

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 23 日現在

機関番号：23803

研究種目：基盤研究(C)

研究期間：2011～2014

課題番号：23500312

研究課題名(和文)階層付きモチーフ・パターン分析によるネットワークの類型化

研究課題名(英文)Classification of Networks using Hierarchical Network Motifs

研究代表者

大久保 誠也 (OKUBO, Seiya)

静岡県立大学・経営情報学部・助教

研究者番号：90422576

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：本研究の目的は大規模ネットワークの特徴抽出技術の構築である。具体的には、1) 現実のネットワークデータの収集、基本統計量の分析、ならびに適切なネットワーク生成モデルの構築。2) 階層付きモチーフ・パターンによるネットワーク類型化法の確立、ならびに人工ネットワークと現実ネットワークに対する計算機実験による評価。3) 時系列データに対する変化点の検出、および主要変化点の編纂に関する各種方法論の確立、を行った。

研究成果の概要(英文)：In this study, we proposed following models and methods. 1) Network generation model of growing networks using network motifs. 2) classification methods of networks using hierarchical network motifs. 3) change-point detection in time series data of reviews.

研究分野：情報セキュリティ、量子計算、行動の類型化、ゲーム情報学

キーワード：情報社会学 ネットワーク類型化

1. 研究開始当初の背景

自然科学や工学、さらには社会科学など多様な分野に共通する重要な分析対象として、例えば、ウェブサイトでのハイパーリンクのネットワークや、企業内での人間関係のネットワークなどのような、時間と伴に変化して成長し続ける複雑ネットワークに着目する。このようなネットワークは、固有のトップダウン設計原理を土台にしつつ、ある種のボトムアップ自己組織メカニズムにより成長すると考えられ、そこには、それぞれのサイトや企業などの本質的な特性が存分に反映されていると想定できる。よって、このようなネットワーク構造を適切に類型化する新分析方法論の探求は、各分野でのトレンドを理解するためにも重要な研究課題である。以下では、ウェブページや個人をノード、ハイパーリンクや人間関係をリンクと呼ぶ。従来、ネットワーク群の構造は、ノード集合の要素間に対し、リンク関係が最も一致する対応関係(グラフ同型)を求め、その一致度合いに基づく類似度で類型化されていた。しかしネットワーク規模が大きくなると、グラフ同型を求める計算コストは指数的に増大するため、実質的に分析が困難となる。一方、ネットワークの基本的統計量に着目しても、例えば、各ノードのリンク数に基づく次数分布は一つの基本統計量であるが、現実の多くのネットワークでスケールフリー性を満たすので、精緻な類型化は一般に困難である。

2. 研究の目的

本研究では、妥当な計算量で精緻な類型化を実現するため、局所的なグラフ同型の積み上げに基づき、ネットワーク構造を特徴付けるモチーフ・パターン分析に着目する(図1参照)。特に、ウェブサイトや企業などのネ

