

## わが国の複合死因データへのネットワーク分析適用

石井太・林玲子・篠原恵美子・別府志海

### はじめに

現在、人口動態統計では、死亡票に記載されている複数の死因から、世界保健機関が勧告する「疾病及び関連保健問題の国際統計分類」に準拠し、直接に死亡を引き起こした一連の事象の起因となった疾病もしくは損傷等を表す単一の「原死因」を用いて死因統計の集計・分析を行っている。これは、一連の病的事象を起こす原因を防止するという公衆衛生的な観点に基づくものであるが、一方で、現在、わが国では生活習慣病が死因の上位を占めるに至り、一人が複数の疾患を抱えることも多くなってきていることから、原死因以外の死因に着目する必要性が高まっている。

このような問題意識の下、社会保障審議会統計分科会疾病、傷害及び死因分類部会は、平成 26 年 11 月に出した報告（「疾病、傷害及び死因に関する分類に係る部会審議の際に出された意見に基づく報告」）の中で、「中長期的には、基礎疾患の情報や介入の状況、合併症、予後等、死亡診断書・死体検案書から得られる複合的な要因を把握できるような分析がなされることが望ましい。」との方向性を打ち出している。しかしながら、人口動態統計死亡票の原死因以外の複合死因情報については、近年、はじめて二次利用が可能となったところであり、わが国ではこのような複合死因データを全人口ベースで取り扱った経験が多いとはいえない状況にあった。一方、諸外国においては、従来から複合死因のデータの活用事例が存在しており、例えば人口学領域においては、原死因と複合死因の関係性を記述するための人口学的指標の構築などの先行研究が行われてきており、特に、国際的な複合死因に関する研究ネットワークである MultiCause Network において様々な研究が蓄積されている。この中で、Egidi et al. (2018) は、イタリアの 2011 年の 65 歳以上の複合死因データを用いて、欧州死因分類をノード、同一診断書に二つの死因分類による死因が記述されているときに二つのノードをエッジで結ぶ無向ネットワークを考え、これにネットワーク分析の手法を適用して死因間の関連分析を行った先駆的な研究である。

筆者らは Egidi et al. (2018) の方法論を 2009 年のわが国の複合死因データに適用することについて、石井他 (2022) で試みたところであるが、本研究では、さらに直近のデータとして、2020 年の複合死因データにネットワーク分析を適用し、複合死因間の関係分析を行うこととしたい。

## 1 先行研究

### 1.1 複合死因分析へのネットワーク分析の応用に関する先行研究

近年、複合死因間の関係分析にネットワーク分析を応用した研究が行われている。これは、各死因分類をネットワークにおけるノード（点）と考え、同一の死亡診断書に二つの死因分類による死因が記述されているときに二つのノードをエッジ（線）で結ぶ無向ネットワークを考えて分析を行う研究である。Egidi et al. (2018) は、イタリアの 2011 年の 65 歳以上の死亡診断書に記述のある死因（原死因と関連死因を区別しない）に対して、欧州死因分類をノード、同一診断書に二つの死因分類による死因が記述されているときに二つのノードをエッジで結ぶ無向ネットワークを考え、これにネットワーク分析の手法を適用して死因間の関連分析を行った先駆的な研究であり、以下、この内容についてレビューする。

最初に、死因  $i$  と  $j$  が同じ死亡診断書に書かれているとき、行列の  $(i, j)$  成分に 1 を計上していくことによって粗隣接行列を作成する。この行列は周辺度数である死因の出現頻度に依存していることから、各死因の出現頻度が同一となるよう、反復比例フィッティング (iterative proportional fitting (IPF)) 法を用いて正規化を行う。さらに、要素が小さい部分についての影響を排除する観点から、正規化された頻度について、最も強い関係があるものとして 95% タイルよりも大きいものだけを選択し、その成分が表す二つの死因の間に関係があると考えて隣接行列を作成し、これに基づくネットワーク構造を考える。

このネットワーク構造に対して、ネットワーク分析の手法を適用しつつ、分析が行われる。最初に、ネットワーク中心性 (centrality) を測定する指標である次数中心性 (degree centrality)、隣接中心性 (closeness centrality)、媒介中心性 (betweenness centrality) を示し、これらの指標に関する男女間比較を行っている。ここで、ネットワーク中心性を示す各指標は以下で定義される。

- 次数中心性  $C_D(n_i)$ : ノード  $n_i$  と他のノードを結ぶリンクの数で定義される。 $n_i$  の次数  $d(n_i)$  と表し、ノード  $n_i$  と  $n_j$  の間に関係があるとき  $x_{ij} = 1$ 、関係がないとき  $x_{ij} = 0$  という変数を用いると、

$$C_D(n_i) = d(n_i) = \sum_j x_{ij} = \sum_j x_{ji}$$

と表される。

- 隣接中心性  $C_C(n_i)$ : 他のノードとの距離が小さいノードほどより中心的であると示す指標であり、 $d(n_i; n_j)$  をノード  $n_i$  と  $n_j$  の距離（最短パスの長さ）として、

$$C_C(n_i) = \left[ \sum_{j=1}^N d(n_i; n_j) \right]^{-1}$$

で定義される。ただし、 $N$  は  $n_i$  が属する連結成分におけるノードの数である。

- 媒介中心性  $C_B(n_i)$ : あるノードが隣接していないノードを結びつける役割の大きさを表す指標であり、 $g_{jk}(n_i)$  をノード  $n_j$  と  $n_k$  を結び  $n_i$  を含む最短パスの長さ、 $g_{jk}$  をノード  $n_j$  と  $n_k$  を結ぶ最短パスの長さとして、

$$C_B(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk}$$

で定義される。

Egidi et al. (2018) ではさらに、密度 (density) や平均次数などのネットワークの構造係数の男女間比較、また、ネットワーク構造をグラフにより視覚化し、ICD-10 の分類と対応付けながら、ネットワークをサブグループに分ける検討を行っている。そして、これらのネットワーク指標やコミュニティ検出などから得られた結果を既存研究等と比較し、死因間の関係性に関する議論を行っている。

