

人工知能学会における研究者ネットワークの分析

Structural Analysis of AI Researchers' Network

安田 雪^{*1*3}
Yuki Yasuda

松尾 豊^{*2*3}
Yutaka Matsuo

^{*1} 東京大学大学院経済学研究科 ^{*2} 産業技術総合研究所 ^{*3} GBRC社会ネットワーク研究所
The University of Tokyo National Institute of AIST Global Business Research Center

Our purpose here is to (1) investigate the structure of the Japanese top AI researchers' network to identify the Invisible College of AI researchers from 2003 to 2005, and (2) to consider the effects of betweenness centrality on future eigenvector centrality. Findings suggest groups of researchers constitute an Invisible College, and that AI researchers' betweenness centrality correlates with their future eigenvector centrality scores.

1. 目的

本報告では、Web 上の情報から人工知能研究者のネットワークを抽出し、一時点の Web 上の情報のマイニングによって後時点での人工知能研究者の現実の関係構造を予測し示す。具体的には、2003 年から 2005 年の人間関係の構造を抽出し、研究者のネットワーク構造の変化と個々の研究者の位置特性の変化を検討する。

既に安田[安田 04]において、わが国の人口知能研究では、先端的な研究を担う一部の優れた研究者の集団が存在し、彼らが「見えざる大学」(Invisible Colleges)[Crane72][Burt 82]を形成していることを確認した。見えざる大学とは、大学や研究所などの制度的な組織に限定されることなく、空間的には拡散しながらも高い生産性を上げ続け、最先端の研究を担うことにより、優れた研究者として認知される人々の集団である。

本報告は、学会での研究者の関係形成支援のための情報支援システムを開発する一連の試み[西村 03]の一部として位置づけられる。関係形成支援ツールは、その無限の可能性とともに常に危惧を伴っている。その危惧とは、潜在的な構造を顕在化させることに対する躊躇である。ある人間を理解するためには、その人本人のみを見つめているよりもむしろ、その人がいかなる人々とかかわりをもっているのかを理解するほうが良い場合がある。とはいえ、大多数の夫婦に典型的なように、個人の持つ人間関係の詳細すべてを相互に知ることが、当事者同士の関係にプラスになるとは限らない場合もある。このことは、潜在的な関係構造を顕在化させるという作業を行う者は常に心しておかねばならない。

およそ学問を職業にする立場にあれば、自らの領域において、見えざる大学の成員や、その階層構造を明らかにすることには少なからぬ抵抗を覚えるかもしれない。だが、現実には、誰が見えざる大学の成員であるのかは、各学会内部ではほぼ周知の事実であることが多いにもかかわらず、専門領域外の間には、ほとんどその実態がわからない。したがって、新しくその分野に入ろうとする者あるいは海外の同領域の研究者にとっては、ある特定の国で誰が見えざる大学を構成しているのかの情報は貴重なのである。もちろん、いかなるネットワークであれ、内部と外部を分ける境界は常に曖昧であり、本報告でも筆者らの判断にそった境界の同定がされている。

本報告では、2003 年から 2005 年にかけての日本の人工知能研究者の協働関係の構造を Web マイニング技術により抽出し、(1)人工知能研究者がもつネットワークの構造特性とその変化を明らかにし、(2)人工知能研究者のネットワーク内での位置特性の変化を検討し、一時点のネットワーク特性からその後のネットワーク上の位置特性の予測が可能であることを検証する。

2. データと抽出方法

分析対象は、松尾ら[松尾 03]が開発したウェブマイニングの手法により、収集・構築された「人間関係ネットワーク」である。この人間関係ネットワークは、人工知能学会に属する研究者のうち、2003 年から 2005 年にかけて継続的に活動していることが明らかな、90 人の研究者をコアメンバーとして抽出し、このコアメンバー間の協働関係を web 上の共起関係から特定したものである。

