(d) 情報

ネットワークモチーフによる共著関係ネットワークの時系列分析法

Analysis of Dynamic Coauthor Networks using Network Motifs

中本　純一　　森 康真　　北上　始

Junichi Nakamoto　　Yasuma Mori　　Hajime Kitakami

広島市立大学大学院情報科学研究科

1. はじめに

現実世界には多くの研究者がおり，様々な分野の研究活動が行われている．これらの研究の中で活発に研究が行われている研究分野はそれだけ社会から注目を浴びているものであり，現実世界に多くの影響をもたらす．注目を浴びている分野についての定量的評価は発表された論文数が挙げられる．ある分野の論文数が増加していればそれだけ注目されているということであり，逆に減少しているということは注目されていなくなってきていることが分かる．多くの研究は単独で行われず，そのほとんどが別の研究者と協力して研究を行っている．研究者の活動を見るに際しては，研究者それぞれの単独的な活動を見るだけでは不十分であり，何らかのつながりの中で研究者間の位置づけを考慮に入れる必要がある．そこで，筆者は”共著”という現象に注目した．ある分野の論文数が増加したということは，その分野について研究する研究者が増加したということである．このとき，増加する以前と比較して共著関係が変化していることが考えられる．

共著関係ネットワーク中の共著関係の変化を観測するためにネットワークモチーフを用いる．ネットワークモチーフはネットワークを特徴づける指標として知られており，任意に選択した分野の論文群の著者間から構成されるネットワークからネットワークモチーフを検出しその検出数の変化を測定する．測定結果から論文数の増減下における共著関係の変化における傾向を分析する．

本研究では，共著関係ネットワーク上で興味ある現象を時系列として把握する方法を提案する．具体的には，一定の間隔で“social network”や“robot”といったキーワードと論文のタイトルと合致した論文を取得する．これらの論文の著者群から形成される共著関係ネットワークから高頻度なネットワークモチーフ検出し，これらのモチーフパターンの発生頻度がどのように時間変化するのかを調べる．

本論文の構成は以下の通りである．2 章では関連研究について説明し，3 章でネットワークモチーフとその検出ツールについて触れ，4 章で共著関係ネットワークの時系列分析について提案手法について説明する．5 章では DBLP (Digital Bibliography & Library Project)のデータを用いた評価実験を行い，最後の 6 章で本研究のまとめを述べる．

2. 関連研究

発表された研究の多くは共著または引用ネットワークのいずれかの特性について考察した．Travis Martin 氏ら[1]は研究者間の共著関係ネットワークと引用ネットワークの 2 つのネットワークにおける特性を検討した．特に彼らは著者が彼ら自身や彼らの協力者を他者より引用する傾向がある範囲や，著者が研究が公表された後により早く自分自身，または彼らの協力者を引用する範囲，そして著者が別の研究者から引用を返される範囲について研究を行った．

R. Milo 氏ら[2]は生化学，神経生物学，生態学，機械工学のネットワーク内でそれぞれのネットワークの特徴を表すネットワークモチーフが検出されことを突き止めた．

本研究では参考文献などの引用ネットワークは用いず，論文の共著からなる共著関係ネットワークからネットワークモチーフを検出する．ネットワークモチーフの発生頻度から，多種多様な学術分野における共著関係と論文発表数との間の相互関係を把握することが目的である．

3. ネットワークモチーフ

3.1 諸定義

ネットワークモチーフは生化学的ネットワーク(例えば, 代謝ネットワーク, 転写調節ネットワーク)や，ソーシャルネットワーク，技術的なネットワークなどの複雑ネットワークを特徴づける指標である．nノードをモチーフパターンに分類するためには，全てのノードが少なくとも 1 以上の入次数または出次数を持つことが条件である．このとき，モチーフはグラフの同型も同時に考慮すると，3 ノードで構成されるモチーフは最終的に 13 パターンとなることが示されている．特に，ID:5 は，生体内の転写制御ネットワークの中で，タンパク質の産生量を安定化させるのに必要なモチーフとして知られており，FFL(フィードフォワードループ)と呼ばれている．同様に 4 ノードのネットワークモチーフは 199 パターンに分類することができる．3 ノードサイズの場合のモチーフパターンを図 1 に示す．

