

ネットワークの応答信号に基づく構造解析法の提案

佐藤 進也^{1,a)}

受付日 2016年9月27日, 採録日 2017年3月3日

概要: 本論文では, ネットワークの構造的特徴をとらえるための新しい手法を提案する. 提案手法では, ネットワークのあるノードから信号を入力し, リンク経由で伝播させ, 入力ノードに到達した応答信号を観測し, そこから構造の特徴を読み取る. 提案手法によって実際に構造を識別, 探索できることを人工的なネットワークを使った実験により示す.

キーワード: 構造解析, 信号処理, フーリエ変換, 潜在トピック, Labeled LDA

Proposal of Structural Analysis Method Based on Response Signals of Networks

SHIN-YA SATO^{1,a)}

Received: September 27, 2016, Accepted: March 3, 2017

Abstract: This paper proposes a new method for grasping structural characteristics of networks. In this method, a signal is input at a node of a given network, and the signal propagates through the network and returns to the input node through various routes. The returning signal at the input node, which is called a response signal, can be thought of as the response of the network to the input signal. The response signal can be used as a representation of the structural characteristics of the network. The paper also describes experiments using artificial networks for verifying the effectiveness of the proposed method. The result shows that the proposed method is actually useful for classifying and searching network structures.

Keywords: structural analysis, signal processing, Fourier transform, latent topic, labeled LDA

1. はじめに

Web や脳あるいは人間社会のような, 互いに関係する複数の要素からなるシステムは, 要素をノード, 関係をリンクで表すことで, ネットワークとしてとらえることができる. そのネットワークの構造的特徴は, システムの振舞いを理解するための重要な手がかりとして利用できる. 研究対象をネットワークとしてモデル化し, その構造を解析するアプローチは, 様々な分野で多くの知見をもたらしてきた [1], [2], [3].

ネットワークの構造的特徴をとらえる代表的な指標に固有パス長や次数分布がある. これらはネットワークの構

造全体をとらえ, そのおおまかな特徴を示すものである. 一方, 一般に実ネットワークの構造は, 複雑ネットワークという名前が示すように, 多くの場合均質ではない. つまり, 注目する部分が違えばその近傍の構造もまた異なっている可能性がある. さらに, いくつかの特定の構造を持った部分がシステムの機能上特別な役割を果たしていることが分かっている [1], [4]. よって, システムを理解する目的でネットワークのつくりを理解するためには, 全体の特徴を把握するとともに, ネットワークを構成するそれぞれの要素 (部分ネットワーク) の局所的構造を把握することが重要である.

部分ネットワークの構造をとらえる観点, 基準は複数ある. たとえば, コミュニティ解析 [5], [6] の場合は相対的な稠密性が基準となる. ネットワークを稠密な部分 (コミュニティ) に分けることが目的であり, コミュニティ内のつ

¹ 日本工業大学情報工学科
Department of Computer and Information Engineering
Nippon Institute of Technology, Miyashiro, Saitama 345-8501, Japan

^{a)} shin-ya.sato@acm.org

なかり方は問わない。一方、ネットワークモチーフに基づく解析 [1] では、リンクの向きも含むつながり方のパターンそれぞれを異なるものとして扱う。さらに、コミュニティ解析とネットワークモチーフ解析の中間に位置づけられる構造のとらえ方もある。たとえば、Web では、重要な情報を提供するページ（オーソリティ）とそれらを多く参照するページ（ハブ）の集合が稠密な 2 部グラフを形成していることが知られている [4]。このような構造を持つ部分ネットワークの探索（ネットワーク全体の中から条件に合致する部分ネットワークを見つけ出すこと）では、ノードが稠密な 2 部グラフを形成しているという条件が重要であり、その規模（ノード数）は問わない。

このように、部分ネットワーク構造のとらえ方は 3 つに大別できる。本論文では、このうち第 3 の観点で部分ネットワークの構造を識別し、探索する手法を提案する。先に述べた稠密な 2 部グラフに関しては、それを探索するアルゴリズムはすでに開発されている [4]。このアルゴリズムは、特定の構造に特化したものであるが、本論文で提案する手法は、複数の構造に適用可能である点が特徴的である。

提案手法では次のようにしてネットワークの構造をとらえる。まず、ネットワークのあるノードから信号（時系列）を入力し、リンク経由で伝播させる。信号はやがて複数の経路を経て入力ノードに到達する。このとき得られる信号を応答信号と呼ぶ。提案手法では、時系列解析の手法を応用し応答信号からネットワークの構造的特徴を読み取る。本論文では、提案手法によって実際に構造を識別、探索できることを人工的なネットワークを使った実験により示す。なお、提案手法は、無向あるいは有効ネットワーク双方に適用可能であるが、本論文では無向ネットワークの場合のみを議論する。

