

人工知能学会におけるネットワーク構造と変化

Analysis on Social Network Structure and Dynamism in AI community

安田 雪^{*1*4}
Yuki Yasuda

松尾 豊^{*2*1*4}
Yutaka Matsuo

武田 英明^{*2*1}
Hideaki Takeda

^{*1} 東京大学 ^{*2} 産業技術総合研究所 ^{*3} 国立情報学研究所 ^{*4} GBRC 社会ネットワーク研究所
University of Tokyo AIST NII Global Business Research Center

This paper investigates the social network structure and its dynamism in the artificial intelligence research community in Japan. The network data through four years is obtained by Web mining techniques. We identify 90 persons as a core member among the AI community, and show how the network of core members evolves. Our analysis supports the importance of academic conferences and also the event space information support in the academic context.

1. はじめに

本稿では、人工知能分野における研究者について、Web マイニング技術によって得られた研究者ネットワークの分析を行う。松尾らの方法[松尾 05]を用い、Web 上の氏名の共起関係による研究者ネットワークを抽出する。以前の報告[安田 04]では、研究者の集団が「見えざる大学」(invisible colleges)[Burt 82]を形成していること、Web マイニングで得られたネットワークが研究者のパフォーマンス(論文数や引用数)と相関があることを確認した。また、昨年の報告[安田 05]では、先端的な研究を担うコアの研究者グループ内では紐帯の数の偏りが年々大きくなり競争が起こっていること、また媒介中心性が次年度以降の固有ベクトル中心性と相関が高く、複数の島をつなぐ研究者が中心的な役割を担うようになる傾向があることを述べた。

本報告は、2003 年から継続的に行われている「イベント空間情報支援プロジェクト」の一環として、研究者の創造的な関係形成支援を目的とする研究のひとつと位置づけられる。本稿では、JSAI2003 から JSAI2006 まで、計 4 年間のネットワークデータを比較することで、人工知能分野における研究者コミュニティの変化の様子を分析する。特に、人工知能分野において、コアのメンバーのネットワークの変化の様子、周辺との関係の変化について調べ、研究者のつながりが知識の構造化に寄与していることを論じる。

2. データと抽出方法

分析対象となるのは、松尾ら[松尾 05]が開発した Web マイニングの手法により収集・構築された研究者のネットワークである。JSAI2003 では 298 人、2004 では 540 人、2005 では 585 人、2006 では 639 人を対象とし、ネットワークが抽出された。2004-2006 年においては、学会発表を行った全著者・共著者が対象である。2003 年に関しては、著者・共著者の約半数である。この中で、データの信頼度を考慮して、Web の検索エンジンでの氏名のヒット件数が 20 件以下の人を除外すると、それぞれ 237 人、428 人、268 人、400 人となる。

2005 年時点での分析[安田 05]では、JSAI2003 から 2005 まで 3 年間継続してデータの取得対象となった 90 人を、人工知能分野のコアメンバーと位置づけ、その分析を行った。JSAI2006 では、この 90 人のうち 89 人までが著者・共著者であったため、本稿ではこの 90 人(今年については 89 人)につい

て継続して分析を行う。

データの抽出方法は、所属組織で同定制限をつけた研究者の氏名を用い、単独の研究者名をクエリー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クエリーにして検索エンジンが抽出するページ数により計算する。そのうえで、シンプソン係数を算出し、これが一定の閾値以上の場合に研究者間に関係が存在するとみなす。また、検索されたページの内容を調べることで、(1)共著論文執筆、(2)同じ研究室に所属、(3)同じプロジェクトに参加という3種のネットワークが抽出する。

3. コアの研究者の動向

3.1 各年ごとのネットワーク

表1は、コアの 90 人の研究者に対して、シンプソン係数、共起件数の変化を示したものである。全体の共起件数は、年ごとに変化しており、それにともなってシンプソン係数の平均も変化している。コアの 90 人に限っては、共起件数、シンプソン係数ともに増加傾向にあり、特に 2006 年には共起件数が非常に大きく増加していることが分かる。

検索エンジンのヒット件数は、検索エンジン自体の仕様変更等によって揺れがあるため、前処理を行っている。具体的には、全体の共起件数が各年ごとに一定になるように正規化を行った上で、シンプソン係数の閾値を設定した。2003 年から 2006 年まで、それぞれ 0.45, 0.19, 0.28, 0.39 としている。つまり、全体のヒット件数が一定であると仮定した上で、相対的な紐帯の強さを測ることになる。表 2 は、各年ごとの密度を示している。密度は、2003 年～2005 年までは、1 年ごとに倍になっていたが、今年はほぼ昨年と同程度である。