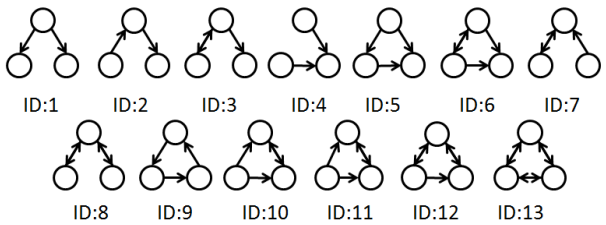


図 1：ノードサイズ 3 のネットワークモチーフ

3.2 モチーフ発見アルゴリズム

ネットワークモチーフを発見するために多くの手法が提案されている．本研究では mfinder[3]というツールを用いる．このツールは有向，無向ネットワーク中でモチーフの探索に利用することができる．モチーフを探索する際に基準となるエッジの集

合をE_sと表記し，E_sの両端のノードとそれらのノードと隣接するノードの集合をV_sと表記する．また，ハッシュテーブルを表す集合をHとする．これらの集合は初期状態として空集合と表記する．探索するモチーフのノード数を n とする．各モチーフの個数M_{ID_NUM}はそれぞれ初期値として 0 を与える．以下に使用したアルゴリズムを示す．

Input ：グラフ G={V, E}, n
E={e_1, e_2, e_3, …, e_N}$(1 \le i \le |E|)$
V={v_1, v_2, v_3, …, v_M}
$(1 \le j \le |V|,\ 1 \le k \le |V|,\ 1 \le l \le |V|)$
Output ：各モチーフの個数M_{ID_NUM}

Step
1. *for each* $e_i \in$ E
2. $i = 1$
3. モチーフを探索する際に基準となるエッジe_i=(v_j, v_k)をE_sに，両端のノードv_j, v_kをV_sに追加する．
4. $E_s = \{e_i\}$, エッジe_i=(v_j, v_k)の両端ノードv_j, v_kと隣接しているノードv_lをV_sに追加する．
$V_s = V_s \cup \{v_l\}$
$E_s = \{e_i\}$, $V_s = \{v_j, v_k, v_l\}$
5. if $|V_s| \neq n$ then
back Step.4
6. if $H = V_s$ then
$E_s = \{e_i\}$, $V_s = \{v_j, v_k\}$
back Step.4
7. ハッシュテーブルに探索したノードの組み合わせを保存する．
$H = H \cup V_s$
$H = \{v_j, v_k, v_l\}$
8. V_sで構成されるサブグラフG'がどのモチーフか判別する．
if モチーフが ID:p だったとき
$M_p = M_p + 1$
9. if $V_s = \{v_j, v_k\}$と隣接しているノードが場合v_l以外にも存在している then
$E_s = \{e_i\}$, $V_s = \{v_j, v_k\}$
back Step.4
10. $i = i + 1$
11. if $i \neq |E|$+1 then
$H = \{\}$, $E_s = \{\}$, $V_s = \{\}$
back Step.3
12. for each $e_i \in E$ loop End
13. 各モチーフ数M_pをnノードの各モチーフに含まれているエッジ数で除算する．
14. return M_{ID_NUM}

簡単なネットワーク G={V, E}として図 2 を示す．エッジE={e_1, e_2, e_3, e_4 }と，ノードは V={v_1, v_2, v_3, v_4 }と定義する．モチーフ探索例として図 2 のネットワーク Gから上記のアルゴリズムを用いて 3 ノードサイズのモチーフの検出例を示す．

