

診療データネットワークのコミュニティ検出 による患者分類法の検討

星 雅丈*¹ 日月 裕*²

われわれは、診療データから構成したネットワークのコミュニティ検出による患者分類手法の開発を試みている。今回、病院情報システムのデータベースにおける患者IDと他のフィールドのデータを頂点としてネットワークを構成した。当該ネットワークはいくつかのコミュニティに分割され、結果としては、患者集合がコミュニティ単位でいくつかの群に分類された。同じデータを用いて、クラスタ分析、およびMDCによる患者分類を行った。われわれは2つの分類の類似性を測る新たな指標としてSIGを作成し、すべての分類間の類似性を計算した。結果として、コミュニティ検出による患者分類は、クラスタ分析による患者分類との類似性が低く、MDCによる分類との類似性がクラスタ分析による患者分類と比較して高かった。このことから、本手法による患者分類は、クラスタ分析と比較して臨床的に意味のある分類である可能性が高い。本手法は、ネットワークの要素となるフィールドの選択を変えることにより、違った患者分類が可能である。応用の範囲も広い。例えば、看護のデータを用いることによって看護の視点での新たな患者分類が可能である。

■キーワード: 患者分類, 診療データ, 複雑ネットワーク, コミュニティ検出, 病院情報システム, オーダエントリシステム

Evaluation of Patient Grouping by Community Detection of the Networks Derived from Clinical Data: Hoshi M*¹, Tachimori Y*²

Our study aims to develop patient grouping by community detection of networks derived from clinical data. We constructed the networks with nodes consisting of patient IDs and the other field's data of the database in the hospital information system. We found that the networks were divided into several communities and as a result, a set of patients was divided into several groups according to the communities (PGN: Patient Grouping by Network community detection). Using the same data, we also tried to group the set of patients by cluster analysis and MDC (Major Diagnostic Categories). And we get the two other patient groupings (PGC: Patient Grouping by Cluster analysis, PGM: Patient Grouping from MDC). We constructed a

*¹成美大学 経営情報学部 医療福祉マネジメント学科
〒620-0886 福知山市西小谷ヶ丘 3370

*²日本福祉大学 社会福祉学部
E-mail: hoshi-m@uv.seibi-gakuen.ac.jp
受付日: 2015年8月27日
採択日: 2016年1月6日

*¹Department of Medical and Welfare Management,
School of Management and Business Administration,
Seibi University
3370 Nishikotanigaoka, Fukuchiyama-shi, Kyoto,
620-0886, Japan

*²Faculty of Social Welfare, Nihon Fukushi University

new index to measure the similarity degree between two patient groupings (SIG: Similarity Index between two Groupings), and we calculated the SIGs for all combinations of groupings. As a result, the similarity degree between PGN and PGC was comparatively low. The similarity degree between PGN and PGM was higher than that between PGC and PGM. This result indicates that the patient grouping by network community detection is clinically more meaningful than that by cluster analysis. We can produce another grouping by a choice of different field as the nodes of the networks. Therefore, our grouping method has wide applications. For example, using nursing data can give us a new patient grouping for nursing care.

Key words: Patient grouping, Clinical data, Complex network, Community detection, Hospital Information Systems, Medical order entry systems

1. はじめに

われわれは、先行研究において、病院情報システムにおける診療行為データを用いてネットワークを構成し、そのネットワークに対してコミュニティの検出を行った。そして、コミュニティ検出が臨床的に利用可能であるかどうかを検討した。結果として、ネットワーク要素を「投薬オーダ」と「患者」に限定して構成したネットワークにおいては、時間的に安定したコミュニティが存在し、従来法とは異なるデータの分類結果を得ることができた¹⁾。

今回、われわれは先行研究において用いたネットワーク構成手法とコミュニティ検出手法について、先行研究とは異なる病院から得たデータを用いて再検討を行った。

さらに、コミュニティ検出、クラスタ分析、DPCデータによる患者分類を行い、それぞれの分類法の類似性について検討した。

2. 目的

急性期医療においてよく用いられているDRGやDPCは主に病名を利用した患者分類の手法であるが、療養型病床などの慢性期医療においては、うまくはいかない場合がある。例えば、高齢者など複数疾患を持つ患者については、病名による分類のみでは特性を捉えきれない可能性がある。

また、わが国の療養型病床で現在行われている

医療区分・ADL区分による患者分類だけでは、捉えきれない患者の特徴もあると考えている。例えば、併存する病名、残存する身体機能、継続的に必要な投薬内容、あるいは患者の家族関係などは高齢患者の病院療養にとって重要な特徴(要素)であるが、現在の医療区分・ADL区分を決定するための評価指標には存在しない。医療区分やADL区分は、基準があらかじめ定められ、それに基づいて行われる患者の分類手法である。急性期医療におけるDRGやDPCも同様である^{2,3)}。

今回、われわれが試みる患者の分類手法は、ある基準により患者を分類するのではなく、患者個々に付随する診療行為などの属性に基づく関係性から患者を分類するものである。この分類手法では、診療行為だけでなく、他の患者との関係性でも分類が決まることがある。ある新たな診療行為や患者が発生することにより、その分類が動的に変化する可能性もある。

つまり、あらかじめ基準を定めずに、患者を分類できるという手法である。例えば、医師による治療(キュア)の視点のみならず、看護師による看護(ケア)の視点で患者を見た結果である「看護のデータ」を加えることにより、分類結果を看護で活用できる可能性がある。療養型病床における高齢の患者については、複数の疾患に対する治療が同時に行われていることが多い。また、診療行為のみではなく、リハビリテーションや看護師によるケアも患者の特徴を捉えるファクターとし