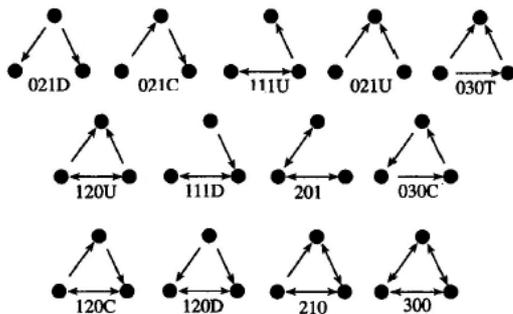


図 1：通常のネットワークモチーフ

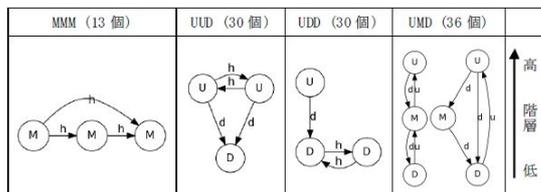


図 2：階層付きネットワークモチーフ

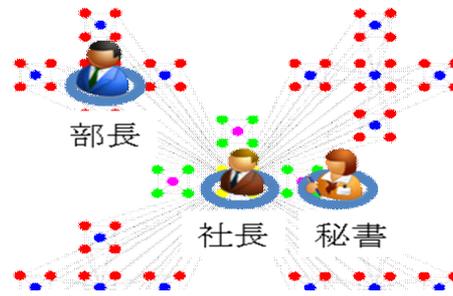


図 3：階層ネットワークのイメージ

ットワークに対しては、トップサイトからの深さや、企業での役職などにより、各ノードに対して自然な階層レベルを付与できる(図3参照)。このような階層性は、多様な社会ネットワークなどにおいても同様に見られ、階層レベルでの上下関係も考慮に入れた「階層付きモチーフ・パターン」での分析法を確立する(図2参照)。

具体的な本研究の目標は、

- 1) 現実のネットワークデータの収集、ネットワークの階層を考慮した基本統計量の分析、ならびに適切なネットワーク生成モデルの構築。
- 2) 階層付きモチーフ・パターンによるネットワーク類型化法の確立、ならびに人工ネットワークと現実ネットワークに対する計算機実験による評価。
- 3) 時系列データに対する変化点の検出、および主要変化点の編纂に関する各種方法論の確立

である。

3. 研究の方法

各目的を達成するために、以下のような方法をとった。1) 与えられたネットワークの基本統計量に基づき、生成モデルの各種パラメータを設定する。2) 設計原理などの異なる生成モデルのそれぞれから複数の人工ネットワークを作成する。3) 生成モデル毎に人工ネットワーク群の階層付きモチーフ・パターン頻度の平均を求める。4) パターン頻度ベクトルに基づき、与えられたネットワークと最類似するモデルを求める。なお、手順4)の類似度については、主にコサイン類似度を土台にする。類型化法の評価には、人工と現実それぞれのネットワークを用いる。人工ネットワークにおいては、その生成モデルが既知なので、類型化の正答率での定量評価を基本とする。また、作成した人工ネットワークに対し、ノードやリンク群のランダムな削除や入替を施し、不完全なネットワークでの類型化能力を定量的に評価することで、階層付きモチーフ・パターンの基本能力やネットワーク類型化法の頑健性を明らかにする。

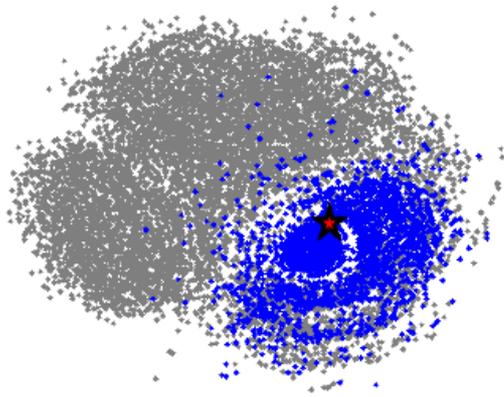


図 4 : twitter のデータ拡散例

また、大規模ネットワークの状況を把握することは困難であるため、各手法に適した可視化方法の検討も行う。

4. 研究成果

1)については、SNS データの収集を中心に、それとともに基本統計量の分析をおこなった。具体的には、twitter データや、本等に対する複数のレビューサイトデータの収集を行った。また、非 SNS データとして、株価推移データや、過去の大学 Web サイトデータの入手を行った。これらのデータからネットワークデータを生成し、基本統計量を解析するとともに、新しい有向ネットワークを生成する、成長モデルを提案した。

2)については、大規模ネットワークデータをネットワークモチーフを用いて分析する方法を複数提案し、実際のデータに適用することで、その有効性の検証を行った。まずはじめに、twitter データに関して、ツイートのタイミングやユーザ集合を特徴量としてキーワードを類型化する手法を提案した。また、レビューサイトデータに対し、ユーザー間の類似度を定式化すると共に、どのように変化していくかについて、統計的な分析を行った。つぎに、ネットワークモチーフ(トライアド)を用いたネットワークの時系列分析方法を提案するとともに、twitter の reply と RT によって作られる 2 つのネットワークに対して適用し、その有効性を示した。また、レビューサイトや株価データを対象として類似度分析を行った。