なお、ネットワーク分析では、ネットワークに関する特性を表す指標を観察するなどの方法に加え、ネットワークのグラフをいくつかのサブグループ (コミュニティ) に分けるコミュニティ検出という手法が存在している。これを複合死因分析に応用すると、原死因だけの観察からはわからない、互いに結びつきの強い死因のグループを抽出することが可能となる。これは、長寿化が進行し、複数の病態を同時に併発することが多いわが国の現在の死亡状況の分析に有効であると考えられる。

まず、矢久保 (2013) に基づき、コミュニティ構造の定義とモジュラリティという指標、また、モジュラリティを利用したコミュニティ検出法について述べる。あるネットワークがその内部で互いに密に繋がった部分グラフに分けられるとき、このネットワークはコミュニティ構造をとるといい、その部分分割グラフのことをコミュニティまたはモジュールと呼ぶ。しかしながら、普遍的に受け入れられているコミュニティの定義は未だに存在せず、与えられた問題に応じていくつもの定義が提案されているとされ、その中で、現在、最も広く受け入れられている方法がモジュラリティという量に基づいたものとされている。これは、ノード数  $N$  のネットワーク  $G$  を、互いに重なり合いのない  $n_c$  個の部分ノード集合族 (コミュニティ)  $C_l (l = 1, 2, \dots, n_c)$  に分割した際、

$$Q = \frac{1}{2M} \sum_{i,j}^N (a_{ij} - \langle t_{ij} \rangle) \delta[C(i), C(j)]$$

で定義される。ここで、 $M$  は  $G$  の全エッジ数、 $C(i)$  はノード  $i$  が属する分割ノード集合、 $\delta[C(i), C(j)]$  は分割集合  $C(i)$  と  $C(j)$  が同一集合であれば 1、そうでなければ 0 となる量である。また、 $t_{ij}$  は  $G$  の各ノード数の次数を保持したまますべてのエッジをランダムに繋ぎ直したグラフ  $G'$  の隣接行列要素であり、 $\langle \dots \rangle$  はエッジ配置に関する平均を表す。すなわち、 $t_{ij}$  は、 $G'$  における  $(i, j)$  間のエッジ本数の期待値を表す。  $Q$  の定義から、 $Q$  が大きな値をとる時、ネットワーク  $G$  は分割  $\{C_1, C_2, \dots, C_{n_c}\}$  をモジュールと

するコミュニティ構造をとっていると考えることができる。

矢久保 (2013) では、モジュラリティの最大値  $Q_{\max}$  を与える分割集合族をコミュニティとする定義が最も広く受け入れられているものの、ノード数が大きくなると分割パターンの組み合わせが非常に大きくなることから、全てのパターンに対してモジュラリティを計算することは事実上不可能となるとされている。このため、 $Q_{\max}$  の近似値を与える分割集合族を見つける方法が多数提案されている。

その一つが、Newman (2004) による貪欲法 (fast-greedy) と呼ばれる方法である。これは、矢久保 (2013) によれば、ネットワーク  $G$  の各ノード一つ一つを分割ノード集合と考え、 $G$  のエッジを介して互いに隣接する分割集合対  $\{C_l, C_{l'}\}$  を合併させた場合のモジュラリティ  $Q'$  を計算し、合併させる前からの増加分  $\Delta Q = Q' - Q$  が最大となる分割集合対を実際に合併させることをネットワークが一つになるまで繰り返し、その過程で現れる最も大きい  $Q$  を  $Q_{\max}$  の近似値とする方法である。なお、この方法は Clauset et al. (2004) により、さらに計算量を削減することが示されている。

貪欲法とは逆に、 $n_C = 1$  の分割から  $n_C = N$  の分割に向かう流れの中で  $Q$  の最大値を探す方法もあり、この場合、分割ノード集合を次々と二分していくことになるが、このためには二分された集合ができるだけコミュニティとしての性質を持つようにする必要が生じる。Newman and Girvan (2004) は媒介中心性を用いてこの分割を行う方法を提案したものである。具体的には、与えられた連結ネットワーク  $G$  の全てのエッジの媒介中心性を計算して最大となるエッジを削除し、削除されたネットワークの全てのエッジの媒介中心性を再計算し、また最大のエッジを削除する、という操作を繰り返し、その過程で現れるネットワークのモジュラリティを求めてその最大値を  $Q_{\max}$  の近似値とする方法である。ただし、この方法は貪欲法に比べて多くの計算量が必要となるとされている。

また、Barabási (2019) によれば、モジュラリティを最適化するアルゴリズムとして、最近、Blondel et al. (2008) らによる Louvain のアルゴリズムがよく使われるとしている。これは、 $N$  個のノードが別々のコミュニティに属しているとする初期状態から出発し、ステップ I として、個々のノード  $i$  をその近接ノード  $j$  の属するコミュニティに加えてみたときのモジュラリティの増加量を計算する。このモジュラリティ増加量が正の値となるもののうち、最大となる近接ノードコミュニティにノード  $i$  を置き直す。もし、どのモジュラリティ増加量も正にならなければ、ノード  $i$  の属するコミュニティは変化しない。このプロセスをどのノードの属するコミュニティも変化しなくなるまで全てのノードについて繰り返す。以上がステップ I となる。次に、ステップ II として、各ノードがステップ I で検出されたコミュニティをノードとする新しいネットワークを構成する。このステップ I と II を繰り返すことによりコミュニティは減っていくが、モジュラリティの増加がなくなり、最大モジュラリティが得られたところで終了する。

なお、ネットワーク分析においては、以上の他にも様々なコミュニティ検出法が提案されており、Luke (2015) では、コミュニティ検出アルゴリズムとして、先述の貪欲法 (fast-

greedy)、Louvain のほか、Walktrap, Edge-betweenness, Spinglass, Label propagation, Leading eigenvector などのアルゴリズムが採り上げられている。

## 2 データと方法

次に、1.1 節で述べた Egidi et al. (2018) の方法を基礎として、わが国の複合死因データにネットワーク分析の適用を試みる。本節ではこれに用いるデータと方法について述べる。