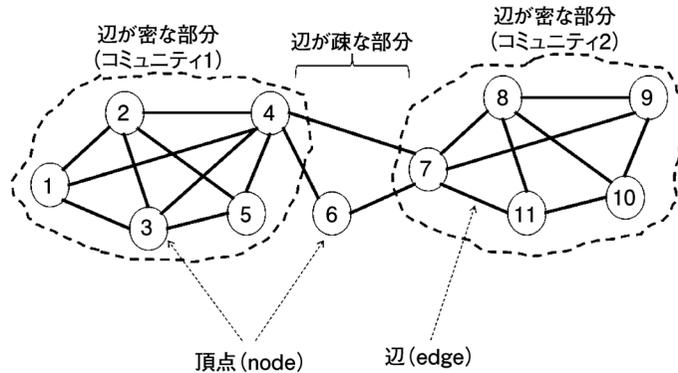


図1 ネットワーク図とコミュニティの検出

ネットワークは、頂点と呼ばれる空でない要素の集合と、辺の集合から構成される。各辺には重み (weight) をつけることもできる。また、図のようにネットワークを辺が密な部分を集めて分割したものをコミュニティと呼ぶ。

て必要である。これらの要因を自由に取り込んで分類することが可能となる手法である。

急性期医療のみならず、慢性期医療に有用な患者分類手法を開発することは重要である。患者を分類できれば、分類ごとの管理が可能になる。診療やケアの過程、成果、消費した医療資源など、同じ分類に含まれる患者を比較することにより、診療の中で何が起きているかを具体的に把握することが可能となる。

そこで、われわれの先行研究に引き続き、急性期・慢性期に関わらず適用可能な患者の新たな分類手法を探ることを目的とした。

3. 方法

1) ネットワークとコミュニティ

ネットワークは図1に示すように、頂点 (node) と辺 (edge) の集合から構成される。辺が結ぶ頂点同士は何らかの関係があることを示す。辺には頂点間の距離や親密度など関係の強さを表す重み (weight) を付けることもできる。ネットワークを用いた分析では、各頂点がつながる辺の数(次数)が多い頂点や、ネットワークの中心性など特徴の導出、あるいはネットワークが部分に分かれているのかなどにより構造を分析する⁴⁾。

ネットワークがいくつかの部分に分かれており、図1のように各部分内の辺が密で、部分同士を結ぶ辺が疎の状態にある場合、辺が密な部分をそれぞれコミュニティという。ネットワークのコミュニティへの分割の良さを示す指標としてモジュラリティ (modularity) がある。モジュラリティは以下のように定義されている。

ネットワークがいくつかの部分に分かれており、すべての辺の重みの総和を m とする。ここで、 e_{ij} を次のように定義する。 $i=j$ の場合は、コミュニティ i (すなわち j でもある) 内の頂点同士を結ぶ辺の重みの和の m に対する割合と定義する。 $i \neq j$ の場合は、辺の重み総和 m に対するコミュニティ i からコミュニティ j に張られている辺の重みの和の割合の $1/2$ と定義する。

ここでモジュラリティ Q は次のように定義される^{5,6)}。

$$Q = \sum_i \left\{ e_{ii} - \left(\sum_i e_{ij} \right)^2 \right\}$$

Q が高いということは、コミュニティ内でのつながりが強く、コミュニティ間のつながりは弱いということの意味する。Newman らは Q が 0.3 以上あれば、得られたコミュニティには意味があ

ると述べている⁶⁾。

この Q が最大となる分割を求めるのが、コミュニティ分割アルゴリズムである。ネットワークにおいて考えられるすべての分割について Q を計算すれば、 Q が最大となる分割を求めることは可能であるが、頂点数や辺数が大きいネットワークにおいては計算時間が大きくなり過ぎるため実行が不可能である⁶⁾。

コミュニティ分割アルゴリズムはいくつか開発されている。本研究では、Aaron Clauset らのアルゴリズムを採用した⁷⁾。このアルゴリズムは、 Q の最大を計算するのではなく、 Q の変化を追跡して極大を計算するものである。最大ではなく極大であるため、さらに良い分割が存在する可能性がある。

このアルゴリズムのプログラムは開発者によって公開されており、われわれはそれを利用してコミュニティの検出を行った⁸⁾。

2) 分類と分類の比較

(1) 分類について

「分類」について以下の通り定義する。

ある要素の集合 $S = \{s_1, s_2, s_3, \dots, s_n\}$ があるとき、この集合 S における分類 A とは、次のような部分集合の組を表す。

$A = \{A_1, A_2, \dots, A_p\}$ ただし $A_j \subseteq S$, $\bigcup_1^p A_j = S$, $A_j \cap A_k = \emptyset$ for $j \neq k$.

