

# エゴセントリックネットワークと形式概念分析を利用した 社会ネットワーク分析の手法の提案

A Method of Social Network Analysis Using Formal Concept Analysis and Ego-Centric Networks

林 宏紀<sup>\*1</sup>    伊東 樹希<sup>\*2</sup>    西尾 典晃<sup>\*2</sup>    武藤 敦子<sup>\*2</sup>    犬塚 信博<sup>\*2</sup>  
Hiroki Hayashi    Tatsuki Itou    Noriaki Nishio    Atsuko Mutoh    Inuzuka Nobuhiro

<sup>\*1</sup>名古屋工業大学工学部情報工学科

Department of Computer Science, Faculty of Engineering, Nagoya Institute of Technology

<sup>\*2</sup>名古屋工業大学大学院工学研究科情報工学専攻

Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology

Social Network Analysis is a field of sociology, where networks, such as the human relations, are analyzed in order to understand roles of parsons in social context. In this paper, we propose a method for this aim using the method of formal concept analysis and ego-centric networks. It clarifies the relationship between social nature and human structure possessed by each individual in the network relationships with friends. We examined the proposal method we used a data acquired by a questionnaire on friendship and social attributes to students. Some groups characterized by the method had statistically significant ratio of students who possess some social attributes.

## 1. はじめに

社会学において、1970年以降、活発に社会ネットワーク分析が研究されてきた。これは行為者の属性ではなく、その関係性に着目して現象を捉える方法論である。人や組織の持つ関係性をネットワークと捉え、どの行為者が中心的であるか、どのような派閥が存在するか、といった研究がされてきた。人間関係、企業間関係、産業間関係、国家間関係などのあらゆるネットワークが対象である [1, 2]。

データマイニングにおいてもグラフやネットワーク上のマイニングが重要になってきている。グラフから頻出する部分グラフを見つけ出し、WWWや分子構造などの局所構造の理解に役立てることができる。

我々は、これに関連して学生の友人関係に注目し、友人関係ネットワークから各個人の行動を考える研究やエゴセントリックネットワークに焦点をあててそのパターンを枚挙する研究を行ってきた [3, 4]。本研究はネットワーク構造を形式概念分析を用いて抽出し、社会的属性との照合を検討する。形式概念分析は対象とその属性データから概念単位を得る方法で、ここではネットワークの構造属性に適用する。

本論文の構成は2節は、社会ネットワーク分析について、3節は形式概念分析について、それぞれの詳細を述べる。4節は提案手法を述べ、5節は提案手法を適用した実験と考察を述べる。6節はまとめと課題を述べる。

## 2. 社会ネットワーク

### 2.1 ネットワークの構造的特徴の分析

金光 [1] は Kapferer [5] の研究を取り上げ、エゴセントリックネットワークが人間の行動分析に有用であることを論じている。Kapferer [5] は、工場でのトラブルの経緯を丹念に分析し、組織上の人間関係など表面上の人間関係では説明できない行動について、それが所属する社会的なネットワークによって説明可能であることを示している。

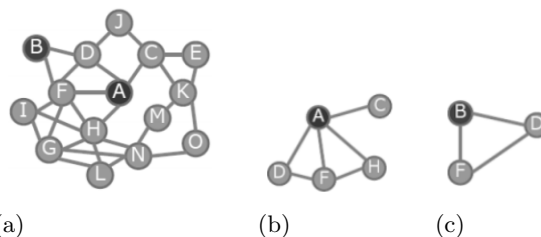


図 1: あるネットワーク (a) とこのネットワーク上の A および B をエゴとしたエゴセントリックネットワーク (b, c)。

このように社会ネットワーク分析において、構造からわかる特徴の詳細な分析が盛んに行われている。これには市場における企業の競争関係、行為者の地位・役割分析、ネットワークの適正規模、株式仲買人や見合い紹介業者などの仲介者の利益に関する分析などがある [1, 2]。

以下ではネットワークの詳細を分析するにあたって、1つの観点であるエゴセントリックネットワークについて説明する。

### 2.2 エゴセントリックネットワーク

ネットワークでの特定の一頂点を定め、そこから直接つながる局所ネットワークをエゴセントリックネットワークという。エゴとは自己を意味し、他者（オルター）との関係を把握するものである。エゴとオルター間、オルター同士の間で作られるグラフがエゴセントリックネットワークである。

