

学生の出席記録を用いた 友人ネットワークの分析

Friendship Network Analysis Using the Attendance Records of Students

松島 裕*1
Hiroshi Matsushima

門阪 祥吾*2
Shogo Kadosaka

山本 修平*2
Shuhei Yamamoto

犬塚 信博*1
Nobuhiro Inuzuka

*1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

*2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

We analyze a friendship network using attendance records of students. In an existing method, in order to make networks, we need time-consuming methods, such as questionnaires, and so analysis for the network formation was high-cost, and was not realistic. In this paper, we calculate friendship-score, the strength of friendship, using attendance records of students and give the directed and weighted graphs by the score. By using these networks, we analyze how friend pairs increase and how common properties of social networks, scale-free, cluster and small-world, change. While the average shortest path length was around six, other properties grow in early stage and are steady afterwards. Also, we compare some roles obtained from questionnaires to high-ranking students of centralities. The centralities using degrees are effective to specify roles concerning leadership.

1. はじめに

社会学において、1970年以降、社会ネットワーク分析の研究が活発に行われてきた。社会ネットワーク分析とは、行為者の属性ではなく関係性に注目して現象を捉えようとする方法論である。人間関係や企業間関係、国家間関係、WWW等のネットワークを対象に、どのようにしてネットワークが形成・発達していくか、どの行為者が中心であるのか、どのような派閥が存在するのか、背後にどのような関係性が存在するのかなどといった研究が行われてきた。

これらのネットワークを得るためにはアンケート等を用いて定性情報を得るといった手間や時間のコストが必要であり、ネットワーク形成のプロセスを調査することは現実的には困難な場合が多い。また、友人関係等のネットワークを重み付きグラフとして考えるには、任意の友人ペアがどれだけ親しいかを調べるための客観的な手法が必要であるといった問題点がある。

本研究では、これらの問題を解決するために、下村ら [1,2] が提案した各学生の授業の出欠を記録した時刻（打刻時刻）から友人関係の強さの度合いである友人スコアを求める手法を用いる。名古屋工業大学ではICチップを用いて授業の出欠管理を行うシステムを採用しており、このデータを用いることにより、期間を指定することで任意の期間の友人関係ネットワークを得ることができ、また学生間の友人関係の強さをより客観的に求めることができる。

本研究では、指定した期間の打刻データを用いて友人スコアを求めることで従来の研究では困難だった各時点での友人ネットワークを生成し、ネットワーク構造や社会ネットワーク共通の性質であるスケールフリー性、クラスタ性、スモールワールド性の時間推移の分析を行い、各性質の推移の法則性について分析する。また、友人スコアから生成したネットワークには各学生の持つ役割がネットワーク構造に反映されている。このこ

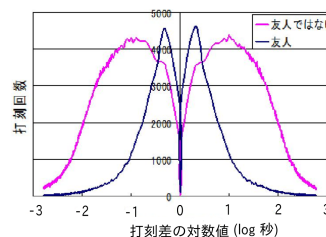


図 1: 打刻時刻差に対するの対人関係別の度数分布

とを検証・分析するために、中心性概念を用いた分析を行った。

2. 出欠記録に基づく友人スコア

名古屋工業大学では授業の出欠管理のために、学生証のICチップを読み込み学生数、授業名、日時、教室名等をデータベースに記録・管理するシステムを採用している。下村らは、各打刻時刻差（以下、打刻差とする）における友人ペア、非友人ペアの打刻数の度数分布（図 1）を活用し、学生間の友人関係の強さの度合いを表す友人スコアを求める手法を提案した [1,2]。本節ではこの方法について述べる。

ある学生ペアに対し、友人ペアであるという事象を f 、打刻差 t からなる集合を $T = \{t_1, t_2, \dots, t_n\}$ とする。また $t \in T$ の発生確率 $p(t)$ 、事象 f に関する条件付き確率 $p(t|f)$ はそれぞれ独立であるとする。このとき、打刻差 T であるとき友人ペアである確率 $p(f|T)$ は次のように表せる。

$$p(f|T) = \frac{p(f) \cdot p(T|f)}{p(T)} = p(f) \prod_{t \in T} \frac{p(t|f)}{p(t)} \quad (1)$$

次に打刻差毎の友人である割合を求める。打刻差 t における友人ペアである割合を r_t としたとき、

$$\begin{aligned} r_t &= \frac{\text{友人ペアに限定したときの打刻差 } t \text{ の総データ数}}{\text{打刻差 } t \text{ の総データ数}} \\ &= \frac{X_f \cdot m_f \cdot p(t|f)}{X \cdot m \cdot p(t)} \end{aligned} \quad (2)$$