A_j ($j = 1, 2, 3, \dots, p$) は分類の上で得られた群(以下、分類群とする)である。

(2) 分類群の類似度

2つの集合の類似度の指標として Jaccard 係数がある。Jaccard 係数(以下、Jaccard Similarity Coefficient: JSC とする)は行政の自治体間分析や生態学研究などで幅広く利用されている⁹⁾。

集合 C と集合 D にそれぞれ含まれる要素の数をそれぞれ a , b とし、集合 C , D に共通に含まれる要素の数を c とすると、 JSC は以下の計算式で示される。

$$JSC = \frac{c}{a+b-c} \quad (0 \leq JSC \leq 1)$$

これは、2つの集合の類似の程度を表している。2つの集合が全く同じであれば、明らかに $JSC = 1$ であり、2つの集合の半分の要素が共通であれば、 $JSC = 1/3$ となる。

2つの分類がある場合、これらの分類が似ている場合には、1つの分類に含まれる各分類群に似た分類群が、もう1つの分類のなかに含まれる可能性が高いであろう。

そこで、われわれは2つの分類の類似の程度を調べるために、それぞれの分類の分類群同士を JSC によって比較した。分類が似ているならば、 JSC の大きい分類群の組が多数存在するであろうと考えたからである。

(3) 分類の類似度

前節の方法は、あくまでも各分類の分類群が集合として似ているかどうかの比較であり、直接、分類が似ているかどうかを比較したものではない。そこで、分類が似ているかどうかを直接比較するために、新しい分類同士の類似度 (SIG : Similarity Index between two Grouping) の指標を作成した。

分類が似ているかどうかの指標を作るためには、分類が似ているとはどういう意味かを改めて考える必要がある。そこで、今回われわれは、「分類が似ている」ことについて、次の通りに考えた。

集合 $S = \{s_1, s_2, \dots, s_n\}$ 上の2つの分類 $A = \{A_1, A_2, \dots, A_p\}$, $B = \{B_1, B_2, \dots, B_q\}$ があつたとき、これらが似ているならば、以下のようなことが起きると考えられる。

「ある2つの要素が分類 A において同じ分類群に含まれるならば、分類 B においても同じ分類群に含まれやすい。逆に2つの要素が分類 B において同じ分類群に含まれるならば、分類 A においても同じ分類群に含まれやすい。」

以上の考察から、 SIG としては、「2つの要素が少なくとも一方の分類において同じ分類群に含まれるという条件の下での、この2つの要素が両方の分類において同じ分類群に含まれる確率」を考えればよいと思われる。

具体的には次のように計算できる。集合 S か

表1 2つの要素が属する分類群の場合の数

分類A \ 分類B	2つの要素が同じAの分類群に属する場合	2つの要素が異なるAの分類群に属する場合	合計
2つの要素が同じBの分類群に属する場合	a	c	a+c
2つの要素が異なるBの分類群に属する場合	b	d	b+d
合計	a+b	c+d	a+b+c+d=M

集合Sから任意の2つの要素の組を作り、それぞれが分類Aと分類Bにおいてどのように分類群に属しているかを調べた際の場合の数を表している。

ら任意の2つの要素の組を作る場合、その場合の数Mは、 $M = \binom{n}{2}$ である。これらの組は次に示す4つの場合に分けられる(表1)。

① 2つの要素が、分類A, Bのいずれにおいても同じ分類群に属する。

② 2つの要素が、分類Aにおいては同じ分類群に属するが、分類Bにおいては異なる分類群に属する。

③ 2つの要素が、分類Aにおいて異なる分類群に属するが、分類群Bにおいては同じ分類群に属する。

④ 2つの要素が、分類A, Bのいずれにおいても異なる分類群に属する。

これら4つの場合、場合の数を順番にa, b, c, dとしたとき、SIGを次のように定義する^{*1}。

$$SIG = \frac{a}{a+b+c}$$

2つの分類が全く同じ場合には、 $b=c=0$, $a+d=M$ となるため、 $SIG=1$ となる。

また、分類Aでは集合S全体を1つの分類、分類Bでは各要素1つで1つの分類を作るとする。この場合、任意の2つは必ず、分類Aでは同じ分類に入り、分類Bでは違う分類に入る。よって、 $a=0$, $c=0$, $b=M$ となり、 $SIG=0$ となる。ただ、この方法では、2つの分類のすべての分類群の要素が1個だけからなる場合には、 $a=b=c=0$ となり、SIGが計算できない。そこで、SIGを用いるために、「少なくとも1つの分類群は2つ以上の要素を持っていないなければならない」という条件を加えた。

われわれは2つの分類の類似の程度を調べるために、前節のJSCに加えてそれぞれの分類同士をSIGによって直接比較した。SIGを使うことによって、分類の類似性(あるいは分類法の類似性)を数値によって直接比較できるようになったと考えている。

3) データ、およびその処理方法

(1) 分析対象データ

今回は急性期型のA病院(837床)より提供を受けた以下のデータを利用した。

① オーダリングシステムにおけるオーダ(投薬(処方・注射)・検体検査・細菌検査オーダ)のデータ

② DPC請求ファイル(Dファイル)

いずれも、2013年3月1日から2013年3月31日におけるデータである。

①は総レコード数33,565件、入院患者延べ711名に対して上記期間中に出された処方・注射・検体検査・細菌検査オーダが記録されたものである。②はDPC参入病院が3カ月に1度厚生労働省に提出するデータのうち、Dファイル(DPC請求データ)と呼ばれるものであり、上記期間における総レコード数は137,229件、入院患者649名分のDPC請求データが記録されていた。

*1当初、われわれはSIGの定義として、 $SIG = \frac{a+b}{a+b+c+d}$ を考えていた。これは、似た分類であるならば、本文にある性質以外に、「片方の分類で異なる分類群に含まれるならば、他の分類でも異なる分類に含まれる」という性質を持っていると考えたからである。しかし、一般に集合Sが大きくなると、dのみが他に比べて圧倒的に大きくなり、SIGが1に近くなりすぎ、分類による差が出にくいことが分かった。そのため、数式からdを除いた。