たとえば、図 1(a) のネットワークからエゴを A または B としたエゴセントリックネットワークが図 1(b) および図 1(c) である。図 1(b) のオルターは  $\{C, D, F, H\}$ 、図 1(c) でのオルターは  $\{D, F\}$  である。

行為者の集合を  $V$  とし、その中の各行為者  $v_1, v_2 \in V$  の間に関係があることを辺  $(v_1, v_2)$  で表したグラフ  $N = (V, E) (E \subseteq V \times V)$  を考える。このとき、1人の行為者（エゴ）に注目して、それを中心としたローカルなネットワーク、つまり、エゴと直接繋がる行為者の集合から誘導される部分グラフをエゴセントリックネットワークという。即ち、次のように定義する。

連絡先: 林宏紀, 名古屋工業大学犬塚研究室, 名古屋市昭和区御器所町, 052-735-5050, hayashi@nous.nitech.ac.jp

**定義 1**  $N$  に対する  $c \in V$  のエゴセントリックネットワーク  $EC(N, c)$  とは, 三項組  $(c, V_c, E_c)$  であり,  $V_c = \{v \in V | (c, v) \in E\} \cup \{c\}$ ,  $E_c = \{(u, v) \in E | u \in V_c \wedge v \in V_c\}$  である.

図 1(a) のグラフ  $N = (V, E)$  に対する頂点  $A \in V$  のエゴセントリックネットワークは次の通りである.

$$\begin{aligned} N_A &= (A, V_A, E_A), \\ V_A &= \{A, C, D, F, H\}, \\ E_A &= \{(A, C), (A, D), (A, F), (A, H), (C, A), (D, A), \\ &\quad (D, F), (F, A), (F, D), (F, H), (H, A), (H, F)\} \end{aligned}$$

エゴ  $e$  のエゴセントリックネットワークでは  $e$  がすべてのオルターと接続するので,  $e$  自身をノードに含めなくても情報は失われない. この場合, エゴセントリックネットワークは次の通りとなる. 以後はこれに従う.

**定義 2**  $N$  に対する  $c \in V$  の (エゴを除いた) エゴセントリックネットワーク  $EC(N, c)$  とは, 三項組  $(c, V_c, E_c)$  であり,  $V_c = \{v \in V | (c, v) \in E\}$ ,  $E_c = \{(u, v) \in E | u \in V_c \wedge v \in V_c\}$  である.

図 1(a) のグラフ  $N = (V, E)$  に対する頂点  $A \in V$  のエゴを取り除いたエゴセントリックネットワークは次の通りである.

$$\begin{aligned} N_A &= (A, V_A, E_A), \\ V_A &= \{C, D, F, H\}, \\ E_A &= \{(D, F), (F, D), (F, H), (H, F)\} \end{aligned}$$

### 3. 形式概念分析

形式概念分析は 1982 年に Wille によって提唱された [6]. 形式概念分析は, 概念データを思考単位として, 概念構造の明確化や事象の分析, データの可視化及び依存関係などを明らかにするものである. 形式概念分析は特にデータ解析の調査においてユーザビリティなどの観点から適している [7].

形式概念分析では, 対象の集合  $G$  とそれが取り得る属性の集合  $M$  を扱う. 例えば, 対象の集合として  $G = \{\text{ハト}, \text{ヒト}, \text{カモノハシ}, \text{ネコ}\}$  を, 属性の集合として  $M = \{\text{卵生}, \text{言葉(をはなす)}, \text{母乳(で育つ)}\}$  を考えると, 表 1 が得られる.  $G, M$  およびこの表を合わせて形式文脈という. 対象のある集合  $X \subseteq G$  と, 属性のある集合  $Y \subseteq M$  に対して,  $X$  のすべての対象が共通してもつ属性の集合が  $M$  であり,  $M$  をすべて有する対象の集合が  $X$  であるとき, 組  $(X, Y)$  を形式概念という. このとき  $X$  (対象の集合) を外延,  $Y$  (対象が持つ属性の集合) を内包という.