これをネットワークに可視化したものが図 1-4 である。最初はばらばらであった行為者が、徐々に紐帯を作り上げ、最終的に密なネットワークを構成していく過程が理解できる。

表 1: シンプソン係数、共起件数の変化

	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	0.0665	0.0908	2.76	3.59
2004	0.00964	0.0471	1.17	2.31
2005	0.0347	0.164	1.73	4.10
2006	0.0352	0.175	2.38	20.22

連絡先: 1,4,安田雪 yyasuda@gbrc.jp, 東京大学 COE ものづくり経営研究センター、〒113-0033 文京区本郷 3-34-3-8F

表 2: ネットワークの密度の変化

	ノード数	エッジ数	密度
2003	90	85	0.0212
2004	90	204	0.0509
2005	90	425	0.106
2006	89	363	0.093

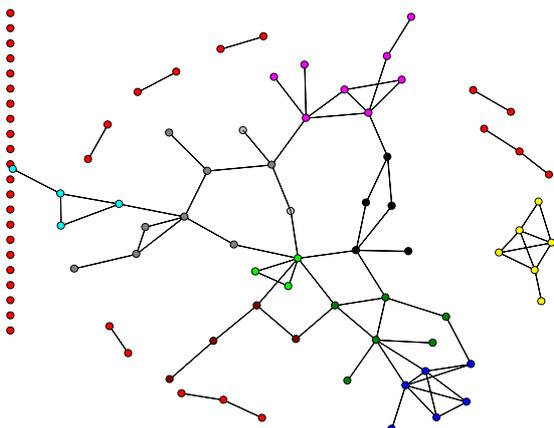


図 1: 2003 年のネットワーク

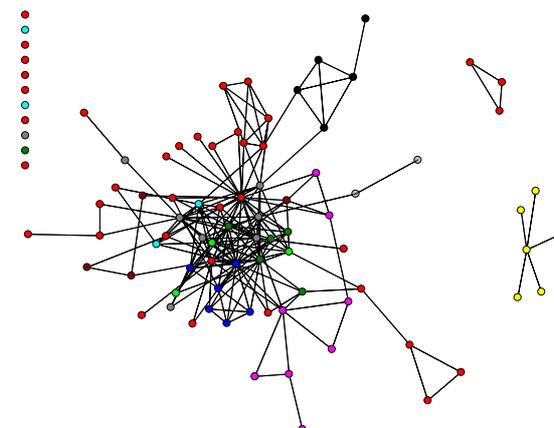


図 2: 2004 年のネットワーク

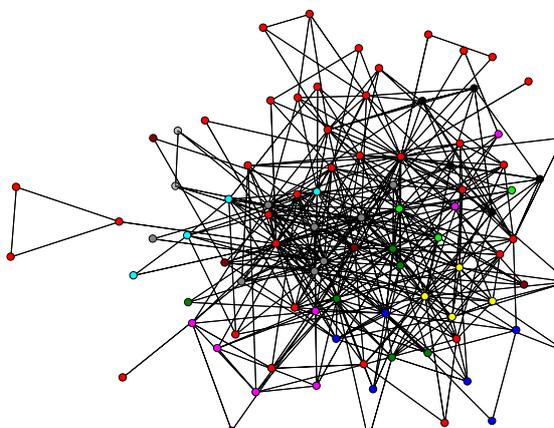


図 3: 2005 年のネットワーク

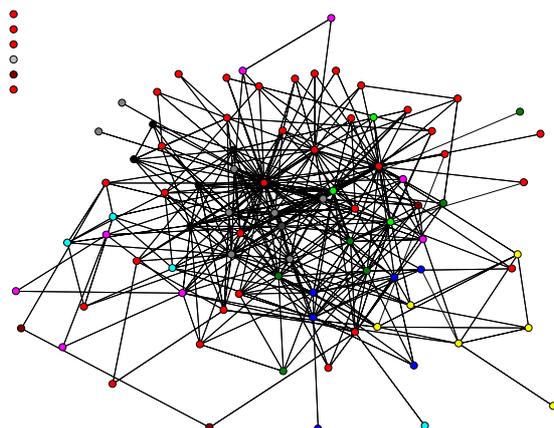


図 4: 2006 年のネットワーク

2003 年には孤立点も多く、分断されたクリークが多数存在しているが、2004 年、2005 年と最大グループが徐々に拡大し、孤立点も減少していく。2003 年にはダイアドが複数存在しているが、時間の経過とともに二者関係がトライアドに成長したり、最大グループに接続していくことも確認できる。