データ抽出手順は、松尾[松尾 05]に従い、所属組織などで同定制限をつけた研究者の氏名を用い、単独の研究者名をキーワードにして検索エンジンが抽出するページ数と、二人の研究者のペアをキーワードにして検索エンジンが抽出されるページ数を特定する。これらの共起数に一連の制約を加え、研究者のダイアドごとにシン普森係数を算出する。シン普森係数が、一定の閾値以上の場合に、研究者間に関係があるとみなす。データ抽出の詳細については、[松尾 04]を参照。この手法により、2003 年、2004 年、2005 年の 3 年間、それぞれについて、研究者間の関係を複数の次元でとらえている。

本報告においては、研究者が相互にとりもつ協働関係を、Web 上の氏名の共起関係をもとに、共起するページの特徴を加味することにより、(1)共著論文執筆、(2)共同研究室所属、(3)共同プロジェクト参加の 3 種のネットワークが抽出されている。2003 年から 2005 年の 3 年間のデータを扱っているため、検索エンジンの変化などにもとない、抽出されるネットワークデータの性質に、ややゆれがあることは否めない。

3. 進展する研究者の相互関係

3.1 ネットワーク密度の変化

まずは、2003 年から 2005 年にかけて、ネットワークの密度がどのように変化したのかを確認する。Web 上での共起回数は、検索エンジンのそのときのアルゴリズム、所属の用い方(検索クエリーへの展開の仕方)などによって異なってしまう。たとえば、

連絡先：安田雪，東京大学大学院経済学研究科，
yasuda@mmrc.e-u.tokyo.ac.jp

全体の共起件数の平均も年ごとに変わり、その中での90人の共起件数の平均も変化している(表1参照)。そこで、Simpson係数の閾値を変えることにより正規化を行う。2003年は0.45、2004年は $0.45 \times 1.17 / 2.7 = 0.19$ 、2005年は $0.45 \times 1.73 / 2.76 = 0.28$ と設定した。全体の平均のヒット件数が一定であるとした場合に、注目している90人のヒット件数がどうなるのかを把握するためである。

表1 研究者全体とコアメンバー90人のWWW共起関係のシンプソン係数および共起関係の平均

	全体のシンプソン係数の平均	90人のシンプソン係数の平均	全体の共起件数の平均	90人の共起件数の平均
2003年	0.0665	0.0908	2.76	3.59
2004年	0.00964	0.0471	1.17	2.31
2005年	0.0347	0.164	1.73	4.10

以上の正規化を行った後に、コアメンバー90人が形成するネットワークの密度を算出した(表2)。

表2 コアメンバー90人のWWW共起のネットワークの紐帯び数と密度の変化

	エッジ数	密度
2003年	189	0.0472
2004年	419	0.105
2005年	850	0.212

表2はコアメンバーが形成するネットワークの密度を示している。研究者の相互関係の密度は毎年約2倍ほど増加しており、コアメンバーの相互の関連性が高まっていることが認められる。表1からも、全体の共起件数に比べて90人の共起件数の増加傾向が確認できる。今回特定している90人のコアメンバーについて、Web上のデータで関連づけていくと、研究者の相互の結びつきが強くなっている様子をうかがうことができる。

3.2 AI研究者ネットワークのべき分布

Webから抽出したデータにおいて研究者間の結びつきが増加しているのであれば、具体的にはいかなる部分で関係が強化されているのであろうか。結びつきの増加には二種類の可能性がある。ひとつは、全体的にコアメンバーが少しずつ紐帯び数を増やしている場合であり、もうひとつは多数のエッジをもち関係の大部分を占める言わば「勝組」と、ほとんどエッジを持たず関係形成に関与しない「負組」との分化が生じる場合である。

エッジの分布状態は、研究者コアメンバー間での競争状態を示すものであるから、きわめて重要な意味を持つ。それは、研究者ネットワークの平等性・不平等性を示唆する。

共著関係、同研究室関係、同プロジェクト関係、同発表関係、Simpson係数による関係のそれぞれについて、度数分布を両対数グラフでプロットした(図一部割愛)。共著(図1参照)、同研究室(図2参照)が比較的きれいな分布をしているのに対し、プロジェクト、同発表はあまり規則性が見えない。両対数で直線というよりは、やや頭打ちになっている形であるが、両対数グラフ上で線形近似してみると、べき分布の乗数 γ は共著の場合、1.3~1.7程度であった。同研究室関係は1.7~0.85という範囲である。べき係数にはゆれがみられ、規則的とは言いがたい。