表 1: 講義毎の再現率・適合率

	講義 K	講義 J	講義 P
再現率	75.3	86.9	65.7
適合率	60.0	88.9	78.9

ここで X は学生ペア数, X_f は友人ペア数, m は 1 組の学生ペアから発生する平均打刻件数, m_f は友人ペアに限定したときの 1 組から発生する平均打刻件数である. すると, 次式が得られる.

$$p(t | f) = \frac{X \cdot m \cdot p(t) \cdot r_t}{X_f \cdot m_f} = \frac{m \cdot p(t) \cdot r_t}{p(f) \cdot m_f} \quad (3)$$

ゆえに, 式 (1) より次を得る.

$$p(f | T) = p(f) \prod_{t \in T} \frac{m \cdot r_t}{m_f \cdot p(f)} = p(f)^{(1-n)} \left(\frac{m}{m_f}\right)^n \prod_{t \in T} r_t \quad (4)$$

ここで, n は T に含まれるデータ数とする.

また, 同様に友人でない確率 $p(\bar{f} | T)$ は次のように得られる.

$$p(\bar{f} | T) = p(\bar{f})^{(1-n)} \left(\frac{m}{m_0}\right)^n \prod_{t \in T} (1 - r_t) \quad (5)$$

ただし, m_0 は友人以外ペアに限定したときの 1 組のペアから発生する打刻データの平均件数である.

$p(f | T)$ にロジット関数を用い, 友人関係の強さの度合いとし, これを友人スコアと呼ぶ. この値が正であるとき, 友人であると推測する.

$$\begin{aligned} \text{logit } p(f | T) &= \log\left(\frac{p(f | T)}{1 - p(f | T)}\right) \\ &= \log(p(f | T)) - \log(p(\bar{f} | T)) \end{aligned} \quad (6)$$

友人スコアを用いた友人推測の再現率・適合率は表 1 の通り報告されている [1,2]. 各講義において再現率, 適合率共に高い数値となり, 友人スコアの精度は信頼性が高い.

3. 友人ネットワーク

2 節で述べた友人スコアから友人ネットワークを生成し, 友人ペア数などの基本的な情報, ネットワーク構造の推移について調べる.

3.1 友人ネットワークの基本的な情報

名古屋工業大学 2007 年度新入生の 2 年間分の打刻データを用いて友人スコアを求め友人ネットワークを生成する. クラスタ係数等を求める際, 重みの値域は $[0,1]$ でなければならないため, 友人スコアにシグモイド関数を用いる. 図 2 は, 実際に生成した友人ネットワークの一部である. また, 表 2 は同ネットワークにおける, 全学生数, 全学生ペア数 (学生ペアとは, 同じ端末に 10 分以内に打刻したことがある学生同士のペアを指す), 友人ペア数によって推測される友人ペア数である.

3.2 ネットワーク構造の時間推移

2007 年度新入生の 1 カ月分の打刻データを 1 日ずつずらしたデータを用い, ネットワーク構造の時間推移について調べる. 図 3 は友人ネットワークの総友人ペア数の時間推移である.

前期では大きく変化することはなかったが, 5 月末までの期間で約 600 ペア増加している. その後は前期・後期ともに安定して推移している.

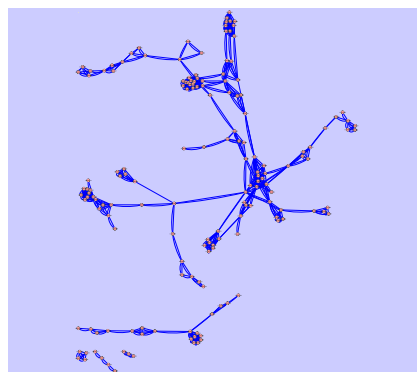


図 2: 友人ネットワーク

表 2: ネットワークの基本情報

項目	人数 (ペア数)
全学生数	948
学生ペア数	579552
友人ペア数	5972

同様の友人ネットワークを用いて, 友人ペアの入れ替わりについて調べる. 1 カ月後のネットワークと比較し, 新たに友人となったペア (NEW) と友人ではなくなったペア (EVER) の数の 1 人あたりの平均を求める (重みの増加量の和の平均と減少量の和の平均で求める). 次の図 4 は NEW, EVER の時間推移である.

その結果, 前期の初期段階では NEW のほうが EVER よりも大きな値となり, そのため 5 月末までの期間で総友人ペアは増加した. その後は小さな増減はあるものの両者とも 1.5-1.7 で安定した. これにより, 友人関係の強さが動的であると考えられるが, さらにスコアの精度も含めた検証が必要である.