これらの成果により、大規模ネットワーク全体が、どのような特徴を持っているかを明らかにできるようになった。加えて、3)の成果と併せて、ネットワーク全体が時間と共にどのように変化していくかと、構造が大きく変化する変化点を、自動的に抽出することが可能となった。

さらに、ネットワーク全体ではなく、ネットワークに参加する個々のオブジェクトの特徴を捉える手法を提案した。具体的には、あるユーザーとその周辺のみからなる部分グラフをモチーフパターン分析することに

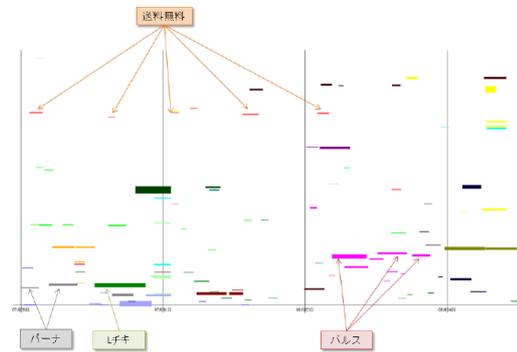


図 5 : タイムライン可視化の例

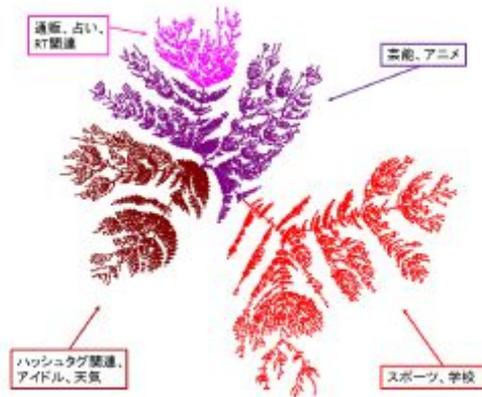


図 6 : 類型化と注釈付加の例

より、あるユーザーがネットワーク上においてどのように活動しているかを特徴付ける手法を提案した。提案手法を Twitter データに適用することにより、どのような行動を行うユーザーがいるかの分析・類型化をおこなった。

3)については、2)の成果と併せて、時系列の変化点を検出する方法を複数提案するとともに、株価データやレビューサイトに対して適用することにより、妥当性の検討を行った。また、ユーザーがレビューサイトにデータを投稿する行動のモデルを提案した。また、データが時間と共に拡散していく様子を効率的に可視化する方法を提案した。加えて、twitter データを対象とした解析等を行い、つぶやきを類型化し可視化する手法を提案した、特に、共通の話題毎に類型化し(図 5 参照) 各集合に適切な注釈単語をつける手法を提案するとともに、実際に Twitter データに適用することにより、その有効性を示した(図 6 参照)。

これらの成果により、大規模ネットワークから特徴的な集団に分類すると共に、各集団がどのように特徴付けられるかの注釈の付加を、自動的に行うことができるようになった。

5. 主な発表論文等

〔雑誌論文〕(計 1 件)

1. 小出 明弘, 齊藤 和巳, 風間 一洋, 鳥海 不二夫, "ネットワーク分析による Twitter ユーザのフォロー形成に関する一考察," 情報処理学会論文誌, 数理モデル化と応用-ネットワークが創発する知能論文特集-, Vol.6, No.2,

〔学会発表〕(計 31 件)

