

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 3 日現在

機関番号：12601

研究種目：基盤研究（B）

研究期間：2009 ～ 2012

課題番号：21300031

研究課題名（和文）

大規模ネットワーク型データからの体系的な信頼度付き情報取得システムの開発

研究課題名（英文）

Development of Systematic Acquisition System of Information Graded by Reliability from Large-scale Linked Data

研究代表者

白山 晋（SHIRAYAMA SUSUMU）

東京大学・大学院工学系研究科・准教授

研究者番号：10322067

研究成果の概要（和文）：

広域分散化する大規模データから有用な情報を取得するための方法・方法論が議論されている。しかし、ネットワーク構造を活かした情報取得の方法・方法論は少なく、また分析結果を理解することが容易ではない。本研究では、大規模な動的ネットワークの構造分析手法と部分的情報からの全体推定法、強化された信頼度ランク算出法、視認性が高くわかりやすい情報提示法を中核とする信頼度付き情報取得システムの開発を試みた。

研究成果の概要（英文）：

The method and methodology for acquiring useful information from the widely distributed large-scale data have been discussed. However, there are few methods of acquiring information which utilizes the nature of the linked data structure. Moreover, it is not easy to understand the results of analysis. In this study, we attempt to develop a new acquisition system with which to reduce information unreliability. The system consists of a structural analysis method for the large-scale dynamical network, an evaluation method of the whole network structure from partial information, the enhanced TrustRank, and an information displaying method in which visibility is high.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	2,600,000	780,000	3,380,000
2010 年度	3,200,000	960,000	4,160,000
2011 年度	1,500,000	450,000	1,950,000
2012 年度	2,300,000	690,000	2,990,000
年度			
総計	9,600,000	2,880,000	12,480,000

研究分野：総合領域

科研費の分科・細目：情報学，メディア情報学・データベース

キーワード：ネットワーク，大規模データ，情報信頼度，コミュニティ構造，集合知，可視化，視線計測，データマイニング

1. 研究開始当初の背景

ネットワーク上には個人や組織が発信する質の高い情報が存在する。情報自体の取得

技術は確立され、情報活用のために信頼性が問題になっている。信頼性に関する研究の多くは意味的内容に関連したものである。一方、

ネットワーク構造を信頼性の評価に利用するという Garcia-Molina らのグループの研究がある。これは意味的内容を補完するという位置付けだが機械処理に適するため、効率化という観点で注目された。しかし、ネットワークの大規模性に起因するネットワーク構造分析の困難さが、そのような構造からの信頼性評価の研究を停滞させている。

ネットワーク構造分析に関する研究は、98,99 年の Watts と Strogatz, Barabasi と Albert による統計物理的手法による理論的解明を契機として、マクロ的な構造から、メソスケール（メソ構造）、ミクロ構造へと移行する。現在ではマクロ的構造を起点として階層的に詳細度を上げる方法が一般的である。それらの分析手法を実用面で利用する場合に、実ネットワークの大規模性と動的変化が問題になっている。この数年の学術雑誌や会議録からも、大規模性に関連して、コミュニティ抽出に基づく階層構造の分析、部分的な情報と統計的手法を用いたネットワーク全体の推定、動的に変化するネットワークの分析、効果的な可視化手法、分析結果の解釈と利用法などが最先端の研究に位置付けられ、未解決な問題として扱われていることがわかる。また、日本からの発信は少なく、欧米での研究が先行していることがわかる。

一方、ネットワーク上のデータや情報の信頼性に関しては、国内のいくつかの学会で研究グループが形成され、欧米に比肩する研究も多くなっている。

しかしながら、意味的内容分析に基づくものの比重が大きく、ネットワーク構造に属する研究は、構造同値、ハブ・オーソリティ、リンクの粗密のような比較的単純な原理を利用したもの限定され、ネットワーク構造を十分に活かしたものは少ない。また、ネット社会の安心を「安全と自由」と考えると、自由が不正行為やネット犯罪を増長するという指摘と自由がインフラ化するネットの持続的な発展を支えているという指摘の双方を配慮する必要があり、意味的内容と分離できるネットワーク構造から情報の信頼度を算出することが非常に重要になっている。