例えば,  $X = \{\text{ハト}, \text{カモノハシ}\}$  において,  $X$  の対象が共通してもつ属性の集合は  $Y = \{\text{卵生}\}$  であり,  $Y$  をすべて有する対象の集合は,  $\{\text{ハト}, \text{カモノハシ}\}$  である. したがって, 組  $(\{\text{ハト}, \text{カモノハシ}\}, \{\text{卵生}\})$  は形式概念である.

二つの形式概念  $(X, Y), (X', Y')$  に対し,  $X \subseteq X'$  によって順序を決めると, 形式概念は束となる. これのハッセ図を図 2(a) に示した. 図 2(b) は (a) の略記法である. 属性のみに注目し, 上から下へ新たに追加される属性のみを記している. 合流したノードは親の属性を併せ持つ.

形式的には形式文脈は 3 つ組  $\mathbb{K} = (G, M, I)$  で, 対象の集合  $G$  と属性の集合  $M$  とその間の二項関係  $I \subseteq G \times M$  から

表 1: 形式文脈の例.

名前	卵生 (a)	言葉 (b)	母乳 (c)
ハト (1)	×		
ヒト (2)		×	×
カモノハシ (3)	×		×
ネコ (4)			×

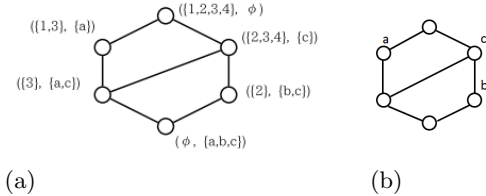


図 2: 表 1 から得た形式概念束のハッセ図 (a) とその略記 (b).

成る. ここで  $(g, m) \in I$  を  $gIm$  と書く. 任意の  $X \subseteq G$  と  $Y \subseteq M$  について以下の写像を定義する.

$$\begin{aligned} X &\mapsto X^I = \{m \in M \mid gIm \text{ for all } g \in X\} \\ Y &\mapsto Y^I = \{g \in G \mid gIm \text{ for all } m \in Y\} \end{aligned}$$

このとき次の通りに形式概念が定義される.

**定義 3 (形式概念)**  $A \subseteq G, B \subseteq M$  のとき, 二つ組  $(A, B)$  が形式文脈  $\mathbb{K} = (G, M, I)$  の形式概念であるとは,  $A = B^I$  かつ  $B = A^I$  であることを言う.  $A$  を外延,  $B$  を内包と呼ぶ.

### 4. 提案手法

形式概念の手法を人間間のネットワークに適用し, 有意味なパターンを得る手法を提案する. 社会ネットワーク分析においては, 各個人が有するエゴセントリックネットワークが, その人の関係的構造を規定すると考えることができ, その人の社会関係資本を規定しているとみることが可能である. そこで, エゴセントリックネットワークの構造と, その人の社会的性質の関係性を結び付けることができると思われる.

一人一人の人間を形式文脈における対象, その人のエゴセントリックネットワークに含まれる部分ネットワークを属性とした形式概念分析を行い, 得た形式概念をパターンとする. これと別に得た個人の社会的属性と形式概念を照合することで, 社会的属性を探るために有用な, ネットワーク構造を探る.

#### 4.1 誘導部分グラフ同型に基づくネットワークの形式文脈

分析したい人とその間の関係に関するネットワーク (グラフ) を  $N = (P, F)$  とする.  $P$  は頂点であり人の集合,  $F$  は辺の集合であり, たとえば友人関係である.

このとき, 次のとおり定義される形式文脈を  $\mathbb{K}_N = (P, M, I_N)$  を考える. 対象集合は  $P$ , すなわちネットワークを構成する人の集合である. 属性集合  $M$  は, ある部分グラフの集合であり, これをここでは属性グラフと呼ぶ.