### 2.1 データ

本研究では、複合死因データとして、厚生労働省『人口動態統計』における調査票の死亡個票データを用いる\*1。なお、死亡個票ではデータがコード化されておらず、死因名がそのまま記載されている。そこで本研究では記載された死因名を ICD-10 ベースに変換し、分析に用いた。死因のコード化についての詳細は篠原他 (2021) を参照されたい。

本稿では、2020 年のデータについて分析を行う。2020 年の日本における日本人の死亡総数は 1,372,755 件であるが、このうち、複合死因を含んでいる死亡個票とのマッチングが可能だったのは 1,354,204 件である。本研究のネットワーク分析においては、年齢を 65 歳以上に限定した分析も行うことから、この中で、年齢不詳がなく、かつ、分類不能等に属さない複合死因が 2 つ以上ある 489,639 件を分析対象とした。一方、ネットワーク分析によるコミュニティ検出結果については、死因関連指標 CDAI との比較を行うが、この CDAI の算出には、複合死因とのマッチングが可能だった 1,354,204 件から、複合死因全不詳や分類不能等を除いた 1,169,490 件を対象とした。

### 2.2 方法

本研究では、MultiCause Network グループが複合死因分析に係る標準的指標の作成方法に関してまとめたガイドラインに従って、死因を取り扱う。ガイドラインでは、最初に、基礎となる死因データについて、以下の条件に従って用意することを求めている。

- 年齢範囲については制限をしない。
- 地理的範囲については特別な死亡パターンをもつ地域については対象外とすることもできる。
- 年次については 2009 年について全ての指標を計算する。ただし、2009 年が利用可

---

\*1 厚生労働省『人口動態調査』における死亡票は、前述の死亡届および死亡診断書等をもとに作成される。この調査票情報（死亡個票）を厚生労働省へ報告するシステムは 2003 年以降、順次オンライン化が進められた。本研究では死亡票情報に加え、オンラインにより提出された死亡個票情報を統計法第 32 条の規定に基づき二次利用している。また、死亡個票情報の独自集計であるため、公表数値とは一致しない場合がある。

能でない場合、最も近い年次を算出する。

- 死因について、星印 (\*) のコードと Z コードを削除する（わが国の死因統計にはこれらは使用されていない。）。
- 一つの死亡データの中に全く同じ ICD-10 コードがあった場合には一つのみを残す。
- 関連死因が死因分類（後述）66,67,68 に当たるものは、分類不能や死亡のメカニズムを表すコードであり、対象から除外する。

また、死因分類としては、表 3 が提案されており、これには 2 つのレベルである 15 分類 (group と表記) と 72 分類 (cause と表記) が存在するが、本研究では 72 分類を用いている。

この死因分類に従ってコード化されたデータを用いて、1.1 節において述べた Egidi et al. (2018) の方法と同様に、粗隣接行列を作成し、IPF 法を用いて正規化を行った後、95% タイルよりも大きい関係だけを取り出してネットワーク構造を作成した。

次に、Egidi et al. (2018) と同様、ネットワーク中心性を示す次数中心性、隣接中心性、媒介中心性を算定した。また、Egidi et al. (2018) では、ネットワーク構造のグラフを ICD-10 の分類と対応付けながらサブグループ（コミュニティ）の検討を行っているのに対し、本研究では、コミュニティ検出に関して提案されている複数のアルゴリズムを適用して、死因間の結びつきが強いコミュニティへの分類を行い、比較を行うこととした。適用するアルゴリズムについては、Luke (2015) で取り上げられているコミュニティ検出アルゴリズムから、Fast-greedy[cfp], Walktrap[cw], Edge-betweenness[ceb], Spinglass[cs], Label propagation[clp], Leading eigenvector[cle], Louvain[cl] の 7 つのアルゴリズムを用いた。

### 3 結果と考察

#### 3.1 ネットワーク構造

本研究におけるネットワーク分析の適用については、性・年齢に関し、男女計・全年齢、男女計・65 歳以上、男性・65 歳以上、女性・65 歳以上の 4 通りについて分析を行った。それぞれのネットワーク構造をグラフで視覚化したものが図 1~4 である。

グラフからは、他のノードと結びついていない孤立点が存在することがわかる。これは、95% タイルよりも大きい関係だけを取り出したことによるものである。

男女計・全年齢では、血液及び造血器の疾患並びに免疫機構の障害 (28\_BLOOD)、脳血管疾患 (45\_Cereb)、その他の筋骨格系及び結合組織の疾患 (60\_OMusc)、男女計・65 歳以上では、血液及び造血器の疾患並びに免疫機構の障害 (28\_BLOOD)、脳血管疾患 (45\_Cereb)、その他の筋骨格系及び結合組織の疾患 (60\_OMusc)、腎不全 (61\_Renal)、男性の 65 歳以上では、子宮の悪性新生物 (17\_Uter)、甲状腺障害 (32\_Thyro)、腎不全

(61\_Renal)、女性の 65 歳以上では、脳血管疾患 (45\_Cereb)、前立腺肥大 (症) (63\_Hyppro)、その他の筋骨格系及び結合組織の疾患 (60\_OMusc) である。

また、男性の 65 歳以上では、認知症 (アルツハイマー病を除く) (36\_Demen)、アルツハイマー病 (39\_Alzh)、パーキンソン病 (40\_Parki)、外的因子による肺疾患 (53\_Lung) が独立した連結成分を形成、女性の 65 歳以上では、ヒト免疫不全ウイルス [H I V] 病 (2\_AIDS) とその他の感染症及び寄生虫症 (6\_OInfec) が、また、前立腺の悪性新生物 (19\_Prost) とその他の神経系の疾患 (41\_ONerv) が独立した連結成分を形成している。