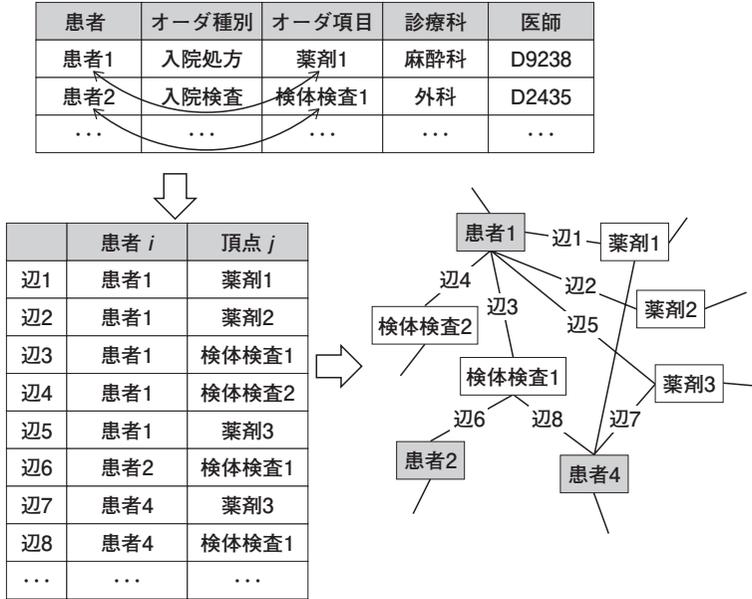


図2 分析に用いたデータの構造とネットワークの構成

オーダリングシステムに関連するテーブルを本表のような形式に再構成した。ネットワーク構成の際に接続するのは、同一レコード内にある患者とオーダ項目の2要素である。患者・薬剤などのフィールド要素がネットワークの頂点 (node) となる。

矢印下の図表は、テーブルからネットワークの頂点の組み合わせを取得し、それらを元に構成したネットワークを示す。

データの出所が異なる①と②においては、それぞれに含まれる患者数が異なっていた。そこで①と②に共通に存在する患者 600 名を抽出した。以下、分析に用いたオーダのデータや D ファイルのデータは、この患者 600 名に関するものである。

なお、今回提供を受けたデータは、当該病院において個人情報の削除、および患者 ID の付け替え処理が行われており、提供データからは個人を特定できないようになっていた。

(2) ネットワーク構成とコミュニティ検出による患者分類

今回、オーダリングデータより、投薬 (処方・注射)・検体検査・細菌検査のオーダが記録されたテーブルをネットワークの元データとして抽出した (図2)。

なお、われわれの先行研究において、テーブル

におけるフィールドをすべて用いたネットワークよりも、患者 (暗号化された患者 ID) とオーダ項目 (薬剤など) のみから構成したネットワークの方が意味のあるコミュニティ分割が得られることが分かっている¹⁾。そこで今回は、テーブルを図2の左下に示すような形式に再構成した。そして、同一レコード内にある、「患者」と「オーダ項目」のみからネットワークを構成した。

図2右下に示すように、患者とオーダ項目を頂点とし、1つのレコード内にある患者とオーダ項目を辺で結んだ。なお、同一の患者に対して複数回、同一のオーダ項目が存在する場合、その回数を辺の重みとした。

今回われわれは、以下の7通りのネットワークを構成した。

- ① 患者—投薬
- ② 患者—検体検査

表2 構成したネットワークからのコミュニティの検出結果

頂点1	頂点2 (オーダ)	ネットワーク名	頂点数	辺数	患者頂点数	モジュラリティ	コミュニティ数
患者	投薬	PN	1,787	7,063	600	0.4750	16
患者	投薬・検体・殺菌	AN	2,033	13,196	600	0.3402	12

- ③ 患者—細菌検査
- ④ 患者—投薬・検体検査
- ⑤ 患者—投薬・細菌検査
- ⑥ 患者—検体検査・細菌検査
- ⑦ 患者—投薬・検体検査・細菌検査

次に, Aaron Clauset らのアルゴリズムを用いてコミュニティの検出を行った⁷⁾. そして, 上記7通りのネットワークのうち, モジュラリティ Q 値が高かった①と⑦について分析を進めた. 以下, 前者のネットワークを PN (Prescription and injection order Network), 後者を AN (All order Network) とする.

PN, AN から得られたコミュニティには, それぞれ患者を表す頂点と, オーダ項目を表す頂点が含まれている. そこで, 各コミュニティに属する患者 (の頂点) の集まりを, 1つの患者群とすることにより患者を分類した. 以下, PN から検出されたコミュニティによる患者分類の名称を PGNP (Patient Grouping from Prescription and injection order Network), AN から検出されたコミュニティによる患者分類の名称を PGAN (Patient Grouping from All order Network) とする.

(3) クラスタ分析および DPC データの処理による患者分類

(2)において構成したネットワーク PN・AN は, 患者相互, オーダ項目相互には接続しない. このような形式のデータ (2部グラフ) は, オーダ項目を患者の属性と考えれば, クラスタ分析が可能である.