4. 社会ネットワークの性質の時間推移

社会ネットワークには, スケールフリー性 [3], クラスタ性, スモールワールド性 [4,5] という 3 つの共通する性質があることが分かっている. スケールフリー性とは「極少数の人は多くの友人を持つが, 他の大多数の人は友人数が少ない」という性質であり, 度数 k の分布 $p(k)$ が $p(k) \propto k^{-\gamma}$ ($\gamma < 0$) に従う. クラスタ性とは「友人同士も友人である割合が多い」という性質であり, この性質を測る尺度としてクラスタ係数がある. スモールワールド性とは「任意の目的地まで比較的少ないステップで達する」という性質であり, 平均最短パス長が短いことを意味する.

本節では, これら 3 つの性質の時間推移について分析する.

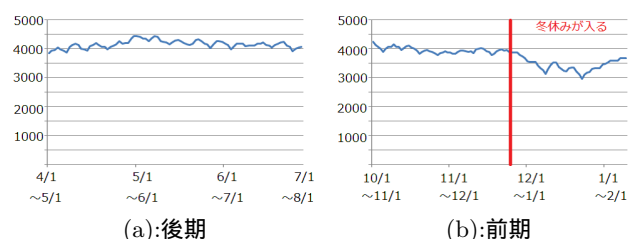


図 3: 総友人ペア数の時間推移

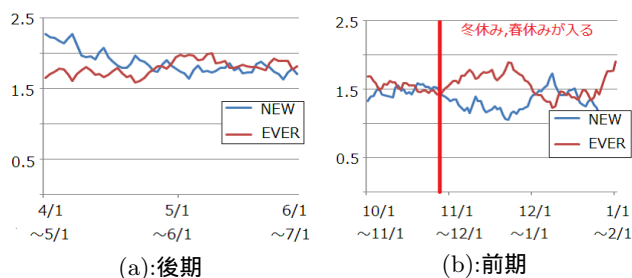


図 4: 新規友人数 (NEW) と友人関係の減少数 (EVER) の時間推移

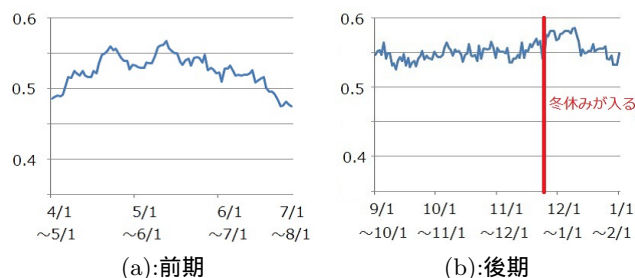


図 6: クラスタ係数の時間推移

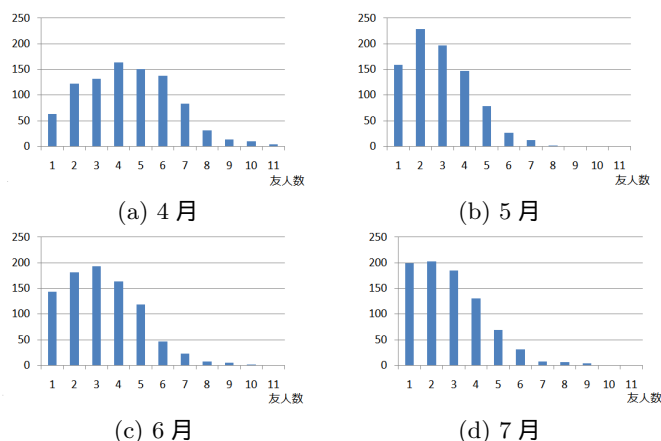


図 5: 度数分布の時間推移

4.1 スケールフリー性の時間推移

図 5 は、名古屋工業大学 2007 年度前期の 1 カ月ごとの打刻データから生成した友人ネットワークの度数分布である。

4 月の時点では、友人数 4 人の度数が最も高くなり、明らかにベキ則を満たしていない。しかし、時間の経過とともに友人数 3 人以下のときの度数が増え、それ以外の度数は減っている。7 月の時点で最もベキ則の形に近いものになった。この後後期では 7 月の度数分布とほぼ同じ形で安定して推移しており、徐々にスケールフリー性を獲得してきたと見ることができる。