これらについては、連結成分でしか計算できない隣接中心性の評価と、コミュニティ検出の対象からは除外している。

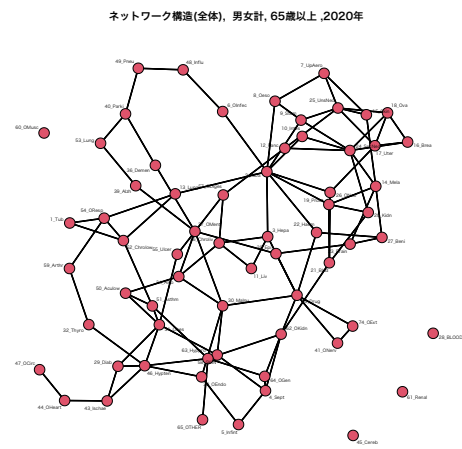
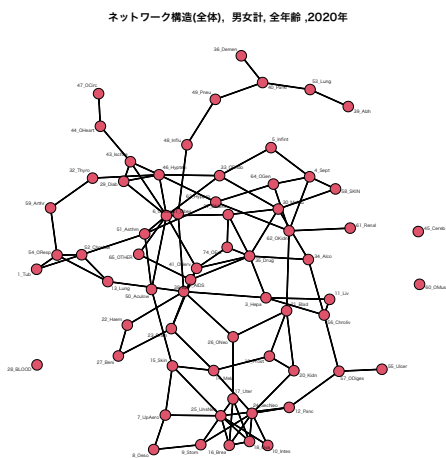


図1 ネットワーク構造 (男女計, 全年齢)

図2 ネットワーク構造 (男女計, 65歳以上)

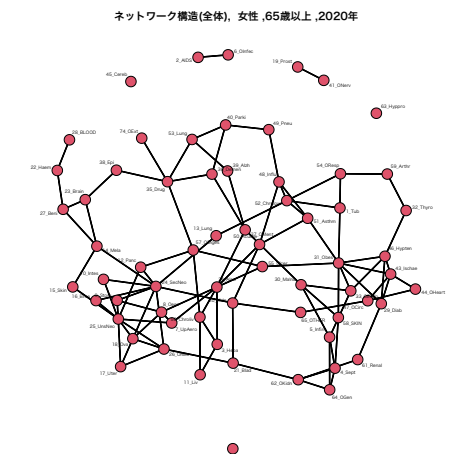
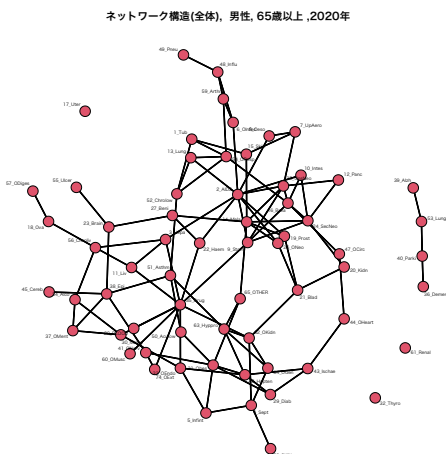


図3 ネットワーク構造 (男性, 65歳以上)

図4 ネットワーク構造 (女性, 65歳以上)

### 3.2 ネットワーク中心性指標

ネットワーク中心性指標の上位3位の死因を示したものが、表1である。

まず、次数中心性を見ると、続発性悪性新生物 (24\_SecNeo) が、男女計・全年齢、男女計・65歳以上、女性・65歳以上で上位3位以内に、また、薬物依存 (35\_Drug) が男女計・全年齢、男女計・65歳以上、男性・65歳以上で上位3位以内に含まれており、次数が高い死因となっている。

次に、隣接中心性を見ると、ヒト免疫不全ウイルス [H I V] 病 (2\_AIDS) が男女計・全年齢、男女計・65歳以上、男性・65歳以上で上位3位以内に、薬物依存 (35\_Drug) が男女計・65歳以上、男性・65歳以上で上位3位以内に含まれているものの、それ以外は全て異なる死因となっており、性別や年齢別で上位が異なっている傾向が見られる。

一方、媒介中心性を見ると、ヒト免疫不全ウイルス [H I V] 病 (2\_AIDS) が男女計・全年齢、男女計・65歳以上、男性・65歳以上で上位3位以内に、薬物依存 (35\_Drug) が男女計・65歳以上、男性・65歳以上で上位3位以内に含まれるなど、隣接中心性と似た傾向が観察される。また、肥満 (症) 及びその他の過栄養<過剰摂食> (31\_Obes) も男女計・全年齢、女性・65歳以上で上位3位以内に含まれている。

表1 ネットワーク中心性指標の上位3位の死因

	男女計、全年齢			男女計、65歳以上			男性、65歳以上			女性、65歳以上			
	略称	次数	死因	略称	次数	死因	略称	次数	死因	略称	次数	死因	
次数中心性	1	31_Obes	9	肥満 (症) 及びその他の過栄養<過剰摂食>	2_AIDS	9	ヒト免疫不全ウイルス [H I V] 病	2_AIDS	10	ヒト免疫不全ウイルス [H I V] 病	24_SecNeo	9	続発性悪性新生物
	2	25_UnsNeo	8	部位不明確、部位が明示されていない、独立した (原発性) 多部位の悪性新生物	24_SecNeo	7	続発性悪性新生物	35_Drug	10	薬物依存	25_UnsNeo	9	部位不明確、部位が明示されていない、独立した (原発性) 多部位の悪性新生物
	3	24_SecNeo	7	続発性悪性新生物	35_Drug	7	薬物依存	14_Mela	7	皮膚の悪性黒色腫	31_Obes	8	肥満 (症) 及びその他の過栄養<過剰摂食>
		35_Drug	7	薬物依存									
隣接中心性		略称	隣接性	略称	隣接性	略称	隣接性	略称	隣接性	略称	隣接性	略称	隣接性
	1	2_AIDS	0.00483	ヒト免疫不全ウイルス [H I V] 病	35_Drug	0.00546	薬物依存	35_Drug	0.00676	薬物依存	34_Alco	0.00588	アルコール使用<飲酒>による精神及び行動の障害
	2	31_Obes	0.00474	肥満 (症) 及びその他の過栄養<過剰摂食>	2_AIDS	0.00541	ヒト免疫不全ウイルス [H I V] 病	9_Stom	0.00637	胃の悪性新生物	50_Aculow	0.00578	その他の急性下気道感染症
	3	41_Onerv	0.00469	その他の神経系の疾患	62_OKidn	0.00535	糸球体疾患及び腎尿管間質性疾患	2_AIDS	0.00633	ヒト免疫不全ウイルス [H I V] 病	55_Ulcer	0.00575	胃潰瘍、十二指腸潰瘍及び胃空洞潰瘍
媒介中心性		略称	媒介性	略称	媒介性	略称	媒介性	略称	媒介性	略称	媒介性	略称	媒介性
	1	2_AIDS	550.557	ヒト免疫不全ウイルス [H I V] 病	2_AIDS	488.975	ヒト免疫不全ウイルス [H I V] 病	35_Drug	439.154	薬物依存	24_SecNeo	372.700	続発性悪性新生物
	2	31_Obes	387.452	肥満 (症) 及びその他の過栄養<過剰摂食>	35_Drug	344.623	薬物依存	2_AIDS	361.696	ヒト免疫不全ウイルス [H I V] 病	31_Obes	332.144	肥満 (症) 及びその他の過栄養<過剰摂食>
	3	6_0Infec	330.000	その他の感染症及び寄生虫症	62_OKidn	279.310	糸球体疾患及び腎尿管間質性疾患	14_Mela	193.610	皮膚の悪性黒色腫	34_Alco	291.574	アルコール使用<飲酒>による精神及び行動の障害