1. Yuki Yamagishi, Seiya Okubo, Kazumi Saito, Kouzou Ohara, Masahiro Kimura, and Hiroshi Motoda, "A Method to Divide Stream Data of Scores over Review Sites," Proc. of the 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI2014), pp.913-919, 2014.
2. 小林 えり, 齊藤 和巳, 大久保 誠也, 池田 哲夫, "凝縮性に基づく注釈単語検出法とその評価," 情報処理学会第 77 回全国大会 (IPSJ2015), Mar. 2015.
3. 藤野 まり菜, 佐藤 杏奈, 伏見 卓恭, 齊藤 和巳, 大久保 誠也, 池田 哲夫, "凝縮性に基づく有用単語検出によるトレンドワードの注釈付き可視化," 情報処理学会第 77 回全国大会 (IPSJ2015), Mar. 2015.
4. 西 可南子, 加藤 翔子, 齊藤 和巳, 湯瀬 裕昭, 武藤 伸明, 大久保 誠也, "アンケート調査による回遊行動モデルの評価," 情報処理学会第 77 回全国大会 (IPSJ2015), Mar. 2015.
5. 加藤 翔子, 小林 えり, 湯瀬 裕昭, 大久保 誠也, 武藤 伸明, 齊藤 和巳, 池田 哲夫, "地域回遊促進に向けたユーザー行動モデル," 第 103 回 知識ベースシステム研究会 (SIG-KBS), Nov. 2014.
6. 西 可南子, 加藤 翔子, 大久保 誠也, 齊藤 和巳, "レビュー, ブログ, マイクロブログサイトの高次数ユーザの自己中心トライアド分析," 第 13 回情報科学技術フォーラム (FIT2014), Sep. 2014.
7. 小林 えり, 齊藤 和巳, 池田 哲夫, "混合距離定義に基づく一般化ピボット法の提案とその評価," 第 28 回人工知能学会全国大会 (JSAI2014), May. 2014.
8. ブイ・スアン・フォン, 榊原 源基, 上野 雄史, 齊藤 和巳, "Mantegna 法に基づく各国株価データの業種同類度分析," 情報処理学会 第 76 回全国大会 (IPSJ2014), Mar. 2014.
9. 山岸 祐己, 齊藤 和巳, "レビューサイトにおけるユーザ間類似度分析," 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM2013), Mar. 2013.
10. 加藤 翔子, 大久保 誠也, 齊藤 和巳, "Twitter の reply と retweet 関係からなる成長ネットワークの比較分析," 第 12 回情報科学技術フォーラム (FIT2013), Sep. 2013.
11. 加藤 翔子, 大久保 誠也, 齊藤 和巳, "トライアド係数を用いた Twitter の reply と RT ネットワークの時系列分析," 第 10 回 ネットワーク生態学シンポジウム (NETECO2013), Sep. 2013.
12. 佐藤 杏奈, 伏見 卓恭, 大久保 誠也, 齊藤 和巳, 風間 一洋, "出現ツイート群の類似度に基づくトレンドワードのタイムライン可視化," 第 10 回 ネットワーク生態学シンポジウム (NETECO2013), Sep. 2013.
13. 加藤 翔子, 齊藤 和巳, "多重有向ネットワーク成長モデルの特性解析," 第 27 回人工知能学会全国大会 (JSAI2013), Jun. 2013.
14. 山岸 祐己, 齊藤 和巳, 池田 哲夫, "レビューサイトにおけるユーザ間の動的類似度分析," 情報処理学会 第 75 回全国大会 (IPSJ2013), Mar. 2013.
15. 木村 唯香, 伏見 卓恭, 齊藤 和巳, "時間減衰ダイナミクスを導入したモデルによる意見形成とユーザ投稿間隔の分析," 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM2013), Mar. 2013.
16. 杉澤 優馬, ブイ・スアン・フォン, 伏見 卓恭, 齊藤 和巳, "変化点検出法に基づく業種別社会的影響分析," 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM2013), Mar. 2013.
17. 山岸 祐己, 齊藤 和巳, "レビューサイトにおけるユーザ間類似度分析," 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM2013), Mar. 2013.
18. 山岸 祐己, 齊藤 和巳, "オンラインレビューサイトにおける投稿間隔分析," 第 9 回 ネットワーク生態学シンポジウム (NETECO2012), Dec. 2012.
19. 木村 唯香, 伏見 卓恭, 齊藤 和巳, "レビューサイトによる時間減衰ダイナミクスを持つ TDV モデルの評価," 第 9 回 ネットワーク生態学シンポジウム (NETECO2012), Dec. 2012.
20. 杉澤 優馬, 伏見 卓恭, 齊藤 和巳, "株価時系列データによる変化点検出法の評価," 第 9 回 ネットワーク生態学シンポジウム (NETECO2012), Dec. 2012.
21. 加藤 翔子, 伏見 卓恭, 小出 明弘, 齊藤 和巳, 元田 浩, "Twitter の 3 機能を用いたネットワーク分析," 第 97 回 知識ベースシステム研究会 (SIG-KBS), Nov. 2012.
22. 小出 明弘, 齊藤 和巳, "マイクロブログ, ウェブログ, レビューサイトのネットワーク構造比較," 第 97 回 知識ベースシステム研究会 (SIG-KBS), Nov. 2012.
23. 小出 明弘, 齊藤 和巳, 風間 一洋, 鳥