3.2 研究者の結びつきの流れ

毎年、紐帯の数が急激に増加し、ネットワークが密になっていく様子は、実際の学会活動をしている立場からは多少、違和感がある。この理由は、次のように説明できるだろう。本稿でコアの研究者と定義したのは JSAI2003 から 2006 までの 4 年間に論文発表を継続して行った研究者である。したがって、例えば、2000 年から 2004 年まで継続して発表した研究者や、2004 年から継続して発表している研究者は含まれない。従って、たまたま 2003 年からの 4 年間、継続した研究者に限定していることになり、この前提条件によって紐帯が急激に張られているように見える。つまり、この期間にコミュニティに継続参加した人という条件が、大きなバイアスを与えている。だが、少なくとも、直近の 4 年間の学会活動への継続的な参加は、研究者相互の紐帯の形成に重要な影響を与え、相互のインタラクションを促進することは確実である。

ここで観察された現象を、より大きく捉えようと、学会活動の全体像が浮かび上がる。外部から新しく入ってくる研究者や学生は、学会への参加を通して紐帯を増加させる。中には、他の学会に活動の場を移す者もいるだろう。孤立点、二者関係、トライアドなどの小グループ形成あるいは、既存のコア中のコアのネットワークに連結をし、コミュニティのコアの一部となっていく。

ここで、学会の全体像を捉えるために、研究者がどのように紐帯を作り、定着していくのかといった面を観察するために、研究者の関係性と研究トピックとの対応を見てみよう。

図 5-7 は、各年のネットワークにおいて、研究者の研究トピックを領域ごとに示したものである。2003 年にはオントロジー、ユビキタス、エージェントなど、典型的な人工知能のトピックがいくつか関連しあっていることが分かる。2004 年になると、いくつかの新しいクラスターが現れている。

例えば、「身体性」のクラスターが現れているが、これは 2003 年度から始まった近未来チャレンジのテーマ「身体知の解明を目指して」が 1 年を経て、参加者同士の関連が徐々に深まっている様子を示している。2005 年には、各トピックの融合がさらに進んでいる。新しく「物語」のクラスターが現れており、これは 2004 年度のオーガナイズドセッション「ことば--コンピュータ--コミュニケーション」により関係者の関連が強まったためであろう。オントロジー、Web、コミュニティなどのトピックは、完全に融合している。2006 年でも同様である。

人工知能学会全国大会では、近未来チャレンジやオーガナイズドセッションといった新しいトピックを集中して発表する場が用意されている。こういった場をきっかけに研究者間での結びつきが強まり、次第に新しいトピックの研究者達が、既存の研究トピックの研究者との関連を形成していく。新しい研究トピックの報告の場を設定する重要性が確認できる。

この傾向を別の角度から見てみよう。JSAI2005 では、研究者の氏名がどのようなキーワードと共起するかによって、研究者を人工知能の下位分野に大まかに分けた[Matsuo06]。これらの下位分野に該当するコアメンバーが、どのように紐帯を増やしているかを示したものが表 3 である。各下位分野ともに増加傾向にあるが、紐帯数が一気に増える年もあれば、多少、減る年もある。紐帯は必ずしも順調に増えていくわけではなく、研究者が学会にただ参加すれば紐帯が増えるというものではないことが分かる。