### 3.3 コミュニティ検出

次にコミュニティ検出の結果について述べる。女性・65歳以上について、Fast-greedy アルゴリズム (貪慾法) によってコミュニティ検出を行って、視覚化したものが図5である。

ここでは、死因が6つのコミュニティ (サブグループ) に分かれている。オレンジ色で示されているのが悪性新生物等を含むコミュニティ A である。赤色で示されているのが肝疾患などの消化器系疾患等を含むコミュニティ B となっている。青色で示されているのがアルツハイマー病などの神経系疾患等を含むコミュニティ C、水色で示されているの



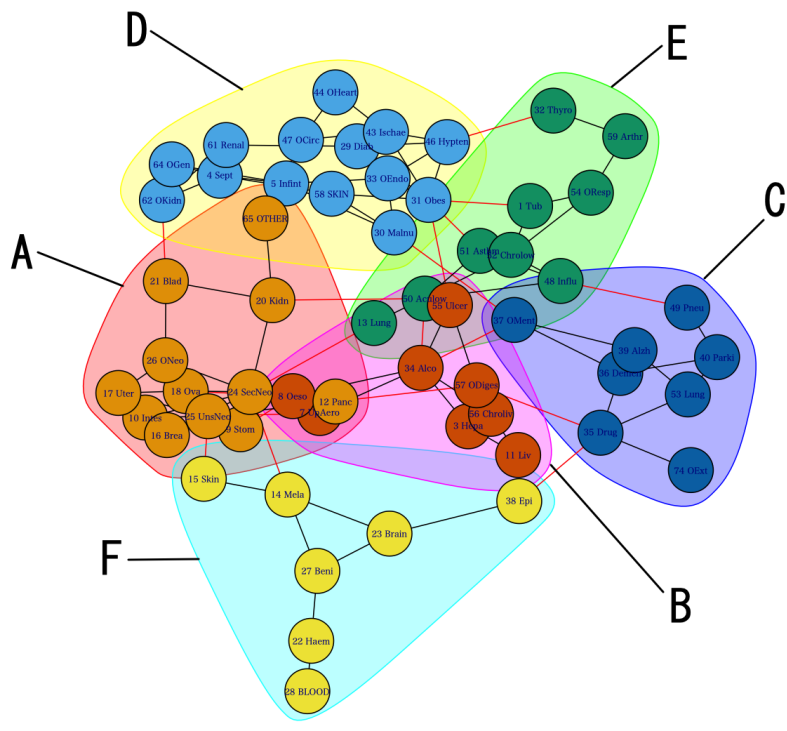


図5 コミュニティ検出 (Fast-greedy), 女性, 65 歳以上

が糖尿病、栄養失調 (症) 及びその他の栄養欠乏症等を含むコミュニティ D、緑色で示されているのが結核、インフルエンザなどの呼吸器系疾患等が含まれるコミュニティ E であり、黄色で示されているのがそれ以外を示すコミュニティ F である。

ここで、コミュニティ A~E を代表する死因として、A は続発性悪性新生物 (24\_SecNeo)、B は肝、肝内胆管、胆のう及びその他の胆道の悪性新生物 (11\_Liv)、C はアルツハイマー病 (39\_Alzh)、D は糖尿病 (29\_Diab)、E は結核 (1\_Tub) を選択し、Fast-greedy 以外のアルゴリズムによる結果についても、これらの代表する死因が含まれるコミュニティを同じ記号で表して比較したものが表 2 である。ただし、Label propagation[clp], Leading eigenvector[cle] では、糖尿病 (29\_Diab) が結核 (1\_Tub) と同じコミュニティに分類されていることから、これらのアルゴリズムについては、D を代表する死因として、栄養失調 (症) 及びその他の栄養欠乏症 (30\_Malnu) を取ることとした。