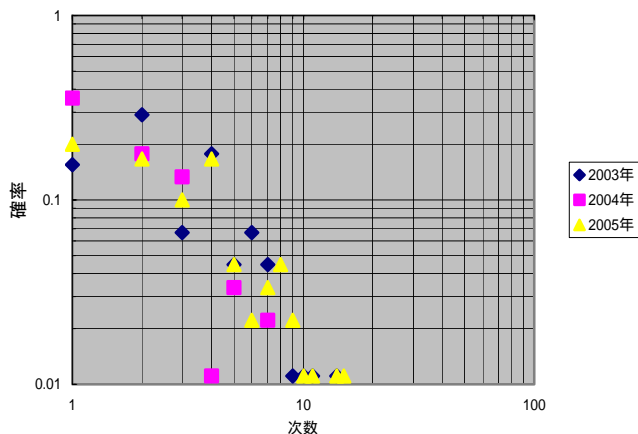


図1 共著関係の度数分布

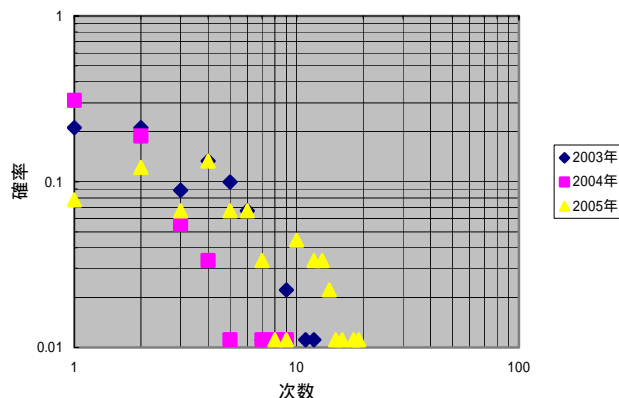


図2 同研究室関係の度数分布

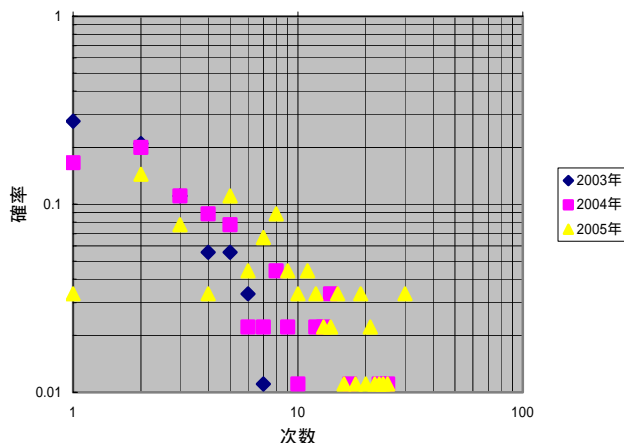


図3 Simpson係数による関係の度数分布

一方、Simpson係数で計測したWeb上の氏名の共起回数では、両対数グラフで比較的直線状の度数分布が得られている(図3参照)。また、2003年から2005年まで、 γ が1.50、1.01、0.62と小さくなっている。すなわち、多くのエッジを持つノードが出現しているということである。

これはWeb上では膨大な研究者の個別データが点在しているのだが、研究者を共起関係によって関連づけていくと、大多数の研究者は少数のエッジを保持し、わずかな研究者が多数のエッジを持っていることを示唆している。Simpson係数で計測した関連付けも、また多くのネットワークのように、べき法則にしたがい、厳しい競争と、そしておそらくその勝敗の決定過程を示しているであろう。