本論文では以下の順序で議論をすすめる。2 章では、関連研究について述べ本研究の位置付けを明らかにする。3 章でネットワークから応答信号を得る方法を示す。4 章で応答信号の特徴について述べ、応答信号から構造的特徴を読み取る方法を 5 章で示す。さらに、提案手法の有効性を確認するために行った 2 つの実験について 6 章で述べ、その結果について 7 章で議論する。

2. 関連研究

2.1 ネットワークモチーフ

ネットワークモチーフとは、解析対象のネットワークにおいて有意に頻出する部分グラフである。ネットワークモチーフを発見するためには、まず、整数 n ($n \geq 3$) を選び、ノード数 n のグラフを列挙する。そして、それぞれのグラフについて、解析対象のネットワークにおける出現頻度を数え上げ、高頻度であるものをモチーフとして採用する。

ネットワークモチーフは、遺伝子転写制御ネットワークや神経ネットワークなどの実ネットワークを解析する手段

として用いられている [1], [2]。遺伝や神経の働きなどをシステムとして理解することを目的としているこれらの研究では、ネットワークモチーフというネットワークの部分的構造とシステムの機能との対応関係が明らかにされてきた。構造と機能の対応関係が分かるということは、構造に関する情報を機能に関する情報に変換できるということである。たとえば、ネットワーク中のある領域に特定の構造が認められることを根拠にその領域は特定の機能を担っている、といった推測ができるようになる。このアプローチは、遺伝や脳の情報処理といった複数の機能が混在している複雑なシステムの理解に寄与すると考えられる。

本論文で提案する手法もネットワークの中に特定の構造を発見するものであり、ネットワークモチーフと同様な「システムの機能の理解」への応用を意図している。ネットワークモチーフでは、単一のグラフで表現される構造に着目している。一方、本研究では、構造クラスとでも呼ぶべき、複数のネットワークで共有される、より抽象度の高い構造的特徴に着目する。そして、与えられたネットワークの中に構造クラスを発見する（より正確に言えば、構造クラスに属する構造を発見する）手段を提供する。これにより、サイズにとらわれることなく構造的特徴に焦点を当てた構造の探索（たとえば、任意サイズの完全グラフの探索）が可能になる。

2.2 ネットワークトモグラフィ

ネットワークトモグラフィとは、通信ネットワークにおける end-to-end のパケットの動きを計測することで、ネットワークの内部状態を推定する技術の総称である [7]。推定対象としては、輻輳地点、損失・遅延などの原因となる異常リンクや、パケットの経路（通信経路）があげられる。複数地点での観測結果を表すベクトルを \mathbf{Y} とし、 \mathbf{X} をリンクの状態を表すベクトルとすると、この 2 者の関係は通信経路を表す行列 \mathbf{A} を使って

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

と表すことができる [7]。輻輳地点や異常リンクの推定は \mathbf{X} を、通信経路の推定は \mathbf{A} を推定することに相当する。

本論文で提案する手法は、ネットワーク上の信号伝達状況を観測しネットワーク解析の手がかりとするものであり、ネットワークトモグラフィとは、解析手段に類似性がある。そして、本研究の解析対象はノード間のつながりであり、上式では \mathbf{A} が相当する。ただし、ネットワークトモグラフィが未知の \mathbf{A} を推定するのに対して、本研究ではノード間のつながりを示す隣接行列は既知である点が異なる。本研究の目的は既知のつながりの中に特定の部分構造を発見することであり、前節で述べたネットワークモチーフ解析やコミュニティ抽出と同じ目的を持つ研究として位置づけることができる。

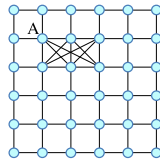


図 1 2次元格子に埋め込まれた完全グラフ

Fig. 1 Complete network embedded in a two dimensional lattice.

2.3 ネットワークの類似性判定

今までにネットワークの類似度を測る指標は複数提案されてきている [8], [9], [10], [11]. Berlingerio らは NetSimile と呼ばれる類似性尺度を提案した. これは, クラスタ係数や次数分布など既存の様々なネットワーク指標を組み合わせて得られる「ネットワークのシグネチャ」を比較することで類似性を判断する [8]. これは, ネットワーク全体の特徴を比較するもので, 本論文の主題とは異なる観点でネットワークの構造をとらえるものである.

2.4 ノードの構造的類似性

ノードの構造的類似性 [12], [13] とは, 2つのノードが類似しているか否かを, 他のノードとのつながり具合の類似性に基いて判断するものである. 定義から, 構造的に類似している2つのノードの近隣は互いに構造的に類似していることになるので, ノードの構造的類似性の判定法は類似構造の探索に寄与すると考えられる. しかし, 混在する異なる構造を識別することは難しい. たとえば, 図 1 に示したような, 2次元格子中に完全グラフが混在している状況を考える. このとき, ノードの構造的類似性の観点からすると, ノード A の近傍は2次元格子とも完全グラフとも異なる第3の構造として認識される. 一方, 6.2 節で示すように, 提案手法では完全グラフを2次元格子から分離して認識することが可能である.