表 2 コミュニティ検出結果の比較, 女性, 65 歳以上

ID	1	3	4	5	7	8	9	10	11	12	13	14	15	16	17	18	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	43	44	46	47	48	49	50	51	52	53	54	55	56	57	58	59	61	62	64	65	74										
Abb	Pub	Hep	Sep	Inf	Pa	Aer	Des	Stom	Int	Col	Panc	Lun	Mea	Skin	Brea	Uter	Ova	Kidn	Blad	Hae	Brain	Cy	Na	ND	Nec	Ben	LOO	Dia	Malnu	Obes	Thyr	End	Alco	Drug	Eme	Mer	Epi	Alzh	Park	Sch	Hae	pt	D	C	Di	Inf	Flu	Pne	Culo	ost	th	rot	lun	g	Res	Jice	phro	D	ipe	SKIN	Art	Renal	Kid	D	Gas	THE	DE	Ext
cf	B	D	D	B	B	A	A	B	A	E	F	F	A	A	A	A	A	A	F	F	A	A	A	F	F	D	D	D	E	D	B	C	C	C	F	C	C	D	D	D	D	E	C	E	E	E	C	E	B	B	B	D	E	D	D	D	A	C										
cw	E	B	G	G	A	A	A	A	B	A	E	F	F	A	A	A	A	H	F	A	A	A	F	H	D	G	D	E	G	B	C	C	C	C	C	F	C	C	D	D	D	D	E	E	E	E	E	C	E	B	B	B	G	E	G	G	D	C										
ceb	E	B	D	D	A	A	A	A	B	A	G	F	F	A	A	A	A	F	F	A	A	A	F	F	D	D	D	E	D	B	C	C	C	F	C	C	D	H	D	H	G	C	G	G	C	E	B	B	B	D	E	D	D	D	H	C												
cs	E	B	G	G	A	A	A	A	B	A	E	F	F	A	A	A	A	F	F	A	A	A	F	F	D	G	D	D	G	B	C	C	C	F	C	C	D	D	D	D	E	C	E	E	E	C	E	B	B	B	G	E	G	G	D	C												
clp	E	B	D	D	A	A	A	A	B	A	G	F	F	A	A	A	A	F	F	A	A	A	F	F	E	D	E	E	E	A	C	C	C	C	C	E	H	E	A	G	G	G	G	C	E	E	B	A	D	E	E	D	D	A	C													
cle	E	B	D	D	A	A	A	A	B	A	A	F	F	A	A	A	I	I	F	F	A	A	A	F	G	E	D	E	D	B	C	C	C	C	C	C	C	E	E	E	H	H	H	H	C	E	B	B	A	D	E	D	D	E	C													
cl	E	B	D	D	B	B	A	A	B	A	E	F	F	A	A	A	A	F	F	A	A	A	F	F	D	D	D	D	B	C	C	C	F	C	C	D	D	D	D	E	C	E	E	C	E	B	B	B	D	D	D	D	A	C														

これを見ると、どのアルゴリズムによる結果を見ても、先の Fast-greedy アルゴリズムによる結果から得られた 5 つのコミュニティに類似したコミュニティが抽出されていることがわかる。しかしながら、その結果を詳細に観察すると、アルゴリズムによって異なる点があることがわかる。まず、抽出されたコミュニティの数は、Fast-greedy[cfg]、Louvain[cl] では 6、Spinglass[cs] では 7、Walktrap[cw]、Edge-betweenness[ceb]、Label propagation[clp] では 8、Leading eigenvector[cle] では 9 と異なっている。また、コミュニティに含まれる死因を観察すると、口唇、口腔及び咽頭の悪性新生物 (7\_UpAero) や食道の悪性新生物 (8\_Oeso) は、Fast-greedy[cfg] と Louvain[cl] では肝、肝内胆管、胆のう及びその他の胆道の悪性新生物 (11\_Liv) を含むコミュニティ B に分類されるのに対して、それ以外のアルゴリズムでは続発性悪性新生物 (24\_SecNeo) を含むコミュニティ A に分類されており、アルゴリズムの選択によってはどちらのグループと近いと判定されるかが分かれる死因が存在することが理解される。

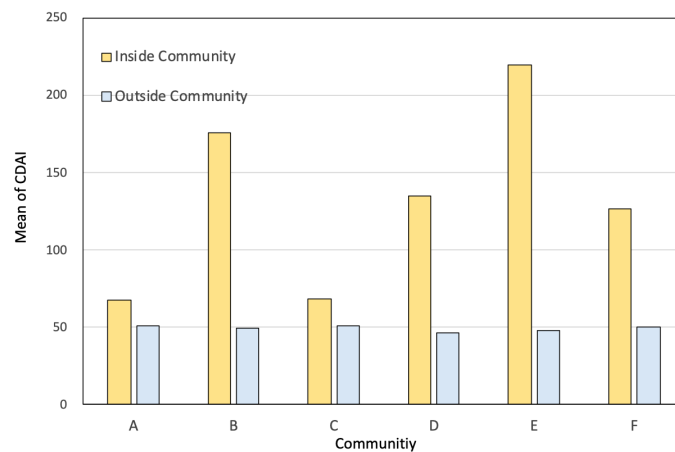


図 6 コミュニティ内外別の CDAI の平均値 (Fast-greedy), 女性, 65 歳以上

次に、このようにして検出されたコミュニティと、原死因と複合死因との関連を表す指標である CDAI との整合性を検証するため、原死因と複合死因が両方ともある同一のコミュニティに含まれる場合の CDAI の平均値 (ただし原死因と複合死因が同じ場合を除く) を「コミュニティ内 CDAI 平均値」、どちらかがそのコミュニティ以外に含まれる場合の平均値を「コミュニティ外 CDAI 平均値」とし、Fast-greedy アルゴリズムによるコミュニティについて、コミュニティ内外別 CDAI 平均値を示したものが図 6 である。これを見ると、どのコミュニティにおいても、コミュニティ内平均値はコミュニティ外平均値を上回っていることが観察され、ネットワーク分析に基づくコミュニティは CDAI 指標と整合性を持っていることが理解できる。

## おわりに

本研究では、複合死因データに関する分析手法について、海外・日本の先行研究を概観するとともに、特にネットワーク分析の応用についての先行研究をレビューした。そして、Egidi et al. (2018) によるネットワーク分析応用に関する方法論を日本の複合死因データに適用することにより、複合死因関連分析へのネットワーク分析応用の可能性を検討した。本研究で行った複合死因データに関する分析手法の概観からは、複合死因データが有する豊富な可能性が理解されたとともに、わが国の複合死因データへのネットワーク分析適用からは、複合死因間の関係を分析する上でのネットワーク分析の有効性が明らかになったといえよう。

しかしながら、残された課題も存在している。本研究では、MultiCause Network グループのガイドラインに従い、2020年の日本の複合死因データについて、表3に示した死因分類を用いて分析を行った。しかしながら、このような分析は、対象とする年次や死因分類によって異なる結果が導かれる可能性も考えられる。したがって、2020年以外のデータへのネットワーク分析の適用や、死因簡単分類など MultiCause Network グループの分類以外の死因分類についても検討することが必要であろう。また、本研究では様々なコミュニティ検出アルゴリズムを適用し、複合死因ネットワークを、相互に関係の深いと考えられるいくつかのコミュニティへと分類することを試みた。しかしながら、詳細に観察すると、分類結果はアルゴリズムによって異なっており、どのアルゴリズムがわが国の複合死因分析に適しているのかということなどについても検討することが必要と考えられる。このような課題については、今後、さらに検討を行っていくこととしたい。

