

column 研究現場の生の声(その①)

ネットワーク構造から見た 学術研究トピックの 分析と学際研究の 可能性発見



内田 誠

(東京大学大学院工学系研究科)

学術研究の領域横断化

最近の学術研究のトピックは広範にわたるものとなり、分野の境目は曖昧になっている。「ネットワーク生態学」のような学際領域の出現と発展は従来からの学術研究に新たな展開をもたらす反面、それを効率的に推進するためには、細分化した専門領域を効率的に結集し、連携を促すための指針を得ることが不可欠である。しかし、広大かつ膨大な学術研究の全貌を把握するのは容易ではなく、これまでに情報工学、図書館情報学、知識工学などの分野でその方法論が研究されてきた。中でも、我々はネットワーク分析を用いるアプローチに注目している。これは、学術研究における学術論文の引用関係、研究者による論文共著やプロジェクトへの共所属の関係といった論理的な関係をネットワークとして捉え、複雑ネットワーク科学に基づく分析手法で特徴を分析することで知見を得るというものである¹⁾。本稿では、「学会」に注目し、ネットワークの観点からその研究交流や対象領域を分析した研究事例を通じて、その可能性と課題について紹介したい。

学会から見た研究活動の分布

《学会間ネットワークの構築》

はじめに、研究開発支援総合ディレクトリ ReaD (<http://read.jst.go.jp>) から Web マイニングによって学会のネットワークを構築した。ReaD には研究者氏名をキーとして、所属、研究分野、キーワード、所属学会などが記されている。研究者レコードが約 12 万登録されており、テキスト処理の結果およそ 1 万 8,000 の学会が抽出された。自由記述のため表記揺れ等も多く、ここでは 20 名以上の研究者が所属する 2,263 学会を分析の対象とした。続いて、所属する研究者の重複が多い学会同士は、

互いに近い研究領域を対象としている、という仮説に基づき、学会をノードとしたネットワークを構築する。ここでは、所属する研究者の cosine 類似度を元に、一定の閾値を超える類似度を持つ学会同士にエッジを定義した。

《ネットワーク構造からのクラスタリング》

ネットワークから互いに関係が密な部分集合を見つける問題はコミュニティ抽出問題として知られており、ネットワーク科学における主要な研究トピックの 1 つである。本稿の事例では、互いに関係が密な学会群は類似の学問分野に属するものである、という仮説に基づいた学会のクラスタリングに用いることができる。そこで、構築した学会間ネットワークに対してコミュニティ抽出アルゴリズム²⁾を適用することで、ネットワークを学会クラスタに分割し、可視化したところ、図-1 の結果を得た。それぞれのクラスタは仮説の通り、類似した研究分野を対象とする学会によって構成されており、表-1 のようにクラスタにラベル付けを行うことができた。

《クラスタを通じた異分野連携の分析》

表-1 は、クラスタに属する学会数、クラスタ内の学会同士のエッジ数も同時に示している。また、表-1 には示していないが、異なるクラスタに属する学会同士をつなぐエッジの数を、クラスタ間エッジとして考えることができる。これらの量は、各クラスタに代表される学術分野の規模や、他分野との連携状況の指標として考えることができる。図-2 では、クラスタを新たにノードとし、ノードの大きさとクラスタに属する学会数を、エッジの濃さ・太さでクラスタ間エッジの数を表現した。ここでは、学会に属する研究者の数や学会間関連度の絶対値を考察していないために定量性の問題は残るが、クラスタ内の学会同士のエッジ数は、当該分野内での研究交流の活発さ、クラスタ間エッジの数は当該分野の異分野連携の活発さの指標であると考えられる。「理

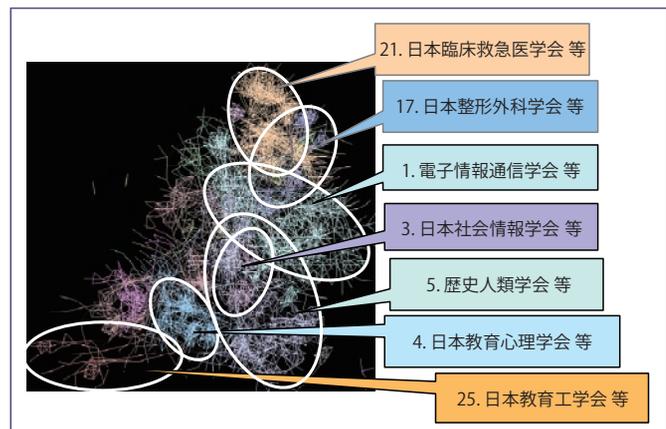


図-1 学会間ネットワークとそのクラスタの可視化図

#	ラベル	学会数	エッジ数
1	理工	413	1,033
3	社会・福祉	72	152
4	心理・経済	365	1,016
5	史学・哲学・宗教	197	650
6	文学	59	181
9	医学(眼科)	14	43
13	映像・写真	6	5
14	農業経済	22	56
16	油脂	10	11
17	医学(整形外科・神経)・薬学	141	450
19	獣医畜産	23	69
21	医学(内科)	322	1,429
22	服飾・繊維	48	70
23	医学(生理学)・細菌	56	89
24	比較文化	51	81
25	文化・言語・教育	218	586
26	図書館	11	27
28	看護	43	142
29	工学(省略)	8	7
30	医学(細胞)	14	43