表3 媒介中心性と固有ベクトル中心性の年ごとの相関

		媒介性03年	媒介性04年	媒介性05年	固有ベクトル03年	固有ベクトル04年	固有ベクトル05年	degree
媒介性03年	Pearson の相関係数	1	.341**	.152	.320**	.421**	.465**	.743**
	有意確率 (両側)	.	.001	.153	.002	.000	.000	.000
	N	90	90	90	90	90	90	90
媒介性04年	Pearson の相関係数	.341**	1	.171	.174	.671**	.427**	.271**
	有意確率 (両側)	.001	.	.108	.100	.000	.000	.010
	N	90	90	90	90	90	90	90
媒介性05年	Pearson の相関係数	.152	.171	1	.063	.240*	.703**	.118
	有意確率 (両側)	.153	.108	.	.552	.023	.000	.268
	N	90	90	90	90	90	90	90
固有ベクトル03年	Pearson の相関係数	.320**	.174	.063	1	.494**	.187	.582*
	有意確率 (両側)	.002	.100	.552	.	.000	.078	.000
	N	90	90	90	90	90	90	90
固有ベクトル04年	Pearson の相関係数	.421**	.671**	.240*	.494**	1	.627**	.377**
	有意確率 (両側)	.000	.000	.023	.000	.	.000	.000
	N	90	90	90	90	90	90	90
固有ベクトル05年	Pearson の相関係数	.465**	.427**	.703**	.187	.627**	1	.351**
	有意確率 (両側)	.000	.000	.000	.078	.000	.	.001
	N	90	90	90	90	90	90	90
degree	Pearson の相関係数	.743**	.271**	.118	.582*	.377**	.351**	1
	有意確率 (両側)	.000	.010	.268	.000	.000	.001	.
	N	90	90	90	90	90	90	90

**：相関係数は 1% 水準で有意 (両側) です。

*：相関係数は 5% 水準で有意 (両側) です。

とはいえ、べき係数にはゆれがあり、必ずしもすべてのネットワークで適合性が高いわけではない。共著関係、研究室関係は、参入障壁も高く時間的継続性を必要とする点で、比較的頑健な関係である。一方、プロジェクトや発表関係は参入も容易であり緩やかな関係である。とはいえ、この4種の関係は、現実社会における研究者の具体的行為(書く、入学する、話す等)を必要とする。一方、Web上の氏名の共起回数は、上記の四種のネットワークとは異なり、現実社会における行為を超えた関係構造をも抽出している。厳密な行為の存在を成立要件に必要とするというネットワークの特性の差異が、べき係数の差異に影響を及ぼしているのであろうか。この点は今なお、曖昧であり、今後、より詳細な検討を必要とする。

3.3 3年間の中心性と全体構造の変化

「人間関係ネットワーク」では、共著ネットワーク、共同研究室所属、共同プロジェクトの3種類のネットワークが抽出されている。この3種のネットワークは、いずれも同じ90人のメンバーとその相互関係によって形成されている。

2004年には、共同プロジェクトネットワークの密度が最も高く、共著ネットワークの密度が最も低いこと、共同プロジェクトは複数参加が比較的容易であるため、多数の重層的な関係が共同プロジェクトネットワーク内には存在し、共著ネットワークよりも多くのクリークが内包されることが確認されている[安田 04]。したがって、共同プロジェクトネットワークや、共同研究室所属のネットワークの柔軟性、共著ネットワークの成立の困難性や頑健性を考えるならば、研究者間の関係性の本質はやはり共著ネットワークにこそ現れるものと考えられる。

そこで、2003年から2005年のコアメンバーの共著ネットワークについて、他者との間接的結合関係の多寡を考慮する固有ベクトルによる中心性を算出したところ、3年間を通じて継続的に上位20位内に存在するのは、ごく少数のメンバーであった。ネットワークの集中度は2005年には54.82%であった。ここでは、コアメンバー90人だけに特定して中心性を算出しているため、コアメンバー90人以外との関係を多数持つような研究者の中心性は低くなるため、より大きなネットワークにおける分析[安田 04][友部 05]とは、中心性の高い研究者が異なる結果になる。

今回はコアメンバーに限定した上で、さらに多数のエッジをもつ研究者とかかわりをもつ人物の中心性が高くなるような指標を用いているため、同定した境界範囲内での関係構造が浮かび上がっていることに留意してほしい。