$P$  の各要素  $e \in P$  からはエゴセントリックネットワーク  $EC(N, e)$  が対応して得られる. すると, 二項関係  $I_N$  は,  $P$  と  $M$  の誘導部分グラフのマッチングを行った結果の集合である. すなわち, あるエゴ  $e \in P$  とある属性グラフ  $S = (V_s, E_s) \in M$

表 2: 図 1(a) と図 4 から得た形式文脈の一部. 各行 A, B, C, ... はエゴ, 各列 1, 2, 3, ... は属性グラフである.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	x	x	x	x	x					x							
B		x															
C	x	x	x	x					x								
D	x	x	x	x	x					x							
⋮																	

に対して, 関係  $eI_N S$  が成り立つのは,  $e$  のエゴセントリックネットワーク  $EC(N, e) = (V_e, E_e)$  ( $V_e \subseteq P, E_e \subseteq F$ ) のある誘導部分グラフが  $S$  と同型になることであるとき, その時に限るとする. つまり, ある  $V' \subseteq V_e$  から  $V_s$  への全単射  $f$  が存在し, 任意の  $(v_1, v_2) \in V' \times V'$  に対し,  $(v_1, v_2) \in E_e$  であることが  $(f(v_1), f(v_2)) \in E_s$  であることの必要十分条件になることである.

上の定義において属性グラフ  $M$  は, エゴセントリックネットワークを分類するのに必要と思われる適当なサイズのグラフを用意するものとする. 上述のとおり定義される形式文脈  $\mathbb{K}_N = (P, M, I_N)$  をネットワーク  $N$  と  $M$  から誘導される形式文脈と呼ぶことにする.  $N$  と  $M$  から誘導される形式文脈を得ることで, これから得られる形式概念束は部分グラフによって特徴付けられる, すなわち社会的関係によって特徴付けられるエゴの集合となることが期待される.

たとえば, 図 1(a) を対象とするネットワーク  $N = (P, F)$  に対して, 図 4 に示した属性グラフの集合  $M$  を用いた場合, 形式文脈  $(P, M, I_N)$  の一部を書き下すと, 表 2 のようになる.

## 4.2 社会的属性を有し, 社会的関係に特徴づけられる形式概念

社会関係資本の理論に倣うならば, エゴセントリックネットワークに含まれる関係的な特徴によって, リーダーシップを有する等の社会的属性が対応づけられる可能性がある. すなわち, 適切な属性グラフが与えられることで, そこから得られる形式概念は, 社会的属性に特徴づけられるエゴの集合と対応することが期待される.

そこで, ネットワーク  $N = (P, F)$  に対してそのエゴの集合  $P$  に社会的属性  $a$  が与えられていると仮定する.  $a$  は二値属性とする. このとき,  $N = (P, F)$  とあるグラフ属性の集合  $M$  から誘導される形式文脈から得られる形式概念  $(P', M')$  ( $P' \subseteq P, M' \subseteq M$ ) が  $a$  に対して有意味であるとは,  $P'$  において社会的属性  $a$  を有する割合が,  $P$  全体で社会的属性  $a$  を有する割合に比べて統計的に有意に高い場合をいう.

図 3 に提案方法をまとめる. ネットワークからそのエゴセントリックネットワークの構造に関する形式概念を求め, 得た形式概念を社会的属性と照合して統計的に有意な概念を得る.

## 5. 実験

### 5.1 実験に利用したデータ

名古屋工業大学のある学年の 3 クラス合計約 170 名の協力を得てアンケートによってデータを得た.

アンケートでは各学生に二群の質問を行っている. 第一群の質問は友人関係に関するものであり, クラスの全学生氏名列挙した用紙を用意し, 友人関係にあるものをマークするよう

手順

入力 一群の人物の集合  $P$  に関するネットワーク  $N = (P, F)$ ,  $P$  に関する社会的属性  $a_1, \dots, a_n$

1.  $P$  の各要素であるエゴ  $e$  に対して, そのエゴセントリックネットワーク  $EC(N, e)$  を得る.
2.  $N$  とあらかじめ定めた属性グラフから形式文脈  $(P, M, I_N)$  を得る. ここで  $I_N$  は  $EC(N, e)$  に対して  $M$  の各属性グラフが誘導部分グラフ同型になるかどうかを表す 2 項関係である.
3. 形式文脈  $(P, M, I_N)$  からすべての形式概念を得る.
4. 社会的属性  $a$  に対して有意味である形式概念  $(P', M')$  を枚挙する.