- 海 不二夫"デンドログラムを用いた震災時における Mention・Hashtag ツイートの利用傾向分析,"第5回 Web とデータベースに関するフォーラム (WebDB2012), Nov. 2012.
24. 山岸 祐己, 齋藤 和巳, 大久保 誠也, "局所改善法によるレビュー変化点検出法,"第11回情報科学技術フォーラム (FIT2012), Sep. 2012.
25. 小出 明弘, 齋藤 和巳, 長屋 隆之, 伊藤 健二, "ネットワーク粗視化による情報拡散過程の可視化法,"第26回人工知能学会全国大会 (JSAI2012), June. 2012.
26. 山岸 祐己, 齋藤 和巳, 大久保 誠也, "レビュー時系列データからの分割統治による変化点検出法,"第26回人工知能学会全国大会 (JSAI2012), June. 2012.
27. 小出 明弘, 齋藤 和巳, 大久保 誠也, 鳥海 不二夫, 風間 一洋, "Twitter の @message で構成される成長ネットワークの分析,"第74回情報処理学会全国大会 (IPSJ2012), Mar. 2012.
28. 加藤 翔子, 伏見 卓恭, 齋藤 和巳, "Twitter の Favorite 機能を用いたユーザー嗜好分析,"第4回データ工学と情報マネジメントに関するフォーラム (DEIM2012), Mar. 2012.
29. 山岸 祐己, 齋藤 和巳, 大久保 誠也, "オンラインレビューサイトの評点時系列データからの異常検出,"第4回データ工学と情報マネジメントに関するフォーラム (DEIM2012), Mar. 2012.
30. 小出 明弘, 齋藤 和巳, 鳥海 不二夫, 風間 一洋, "フォロワー類似度による Twitter 高次数ユーザの関係分析,"第3回集合知シンポジウム (SIG-NLC), Feb. 2012.
31. Akihiro Koide, Kazumi Saito, Masahiro Kimura, Kouzou Ohara, and Hiroshi Motoda, "Estimating Diffusion Probability Changes for AsIC-SIS Model from Information Diffusion Results," Proc. of the third Asian Conference on Machine Learning (ACML2011), pp.297--313, 2011.

6. 研究組織

(1) 研究代表者

大久保 誠也 (OKUBO Seiya)
静岡県立大学・経営情報学部・助教
研究者番号: 90422576

(2) 研究分担者

武藤 伸明 (MUTOH, Nobuaki)
静岡県立大学・経営情報学部・准教授
研究者番号: 40275102

齋藤 和巳 (SAITO, Kazumi)
静岡県立大学・経営情報学部・教授
研究者番号: 80379544

風間 一洋 (KAZAMA, Kazuhiro)
和歌山大学・システム工学部・教授
研究者番号: 60647204