関係パターンの類似度を測る QAP 相関を算出すると、共著ネットワークの2003年と2004年では、0.452、2003年と2005年は0.349、2004年と2005年は0.329である。したがって、共著ネットワークの紐帯分布の類似度は、2003年と2004年のほうが、2004年と2005年よりも高くなっており、2004年から2005年にかけての変化のほうが、その前の一年よりも大きいことになる。

QAP 相関の係数はいずれも、0.3を上回っており、関係類似の水準自体は低くは無いことがわかる。また、上位20名の変化をみても、QAP 相関を見ても研究者の共著ネットワークにおいては、03年から04年よりも04年から05年における変化が著しいことが認められる。

次に、Simpson 係数により Web 上での氏名の共起関係が織り成すネットワークを見てみよう。表3に、年ごとの媒介中心性と固有ベクトル中心性の相関図を示す。03年の媒介中心性と、04年、05年の固有ベクトル中心性は相関が高い。04年の媒介中心性も、05年の固有ベクトル中心性と相関が高い。さらに、03年の媒介性と03年の固有ベクトルの相関.320よりも、03年媒介性と04年固有ベクトルの相関係数.421のほうが大きくなっていることが分かる。また、03年媒介性と05年固有ベクトルの相関係数0.465はさらに大きい。

媒介性という指標は、異なるクリーク(グループ)を結びつける機能の程度を測定しているため、一連の相関係数は、前年、あるいは前々年の媒介性の高さが、その翌(および翌々)年の固有ベクトルの高さに関連があることを示している。同一指標の経年変化で一定の相関が見られるのは当然だとしても、前年の媒介性が次年度の固有ベクトルによる中心性をあげるの、注目に値する。なぜならば、Simpson 係数の共起ネットワークにおいては、早い時点で異なるグループの関係を媒介する位置にいた者が、その後、関係している他者が求められる程度までを考慮にいれた中心性が高くなるのである。つまり、Simpson 係数の共起ネットワークにおいて媒介性の高い者は、次年度、次々年

係数^a

モデル	非標準化係数		標準化係数		有意確率
	B	標準誤差	ベータ	t	
1 (定数) 媒介性03年	6.523E-02 2.746E-04	.007 .000	.465	9.150 4.930	.000 .000

a. 従属変数: 固有ベクトル05年

モデル集計

モデル	R	R2 乗	調整済み R2 乗	推定値の標準誤差
1	.465 ^a	.216	.208	.06044058

a. 予測値: (定数)、媒介性03年。

表 4 2003 年媒介性と 2005 年固有ベクトル中心性の線形回帰

度の Simpson 係数の共起ネットワークにおいて、固有ベクトルによる中心性が高いのである。

この結果は、異なるグループや異質な人々との関係を結び、ネットワークにおける優位性を導くという、Structural Holes[Burt 02]の概念)と整合的である。早期に異質なグループの他者を結ぶ者は、後になって、他者と関係を多数取り結び人々と連結しつつ中心性の高い位置を占めるようになることがうかがえる。

なお、03 年の媒介性から 05 年の固有ベクトルにもとづく中心性について、線形回帰を行った結果は表 4 のとおりである。一時点の媒介性によって、後日の固有ベクトルによる中心性の予測が可能になる。

回帰分析の結果は、Web上の情報をもとに関連づけを行って作成したネットワークにおける位置の特性から、将来の共著関係における研究者の位置特性が、ある程度、予測できることを示している。すなわち、一見、規則性も多様な立場に独立して点在する、Webページの情報を、相互に関連づけてネットワーク化してみると、後の時点でのWebページに登場する行為者(=研究者)の関係を一定程度、予測することが可能になるのである。相互に関連や脈絡を持たず、独立して存在するWebページの作成者たちの個別webページから松尾らのマイニングの技術が抽出するネットワークは、偶然性に左右されそうであるにもかかわらず、後の時点における、(たとえば共著関係のごとき)、頑健なネットワークの構成要素とその位置の予測を可能とするのである。松尾らのWeb情報の関連づけは、実は、その後の現実の協働(コラボレーション)ネットワークの前身を、恐ろしいほどに的確に抽出している。