図 3: 提案方法の流れ.

表 3: 人格適応論 [8] の人格タイプを参考とした社会的属性のためのアンケート質問項目.

A	場を盛り上げるのが好きで, 人に気をを使う
B	責任感が強く, 中途半端は嫌い
C	用心深く, 注意深く考えて行動する
D	控えめで, 他人の考えに譲る
E	周りに流されず, 自分自身の考えで行動する
F	自分を周りにアピールするのがうまい

求めた. ここでは, 友人関係の度合いをそろえるため「授業中に限らずよく行動を共にする」, 「授業などで顔を合わせると話をする」, 「話はあまりしないが顔と名前は一致する」, 「知らない」の四段階から答えることとし, この内「授業中に限らずよく行動を共にする」と答えたものを友人関係にあるものとした. ただし, 二名の学生に対して, その一方が他方を友人とした場合, その逆が答えられていなくとも, 二名の学生間に対称的に友人関係があるものとした. 学生をノード, 友人関係を辺とするネットワークを分析対象とした.

第二群の質問として, 社会的属性を得る質問を同じ学生に行った. この質問項目を表 3 に挙げる. 質問は V. Jains らの提唱する人格適応論 [8] で分類する人格タイプ, 演技型, 強迫観念型, パラノイド型, スキゾイド型, 受動攻撃型, 反社会型に対応する. これらの名称はネガティブなイメージがあるが, それぞれプラスとマイナスの両面を持つとされており, 一人の人物が複数の属性を有することは一般的であるとされる. 各項目に対して, 自分がそれに当てはまる度合いを, 「よくあてはまる」, 「まああてはまる」, 「あまりはてはまらない」, 「まったくあてはまらない」から選択してもらった. これらの A~F の各属性について, 「よくあてはまる」, 「まああてはまる」としたものを  $A \uparrow, B \uparrow$  等と記し, 「あまりはてはまらない」, 「まったくあてはまらない」を  $A \downarrow, B \downarrow$  等と記す. 即ち, 計 12 個の社会的属性を用いた.

### 5.2 実験の内容と結果

上述の友人関係ネットワークを対象として提案方法で述べた通りの方法を実施した. 形式概念生成に用いる属性グラフの集合は図 4 の 17 個の部分グラフとした. これはサイズ 4 までの

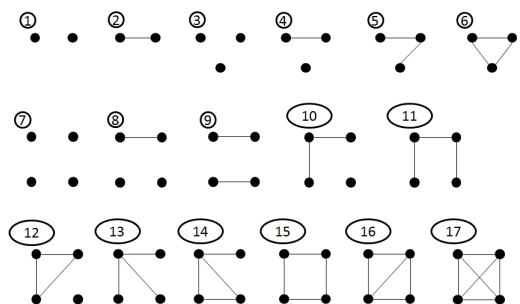


図 4: 実験で使用した形式文脈生成のための属性グラフ.

すべての可能なグラフを枚挙したものである.

社会的属性の有意性を見るため2項検定を実施し, p 値1%未満の場合に有意とした. 形式概念分析の結果得られた形式概念束を図5に示す. 頂点1つ1つが形式概念である. 頂点に付される数字は属性グラフの番号であり, 図2(b)の概念束で示したように属性集合に属性が加わるときのみそれを表記する省略法を用いている.

図5は有意な社会的属性も示している. あるノードに  $A \uparrow$ ,  $A \downarrow$  等が付されているのは, その形式概念で社会的属性  $A$  またはその否定的属性が p 値1%未満となったことを意味する. また,  $BC \downarrow$  は,  $B \downarrow$  と  $C \downarrow$  の両方を意味する.

実験においては  $\uparrow$  の属性で有意であったものは1つもなかったが,  $\downarrow$  で有意であったものが多数確認された. たとえば図5の上から2段目最右の頂点は属性グラフ1と2を誘導部分グラフとして持ち, 他を持たない友人関係を持つ学生のグループであり, そこに  $B \downarrow$  が付されているのはその学生が「用心深く, 中途半端は嫌い」に「あてはまらない」と答えた割合が, 学生全体に比して有意に大きいということの意味する.