そこで, PN・AN を構成したデータをそのまま用いて, K-means 法および Ward 法によるクラスタ分析を行った^{10,11)}. クラスタ数は PN・AN において検出されたコミュニティ数をそれぞれ用

いた. 以下, K-means 法による患者分類をそれぞれ PGPk (Patient Grouping of Prescription and injection order by k-means method) と PGAk (Patient Grouping of All order by k-means method), および Ward 法による患者分類を PGPw (Patient Grouping of Prescription and injection order by Ward's method) と PGAw (Patient Grouping of All order by Ward's method) と呼ぶ.

また, 3) (1)-②に示した DPC データにおける D ファイルを患者ごとに集計し, さらに MDC 別に分けた. この患者分類を PGMDC (Patient Grouping from Major Diagnostic Category) と呼ぶ.

4. 結果

1) コミュニティの検出および患者分類の結果

PN および AN からコミュニティ検出を行った結果を表 2 に示す. 共にモジュラリティは 0.3 以上であり, 十分意味のある検出が得られた.

また, コミュニティ検出による患者分類 (PGNP・PGAN) により得られた分類群それぞれに含まれる患者数, クラスタ分析による患者分類 (PGPk・PGAk・PGPw・PGAw) により得られた分類群それぞれに含まれる患者数, および MDC による患者分類 (PGMDC) により得られた分類群それぞれに含まれる患者数を表 3 に示す.

ここで, 分類群の番号は各分類別につけており, 分類間の共通性はない.

表 3 に示す通り, コミュニティ検出による分類では, 患者 1 人からなる分類群が存在するが, 他の分類ではそのような分類群は存在しなかった.

表3 コミュニティ検出, クラスタリング, および MDC による分類群別の患者数

分類群番号	コミュニティ検出		クラスタ分析 (K-means 法)		クラスタ分析 (Ward 法)		PGMDC
	PGPN	PGAN	PGPk	PGAk	PGPw	PGAw	
01	15	24	25	51	42	42	18
02	79	96	154	30	29	29	22
03	19	144	60	79	22	56	39
04	146	69	65	44	30	51	68
05	86	123	29	6	34	55	68
06	54	89	23	43	21	41	119
07	58	37	17	24	23	60	27
08	42	12	27	62	32	37	14
09	73	1	6	114	41	65	3
10	21	2	28	72	27	117	16
11	2	2	24	37	33	41	55
12	1	1	31	38	37	6	53
13	1		41		65		6
14	1		25		117		17
15	1		26		41		30
16	1		19		6		40
17							
18							5
患者数合計	600	600	600	600	600	600	600

クラスタ分析については、それぞれ検出されたコミュニティ数をクラスタ数として定めて実施した。

2) 分類群の比較

まず、今回得られた患者分類の類似性について、各分類内の分類群同士を、JSC を用いて比較した。結果の一部を表 4~7 に示す。

各表において太枠で囲んだ箇所は、JSC が 0.3 (各群の約半分の要素が一致) を超えたものである。

表 4 はコミュニティ検出という同様の手法から得られた分類群 (PGPN, PGAN) の比較である。半分以上の要素が一致する患者群の組み合わせが 8 つ、さらに JSC が 0.2 (各群の 3 分の 1 の要素が一致) 以上 0.3 未満の組み合わせも 4 つ見られ、似た分類群が比較的多いことが分かった。

表 5 はコミュニティ検出により得られた分類群

とクラスタ分析 (K-means 法) により得られた分類群の比較であるが、最大の JSC が 0.212 であり、0.3 を超える組み合わせが見られず、分類群同士の比較としては似ていないことが分かった。

表 6, 表 7 からは、コミュニティ検出による分類群とクラスタ分析による分類群のどちらが、MDC による分類群と似ているかを見ることがができる。MDC による分類群とコミュニティ検出による分類群との比較では、JSC が 0.3 以上の組み合わせが 4 つ見られたのに対し、クラスタ分析による分類群との比較では、0.3 以上の組み合わせは見られなかった (最大の JSC は 0.143)。これらのことから、コミュニティ検出による患者群は

表 4 患者群の比較結果の例 (PGAN と PGPN)

	PGPN01	PGPN02	PGPN03	PGPN04	PGPN05	PGPN06	PGPN07	PGPN08	PGPN09	PGPN10	PGPN11	PGPN12	PGPN13	PGPN14	PGPN15	PGPN16
PGAN01	0	0	0.344	0	0.028	0.068	0.012	0	0	0.023	0	0.042	0	0	0.042	0.042
PGAN02	0	0.207	0	0.017	0	0.136	0.387	0.007	0	0	0	0	0	0	0	0
PGAN03	0.006	0.037	0.006	0.526	0.009	0.158	0.010	0	0.005	0.012	0	0	0	0	0	0
PGAN04	0	0.007	0	0	0.582	0.008	0.067	0.018	0	0	0	0	0	0	0	0
PGAN05	0.007	0.202	0	0.116	0.094	0.006	0	0.299	0.010	0.007	0	0	0	0	0	0
PGAN06	0	0.018	0.009	0.054	0	0	0.014	0.008	0.742	0	0	0	0	0.011	0	0
PGAN07	0.333	0	0	0.011	0.008	0.011	0.022	0	0.009	0.415	0	0	0	0	0	0
PGAN08	0	0	0.240	0	0.054	0	0	0	0	0	0	0	0.083	0	0	0
PGAN09	0	0.013	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PGAN10	0	0.025	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PGAN11	0	0	0	0	0	0	0	0	0	0	1.000	0	0	0	0	0
PGAN12	0	0	0	0	0	0.019	0	0	0	0	0	0	0	0	0	0