2. 研究の目的

信頼度付き情報取得システムの構築のために、これまで明らかにされていない大規模で動的に変化するネットワーク構造の性質を系統立てて整理すること、それらを信頼度算出法に利用し、大規模動的ネットワークから信頼度付きの情報を取得するための方法・方法論を体系的に示すこと、視認性が高くわかりやすい情報提示法までを検討することが本研究の目的である。

3. 研究の方法

部分情報からのネットワーク全体推定法、ネットワーク構造を利用した信頼度算出法、

視認性の高い情報提示法の三つの手法を統合することでシステムを構築する。以下にそれぞれの方法を示す。

(1) ネットワークモデル群を利用した機能からの系統的ネットワーク構造探索、ノード予測とリンク予測によって、観測された部分情報からネットワーク全体を推定する。

① はじめに、ネットワークモデルから形成されたネットワークを $G(V, E)$ とし、ネットワーク上の現象を局所ルールでモデル化し、図 1 に示すように入力 i と出力 o の関係性を求める。次に、ネットワーク上の現象がモデル化できると仮定し、未知のネットワーク上の現象に対して観測された入力 i と出力 o の関係から、未知のネットワークを説明しうるネットワークモデルを推定する。これによりネットワーク構造探索を行う。

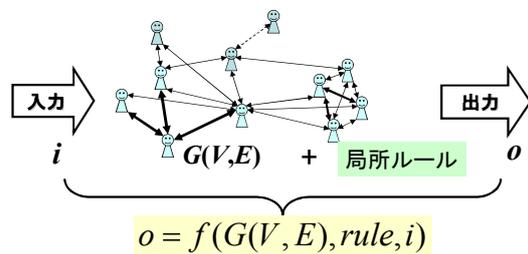


図 1 機能からの系統的ネットワーク構造探索

② 観測された部分情報からノード予測とリンク予測により全体を推定する。

(2) 信頼度算出は、Garcia-Molina らのグループが提案している TrustRank と BadRank に基づく。彼らの方法には、良いノードと悪いノードの与え方、信頼度の局在化（ノード間の信頼度の不完全な伝播）、ノードとリンク情報の不完全性という問題があり、改良が必要である。そこで、以下の改良法を提案した。良いノードと悪いノードの与え方、信頼度の不完全な伝播に対して、図 2 のようなコミュニティグラフという階層的構造を考慮した方法を提案し、良いノードと悪いノードの与え方の効率化、および効率的な信頼度伝播を実現する。また、ノード予測とリンク予測によって情報の不完全性を補う。

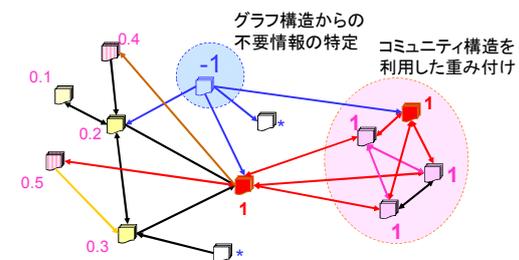


図 2 コミュニティによる信頼度の効率的算出

(3) 視認性の高い情報提示のためには、人間の視覚認知機構を明らかにする必要がある。視覚情報に対して、人間にとっての重要度を算出する方法を、視覚顕著度マップを用いた領域分割と視線データのネットワーク分析を用いて構築した。また、ネットワーク可視化におけるグラフィアウト作業時の操作と視線を対応付けることで視認性の高い情報の特性を明らかにする。

4. 研究成果

(1) 部分情報からのネットワーク全体推定法に関する成果を以下に述べる。

① 機能からの系統的ネットワーク構造探索においては、ネットワークがもたらす機能とネットワーク構造の異質性（次数分散とベキ乗指数）が密接に関係すること、進化ゲーム理論に基づくモデルの階層化によって機能からの構造推定の効率化に寄与する素過程が顕在化されることを示した。図3に一例を示す。局所ルール、協調の進化、次数分散には図のような関係性がある（これは世界で初めて示されたものである）。この関係性は局所ルールとネットワーク構造に依存するため、協調者の割合が未知のネットワークを説明する指標であり、これによりネットワークモデルが推定できると考えられる。