## 参考文献

- 石井太・林玲子・篠原恵美子・別府志海 (2022) 「複合死因データに関する分析手法とその応用—わが国データへのネットワーク分析適用の試み—」, 『人口問題研究』, 第78巻, 第1号, pp.56–77.
- 篠原恵美子・別府志海・林玲子・石井太 (2021) 「死亡個票における「死亡の原因」欄の記載文字列の分析」, 『医療情報学』, 第41(Suppl.)巻, pp.733–737.
- 矢久保孝介 (2013) 『複雑ネットワークとその構造』, 共立出版.
- Barabási, A.-L. (2019) 『ネットワーク科学: ひと・もの・ことの関係性をデータから解き明かす新しいアプローチ (池田裕一他監訳)』, 共立出版.
- Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008) “Fast unfolding of communities in large networks”, *Journal of statistical mechanics: theory and experiment*, Vol. 2008, No. 10, p. P10008.
- Clauset, A., M. E. Newman, and C. Moore (2004) “Finding community structure in

- very large networks”, *Physical review E*, Vol. 70, No. 6, p. 066111.
- Egidi, V., M. A. Salvatore, G. Rivellini, and S. D’Angelo (2018) “A network approach to studying cause-of-death interrelations”, *Demographic Research*, Vol. 38, pp. 373–400.
- Luke, D. A. (2015) *A User’s Guide to Network Analysis in R*: Springer.
- Newman, M. E. and M. Girvan (2004) “Finding and evaluating community structure in networks”, *Physical review E*, Vol. 69, No. 2, p. 026113.
- Newman, M. E. (2004) “Fast algorithm for detecting community structure in networks”, *Physical review E*, Vol. 69, No. 6, p. 066133.

表3 使用する死因分類

Group	Cause	略称	Disease or external cause	死因名	ICD-10 code	
1		INF	Infectious and parasitic diseases	感染症及び寄生虫症		
	1	Tub	Tuberculosis	結核	A15-A19, B90	
	2	AIDS	AIDS (HIV-disease)	ヒト免疫不全ウイルス [HIV] 病	B20-B24	
	3	Hepa	Viral hepatitis	ウイルス肝炎	B15-B19, B94, 2	
	4	Sept	Septicaemia	敗血症	A40-A41	
	5	Infint	Intestinal infectious diseases	腸管感染症	A00-A09	
	6	OInfec	Other Infectious and parasitic diseases	その他の感染症及び寄生虫症	(A00-B99) + U01-U49 + U80-U89 - Supra 'A' and 'B' codes	
2		NEO	Neoplasms	新生物		
	7	UpAero	Malignant neoplasm of lip, oral cavity, pharynx	口唇、口腔及び咽頭の悪性新生物	C00-C14	
	8	Oeso	Malignant neoplasm of oesophagus	食道の悪性新生物	C15	
	9	Stom	Malignant neoplasm of stomach	胃の悪性新生物	C16	
	10	Intes	Malignant neoplasm of small intestine, colon, rectum and anus, and other/ill-defined digestive organs	小腸、結腸、直腸 S 状結腸移行部、直腸、肛門、肛門管、その他及び部位不明確の消化器の悪性新生物	C17, C18, C19, C20, C21, C26	
	11	Liv	Malignant neoplasm of liver, the intrahepatic bile ducts, gallbladder and other unspecified parts of biliary tract	肝、肝内胆管、胆のう及びその他の胆道の悪性新生物	C22, C23, C24	
	12	Panc	Malignant neoplasm of pancreas	膵の悪性新生物	C25	
	13	Lung	Malignant neoplasm of larynx and trachea/bronchus/lung	喉頭、気管、気管支及び肺の悪性新生物	C32-C34	
	14	Mela	Malignant melanoma of skin	皮膚の悪性黒色腫	C43	
	15	Skin	Malignant neoplasm of skin	皮膚のその他の悪性新生物	C44	
	16	Brea	Malignant neoplasm of breast	乳房の悪性新生物	C50	
	17	Uter	Malignant neoplasm of cervix uteri and other parts of uterus	子宮の悪性新生物	C53, C54, C55	
	18	Ova	Malignant neoplasm of ovary	卵巣の悪性新生物	C56	
	19	Prost	Malignant neoplasm of prostate	前立腺の悪性新生物	C61	
	20	Kidn	Malignant neoplasm of kidney	腎盂を除く腎の悪性新生物	C64	
	21	Blad	Malignant neoplasm of bladder	膀胱の悪性新生物	C67	
	22	Haem	Malignant neoplasm of lymph./haematopoietic tissue	リンパ組織、造血組織及び関連組織の悪性新生物	C81-C96	
	23	Brain	Malignant neoplasm of eye, brain and other parts of central nervous system	眼、脳及びその他の中枢神経系の部位の悪性新生物	C69-C72	
	24	SecNeo	Secondary malignant neoplasm	続発性悪性新生物	C77-C79	
	25	UnsNeo	Malignant neoplasm of ill-defined/unspecified/independent (primary) multiple sites	部位不明確、部位が明示されていない、独立した(原発性)多部位の悪性新生物	C76, C80, C97	
	26	ONeo	Other malignant neoplasms	その他の悪性新生物	(C00-C99) - Supra 'C' codes	
	27	Beni	Benign neoplasms, In situ neoplasms and neoplasms of uncertain or unknown behaviour	その他の新生物	D00-D48	
	3	28	BLOOD	Diseases of the blood(-forming organs), immunol. disorders	血液及び造血器の疾患並びに免疫機構の障害	D50-D89
	4		ENDOC	Endocrine, nutritional and metabolic diseases	内分泌、栄養及び代謝疾患	
		29	Diab	Diabetes mellitus	糖尿病	E10-E14
		30	Malnu	Malnutrition and other nutritional deficiencies	栄養失調(症)及びその他の栄養欠乏症	E40-E64
		31	Obes	Obesity	肥満(症)及びその他の過栄養<過剰摂食>	E65-E68
32		Thyro	Disorders of thyroid gland	甲状腺障害	E00-E07	
33		OEndo	Other Endocrine, nutritional and metabolic diseases	その他の内分泌、栄養及び代謝疾患	(E00-E90) - Supra 'E' codes	
5		MENT	Mental and behavioural disorders	精神及び行動の障害		
	34	Alco	Alcoholic psychosis/chronic alcohol abuse	アルコール使用<飲酒>による精神及び行動の障害	F10.1-F10.9, G31.2	
	35	Drug	Drug dependence, toxicomania	薬物依存	F11-F16 except F11.0, F12.0, F13.0, F14.0, F15.0, F16.0, F18-F19 except F18.0, F19.0	
	36	Demen	Dementias (excluding Alzheimer)	認知症(アルツハイマー病を除く)	F01, F03, G31.0, G31.1, G31.8, G31.9	
	37	OMent	Other Mental and behavioural disorders	その他の精神及び行動の障害	(F00-F99) except F10.0, F11.0, F12.0, F13.0, F14.0, F15.0, F16.0, F17.0, F18.0, F19.0 - Supra 'F' codes	