全オランダのネットワークから抽出した患者群 (PCAN01~12) と投薬オランダのネットワークから抽出した患者群 (PCPN01~16) の比較結果。表中の数値は JSC (Jaccard Similarity Coefficient) である。表中の太枠で囲んだ箇所は、JSC が 0.3 を超えている。

表 5 患者群の比較結果の例 (PGAK と PGAK)

	PGAK01	PGAK02	PGAK03	PGAK04	PGAK05	PGAK06	PGAK07	PGAK08	PGAK09	PGAK10	PGAK11	PGAK12
PGAN01	0	0	0.040	0	0	0	0.021	0.024	0.113	0.021	0	0.016
PGAN02	0.043	0.041	0.094	0.037	0	0.045	0.026	0.082	0.071	0.098	0.064	0.055
PGAN03	0.096	0.018	0.093	0.062	0.020	0.056	0.012	0.067	0.084	0.096	0.077	0.083
PGAN04	0.034	0.125	0.021	0.046	0	0.018	0.057	0.008	0.212	0.014	0.010	0.029
PGAN05	0.074	0.041	0.116	0.071	0	0.051	0.043	0.063	0.087	0.071	0.053	0.052
PGAN06	0.045	0.026	0.050	0.081	0.033	0.091	0.037	0.119	0.052	0.066	0.041	0.024
PGAN07	0.073	0	0.074	0.013	0	0.039	0.017	0.053	0.013	0.079	0.028	0.014
PGAN08	0	0.050	0.011	0.018	0	0.038	0.059	0	0.016	0.012	0	0.020
PGAN09	0	0	0	0	0	0	0	0	0	0.014	0	0
PGAN10	0	0	0	0	0	0	0	0.016	0.009	0	0	0
PGAN11	0	0	0	0	0	0.023	0	0	0	0.014	0	0
PGAN12	0	0	0	0	0	0	0	0.016	0	0	0	0

全オランダのネットワークから抽出した患者群 (PCAN01~12) と K-means 法による患者群 (PCAk01~12) の比較結果 (表中の数値は JSC)。JSC が 0.3 を超える部分が 1カ所も見当たらない。

表 6 患者群の比較結果の例 (PGAN と PGMDC)

	PGMDC01	PGMDC02	PGMDC03	PGMDC04	PGMDC05	PGMDC06	PGMDC07	PGMDC08	PGMDC09	PGMDC10	PGMDC11	PGMDC12	PGMDC13	PGMDC14	PGMDC15	PGMDC16	PGMDC17	PGMDC18
PGAN01	0	0	0	0.011	0	0.007	0.020	0	0	0	0.068	0	0	0.640	0	0	/	0
PGAN02	0.027	0.009	0.015	0.079	0.439	0.044	0	0.019	0	0.028	0.027	0	0.041	0	0.024	0.015	/	0.010
PGAN03	0.025	0	0.011	0.014	0.029	0.124	0.036	0.013	0	0	0.178	0.247	0	0	0.012	0.115	/	0.014
PGAN04	0.074	0	0.038	0.171	0	0.080	0	0.012	0	0	0	0	0.014	0.012	0.286	0	/	0
PGAN05	0.014	0	0	0.098	0.027	0.322	0.034	0	0.024	0.015	0.047	0.080	0	0	0.013	0.038	/	0.008
PGAN06	0.009	0.018	0.219	0.083	0.033	0.020	0.137	0.096	0	0.061	0.007	0	0.011	0	0	0.093	/	0
PGAN07	0	0.475	0	0.010	0.019	0.013	0.016	0	0	0.104	0.057	0	0	0	0.015	0	/	0.024
PGAN08	0.034	0	0.159	0	0	0.008	0	0	0	0	0.031	0.016	0	0	0	0	/	0
PGAN09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.025	/	0
PGAN10	0.053	0	0.025	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0
PGAN11	0	0	0	0.029	0	0	0	0	0	0	0	0	0	0	0	0	/	0
PGAN12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.025	/	0

全オーダーのネットワークから抽出した患者群 (PGAN01~12) と MDC による患者群 (PGMDC01~18) の比較結果 (表中の数値は JSC)。

表 7 患者群の比較結果の例 (PGAK と PGMDC)