2.5 ネットワークの可制御性

一般に, システムを任意の初期状態から目的の状態に有限時間内で遷移させることができるとき, そのシステムは可制御であるという. 可制御性を調べるための問題設定の1つとして, ネットワークとして抽象化したシステムにおける信号の伝播を調べたものがある [14]. この設定では, 「ある時刻において各ノードがどのような信号を受け取っているか」ということが状態に対応する. そして, 可制御性であるとは, 適当なノードとそこから入力する信号を選び, それを伝播させ, 有限時間内に各ノードで所望の信号を受け取るようにできる, ということに相当する. これは, 応答信号から入力信号を決定するという問題であり, 本論文で議論する内容の逆問題にあたる.

3. ネットワークの応答信号

1章で述べたように, 提案手法では, ネットワークへの信号入力に対する応答から構造的特徴を読み取る. 本章では, この応答信号を定義する.

3.1 ネットワークにおける信号の伝播

本論文では, ネットワークにおける信号の伝播を以下のように定式化する. N を無向ネットワークとし, そのノードの集合を $\{v_i\}$ とする. また, $A = (a_{ij})$ を隣接行列とする. 信号は時系列 (すなわち数量の列) として表現する. 信号の伝播とは, この数量が, 時間経過にともないリンクを介して伝わっていくことである. より具体的には, 時刻 $t-1$ にあるノードが受け取った数量は, 時刻 t にその近傍にある他のノードたちに分配される. 各ノードで時刻 t に受け取る数量を表すベクトルを q_t とし, 確率行列 B を

$$B = (b_{ij}), b_{ij} = a_{ij} / \sum_i a_{ij}$$

とすると, 信号の伝播は

$$q_t = Bq_{t-1}$$

と書ける.

いま, ノード v_k から信号 $s = \{s_t\}$ が入力されるとしよう. このとき, v_k は, 近傍ノードから受け取った数量とともに入力 s_t を, 次時刻に分配する. これは, k 番目の要素だけが1である単位ベクトル e_k を使って,

$$q_t = B(q_{t-1} + s_{t-1}e_k)$$

と書ける. ノード v_m で観測される信号を $r = \{r_t\}$ で表すと, 定義から r_t は

$$r_t = e_m q_t$$

と書ける. r はネットワーク N , 入力信号 s , 入力ノード v_k , そして観測ノード v_m に依存するので, 厳密に言えば, $r(N, s, v_k, v_m)$ と書くべきである. ノード v_k で観測される, 入力 s に対する応答信号 $\epsilon(N, s, v_k)$ は, 上記の式を使って,

$$\epsilon(N, s, v_k) = r(N, s, v_k, v_k)$$

と定義される.

4. 応答信号の特徴

本章では, いくつかの具体例を示しながら応答信号の性質・特徴について述べる. なお, 本論文では, ネットワークを, 構造を示す記号とノード数を組み合わせて表記する. ノード数4の完全グラフ (complete graph) の場合, 完全グラフを示す「C」と4を組み合わせ「C₄」と表記する.

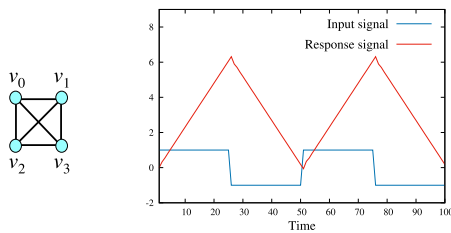


図 2 C_4 の構造と入力信号およびその応答信号
Fig. 2 Structure of C_4 and its response signal.

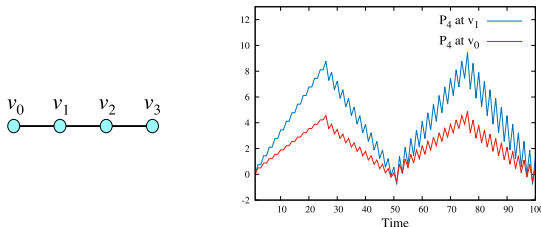


図 3 P_4 の構造と v_0, v_1 それぞれで観測した応答信号
Fig. 3 Structure of P_4 and its response signals observed respectively at v_0 and v_1 .