Group	Cause	略称	Disease or external cause	死因名	ICD-10 code
6		NERV	Diseases of the nervous system	神経系の疾患	
	38	Epi	Epilepsia	てんかん及びびてんかん重積	G40-G41
	39	Alzh	Alzheimer' s disease	アルツハイマー病	G30
	40	Parki	Parkinson' s disease	パーキンソン病	G20, G21
	41	ONerv	Other Diseases of the nervous system	その他の神経系の疾患	(G00-G98) except (G31.0, G31.1, G31.2, G31.8, G31.9) - Supra 'G' codes
7		CIRC	Diseases of the circulatory system	循環器系の疾患	
	43	Ischae	Ischaemic heart diseases	虚血性心疾患	I20-I25
	44	OHeart	Other heart diseases	その他の心疾患	I00-I09, I30-I33, I39-I45, I47-I48, I49 except I49.0, I50-I52
	45	Cereb	Cerebrovascular diseases	脳血管疾患	I60-I69
	46	Hypten	Hypertensive diseases	高血圧性疾患	I10-I15
	47	OCirc	Other Diseases of the circulatory system	その他の循環器系の疾患	(I00-I99) except (I46, I49.0, I95.9, I99) - Supra 'I' codes
8		RESP	Diseases of the respiratory system	呼吸器系の疾患	
	48	Influ	Influenza	インフルエンザ	J10-J11
	49	Pneu	Pneumonia	肺炎	J12-J18
	50	Aculow	Other acute lower respiratory diseases	その他の急性下気道感染症	J00-J09, J19-J22
	51	Asthm	Asthma	喘息	J45-J46
	52	ChroLow	Other Chronic lower respiratory diseases	その他の慢性下気道感染症	J40-J44
	53	Lung	Lung diseases due to external agents	外的因子による肺疾患	J60-J70
	54	OResp	Other Diseases of the respiratory system	その他の呼吸器系の疾患	(J00-J99) except (J96.0, J96.9) - Supra 'J' codes
9		DIGES	Diseases of the digestive system	消化器系の疾患	
	55	Ulcer	Ulcer of stomach, duodenum and jejunum	胃潰瘍、十二指腸潰瘍及び胃空腸潰瘍	K25-K28
	56	ChroLiv	Chronic liver disease	慢性肝疾患	K70, K73-K74
	57	ODiges	Other Diseases of the digestive system	その他の消化器系の疾患	(K00-K93) - Supra 'K' codes
10	58	SKIN	Diseases of the skin and subcutaneous tissue	皮膚及び皮下組織の疾患	L00-L99
11		MUS	Diseases of the musculoskeletal system/connective tissue	筋骨格系及び結合組織の疾患	
	59	Arthr	Rheumatoid arthritis and osteoarthritis	関節リウマチ及び関節症	M05-M06, M15-M19
	60	OMusc	Other Diseases of the musculoskeletal system/connective tissue	その他の筋骨格系及び結合組織の疾患	(M00-M99) - Supra 'M' codes
12		GEN	Diseases of the genitourinary system	泌尿器系の疾患	
	61	Renal	Renal Failure	腎不全	N17-N19
	62	OKidn	Other Diseases of kidney and ureter	糸球体疾患及び腎尿管間質性疾患	N00-N16
	63	Hyppro	Hyperplasia of prostate	前立腺肥大(症)	N40
	64	OGen	Other Diseases of the genitourinary system	その他の泌尿器系の疾患	(N00-N99) - Supra 'N' codes
13	65	OTHER	Other diseases	その他の疾患	000-099, (P00-P96) except P28.5, Q00-Q99, H00-H95, R95
14		ILLDEF	Other Symptoms, signs, abnormal findings, ill-defined causes	症状、徴候及び異常臨床所見・異常検査所見で他に分類されないもの	
	66	Senil	Senility	老衰	R54
	67	Mecha	Mechanisms of the death*	死亡のメカニズム	I46, I49.0, R09.2, R40.2, R57
	68	OIlldef	Other Symptoms, signs, abnormal findings, ill-defined causes**	その他の症状、徴候及び異常臨床所見・異常検査所見で他に分類されないもの	(R00-R94), (R96-R99), I95.9, I99, J96.0, J96.9, P28.5, U00 - Supra "R" codes
15		EXT	External cause	傷病及び死亡の外因	S, T, V, W, X, Y codes + F10.0, F11.0, F12.0, F13.0, F14.0, F15.0, F16.0, F17.0, F18.0, F19.0
	70	Trans	Transport accidents	交通事故	V01-V99, Y85
	71	Fall	Accidental falls	転倒・転落	W00-W19
	72	Pois	Accidental poisoning	有害物質による不慮の中毒及び有害物質への曝露	X40-X49
	73	Suic	Suicide/intentional self-harm	自殺	X60-X84
	74	OExt	Other external causes	その他の外因	S, T, V, W, X, Y codes + F10.0, F11.0, F12.0, F13.0, F14.0, F15.0, F16.0, F17.0, F18.0, F19.0 - Supra codes