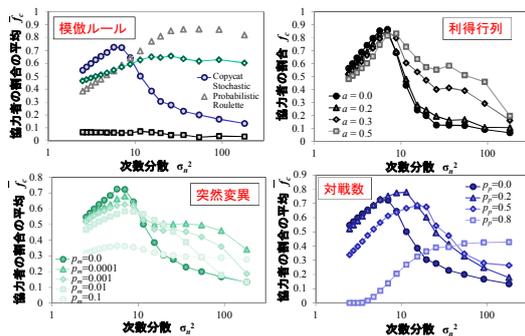


図3 次数分散と協調の進化の関係性

また、情報伝播において、陽のコミュニティ、および潜在リンクを考慮する必要があることを示し、それらが制御できるネットワーク成長モデルを提案した。

さらに、最適化の考えを用いて、機能と構造の関係性から、所与の機能を実現化するためのネットワーク生成法を構築した。この生成法はこれまで知られていなかったものであり、ネットワークモデル群に加えることができる。これにより未知のネットワークをより正確に推定することが可能になる。これは、大規模ネットワーク型データをネットワークモデルのパラメータ空間で表現できることにつながる。結果として、大幅な次元縮約につながり、情報取得が容易になることが期待できる。

この研究と平行して、現象の特徴とネットワークの統計的指標との関係を分析するた

めのニューラルネットワークと決定木を用いた新たな手法を開発し、得られた関係性からネットワークの生成機構が推定できることを示した。はじめに、多数の数値シミュレーションを行い、統計的指標を説明変数とし、現象の特徴量を被説明変数として、説明変数と被説明変数の組をデータセットとして求める。次に、データセットの一部を用い、入力層を説明変数のすべての組み合わせにしたニューラルネットワークによって学習する。残りのデータセットの説明変数を用いて被説明変数を予測し、予測精度を求める。決定木によって予測精度が高くなる説明変数（統計的指標）の組み合わせを求め、それを現象の特徴を表すものとする。複数の生成機構の組合せによって生成されたネットワークに対してこの方法を適用する。これより、生成機構と現象の関係性をデータマップとして得ることができた。

また、ネットワーク上での情報伝播に対するコミュニティの影響を明らかにした。これにより、コミュニティ構造の類型化によるメソ構造の抽出と推定によるネットワーク構造探索の可能性が示唆された。

② はじめに、観測可能な情報のみでネットワークの統計的指標を算出し、統計的指標とリンク予測指標の関係を導いた。その関係性を前処理に用いたリンク予測手法を開発し、部分情報から効果的に欠損情報が補えることを示した。

以上によりネットワーク型のデータ、および情報の背後にあるネットワーク構造をより深く分析でき、部分情報からのネットワークの全体像を従来手法よりも効率的、かつ効果的に明らかにできることを示した。

(2) 我々のグループでは、TrustRankとBadRankを利用した信頼度算出におけるノード間の信頼度の伝播アルゴリズムの不備の問題に対して、Leichtらのノード類似度算出を拡張したコミュニティ抽出から得られたコミュニティグラフの利用による解決法を提案してきた。図4左図がシード、右図がコミュニティグラフの信頼度分布の一例である（グラフィアウトはPajekによる）。

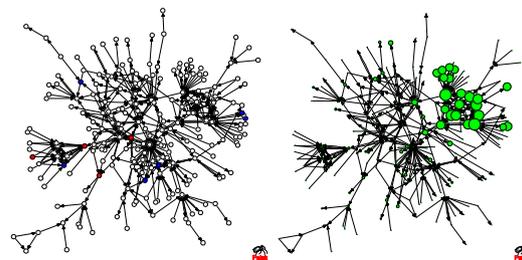


図4 コミュニティグラフの信頼度分布

この例では、カバー率（伝播によって信頼度が与えられたノードの比率）が従来法の

2.3%に対して、提案手法では98.5%であり、大幅に向上した。しかしながら、コミュニティ構造が伝播に与える影響が把握できておらず、メカニズムが不明であった。この点に関しては、他の研究グループも明らかにしていない。そこで、信頼度に影響する情報伝播がコミュニティを有するネットワーク構造から受ける影響について数値シミュレーションによって調べた。得られた結果の網羅的探索により、コミュニティの与える影響がある程度ではあるが定量的に評価できることを示した。