4.1 入力信号

入力信号 s には任意の時系列を用いることができるが、本論文では矩形波を使った場合について議論する。矩形波は周期 p をパラメータとして以下のように定義できる：

$$s_t = \begin{cases} 1 & t \bmod p < p/2 \text{ である場合} \\ -1 & \text{その他の場合} \end{cases}$$

以降、本論文では $p = 50$ として議論をすすめる。

4.2 完全グラフ

まず、4 ノードからなる完全グラフ C_4 と、その応答信号の波形を入力信号とともに図 2 に示す。完全グラフでは、すべてのノードが相互につながっており、入力信号はそれぞれに等しく分配される。そして、波形は、矩形波を積分して得られる三角波に類似したものとなっている。ここでは例としてノード数が 4 の場合を示したが、 C_k の応答信号は k の値にかかわらず三角波様の波形を示す。

4.3 パスグラフ

次に、4 ノードのパスグラフ (path graph) P_4 と、その応答信号を図 3 に示す。応答信号は、ノード v_0 と v_1 それぞれで観測したものを示した。観測ノードに依存して異なる波形を示しているが、三角波様の変動に高周波の変動を組み合わせた特徴的な変動パターンが共通している。

4.4 異種グラフの組合せ

第 3 の例として、図 4 に示した、パスグラフ P_{10} と完全グラフ C_3 を組み合わせたネットワークをとり上げる。

図 5 は、それぞれノード v_0 と v_8 で観測した応答信号である。共通してパスグラフの例で見られた高周波変動が認

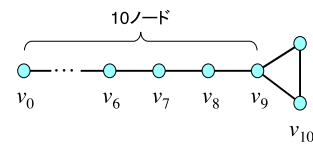


図 4 P_{10} と C_3 を組み合わせたネットワーク
Fig. 4 Network made up of P_{10} and C_3 .

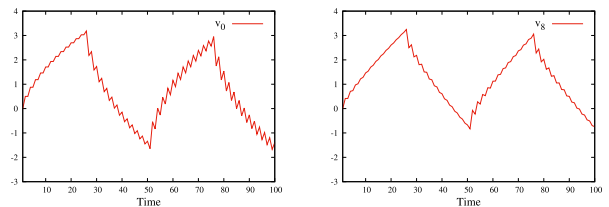


図 5 P_{10} と C_3 を組み合わせたネットワークの応答信号
Fig. 5 Response signal of the network made up of P_{10} and C_3 .

められるが、振幅の大きさに差があるのが分かる。 v_0 に比べ、完全グラフに近い v_8 では振幅が弱くなっており、完全グラフの応答信号により近い波形となっている。

5. 応答信号のスペクトル解析による構造解析

前章で示した例は、ネットワークの構造と応答信号の変動パターンとの間に関連性があることを示唆している。本章では、この知見をもとに考案した、応答信号からネットワークの構造を推定する方法について述べる。

5.1 構造解析のアイデア

本論文で提案する構造解析の基本的なアイデアを図 6 を使って説明する。この例では完全グラフの構造をとらえることを目標とする。そのために、完全グラフの構造を持つネットワーク (A), (B) と、異なる構造を持つネットワーク (C), (D) を用意し、それぞれの応答信号を観測しその周波数分布 (スペクトル) を計算する。これらのスペクトルにおいて、(A) と (B) に共通して認められるが (C) や (D) には見られない分布のパターンが、完全グラフの構造に対応するものであると考える。実際にこのような分布のパターンを見つけるためにはトピックモデルを利用する。すなわち、トピックモデルによって発見される潜在トピックを目的の構造に対応するものとする。任意のネットワークの構造を推定は、その応答信号のスペクトルを学習により得られたモデルに照合することで行う。

5.2 LDA による構造推定

提案手法では、潜在トピックの学習と推定に Latent Dirichlet Allocation (LDA) [15] あるいは Labeled Latent Dirichlet allocation (L-LDA) [16] を用いる。LDA はトピックモデルの 1 つであり、文書に潜在するトピックの推定などに用いられる。LDA は、文書を、そこに出現する語の集合 (bag-of-words) として抽象化する。そして、その

bag-of-words は、文書に潜在する複数のトピックがそれぞれ固有の確率分布に従い語を生成した結果としてとらえる。このモデルに基づき、文書集合において観測される語の出現状況から潜在トピックを推定する。推定の結果、各文書を構成する潜在トピックの割合が、潜在トピック上の確率分布 $\theta(\tau)$ として得られる。 $\theta(\tau)$ が大きい τ がその文書の主要なトピックである。

一方、一般に、時系列をフーリエ変換することにより周波数分布が得られるが、周波数分布は、周波数成分の強度を頻度と見なすことにより、bag-of-frequencies (BoF) と見なすことができる。BoF に対して LDA を適用することにより、周波数成分を生成する潜在トピックを推定することができる。実際に、日常生活で観測される環境音の周波数分布 (あるいはそこから導出される情報) に LDA を適用し、特徴的な音を生み出している状況や原因を潜在トピックとして抽出するという取り組みがある [17], [18], [19]。本研究では、応答信号の特徴はネットワークの構造に由来するという仮説をおき、構造抽出のためにこれらの取り組みに倣って LDA を利用する。応答信号に対して LDA を適用して得られた潜在トピックが、前節で述べた、ネットワーク構造に対応していると考えられるトピックである。本論文ではこのトピックを「構造トピック」と呼ぶことにする。 $\theta(\tau)$ は、構造トピックの確率分布である。与えられたネットワークの構造は、その応答信号を観測し LDA により $\theta(\tau)$ を計算することにより推定できる。 $\theta(\tau)$ の値が大きな構造トピック τ が、そのネットワークの主要な構造であると考えられる。