4.2 クラスタ性の時間推移

クラスタ性を測る尺度としてクラスタ係数を求める。

重み付き有向グラフ $G = (V, E, c)$ (V : 頂点集合 E : 辺集合 c : 重み) のクラスタ係数 C は、頂点 v_i の隣接頂点の集合を $E(v_i)$ としたとき、次の式で求める [6]。

$$C = \text{avg}_{v_i \in V} (C(v_i))$$

$$C(v_i) = \frac{\sum_{v_j \neq v_k \in E(v_i)} \frac{c(v_i, v_j) \cdot c(v_j, v_k) + c(v_i, v_k) \cdot c(v_k, v_j)}{c(v_i, v_j) + c(v_i, v_k)}}{k|E(v_i)|P_2}$$

次の図 6 は、2007 年度の名古屋工業大学の新生の前期・後期の 1 カ月分の打刻データを 1 日ずつずらしながら生成した友人ネットワークのクラスタ係数の推移である。前期では、6 月上旬までの期間で上昇する傾向にあったが、その後は減少する傾向にあり、後期では、冬休みの影響を受ける期間を除き 0.55 前後で推移し続けた。また、前期の終わり後期の最初では、値が約 0.06 増加している。一般的にクラスタ性が高いといわれる mixi のネットワークのクラスタ係数が約 0.36 であり、

表 3: 平均最短パス長

	4 月	5 月	6 月	7 月
頂点数	895	909	898	885
平均最短パス長	6.271	5.982	5.831	5.992
	10 月	11 月	12 月	1 月
頂点数	892	866	822	802
平均最短パス長	6.179	6.477	7.120	7.395

友人ネットワークのクラスタ性は高い。クラスタ性の変化については、その精度を含めさらに検証が必要である。

4.3 スモールワールド性の時間推移

平均最短パス長を求めるために、ネットワークを強連結性成分ごとに分割する。友人ネットワークは、少数の頂点を含む強連結成分と残り大多数の頂点 (900 弱) を含む 1 つの強連結成分で構成されており、この性質は 1 年間の打刻データを 1 カ月ごとに分割して生成したネットワークでも成り立つ。また最大の強連結成分の頂点数はほぼ一定の値で落ち着いた。表 3 は、2007 年度名古屋工業大学新生の 1 カ月毎のネットワークに含まれるもっとも大きい強連結成分の頂点数と平均最短パス長である。

平均最短パス長は大きく変化することはなかったが、常に 6 前後の値である。これは、1960 年代に Milgram が行った実験 [4] のときの値とほぼ同様になった。また、12 月、1 月のみ他に比べて頂点数が小さく平均最短パス長が大きくなった。長期休暇により打刻データの件数が少なく、友人スコアの精度にも影響を与える可能性があり、これらの値の有意性は今後検証が必要である。

5. 学生の役割の分析

友人ネットワークが学生の社会的役割を反映していることを示すため、中心性概念を役割と比較・検証した。

表 4 は、本研究で比較・検証のために用いた、頂点 $v_i \in V$ における各中心性の定義である。

離心中心性、近接中心性はネットワークが強連結でなければ求めることができない。そこで各強連結成分を求め、各成分ごとにサイズに比例した各中心性上位者を 54 名抽出した。これをグラフ論的中心人物と呼ぶ。

また、アンケートは、2007 年度入学生 2 年生のある学科を対象とし、次の項目について実施した。各項目は自己評価 (自分自身に対して回答したもの) と他者評価 (自分以外の学生を評価したもの) に分類される。

表 4: 中心性定義

中心性	定義
次数中心性	$c_D(v) = deg(v)$ (deg : 次数)
Pagerank	$c_{PR}(v) = (1 - d) + d \sum_{i=1}^n \frac{c_{PR}(T_i)}{d^-(T_i)}$ ($d^- (v)$: v の入次数, T_i : 頂点 v からリンクされている頂点)
媒介中心性	$c_B(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \delta_{st}(v)$ ($\delta_{st}(v)$: v を含む st 間の最短パスの割合)
離心中心性	$c_E(v) = \frac{1}{\max d(u,v)}$ ($d(u,v)$: u, v 間の最短パス長))
近接中心性	$c_C(v) = \frac{1}{\sum_{v \in V} d(u,v)}$

表 5: アンケート回答者数

対象となる 学生数	自己評価の 回答数	他者評価の 回答数
162	137	112

- 自己評価
 - 自分は社会的である
 - 自分はリーダーシップがある
 - 友人と行動するとき, 提案するタイプだ
- 他者評価
 - その人を頼りにしている
 - その人は自分より格上だ
 - その人はリーダーシップがある
 - その人は顔が広い

自己評価は a から d の 4 段階で回答させ, a と b と回答した学生を自己評価における中心人物とした. 他者評価は 1 人でも支持している学生を中心人物とした. ただし, 他者評価の「頼りにしている」は中心人物が 100 人となったため, 2 人以上から支持されている学生を中心人物とした. また, 表 5 は対象となる学生数, 自己評価の回答数, 他者評価の回答数, 表 7 はアンケートの各項目で中心人物である学生の人数である.