一方、Leichtらのノード類似度算出を拡張して得られたコミュニティ構造では、信頼度に与えるネットワーク構造の局所的な特徴が十分に捉えられないことが明らかになった。特に、時間変化するリンク構造をもつネットワーク型データに対しては、信頼度の局在化の問題はより大きなものになること、その原因の特定が難しいことが明らかになった。そこで、自己組織化マップを用いた新たなクラスタリング手法によって時間的にコミュニティ間を移動するノードを顕在化するという方法によってこの問題の解決を試みた。はじめに、ノード属性を多変量変数とし、自己組織化マップを用いてノードを分類し、コミュニティ(クラスター)を抽出する。図5に一例を示す。

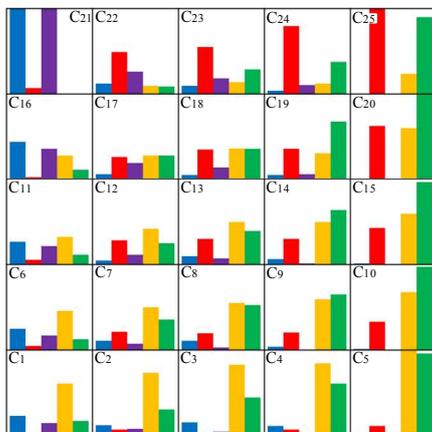
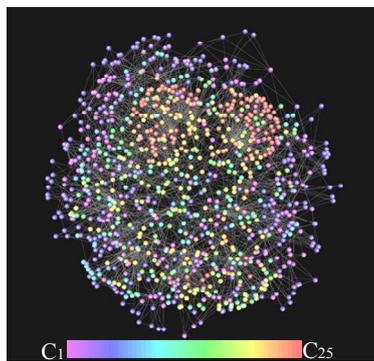


図5 自己組織化マップによるコミュニティ抽出

図5上図は、ネットワーク全体をコミュニティによって色分けして示したものである。下図は自己組織化マップ上でのそれぞれの特徴ベクトルを成分毎に棒グラフで示したものである。このように、ノードの属性によってコミュニティに分割できている。この方法は、上図に示すグラフィックアウトを必要としない。このため、大規模ネットワークへの適用が容易という利点があり、この点に関しての内外での評価が高い。

また、時間変化するネットワークに関しては、自己組織化マップの時刻間の特徴ベクトルを利用することでコミュニティを移動するノードを顕在化できる。航空ネットワークにこの方法を適用した例を図6に示す。左図は2000年、右図が2010年のものである。コミュニティ抽出は、2000年のネットワークから得られた特徴ベクトルによって行っている。指定したノード(右上角のコミュニティに属している)から各コミュニティ内のノードへのリンク数の変化を図示したものである。このノードを中心としたメソ構造の変化が健在化できている。信頼度算出への応用は今後の課題であるが、信頼度が十分に伝播するノード、あるいは伝播しないノードに対するリンクの時間変化の影響が、コミュニティ構造の観点から明らかにできると考えている。

0	3	0	0	0	0	16	0	0	0	0	0	19
0	0	0	0	0	0	3	0	0	0	0	0	8
0	0	0	0	0	1	12	0	0	0	0	0	2
0	1	0	0	0	0	4	0	0	0	0	0	1
0	0	0	0	0	0	5	0	0	0	0	0	2
1	0	0	0	0	2	2	0	0	0	0	0	0
2	3	1	4	1	6	0	0	0	0	0	0	0

図6 メソ構造の時間的変化

リンク情報の不完全性という問題に対しては、部分情報から効果的に欠損情報を補うために開発した前処理付きリンク予測手法を用いた。また、ノード情報の不完全性に対しては、ライングラフの考えを用いてノード予測へと拡張できることを示した。