	PGMDC01	PGMDC02	PGMDC03	PGMDC04	PGMDC05	PGMDC06	PGMDC07	PGMDC08	PGMDC09	PGMDC10	PGMDC11	PGMDC12	PGMDC13	PGMDC14	PGMDC15	PGMDC16	PGMDC17	PGMDC18
PGAR01	0.015	0.074	0.023	0.017	0.044	0.090	0.040	0.016	0.019	0.031	0.019	0.072	0	0	0.025	0.046	/	0
PGAR02	0.021	0	0.015	0.054	0	0.042	0.018	0.048	0	0	0.037	0.051	0.029	0.022	0.071	0.014	/	0
PGAR03	0.043	0.063	0.026	0.035	0.105	0.088	0.019	0	0.012	0.022	0.098	0.056	0.012	0	0.009	0.044	/	0
PGAR04	0.016	0.015	0.038	0.037	0.047	0.065	0.060	0.018	0.022	0.034	0.021	0.032	0	0	0	0.077	/	0.021
PGAR05	0	0	0.023	0	0	0.008	0.065	0	0	0	0.034	0	0	0	0	0	/	0
PGAR06	0.034	0.016	0.051	0.028	0.047	0.038	0.077	0.036	0	0.035	0.065	0.021	0	0	0	0.064	/	0
PGAR07	0	0	0.033	0.082	0.011	0.036	0.041	0	0.026	0.026	0.026	0.013	0	0	0.019	0.032	/	0
PGAR08	0.053	0.024	0.063	0.092	0.083	0.065	0.047	0	0.013	0.035	0.035	0.036	0.015	0.013	0	0.030	/	0
PGAR09	0.008	0.023	0.048	0.096	0.046	0.069	0.014	0.049	0	0.016	0.037	0.071	0	0.120	0.143	0.020	/	0.017
PGAR10	0.034	0.033	0.067	0.029	0.120	0.098	0	0.012	0	0.035	0.050	0.059	0.013	0	0.010	0.028	/	0.013
PGAR11	0	0.017	0.013	0.050	0.040	0.068	0.016	0	0.019	0.057	0.011	0.024	0	0	0	0.085	/	0.024
PGAR12	0.018	0	0.027	0.060	0.010	0.054	0.016	0.020	0	0	0.057	0.071	0.023	0.019	0.046	0.026	/	0

全オーダーのクラスタ分析 (k-means 法) により抽出した患者群 (PGAR01~12) と MDC による患者群 (PGMDC01~18) の比較結果 (表中の数値は JSC)。

表8 ネットワークコミュニティ検出, クラスタ分析, および MDC による患者分類の SIG

分類法		ネットワーク コミュニティ検出		クラスタ分析				MDC
				K-means 法		Ward 法		
		PGPN	PGAN	PGPk	PGAk	PGPw	PGAw	PGMDC
ネットワーク コミュニティ検出	PGPN		0.316	0.071	0.067	0.060	0.067	0.125
	PGAN			0.074	0.075	0.063	0.072	0.161
クラスタ 分析	K-means 法	PGPk			0.466	0.594	0.599	0.057
		PGAk				0.567	0.625	0.058
	Ward 法	PGPw					0.838	0.051
		PGAw						0.056
MDC		PGMDC						

表中の数値は SIG (Similarity Index between two Grouping) である。

コミュニティ検出による各分類・クラスタ分析による各分類と、MDC による分類の比較においては (表中の太枠内)、コミュニティ検出による各分類の方が、クラスタ分析による各分類と比較して、約 2~3 倍、MDC による分類との SIG が高かった。

クラスタ分析に比べて、より MDC による分類群に似ていると考えられる。

3) 分類類似度による比較

次に、今回われわれが行った 7 つの分類について、すべての組み合わせ (21 通り) における SIG を計算した。その結果を表 8 に示す。

コミュニティ検出による分類同士では $SIG = 0.316$ 、クラスタ分析による分類同士の SIG は最小 0.466、最大 0.838 と、同様の方法による分類同士では高い値を示した (表中のグレーの部分)。

コミュニティ検出による分類とクラスタ分析による分類においては (表中の点線で囲んだ部分)、最大でも $SIG = 0.075$ とかなり低い値を示した。

コミュニティ検出による各分類・クラスタ分析による各分類と、MDC による分類の比較においては (表中の太枠内)、コミュニティ検出による各分類の方が、クラスタ分析による各分類と比較して、約 2~3 倍、MDC による分類との SIG が高かった。

前節の JSC による比較と、本節における SIG による比較は同様の結果となった。これらの結果をまとめると次のようになる。

- ・同様の方法による分類では類似した分類が得

られる。

- ・コミュニティ検出による患者分類とクラスタ分析による患者分類はあまり似てはいない。
- ・クラスタ分析に比べて、コミュニティ検出による方法では、より MDC による分類に近い分類を得ることができる。

5. 考察

今回、われわれは病院情報システムに蓄積されている診療行為オーダのデータを用いて、ネットワークを構成しコミュニティの検出を行った。検出した各コミュニティから得られた患者分類と同じデータを用いたクラスタ分析による患者分類、および MDC による患者分類とを、JSC と新たに開発した SIG を使って比較した。

結果として、ネットワークのコミュニティ検出により、従来からあるクラスタ分析とは異なる患者の分類を行うことが可能であることが分かった。また、コミュニティ検出により得られた患者分類は、クラスタ分析により得られた患者分類と比較して、MDC による患者分類に似たものであった。

コミュニティ検出による患者分類においては、

診療データからネットワーク構成に必要な形式のデータを抽出する必要がある。今回われわれは、患者とオーダ項目という2つのフィールドのみを選択しネットワークを構成したが、診療データにはさらにさまざまなフィールドが存在する。また、医師の診療データ以外にも、例えば看護師による各種ケアのデータなどを利用できる可能性がある。看護支援や看護記録の情報システムが運用されている病院においては、データベース内に看護行為や看護の観察結果などのデータが存在する。利用したいデータを、図2において示したような、1レコードに「患者」とその患者に対して行われた「医療行為」が掲載される一般的な診療オーダのフィールド構成に整えられれば、ネットワークの構成要素に加えることが可能であり、コミュニティ検出による患者分類が可能である。このような看護行為や看護師の患者観察という視点での患者分類は未だかつて行われていない。