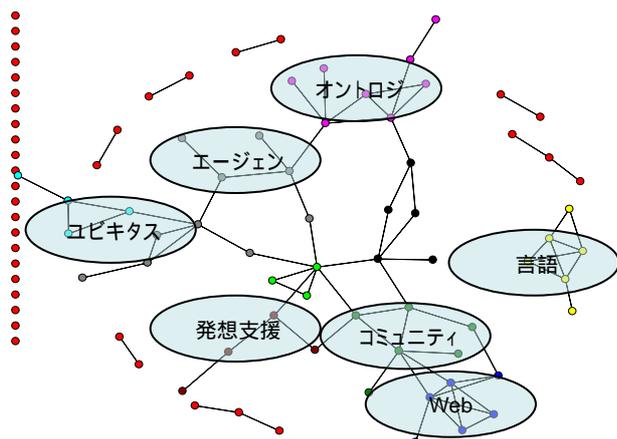


図 5：2003 年における研究トピック

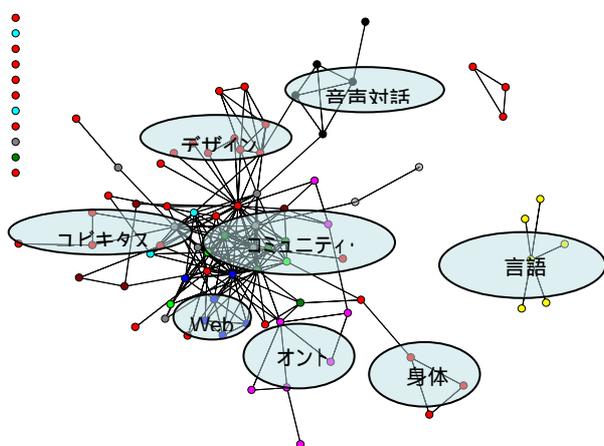


図 6：2004 年における研究トピック

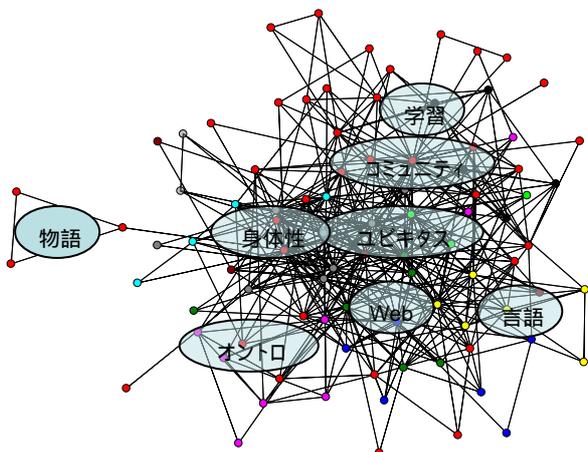


図 7：2005 年における研究トピック

とりわけ、2003 年で分断されている関係にある、コアメンバーとそれ以外の参加者はどのような関係を形成しているのだろうか。表 4 では、コアメンバー内での紐帯だけでなく、周辺のメンバーとどのように紐帯が形成されているかを示している。(シンプソン係数のネットワークで、各メンバーの次数平均を示している。)コアメンバーは、コア内、周辺ともに紐帯を増やしている一方で、周辺メンバーは 2003 年時点ではコアメンバーへの紐帯の割合の方がやや高いが、徐々に他の周辺への紐帯の割合が高くなっている。つまり、新しく紐帯を獲得する研究者、ひいては他との関連が構築される研究トピックは、年とともに徐々に移り変わっているのである。

最後に、表4では、参加者全員のネットワークにおける固有ベクトル中心性の上位が、コアメンバーにどのくらい含まれるかを示したものである。4 年間継続参加しているコアメンバーが、必ずしも参加者全体のネットワークで中心的な位置を占めているわけではないことに注目してほしい。中心的なメンバーの割合も年々変化している。共起件数によるネットワークの方が、比較的著名な研究者の紐帯が多くなる傾向があるので、より安定しているものの、新しい研究トピックが次々に生まれ、その中で中心性を保つのは容易ではない。

このように、人工知能分野では、新しいトピックや研究者が参入し、研究者グループをつなぐ紐帯が作られ、それが徐々に密になって中心的な研究者集団を形成すると同時に、常に新しいトピックが現れ、中心も移り変わっていくことが理解できる。

ここで、本分析の対象データである Web 上の共起関係の本質について考察する必要がある。まず、Web 上での氏名の共には、さまざまな理由が考えられる。例えば、同じ学会への参加や同じ学会誌への掲載、プロジェクトなどにおける協働関係、関連した研究としての共引用、関連した研究者としての並記などである。研究トピックやキーワードなどと同様に、Web ページの作成者による、共通関心に導かれて、研究者の氏名は同一ページ上に前後して記述される。