① $i = 1$ (Step.2)
② 探索の基準となるエッジe_1={v_1, v_2}をE_sに，その両端であるノードv_1, v_2をV_sに追加する．
$E_s = \{e_1\}$, $V_s = \{v_1, v_2\}$ (Step.3)
③ V_sと隣接しているノードはv_3, v_4が存在する．v_3をV_sに追加する．
$E_s = \{e_1\}$, $V_s = \{v_1, v_2, v_3\}$ (Step.4)
④ $|V_s|$= 3となったので，V_sの組み合わせをHに保存する．
$H = \{v_1, v_2, v_3\}$ (Step.7)
⑤ V_sで構成されるモチーフの形は図 1 の ID:1 のモチーフの形と同じである．
⑥ $M_1 = M_1 + 1 = 1$ (Step.8)
⑦ V_sと隣接しているノードはv_3の他にv_4が存在する．V_sからv_3に取り除く．
⑧ $E_s = \{e_1\}$, $V_s = \{v_1, v_2\}$ (Step.9)
⑨ 次に，v_4をV_sに追加する．
$E_s = \{e_1\}$, $V_s = \{v_1, v_2, v_4\}$ (Step.4)
⑩ $|V_s|$= 3となったので，V_sの組み合わせをHに保存する．
$H = \{(v_1, v_2, v_3), (v_1, v_2, v_4)\}$ (Step.7)
⑪ V_sで構成されるモチーフの形は図 1 の ID:1 のモチーフの形と同じである．
$M_1 = M_1 + 1 = 2$ (Step.8)
⑫ v_1, v_2と隣接するノードはもう存在しない．これでe_1を基準としたモチーフの探索を終了する．
(Step.9)
⑬ $i = i + 1$=2 (Step.10)
⑭ $i \neq |E|$+1 となりループの終了条件を満たしていないのでエッジe_2を探索の基準とし，探索を続行する．
$H = \{\}$, $E_s = \{\}$, $V_s = \{\}$ (Step.11)
⑮ 探索の基準となるエッジe_2={v_2, v_3}をE_sに，その両端であるノードv_2, v_3をV_sに追加する．
$E_s = \{e_2\}$, $V_s = \{v_2, v_3\}$ (Step.3)
⑯ 以上のようにネットワーク内のすべてのエッジを探索の基準としてモチーフの探索を進める．図 2 における全てのエッジの探索が終了したときの各M_{ID_NUM}は以下のようになっている．
$M_1 = 4$
$M_5 = 3$
また，残りのM_2, M_3, M_4, M_6, …, M_{13}の値は 0 である．
⑰ 最後の処理としてM_1, M_5の各値を各モチーフが含むエッジ数で除算する．図 1 における ID:1 のエッジ数は 2 本，ID:5 のエッジ数は 3 本である．
$M_1 = M_1 \div 2 = 2$
$M_5 = M_5 \div 3 = 1$ (Step.13)
⑱ 上記より図 2 に記されたグラフ Gからは図 1 に記された ID:1 のモチーフが 2 個，ID:5 のモチーフが 1 個検出された．
(Step.14)

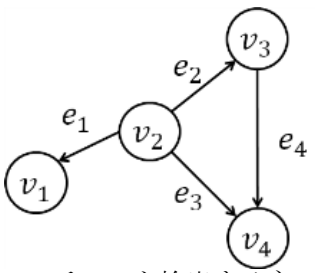


図 2：モチーフを検出するネットワーク G

4. 提案手法

本研究ではネットワークを構成する基本的な構成要素であるネットワークモチーフを用いる．著者

間において構成される共著関係ネットワークからネットワークモチーフを検出し，モチーフの発生頻度がどのように時間変化するのかを調べ，その発生頻度と論文発表数との間の相互関係を考察する．提案手法による処理手順は以下の通りである．

（1） 論文データの取得
一定の間隔で論文の情報(論文タイトル，著者名，発表年数)を取得する．

（2） キーワードから論文を選択
使用する論文データからタイトルに任意に選んだキーワードを含む論文のみを選択する．

（3） 共著関係ネットワークの生成
本研究において“論文の共著関係ネットワーク”とは各論文間の共著者関係を表現したものである．1 つの論文を作成するに当たり著者が複数人存在する場合がある．先頭に記載してある著者を筆頭著者，筆頭著者の後に記載してある著者を共著者と定義し，ネットワーク中のノードをこれらの著者とする．例えば，3 名の研究者が連名になっている論文について考えてみる．筆頭著者を A，2 名の共著者を B および C とすると，ノード間のエッジを図 3 のように定義する．