4. まとめ

2003 年から 2005 年の間に存在する Web 上の情報から、わが国の人工知能研究者 90 名をコメンターとして特定し、彼らが形成する重層的なネットワークを抽出して分析したところ、(1) コメンターが形成するネットワークの密度が一年ごとにほぼ倍になりつつあること、(2)一部のネットワークにおいては、研究者の保持するエッジ数がべき分布にしたがうこと、すなわちネットワーク内部で競争原理が働き不平等が生じつつあること、(3)3 年間で研究者の中心性には移動がみられ、それも直近になるほど変化が著しいこと、さらには(4)Web マイニング技術で抽出した Web 上での共起関係ネットワークにおける研究者の媒介性により、その後の時点での研究者の共著関係を、ある程度予測することが確認された。

研究者に限らず、個人が形成するネットワークの特徴とその内部における個別の位置を明示化することには、長所、短所がともにつきまとう。本報告も範囲の同定法や、係数の閾値の設定など筆者らの判断により多くの情報がフィルタリングされている。研究者や研究に対する評価は、この制約の範囲におけるものである。実際の現実世界に対する人間の認知に限界があるのと同様に、Web のマイニング技術とネットワーク分析とが抽出、記述できる構造にも限界はある。

だが、AI 研究者の様々なネットワークにおいても、おそらく一部では働いているであろう、べき法則、そして入れ替わる中心性と、年ごとに大きく変化するネットワーク集中度は、最先端の研究領域における競争の厳しさや熾烈さを明らかに示している。研究者のネットワークにおける競争の厳しさは、学問の発展と健全性を示しており、人工知能学会における競争は、今後もさらにネットワークの複雑さを増加させつつ変化していくであろう。

Web の情報は個々のページの作成者が個別に独立に、散在しながら発信しているにもかかわらず、それらに関連づけ、総体としてできるネットワークの構造と登場する行為者(actor)の位置を時系列に分析すると、抽出されたネットワークの特性は、後時点において行為者がリアルに形成する関係と密接にかかわっていることが明らかになった。筆者らが強調したいのは、Web上に点在する関係情報は、後日の現実的な行為とその関係を予測する指標になりえるという事である。

個別の独立したWebページ作成情報が、結果として見えざる大学の全体像とその構成を予測しうる情報となっている。これはWebの情報発信機能の再評価にもつながる重要な特性といえよう。個別最適が必ずしも全体最適につながらないことは、パレートが久しく以前に指摘しているが、個別情報であれ、その蓄積が関連づけによって、全体構造とその後の変化・発展の予測を可能にすることが本分析によって認められた。

なお、個別のネットワークにおける個々の研究者のネットワークポジションの移動等は、報告当日に述べる。

参考文献

- [Burt 82] Burt, Ronald S. Toward a Structural Theory of Action, Academic Press, 1982 .
- [Burt 02] Burt, Ronald S. The Social Capital of Structural Holes, pp. 148-192 in The New Economic Sociology, Russell Sage Foundation, 2002
- [Crane72] Crane, Diana: Invisible Colleges, University of Chicago Press, 1972 .
- [友部 05] 友部博教、松尾豊、武田英明、安田雪、橋田浩一、石塚満: Semantic Web のための人の社会ネットワーク抽出と利用、情報処理学会論文誌、2005(印刷中)
- [松尾 03] 松尾豊、友部博教、橋田浩一、石塚満: Web から人間関係ネットワークの抽出と情報支援、第 17 回人工知能学会全国大会、1F1-02、2003 .
- [松尾 05] 松尾豊、友部博教、橋田浩一、中島秀之、石塚満: Web 上の情報から人間関係ネットワークの抽出、人工知能学会論文誌、Vol. No.1E, pp.46-56、2005
- [西村 03] 西村拓一、濱崎雅弘、松尾豊、大向一輝、友部博教、武田英明: 2003 年人工知能学会全国大会支援統合システム、人工知能学会誌解説記事、Vol.19 No.2.43-51、2004 .
- [安田 01] 安田雪: 「実践ネットワーク分析」、新曜社、2001 .
- [安田 04] 安田雪: 「人脈づくりの科学」、日本経済新聞社、