共起する氏名は、必ずしも似通った評価を与えられるとは限らず、一方に対しては肯定的、他方に対しては否定的な評価やコメントが記述されていることもありうる。もちろん、Web ページの上では、双方をともに肯定的、あるいは否定的に評価することもあろう。また、一見、両者のあいだに、直接的な関係がないように、箇条書きのように併記がなされている場合もあるだろう。

以上のような、Web 上の共起関係の本質は、文脈の共通性である。研究者の研究内容やその背景、目的、それが与える意義など、さまざまな要因が似通っていることが、Web での共起の原因となる。氏名を始め、Web 上で共起する物事には、その両者をともに位置づける文脈が背後には必ず存在する。

この、文脈の共有、あるいは文脈の共通性を、我々は、Web 上の共起関係の本質と考える。

この視点から、学会内でのネットワークの変化を考えると、新しい研究トピックが立ち現れ、徐々に他の研究とつながっていくのは、すなわち文脈の共有が進むためであろう。文脈を共有することで、互いの研究が的確に位置づけられ、それが知識の体系化、構造化につながっていく。

表 3: 分野と紐帯数の変化

	人数	2003	2004	2005	2006
Webと言語	23	5.9	7.4	21.4	25.7
構造と分析	6	7.9	3.9	22.8	31.3
AI応用	18	8.7	10.1	25.5	31.3
ロボット・センサ処理・身	5	5.7	4.2	14.6	8.5

体性					
インタラクシ ョンと知識共 有	5	11.6	10.9	40.8	24.1
推論とセマ ンティクス	13	8.3	7.5	27.6	35.0
コピキタス・ エージェント	15	7.5	5.1	24.6	30.3

表 4: コアと周辺との紐帯

	コア		周辺	
	コアへの 紐帯	周辺への 紐帯	コアへの 紐帯	周辺への 紐帯
2003	1.9	10.2	4.4	3.9
2004	4.5	14.9	3.0	3.6
2005	9.4	53.1	9.6	9.5
2006	7.9	33.6	5.5	6.7

表4: コアメンバーのうち、全体の中心性 100 位以内
に入っている人数

	共起のネットワークにお ける固有ベクトル中心性	シンブソン係数のネットワ ークにおける固有ベクト ル中心性
2003	43	37
2004	49	53
2005	23	17
2006	41	36

4. おわりに

本稿では、4年間に渡るコア研究者のネットワークの変化を調べ、その中で研究トピックの結びつきの変化に着目した。孤立点、ダイアドなどの少数の関係保持状況から、巨大グループへの結合、あるいは局所的な関係形成を経て、研究者が相互に連結を深めていく過程が確認できた。

この過程は同時に、知識の体系化の過程でもあり、比較的、新しい研究トピックが生じやすく、またトピック相互の位置づけに揺れがみられる人工知能という学問の特徴を反映しているのであろう。知識が構造化・体系化されていく上で、異なる視点のぶつかり合いによって、結果的に文脈が共有され、学術的に互いの位置づけを決定していく過程が重要である。

この意味では、人工知能学会における近未来チャレンジやオーガナイズドセッションの意義は大きいと考えられるし、またイベント空間情報支援により、異なるクラスターの研究者同士が交流することは、学会全体の知識の構造化の上で有益であろう。

知識を構造化するためには、異なる分野に属する研究者の相互交流が不可欠であり、この交流によってはじめて、新概念や知識を、既存の学術的文脈のなかに位置づけ、体系化することが可能になるのである。まさしく、科学は「巨人の肩の上に」[Merton, 1979]という、過去の学術的蓄積に対して敬意を払いつつ連携するという謙虚さの必要性が示唆するものである。

参考文献

- [安田 04] 安田雪: イベント空間情報支援システムが抽出するヒューマン・ネットワークの構造、第 18 回人工知能学会全国大会、2004
- [安田 05] 安田雪、松尾豊: 人工知能学会における研究者ネットワークの分析、第 19 回人工知能学会全国大会、2005
- [松尾 05] 松尾豊、友部博教、橋田浩一、中島秀之、石塚満: Web 上の情報からの人間関係ネットワークの抽出、人工知能学会論文誌、Vol.20, No.1E, pp.46-56, 2005
- [Burt82] Burt, Ronald S. Toward a Structural Theory of Action, Academic Press, 1982
- [Matsuo06] Y.Matsuo et.al: POLYPHONET: An advanced social network mining from the Web, Proc. WWW2006, 2006
- [Merton,1979] Merton, Robert K. The Sociology of Science University of Chicago Press 1979.