L-LDA は LDA を教師あり学習にしたものである。学習対象のデータ (本論文での議論の場合、BoF) に潜在トピックの存在を示すラベルを付与することで、間接的に潜在トピックを定義することができる。図 6 の例でいえば、(A), (B) の BoF だけに完全グラフ構造を示すラベル (たとえば C) を付与することにより、これらに固有な潜在トピック τ_C を見つけ出すことができる。(A), (B) という、構造的

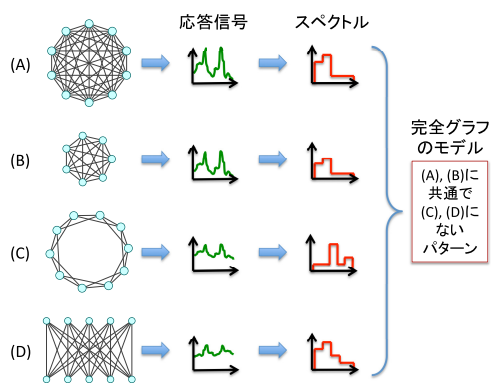


図 6 応答信号に基づく構造解析の考え方

Fig. 6 Idea of the structural analysis based on response signals.

特徴は共通している大きさが異なるネットワークに同一のラベルを付与したデータか抽出された τ_C は、ネットワークの大きさに非依存な構造的特徴に対応していると考えられる。そして、あるネットワークの応答信号から計算した構造の確率分布 $\theta(\tau)$ において、 $\theta(\tau_C)$ の値は、大きさを度外視した「完全グラフらしさ」の度合いを表していると考えられる。

6. 評価実験

6.1 構造の推定

提案手法の有効性を確認するために構造推定の実験を行った。本節では、その内容と結果について述べる。

本実験では、完全グラフやパスグラフなど 7 種類の構造を選び、それぞれの構造を持つ複数のネットワークを作成してそれらの構造を推定した。構造の推定はあらかじめ決められている 7 種類の中からの選択なので、クラス分類問題である。実験に使用した構造とネットワークを表 1 に示す。また、いくつかのネットワークの具体例を図 7 に示す。

構造の特徴を学習させるため、各構造ごとに 10 程度の応答信号が得られるよう複数のネットワークを用いた。たとえば、表の完全グラフの行にノード数が 6~15 とあるが、これは完全グラフの構造を持つネットワークとして、 $C_6 \sim C_{15}$ の 10 個を実験に使用したことを示している。また、4.3 節で示した例のように、ノードの選び方によって応答信号が異なる場合がある。観測ノードに非依存な構造の

表 1 構造推定実験に使用した構造

Table 1 Structures used in the structure estimation experiment.

構造	記号	ノード数
完全グラフ	C	6~15
パスグラフ	P	4~7
リンググラフ	R	5~14
スターグラフ	S	4~8
完全 2 部グラフ	CB	6~24 (偶数)
1 次元格子	$1DL$	6~15
2 次元格子	$2DL$	9(3×3)~25(5×5)

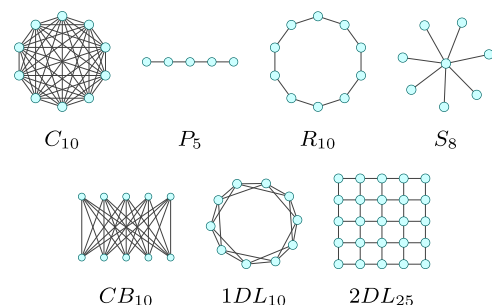


図 7 構造推定実験に使用したネットワークの例

Fig. 7 Some networks used in the structure estimation experiment.

表 2 構造ごとの $\eta(\tau)$ の平均値
Table 2 Average of $\eta(\tau)$ for each structure.

	C	P	R	S	CB	1DL	2DL
C	1.0000	0.0315	0.0840	0.0445	0.0424	0.2377	0.0273
P	0.5948	0.7220	0.6752	0.6919	0.6716	0.6532	0.8199
R	0.2267	0.3474	0.5797	0.2019	0.1490	0.4538	0.4787
S	0.5152	0.2517	0.3929	0.9450	0.8906	0.4645	0.2221
CB	0.4794	0.2336	0.3694	0.9855	0.8357	0.4423	0.2113
1DL	0.3620	0.0330	0.1981	0.0186	0.0218	0.9195	0.0341
2DL	0.2834	0.6323	0.4429	0.3201	0.2436	0.3525	0.9075