表-1 学会クラスタのラベルと規模

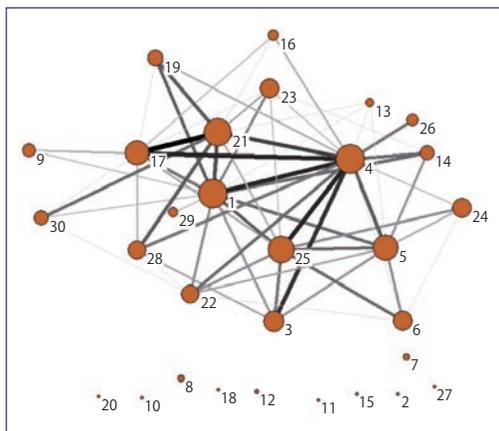


図-2 学会クラスタ相互間の関係

工」(#1),「心理・経済」(#4),「医学(内科)」(#21),「文化・言語・教育」(#25)といった大きな分野が存在するが、他分野との連携の傾向は大きく異なり、たとえば「心理・経済」(#4)クラスタの他とのつながりの数が際だつ。また、「医学(内科)」(#21)クラスタは、学会の数に比較して内部での連携が特に密である、このように、ネットワーク分析によるクラスタリングを用いて粗視化することで、現在の日本の学術研究の分布と、分野間連携の状況を俯瞰し、クラスタの規模やクラスタ内/クラスタ間エッジの数といった定量的な値によって考察、議論することが可能となった。

キーワードによる学際領域の発見

これらの結果は、テキスト・キーワードによるコンテンツ分析と組み合わせることで、より深い知見を得ることができる。そこで、あるクラスタの学会に所属する研究者に付与されている研究キーワードを学会クラスタ内の研究キーワード、複数のクラスタの学会にまたがって所属する研究者に付与されている研究キーワードを、当該クラスタ間のクラスタ間キーワードとして定義、分析した。例として、「理工」(#1)と「医学(整形外科・神経)・薬学」(#17)では、「有機化学」「分子生物学」など、化学・生物系のキーワードがクラスタ間キーワードとなっており、これらのテーマで理工学と医学・薬学の学際領域を形成していることが分かる。一方で、「知能情報学」「機能材料」「機械システム」などは、双方のクラスタ内キーワードでありながらクラスタ間キーワードではない。これらの語に代表されるテーマ、たとえば医用デバイス・

医用マテリアルの領域では、両者の分野間連携を促すことで学際研究の新たな展開の可能性があることが、この分析によって示唆される。

ネットワーク分析の強みと課題

ネットワークに着目した手法は、膨大かつ複雑に入り組む研究者データベースからのデータマイニングにより、学術研究トピックを俯瞰、分析するための強力なツールとなる。一方で、課題も少なくない。そもそも、ネットワークに着目するにあたっては「何のネットワークなのか」を明確に定義する必要がある。本稿ではコミュニティ抽出による汎用的なクラスタリング手法を用いたが、ここで得られるクラスタの意味や、その解釈はネットワークの定義に強く依存するため、ネットワークデータの構築、手法の選定、結果の解釈と評価までを、事例に合わせてその都度検討する必要がある。また、これまでに蓄積されてきたデータベースはネットワーク型データとして利用されることを想定していないものも多く、ネットワーク分析の手法を適用するために前処理を行う必要があるなど、データ収集と構築の難しさという問題も残る。そのため、情報工学の他分野で研究が進むデータの記述モデルや蓄積、構造化手法などを効果的に取り入れ、適切に組み合わせることが、複雑ネットワーク分析を発展させ、新たなツールとして実問題に広く適用していく上で重要であろう。

参考文献

- 1) Newman, M. E. J.: PNAS, 101, pp.5200-5205 (2004).
- 2) Newman, M. E. J.: Phys. Rev. E, 69, 066133 (2004).

(平成 19 年 12 月 12 日受付)

内田 誠(学生会員)
uchida@race.u-tokyo.ac.jp

2006 年東京大学大学院工学系研究科修士課程修了。同大人工工学研究センター学術研究支援員を経て 2006 年 10 月より現所属。複雑ネットワークの構造およびダイナミクスの分析・可視化に関する研究に従事。