前処理付きリンク予測手法、およびノード予測手法と、TrustRank、およびBadRank法を融合し、予測されたリンクとノードの追加によって信頼度が大きく変化するノードを発見する方法を構築した。これにより、欠損したノードとリンク情報の推定とともに、注目すべきノードが限定され、効率的にノード上の情報が精査できることを示した。

(3)人間の認知特性を考慮した情報提示の基礎として、視線計測実験によって解釈までを考慮した可視化の在り方を考察した。はじめに、視覚認知モデルを用いた視覚情報の領域

分割法を構築した。トラックバックをリンクとするプログネットワークの可視化図に適用した例を図7に示す。左図に対する分割例が右図である。

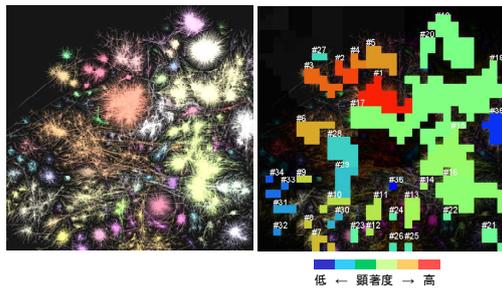


図7 視覚認知モデルを用いた領域分割

また、人手によるグラフィアウト変更操作時の視線変化を計測し、注目点とネットワークの特微量の関係を明らかにすることで、視認性の高いグラフィアウトに必要とされる要件を明らかにした。表1に結果の一例を示す。20ノードのネットワークに対して、手作業によるレイアウトの変更を行い、その際の視線データを分析し、ノードに対する視覚認知的な重要度を求めている。このような方法で人間の認知特性を考慮したネットワーク可視化法を検討した研究はなく、可視化研究としての貢献も大きいと考えられる。

表1 ノードに対する中心性指標と視線データ分析から得られる特微量

No.	統計的指標			クラスター数	視線ページランク			
	次数	媒介性	近接		円形		力学	
1	0.21	0.01	0.45	0.24	0.058	0.126	0.162	0.076
2	0.37	0.27	0.54	0.21	0.104	0.046	0.068	0.039
3	0.26	0.09	0.48	0.23	0.031	0.016	0.036	0.041
4	0.16	0.02	0.4	0.29	0.155	0.117	0.112	0.053
5	0.42	0.32	0.58	0.29	0.047	0.051	0.043	0.052
6	0.21	0.02	0.4	0.14	0.026	0.01	0.046	0.023
7	0.16	0.01	0.39	0.14	0.108	0.11	0.009	0.044
8	0.37	0.28	0.58	0	0.102	0.054	0.06	0.152
9	0.32	0.26	0.54	0.14	0.034	0.126	0.021	0.046
10	0.21	0.02	0.42	0.14	0.042	0.02	0.086	0.103
11	0.26	0.03	0.44	0.14	0.052	0.063	0.003	0.006
12	0.21	0.02	0.41	0.29	0.07	0.061	0.007	0.027
13	0.21	0.02	0.46	0.1	0.031	0.047	0.071	0.032
14	0.16	0.01	0.42	0.38	0.011	0.021	5E-04	0.003
15	0.21	0.01	0.4	0.29	0.038	0.011	0.003	0.018
16	0.11	0.01	0.35	0.19	0.035	0.024	0.019	4E-04
17	0.16	0.03	0.4	0.19	0.024	0.018	0.076	0.068
18	0.16	0.01	0.4	0.29	0.017	0.014	5E-04	0.005
19	0.21	0.02	0.42	0.19	0.009	0.042	0.033	0.071
20	0.16	0.01	0.4	0.14	0.006	0.022	0.093	0.141

さらに、ネットワークの統計的指標に基づく新しい力学モデルによるグラフィアウト手法を構築し、コミュニティ構造の表出化に成功した。

要素技術に関しては、当初計画以上の成果を得たと考えている。一方、全体システムの構築に関しては、情報取得システムの基礎となるデータ操作やネットワーク可視化など

の手法を Java による基本ライブラリ群として整備し、Processing を用いて統合を試みたが、プロトタイプシステムの試作に留まっている。研究を深化させていく中で、当初計画では考えていなかった要素技術の開発が必要になったためである。このため、研究期間内で全体システムが構築できたとはいえない。今後の課題である。