特徴を学習させるため、このような場合には、それぞれの応答信号を学習に用いた。たとえば、 P_4 というネットワークの場合、異なる 2 つのノード (図 3 の v_0, v_1) それぞれで観測した応答信号を使用した。このようにして、表 1 にあげたネットワークから、総計で 74 の応答信号を得た。これら各信号から BoF を計算し、構造のラベルを付与して評価用データを作成した。応答信号にはネットワーク構造の特徴とともに入力信号の特徴も反映されていると考えられる。そこで、構造のラベルに加え入力信号に対応するラベルを導入し、すべての応答信号に付与した。L-LDA のトピック数はラベル数に等しい。この実験では、構造に対応するものと入力信号に対応するものを合わせて 8 の潜在トピックを推定する。評価は、74 のデータのうち、73 個のデータから構造のモデルを学習し、残りの 1 個の構造を推定する leave-one-out 交差検証により行った。

L-LDA による推定の結果得られる $\theta(\tau)$ の値で降順にソートし、第 1 位となった構造トピック τ を推定結果として採用した場合、正解率は $38/74 = 0.5135$ であった。この評価方法では 1 位以外を切り捨てているが、たとえば 2 位となった構造トピックが正解で $\theta(\tau)$ の値が 1 位と僅差、というような場合もありうる。後で議論するように、識別すべき 2 つの構造に類似性がある場合 (たとえば、一方が他方を包含するなど) には特にこのような状況が発生する。そもそも構造どうしの境界は曖昧で明確な分離が難しい。このことを考慮し、2 位以下も適切に評価すべきである。そこで、以下で定義される $\eta(\tau)$ で評価することを考える。

$$\eta(\tau) = \frac{\theta(\tau)}{\max_{\tau} \theta(\tau)}$$

これは、最も可能性があると推定された 1 位の構造を基準として推定の確からしさがどの程度であるかを評価する尺度であり、 $0 \leq \eta(\tau) \leq 1$ である。推定された (すなわち 1 位になった) 構造トピックが正解である場合に $\eta(\tau) = 1$ となる。すべての推定結果それぞれから算出した $\eta(\tau)$ の平均は 0.8426 であった。

表 2 は、構造ごとに分けて $\eta(\tau)$ の平均をとったものである。各行ごとに、与えられた構造を持つネットワークが、提案手法によりどのような構造として推定されたのかを

$\eta(\tau)$ の平均値で示している。正解には下線を施し、値が最も大きいものをボールド体で表した。この表から、正答がどの程度得られているか、それぞれの構造が他のどの構造と混同されやすいのか、といった情報を読み取ることができる。たとえば、 $C_6 \sim C_{15}$ の構造を推定した場合には、いずれも正解である完全グラフが第 1 候補に上がり、その 1/4 ほどの確からしきで 1 次元格子が第 2 候補となっている。一方、バスグラフ構造の行を見ると、最低でも 0.5948 となっており、その他の構造と間違われやすいことが分かる。

6.2 構造の探索

次に、提案手法によって、構造が混在している状況下で目的とする構造を識別できるかを実験により確認した。

6.2.1 完全グラフの探索

まず、2.4 節で言及したような、2 次元格子中の完全グラフを見つけ出すという問題を提案手法を用いて解き、正答数を評価した。問題の設定と解法の手順は次のとおりである。

- (1) 2 次元格子 L を選ぶ。
- (2) L に属している近接する k 個のノードを選ぶ。このノードの集合を K とする。 K に属するノードが完全グラフとなるように L にリンクを追加する。得られた完全グラフを含む 2 次元格子のネットワークを G とする。
- (3) 前節で使った 74 種類のラベル付きデータから探索対象である C_k を除いたデータで構造のモデルを学習する。 C_k を除くのは、交差検証において学習用データと評価用データを分離することと同じ理由である。
- (4) G の各ノード v で応答信号を観測し、(3) で作成したモデルに基づき確率分布 $\theta(\tau)$ を計算する。
- (5) 完全グラフに対応する構造トピックを τ_C とする。 $\theta(\tau_C)$ は v の周辺の「完全グラフらしき」を表していると考え、 $\theta(\tau_C)$ の値が大きい上位 k ノード M を解 (完全グラフの構成要素と推定されたノードと集合) とする。このとき、正解ノード数は $|K \cap M|$ である。

実際の実験には、 L として 6×6 の 2 次元格子 $2DL_{36}$ を使い、 k は 6 とした。このとき、 G としては構造が異なる 6 つのネットワークが得られる (図 8, 図 1 に示したネッ

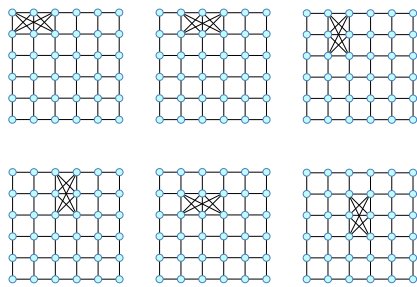


図 8 実験に使用した 6 種類のネットワーク G

Fig. 8 Six types of network G used in the experiment.