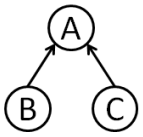


図 3：著者 3 名に対するノード間のエッジ

これは 1 本の論文を作成するにあたり共著者 B, C が筆頭著者 A に協力したことを示している．

これによって作成される共著関係ネットワークは著者間の協力関係を表したネットワークとなる．選択した論文から各時系列全てのネットワークを作成する．論文の共著関係の例を図 4 に示す．

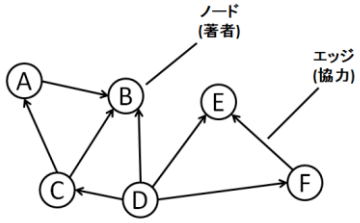


図 4：論文の共著関係ネットワークの例

（4） ネットワークモチーフの検出
各時系列間隔に対して，共著関係ネットワーク内の 3 ノードのネットワークモチーフ mfinder のを用いて検出する．

（5） ネットワークモチーフの検出
検出されたそれぞれの 3 ノードのネットワークモチーフの発生頻度と論文発表数との間に相互関係があるかどうか考察を行う．

5. 評価実験

本提案手法の有効性を検討するために，評価実験を行う．評価実験を行う際に使用するデータは国際会議録誌や学術論文誌に掲載された論文の書誌情報が記録されている DBLP (Digital Bibliography & Library Project)[4]のデータを利用する．

5.1 実験方法

DBLP から 2003 年から 2012 年に発表された論文を取得し，1 年間隔で書誌情報を取得する．

書誌情報のキーワードを“social network”，“robot”，“neural network”とし，それぞれの検索結果を用いて，著者間において構成される著者関係ネットワークを作成した．

以下に各キーワードから抽出した論文群の情報を表 1，表 2，表 3 として，各キーワードを含む論文数の時間的推移を図 5 として以下に記載する．

5.2 実験結果と考察

図 5 の論文数を見ていくとキーワードを“robot”としたときのネットワークは 2011 年までは一定の割合で増加していることが分かった．一定の割合で増加している“robot”と異なり，“social network”では 2009 年から大幅に論文数が増加していた．次に，キーワードを“neural network”としたときは 2006 年までは増加しているがそれ以降は大きな変化は見られず，2009 年以降は減少傾向が見て取れる．これらの結果をみていくと，キーワードを“robot”，“social network”としたときの論文数は増加傾向，キーワードを“neural network”としたときの論文数は減少傾向となっていることが分かった．

表 1，2，3 は DBLP から取得したデータに関する基本的な情報をまとめたものであり，各キーワードを“robot”，“social network”，“neural network”としたものである．表 1 を見ていくと各値が表 2，表 3 と比較して大きいことが分かった．ここから“robot”に関する論文を書く著者が他の 2 つのキーワードに関する論文を書く著者よりも多いことが分かった．

表 4 は，この 3 つのネットワークから抽出された 6 種類のネットワークモチーフ(ID:1，ID:2，ID:4，ID:5, ID:7, ID:11)のそれぞれについて，時系列パターン(増加，減少，増加と減少の混在，概ね一定)検出数(最大検出数 Max，最少検出数 Min)をまとめたものである．なお，3 ノードのネットワークモチーフは 13 種類存在するが，ID:3，ID:6，ID:8，ID:9，ID:10, ID:12, ID:13 の 7 種類のネットワークモチーフは 10 年間で最大検出数が 50 未満であったため本実験の結果からは除外した．

この 3 つの共著関係ネットワークのどれをみても，ID:4 の検出量が他のモチーフに比べて多いことが分かった．ID:4 は 1 名の筆頭著者に向けて 2 名の共著者からエッジが向けられているモチーフである．書誌情報から共著関係ネットワークデータを作成する際，筆頭著者に向けて 2 名の共著者からエッジを張るデータのみを作成したため，より多く検出されたのだと考えられる．このような事情により，ID:4 は論文発表数の推移に比例している傾向が強くみられた．ID:1, ID:2 に関しても ID:4 ほどではないがこの傾向が見られた．これはノード間のエッジが 2 本しか必要とせず他のモチーフと比べて共著関係ネットワークの一部として生成されやすいからだと考えられる．