しかしながら、開発した要素技術の中には単独で利用できるものも多い。プロトタイプシステムを部分的に利用したものではあるが、効率的、かつ頑健な電力網の設計への応用、TrustRank ではなく PageRank による重要度を用いたものであるが、視線分析への応用を試み、提案手法の有用性を確認している。これにより、要素技術に関しては普及の目処がつかないと考えている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 30 件)

- ① 大森崇弘, 白山 晋, 視線計測を用いたグラフィアウトの視認性評価, 第9回ネットワーク生態学シンポジウム予稿集, 査読無, 2012, pp.108-113
- ② Nagata, K., Shirayama, S., Method of Analyzing the Influence of Network Structure on Information Diffusion, Physica A, 査読有, Vol.391, 2012, pp. 3783-3791
- ③ Yuasa, T., Shirayama, S., A New Analysis Method for Simulations Using Node Categorizations, Social Network Analysis and Mining, 査読有, Vol.2, No.2, 2012, pp.189-196
- ④ 湯浅友幸, 白山 晋, 複雑ネットワークに対するノード分類法とその応用, 人工知能学会論文誌, 査読有, Vol.27, No.3, 2012, pp.111-120
- ⑤ Egawa, A., Shirayama, S., A Method to Construct an Importance Map of an Image Using the Saliency Map Model and Eye Movement Analysis, Proceedings of 2012 Eye Tracking Research & Applications, 査読有, 2012, pp.21-28
- ⑥ Tsukamoto, E., Shirayama, S., A study of the relationship between scale-freeness and evolution of Cooperation, Int. J. Bio-Inspired Computation, 査読有, Vol.3, No.3, 2011, pp.142-150
- ⑦ 小林直樹, 白山 晋, ネットワーク構造の局所的な変化と同期現象の関係, コンピュータソフトウェア, 査読有, Vol.8, No.4, 2011, pp.333-343
- ⑧ Nagata, K., Shirayama, S., Analysis

Method of Influence of Potential Edge on Information Diffusion, Procedia CS, 査読有, Vol. 4, 2011, pp. 241-250

- ⑨ 塚本 鋭, 白山 晋, 協調を促進させるベキ乗指数について, 人工知能学会論文誌, 査読有, Vol. 26, No. 1, 2011, pp. 34-41
- ⑩ 塚本 鋭, 内田 誠, 白山 晋, 協調の進化に与える初期ネットワーク構造の影響人工知能学会論文誌, 査読有, Vol. 24, No. 5, 2009, pp. 397-404
- ⑪ Tsukamoto, E., Shirayama, S., Influence of the variance of degree distributions on the evolution of cooperation in complex networks, Physica A, 査読有, Vol. 389, 2010, pp. 577-586
- ⑫ Egawa, A., Shirayama, S., Generation of an Importance Map for Visualized Images, Lecture Note in Computer Science, 査読有, No. 5875, 2009, pp. 135-146

[学会発表] (計 10 件)

- ① 白山 晋, ノードの類似性が情報伝播に与える影響について, 人工知能学会第 24 回全国大会, 2011 年 6 月 1 日~3 日, 岩手県盛岡市・アイーナ岩手県民情報交流センター
- ② 白山 晋, 複雑ネットワークに対する新たなノード分類法とその応用, 第 8 回ネットワーク生態学シンポジウム, 2012 年 3 月 15 日~16 日, 神奈川県藤沢市・慶応大学湘南キャンパス
- ③ 白山 晋, 陽なコミュニティの情報伝播に与える影響, 人工知能学会第 24 回全国大会, 2010 年 6 月 9 日~11 日, 長崎県長崎市・ブリックホール
- ④ 白山 晋, ネットワーク上の現象における素過程について, 日本行動計量学会第 38 回大会, 2010 年 9 月 22 日~25 日, 埼玉県さいたま市・埼玉大学

6. 研究組織

(1) 研究代表者

白山 晋 (SHIRAYAMA SUSUMU)
東京大学・大学院工学系研究科・准教授
研究者番号: 10322067

(2) 研究分担者

なし

(3) 連携研究者

なし