表 6 は, アンケートの中心人物の集合と各グラフ論的中心人物の集合に対して再現率, 適合率を求めたものである. アンケートの他者評価の項目の「リーダーシップがある」「自分より格上である」という項目で, 次数を用いた中心性 (即ち, 次数中心性, Pagerank) との相関が見られた.

表 6: 中心性指標とアンケートとの比較

自己評価					
社会的だ	次数中心性	PageRank	媒介中心性	離心中心性	近接中心性
再現率	43.50%	47.80%	39.10%	39.10%	43.50%
適合率	37.00%	40.70%	34.60%	33.30%	37.00%
リーダーシップがある					
再現率	44.00%	40.00%	36.00%	48.00%	40.00%
適合率	20.40%	18.50%	17.30%	23.50%	19.60%
提案するタイプだ					
再現率	42.00%	38.00%	30.00%	32.00%	38.00%
適合率	38.90%	35.20%	28.80%	29.60%	35.20%
他者評価					
頼りにしている	次数中心性	PageRank	媒介中心性	離心中心性	近接中心性
再現率	39.60%	49.10%	45.30%	41.50%	45.30%
適合率	38.90%	48.10%	46.20%	43.10%	47.10%
自分より格上だ					
再現率	57.40%	51.90%	40.70%	44.40%	46.30%
適合率	57.40%	51.90%	42.30%	47.10%	49.10%
リーダーシップがある					
再現率	55.90%	50.00%	41.20%	44.10%	47.10%
適合率	35.20%	31.50%	26.90%	29.40%	31.40%
顔が広い					
再現率	36.70%	34.70%	32.70%	32.70%	34.70%
適合率	33.30%	31.50%	30.80%	31.40%	33.30%

表 7: 項目別中心人物の人数

自己評価	
質問の項目	中心人物の人数
社会的だ	46
リーダーシップがある	26
自分は提案するほうだ	50
他者評価	
質問の項目	中心人物の人数
頼りにしている	53
自分より格上だ	54
リーダーシップがある	34
顔が広い	49

6. まとめ

本研究では, 下村らが提案した学生の打刻データから友人関係の強さの度合いを表す友人スコアを求める手法 [1,2] を用い, 友人スコアから生成された友人ネットワークを用いて, 社会ネットワークの性質・ネットワーク構造の時間推移の分析, 中心性概念を用いた役割の分析を行った.

ネットワーク構造の時間推移の分析では, 総友人ペア数は 5 月末までの期間で増加し, ある程度早い段階で落ち着くこと, 友人の強さは常に変化があったためその構成は動的であることが分かった. 社会ネットワークの性質の推移の分析では, 次数分布は前期の段階でベキ則に落ち着くこと, クラスタ係数は 5 月末までの期間で増加し, 落ち着くまで減少すること, 平均最短パス長はあまり変化がないが, Milgram の実験のときと同様に 6 前後であることがわかった.

中心性指標を用いた学生の役割の分析では, 次数中心性や Pagerank など次数に関する中心性指標を用いることがリーダーシップに関する役割の分析に有効である可能性が示された.

以上より, 友人スコアを用いた友人ネットワークの分析は, これまでに知られている社会ネットワークの諸性質とよく対応し, また時間推移の分析や中心性分析に有用であることが分かった.

今後の課題としては, これらの結果の詳細な分析, 友人推定と精度との関係, 中心性を用いた時間推移の分析, 新たな時間推移の性質の発見が挙げられる. また, 本研究の時間推移の結果が他の友人ネットワークでも成り立っているかを検証する必要がある. そして, これらの性質を満たすグラフモデルの提案が挙げられる.

参考文献

- [1] N.Inuzuka, T.Nakano, K.Shimomura. Friendship Analysis Using Attendance Records to University Lecture Classes. IASK International Conference Teaching and Learning, 478-486, 2008.
- [2] 下村 幸作, 中野 智文, 犬塚 信博, 学生の出欠時間を活用した学生の友人関係分析. 第 6 回データマイニングと統計数理研究会 (SIG-DMSM), 2008.
- [3] R.Albert, A.-L.Barabasi. Statical mechanics of complex networks. Review of Modern Physics, 74, 47-97, 2002.
- [4] S.Milgram. The small world problem. Pycology Today, 1:67, 1967.
- [5] D.J.Watts. Small Worlds. Princeton University Press, 1999.
- [6] 鈴木 智也. 有向重み付き複雑ネットワーク解析. ネットワークが創発する知能研究会 (JWEIN06), 2006.