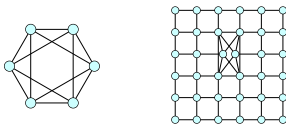


図 9 1次元格子 $1DL_6$ とその $2DL_{36}$ への埋め込み

Fig. 9 One dimensional lattice $1DL_6$ and its embedding to $2DL_{36}$.

トワークは、これらのうちの 1 つである)。それぞれについて上記手順で M を選出して正解ノード数を調べた。

実験の結果、正解ノード数の平均は 4.1667 であった。一方、36 ノードから無作為に 6 ノード選んだ場合、その中に含まれる正解ノード数の期待値は 1.0 である。本手法で得られた結果と無作為抽出との差が有意であることを確認するため、Wilcoxon の順位検定を行った。その結果、 p 値は 0.0013 であり、危険率 1% で有意に多くの正解が得られていることが確認された。

6.2.2 1次元格子の探索

さらに、完全グラフの代わりに 1次元格子 $1DL_6$ を $2DL_{36}$ に埋め込み同様な実験を行った。 $1DL_6$ と、それを $2DL_{36}$ に埋め込んで得られるネットワークの例を図 9 に示す。 $1DL_6$ は完全ネットワーク C_6 から 3 本リンクを除いて得られるネットワークである。 $1DL_6$ を C_6 と比較すると、密度の点では、より $2DL_{36}$ に近い構造である。この 1次元格子を探索する実験を行った結果、正解ノード数の平均は 5.1667 であり、完全ネットワークを探索した場合よりも多くの正解が得られた。

7. 考察

5.2 節で述べたように、表 2 に示した数値は、構造の「らしさ」の度合いを表している。7 種類の構造のうち、パスグラフ構造、完全 2 部グラフ構造を除く 5 種類については正解の値が最も高く、構造を正しく把握できている。パスグラフ構造については 2次元格子らしさの値が最も高くなっている。加えて、他の構造らしさも比較的高い値を示している。これは、パスグラフの構造が単純であり、他の構造の構成要素となっていることが要因の 1 つと考えられる。完全 2 部グラフ構造の推定では、スターグラフと完全 2 部グラフらしさがともに高く、この 2 種類の構造を混同して

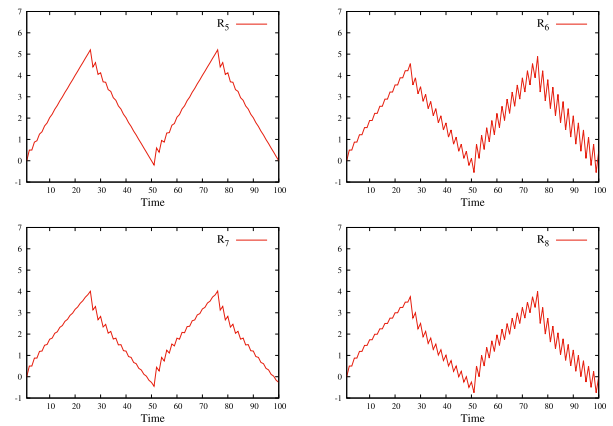


図 10 リンググラフにおける応答信号変動の違い

Fig. 10 Difference in fluctuation patterns of ring graphs.

いる状態と考えられる。実際、2 部グラフの 1 つのノードの近傍はスターグラフ構造をしており、このことが混同の一因と考えられる。パスグラフ構造、完全 2 部グラフ構造の結果は、推定対象構造に包含関係がある場合に、提案手法では推定の性能が低下することを示している。

一方、リンググラフ構造の推定結果をみると、正解の値は他より相対的に高いものの、値自体は 0.6 弱にとどまっている。この原因の 1 つに応答信号の多様性が考えられる。図 10 は R_5, R_6, R_7, R_8 の応答信号を示したものであるが、ノード数が偶数と奇数の場合で高周波変動の違いが認められる。ノード数が奇数、偶数それぞれの場合に共通する特徴がありながら、リンググラフ構造全体ではその特徴は共有できないという状況が、リンググラフ構造のモデル構築を難しくしていると考えられる。

6.2 節で述べた構造の探索で、完全グラフを探索対象とした場合、発見できた正解ノード数は平均 4.1667 個であり、無作為抽出の期待値の 4 倍を上回っているが、見つけ出すべき 6 ノードの 69% にとどまった。一方、1次元格子を探索対象とした場合には、正解数の平均は 5.1667 個であり、正解率は 86% を超えた。完全グラフは最も稠密な構造であり、その構造の発見は他の場合に比べて容易であることを期待していたが、その予想に反する結果が得られた。今回の実験で、探索の正解率は探索対象の構造によって異なるという事実を確認したが、さらに、探索対象が埋め込まれている周りの構造（この実験の場合は 2次元格子）、あるいは探索対象とその周りの構造の相対的關係にも依存している可能性がある。探索精度のこれらの構造への関連性については、さらに詳細な解析が必要である。