表 1：キーワードを“robot”としたとき取得した論文数と著者数

総論文数	23224
10 年間における最少論文数	1223
10 年間における最大論文数	3126
全著者数	34743
論文 1 本あたりの平均著者数	3.58
著者 1 人あたりの平均論文数	2.40

表 2：キーワードを“social network”としたとき取得した論文数と著者数

総論文数	4424
10 年間における最少論文数	25
10 年間における最大論文数	1198
全著者数	9923
論文 1 本あたりの平均著者数	3.35
著者 1 人あたりの平均論文数	1.59

表 3：キーワードを“neural network”としたとき取得した論文数と著者数

総論文数	15798
10 年間における最少論文数	815
10 年間における最大論文数	2029
全著者数	28986
論文 1 本あたりの平均著者数	3.15
著者 1 人あたりの平均論文数	1.71

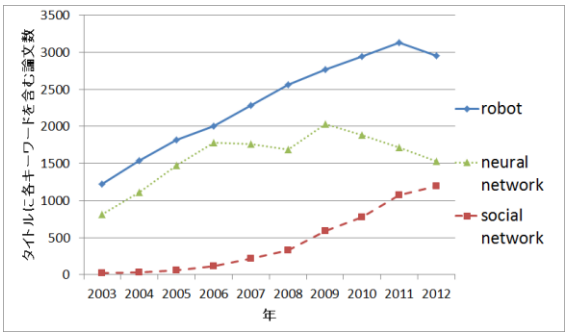
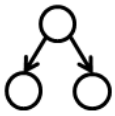
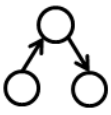
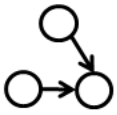
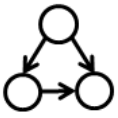
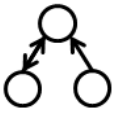


図 5：タイトルにキーワードを含む論文数

表 4：それぞれのネットワークから検出されたネットワークモチーフとその傾向

ネットワーク / モチーフ / キーワード	 ID:1	 ID:2	 ID:4	 ID:5	 ID:7	 ID:11
robot	Min:616 Max:2562	Min:569 Max:1432	Min:3717 Max:11452	Min:112 Max:433	Min:245 Max:455	Min:80 Max:169
social network	Min:0 Max:202	Min:0 Max:192	Min:40 Max:3612	Min:0 Max:57	Min:0 Max:84	Min:0 Max:73
neural network	Min:119 Max:426	Min:98 Max:417	Min:1255 Max:4045	Min:24 Max:68	Min:34 Max:162	Min:13 Max:50

6. まとめ

本研究では，ネットワークモチーフを利用して，共著関係ネットワークの時系列データを分析する手法の開発をめざし，3 種類の共著関係ネットワークからネットワークモチーフを検出した．ネットワークモチーフ ID:1, ID:2, ID:4 は論文発表数に強く影響を受けるということが分かった．また，ID:5, ID:7, ID:11 の 3 つに関しては論文発表数に強く影響を受けてはいなかった．その中に，FFL が含まれていたことは興味深い．今後の課題として，ID:5, ID:7, ID:11 の 3 つのモチーフが強い影響を受けていない理由についての調査が挙げられる．

参考文献

[1] Travis Martin, Brian Ball, Brian Karrer, and M. E. J. Newman, "Coauthorship and citation in scientific publishing", Phys. Rev. E 88, 012814,2013.

[2] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D.Chklovskii and U. Alon, "Network motifs: simple build-ing blocks of complex Networks" Science 298, pp. 824-827,2002.

[3] mfinder http://www.weizmann.ac.il/mcb/UriAlon/

[4] DBLP (Digital Bibliography & Library Project) http://www.informatik.uni-trier.de/~ley/db/