## 6. おわりに

形式概念を人間間のネットワークに適用し, 有意なパターンを得る手法を提案した. アンケートで得た友人関係ネットワークと社会的属性に適用し, 社会的属性を探るために有用なネットワークの部分構造を探った.

ここでの実験では社会的属性のうち, ネガティブなもの ( $A \downarrow$ ,  $B \downarrow$  等) のみが有意として抽出されたが, これへの合理的理由は判明していない. また形式概念の中には対象となる学生数が小さなものもあるため2項検定を用いたが, 有意な傾向を見るためには更にサンプリング数を増やして実験を行う必要がある. 他に, 属性グラフの選択方法, 社会的属性の妥当性, 計算効率に関する検討も今後の課題である.

## 参考文献

[1] 金光淳, 社会ネットワーク分析の基礎 — 社会関係資本論にむけて, 勁草書房, 2003.

[2] 安田雪, 実践ネットワーク分析 — 関係を解く理論と技法, 新曜社, 2001.

[3] N. Inuzuka, T. Kondo and S. Yamamoto, Analysis of Asymmetric Friendship among Students from Class Attendance Records, *New Advances in Intelligent Decision Technologies, Studies in Computational Intelligence*, **199**, pp. 393-404, Springer, 2009.

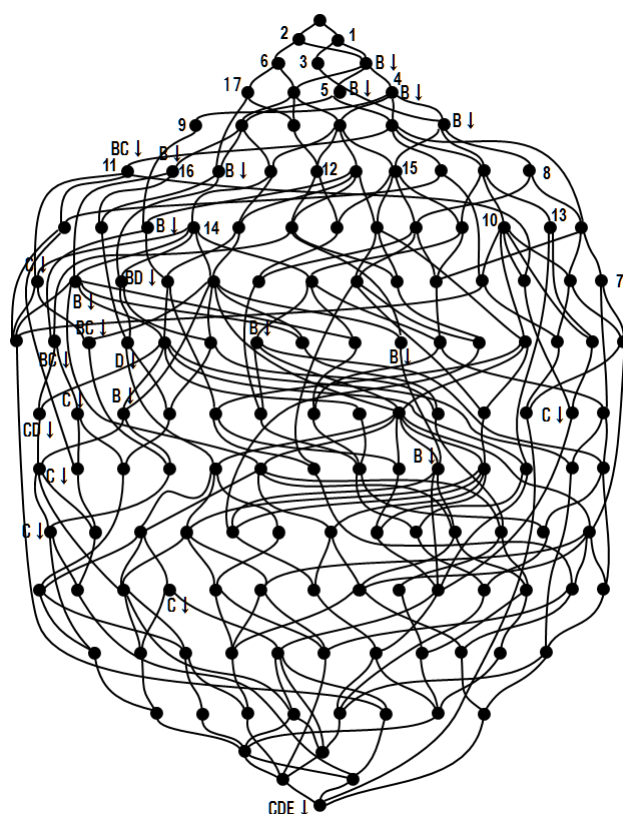


図 5: 友人関係ネットワークから得た形式概念束. 各形式概念に有意な社会的属性がある場合は記号を付している.

[4] N. Inuzuka, S. Takeuchi and H. Matsushima, Pattern Mining on Ego-Centric Networks of Friendship Networks, *Knowledge-Based and Intelligent Information and Engineering Systems, LNCS*, **6884**, Springer, pp. 89-97, 2011.

[5] K. Bruce. Norms and the Manipulation of Relationships in a Work Context, in *Social Networks in Urban Situations*, Manchester Univ. Press, 1969.

[6] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, In *in: I. Rival, ed., Ordered sets*, D. Reidel, pp. 445-470, 1982.

[7] J. H. Correia, G. Stumme, R. Wille, and U. Wille, Conceptual knowledge discovery — a human-centered approach, *Applied Artificial Intelligence*, Vol. 17, No. 3, pp. 281-302, 2003.

[8] V. Joines and I. Stewart, *Personality Adaptations: A New Guide to Human Understanding in Psychotherapy and Counselling*, Lifespace Publishing, 2002 (邦訳: 白井他, 誠信書房, 2007).

[9] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*, Springer, 1998.