8. むすび

本論文では、ネットワークの構造的特徴をとらえるための新しい手法を提案した。提案手法では、ネットワークのあるノードから信号を入力し、リンク経由で伝播させ、入力ノードに到達した応答信号を観測し、そこから構造の特

徴を読み取る。構造のクラス分類と、構造が混在している状況下で目的とする構造の識別という2つの問題に提案手法を適用した結果、識別すべき構造間に類似性がある場合には正解率が低下するが、そのような関係がない場合には適切に識別できることを確認した。

類似構造は、言葉における同義語のようなものと考えられる。情報検索においてより適切な結果を得るために同義語に対処する処理が行われるように、提案手法による構造解析においても、識別すべき構造の類似性を考慮することにより、より精度の高い構造解析が可能になると考えられる。

参考文献

- [1] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network motifs: Simple building blocks of complex networks, *Science*, Vol.298, No.5594, pp.824–827 (2002).
- [2] Sporns, O. and Kötter, R.: Motifs in Brain Networks, *PLoS Biology*, Vol.2, No.11, pp.1910–1918 (2004).
- [3] Castillo, C., Donato, D., Gionis, A., Murdock, V. and Silvestri, F.: Know your neighbors: Web spam detection using the web topology, *Proc. 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.423–430 (2007).
- [4] Kleinberg, J.: Authoritative sources in a hyperlinked environment, *J. ACM*, Vol.46, No.5, pp.604–632 (1999).
- [5] Palla, G., Derényi, I., Farkas, I. and Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society, *Nature*, Vol.435, No.7043, pp.814–818 (2005).
- [6] Girvan, M. and Newman, M.E.J.: Community structure in social and biological networks, *Proc. National Academy of Sciences*, Vol.99, No.12, pp.7821–7826 (2002).
- [7] Castro, R., Coates, M., Liang, G., Nowak, R. and Yu, B.: Network tomography: Recent developments, *Statistical Science*, Vol.19, No.3, pp.499–517 (2004).
- [8] Berlingerio, M., Koutra, D., Eliassi-Rad, T. and Faloutsos, C.: Network similarity via multiple social theories, *Proc. 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp.1439–1440, ACM (2013).
- [9] Przulj, N.: Biological network comparison using graphlet degree distribution, *Bioinformatics*, Vol.26, No.6, pp.853–854 (2010).
- [10] ElGhawalby, H. and Hancock, E.R.: Measuring Graph Similarity Using Spectral Geometry, *Proc. 5th International Conference on Image Analysis and Recognition*, pp.517–526 (2008).
- [11] Kelmans, A.: Comparison of graphs by their number of spanning trees, *Discrete Mathematics*, Vol.16, No.3, pp.241–261 (1976).
- [12] Leicht, E.A., Holme, P. and Newman, M.E.J.: Vertex similarity in networks, *Phys. Rev. E*, Vol.73, pp.1–10 (2005).
- [13] Xu, X., Yuruk, N., Feng, Z. and Schweiger, T.A.J.: SCAN: A Structural Clustering Algorithm for Networks, *Proc. 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.824–833, ACM (2007).
- [14] Liu, Y.-Y., Slotine, J.-J. and Barabasi, A.-L.: Controllability of complex networks, *Nature*, Vol.473, No.7346, pp.167–173 (2011).
- [15] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet allocation, *Journal of Machine Learning Research*, Vol.31, pp.993–1022 (2003).
- [16] Ramage, D., Hall, D., Nallapati, R. and Manning, C.D.: Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora, *Proc. 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1*, pp.248–256, ACL (2009).
- [17] Kim, S., Sundaram, S., Georgiou, P.G. and Narayanan, S.: Audio Scene Understanding using Topic Models, *Proc. Neural Information Processing Systems (NIPS) Workshop* (2009).
- [18] Sato, S., Takahashi, M. and Matsuo, M.: Giving Context to Sounds Through Mediation of Physical Objects, *Proc. 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, pp.91–94, ACM (2013).
- [19] Kalmbach, A., Girdhar, Y. and Dudek, G.: Unsupervised environment recognition and modeling using sound sensing, *2013 IEEE International Conference on Robotics and Automation*, pp.2699–2704, IEEE (2013).



佐藤 進也 (正会員)

1986年東北大学理学部数学科卒業。1988年同大学大学院修士課程修了。同年日本電信電話株式会社入社。NTT未来ねっと研究所等を経て、2014年より日本工業大学情報工学科教授。博士(情報理工学)。協調作業支援、Web情報検索・マイニング、複雑ネットワーク等の研究に従事。訳書『スモールワールド』(ダンカン・ワッツ著、東京電機大学出版局、共訳)。ACM、ISOC、電子情報通信学会、人工知能学会各会員。