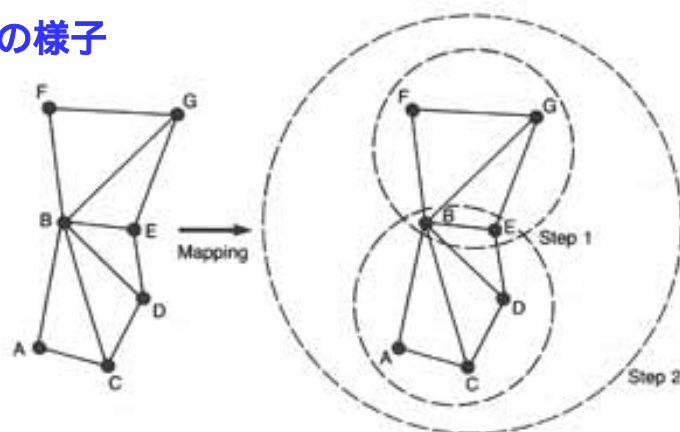
Clique の定義の2つの拡張

- K-plex を考える上で重要なのは、研究者が k の範囲を十分考慮すること。
- 特に得られる k -plex のサイズに最小値を定めないと、 k を大きくしたときに当たり前の結果になってしまう。
- 経験則として、 k -plex のサイズの最小値は $k+2$ とするのが良い。
- 多少の問題はあるものの、 k -plex の概念は、2より大きな n を用いた n -clique より、多くの知見を与えてくれるように思われる。
- k -plex も n -clique と同様に 重み付きグラフにおいては閾値を与えることで、nested k -plex を考えることができる。

Circle の考え方

- Alba & Kadushin (1976), Alba & Moore (1978), Kadushin (1966, 1968) => 'Social Circle' の提案
- Alba の基本アイデア
「clique 同士がある程度メンバーを共有していれば、それらをより大きな1つのまとまり 'circle' として見なせるのではないか」
- Step1) size3 の 1-clique (triad) を特定。2/3 を共有する他の clique と結合させる。
- Step2) 残った clique をより低いレベルでの重なりでくっつける。(彼は 1/3 の共有があれば結合させる、とした)

その様子



1-cliques: {A,B,C} {B,C,D} {B,D,E} {B,F,G} {B,G,E}
 1st step circles: {A,B,C,D,E} {B,F,G,E}
 2nd step circles: {A,B,C,D,E,F,G}

Figure 6.14 *Intersecting social circles*

むむ？

5. Components and Citation Circles

- Sociology of Science は social network の考え方を思い起こさせる多くの研究が成された重要な分野の1つ。
- Crane(1972) 'invisible college'
- Mullins(1973) 失敗
- Gattrell(1984)
 - ・1960 ~ 1978 の地理学の論文を集め、引用関係を調査、citation network の component structure を調べるのが目的。
 - ・例の閾値の方法で、様々なレベルでの component の nesting をしらべた。
 - ・一定の成果を得た。