選択するフィールドにより患者分類がどのように変化するか、およびその利用方法などは本手法における今後の検討課題である。

今回、われわれが試みたネットワークのコミュニティ検出による患者の分類方法は、従来のような観察者の価値観が入った分類（前もってある基準による分類）とは異なり、医療者、患者、その他の環境の間の相互作用によって、自然に生まれてきた（創発された）分類を観察しているとも考えられる。日月らが指摘しているように、医療データのネットワークは、スモールワールド、スケールフリー、階層構造という特徴をもつ複雑系ネットワークであることが分かっている^{12~15}。これは、医療そのものが医療者、患者、環境などの間の相互作用によって秩序が創発される複雑系であることを意味している。コミュニティもそのような創発されてきた秩序であると考えられるならば、今回の患者分類の手法は、医療の複雑系としての特徴を調べる一つの方法であるとも考えられる。

しかし、今回われわれが患者分類に用いたコミュニティ検出手法には、大きな欠点がある。例えば、多くの患者に共通に実施される診療行為や

投薬が存在すると、ネットワークがうまくコミュニティに分割されない。今回用いた患者とオーダ項目のネットワークにおいても、ある種の輸液や血液検査などは多くの患者に対して実施されているため、そのような頂点がどのコミュニティに属するかによって患者分類が変化する可能性がある。前述したように、医療データのネットワークは、いわゆる複雑系ネットワークを構成する¹⁵。そのようなネットワークは非常に多数の頂点と繋がった頂点（ハブ）を多数持つことが分かっている。ハブは非常に多くの頂点と繋がるため、複雑系ネットワークは先に述べたうまく分割ができないネットワークに該当する。この問題への対応として、複数のコミュニティに同時に属する頂点の存在を前提としてコミュニティ（Overlapping Community）の検出を行う手法や^{16,17}、それを拡張したファジーなコミュニティ検出手法¹⁸が開発されつつある。

われわれが現在試みている患者分類では、ネットワークのコミュニティ検出が、患者集合を分割するためのキーとなる手法である。より良いコミュニティへの分割が期待できるのであれば、現在の手法に固執せず、さらに新たなコミュニティ検出手法の利用を検討すべきであると考えられる。

6. 結論

今回、われわれは病院情報システムに蓄積されている診療行為オーダのデータを用いて、ネットワークを構成し、コミュニティ検出から患者分類を行った。その患者分類とクラスタ分析、およびMDCによる患者分類との比較を行った。結果として、本手法では、クラスタ分析とは異なり比較的MDCと似た分類が得られた。

本手法は、ネットワークの要素となるフィールドの選択のみで、患者の分類が可能である。医師による診療データのみならず、看護のデータなどを自由に取り込んで分析を行い得る手法であるとわれわれは考えている。

謝辞

本研究のデータをご提供頂きました, 財団法人竹田綜合病院理事長 竹田秀先生以下病院職員の皆様のご協力を深く感謝いたします。また, 研究手法および論文執筆について, 丁寧にご指導, ご助言を頂きました医療福祉基盤研究所 田原孝先生, 岩永浩明先生に厚く御礼申し上げます。

参 考 文 献

- 1) 星 雅丈, 日月 裕. 診療情報から構成したネットワークのコミュニティ分析の検討と患者分類への応用. *医療情報学* 2014 ; **34**, 1 : 3-15.
- 2) 川淵孝一. DRG/PPS の全貌と問題点. 薬業時報社, 1997.
- 3) 松田晋哉. 基礎から読み解く DPC : 実践的に活用するために. 医学書院, 2011.
- 4) Newman EJM. *Networks an Introduction*. OXFORD UNIVERSITY PRESS, 2010.
- 5) Newman EJM, Giran M. Finding and evaluating community structure in networks. *Physical Review E* 2004 ; **69**, 026113.
- 6) Newman EJM. Fast algorithm for detecting community structure in networks. *Physical Review E* 2004 ; **69**, 066133.
- 7) Clauset A, Newman EJM, Moore C. Finding community structure in very large networks. *Physical Review E* 2004 ; **70**, 066111.
- 8) Clauset A. *Structure & Strangeness*, 2005. [<http://www.cs.unm.edu/~aaron/research/fastmodularity.htm> (cited 2013-Aug-29)].
- 9) 大垣俊一. 多様度と類似度, 分類学的新指標. *Argonauta* 2008 ; **15** : 10-22.
- 10) 佐藤義治. 多変量データの分類—判別分析・クラスター分析. シリーズ・多変量データの統計科学. 朝倉書店, 2009.
- 11) 中村永友, 金 明哲. 多変量データ解析法. R で学ぶデータサイエンス. 共立出版, 2009.
- 12) Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* 1998 ; **393** : 440-442.
- 13) Barabási AL, Albert R. Emergence of scaling in random networks. *Science* 1999 ; **286** : 509-512.
- 14) Ravasz E, Barabási AL. Hierarchical organization in complex networks. *Phys. Rev. E* 2003 ; **67** : 026112-1-026112-7.
- 15) Tachimori Y, Iwanaga H, Tahara T. The networks from medical knowledge and clinical practice have small-world, scale-free, and hierarchical features. *Physica A* 2013 ; **392**, 23 : 6084-6089.
- 16) Gergely Palla, Imre Derényi, Illés F, Tamás V. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 2005 ; **435** : 814-818.
- 17) Andrea L, Santo F, János K. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics* 2009 ; **11** : 20.
- 18) Tamás N, Andrea P, László N, Fülöp B. Fuzzy communities and the concept of bridgeness in complex networks. *Physical Review E* 2008 ; **77** : 016107.