

# 最適化に基づく複雑ネットワークの指標の理解

Understanding topological metrics in complex networks based on optimized networks

土屋裕一 鈴木麗璽 有田隆也  
Yuichi Tsuchiya Reiji Suzuki Takaya Arita

名古屋大学 大学院情報科学研究科  
Graduate school of Information Science, Nagoya University

A number of metrics that characterize the topology of complex networks have been introduced. This paper aims to comprehensively understand the space of network topology by visualizing the two-dimensional view of the networks each of which has a maximized or minimized metric value. We use a genetic algorithm to optimize the network in terms of one of the metrics whose initial topology is a random network, and then use the self-organizing map to visualize the space of the optimized networks characterized by all of the metric values. We show several interesting properties of the metrics based on the generated map, and also examined the correlation between metrics.

## 1. はじめに

近年、盛んに研究されている複雑ネットワーク研究において、ネットワークの特にトポロジーに関する性質の定量化のために、最短経路長、クラスター係数、次数などに基づく様々な指標が導入されてきた。一例として、Newman は、共著者ネットワークは正、インターネット接続は負の次数相関を持っており、トポロジー的には異なった性質を持っていることを明らかにした [Newman 02]。

しかし、指標の中にはその値がネットワーク構造にいかに対応し反映しているかということが明確でないものもあり、また、多数の指標間の関係性は明確になっているとは言いがたい。ただし、従来研究として、Jamakovic らが電力網等の実世界のネットワークのデータを用いて指標間の相関関係を求め、実世界のネットワークに関するトポロジー解析に有用な指標のサブセットの提案を行っている [Jamakovic 08] ことは注目される。ここで、より根本的に言うならば、それら多数の指標がそれぞれ断片的に計測するネットワーク構造の全空間がいかなるものかということに関する我々の理解はまったく不十分である。

そこで、我々は、まず各指標について最大化あるいは最小化した特徴的構造を有するネットワーク群を人工的に作成してデータとして扱うことにした。現実世界に存在するネットワークのトポロジカルなデータは数が多くなく、また、それが現実に機能している場所に応じたバイアスが強くかかっているためである。次に、それぞれの特徴ネットワークの持つ全指標値をそれを特徴づけるベクトルと考え、それを2次元に削減して可視化する。また、同時に、個別の指標間の相関関係も計算する。これらを通じて、可能なネットワークが構成する構造空間の包括的な理解のための第一歩とする。

## 2. 手法

### 2.1 基本方針

まず、最適化を行う指標に関しては、最近の提案も含め、可能な限り多く扱うことを目指す。ノード単位などのマイクロな指標

連絡先: 連絡先: 土屋裕一、名古屋大学大学院情報科学研究科、〒464-8601、名古屋市千種区不老町、e-mail: yuichi.t@alife.cs.is.nagoya-u.ac.jp

に関しては、ネットワーク全体の平均、及び分散の両者を対象とする(一定値の場合は除く)。それを遺伝的アルゴリズムを用いて、それぞれを最大化、及び最小化した特徴ネットワークを作成する。作成されたネットワーク構造の可視化のために、自己組織化マップを用いることにした [Kohonen 82]。それぞれの特徴ネットワークが持つ全指標値を入力とするものである。このデータを用いて、同時に指標間の関係も計算する。得られたマップを基にして、随時、作成されたネットワークを個別に可視化しながら、各指標の持つ位置づけや互いの関係を理解する。

### 2.2 ネットワーク構造の各種指標による最適化

遺伝的アルゴリズムにおいて、遺伝子情報として、リンクの有(無)を遺伝子1(0)により、ノード数  $N$ 、リンク数  $L$  が一定のネットワークを表現する。世代交代は Minimal Generation Gap [佐藤 96] を用い、交叉、突然変異は花田らが導入したものをを用いる [花田 07]。交叉は親個体に共通するリンク、及び、規定リンク数に達するまでは非共通のリンクからランダムに選び、子個体に継承させるものであり、突然変異はランダムなリンクの付け外しを行い、リンク数が増(減)した場合は規定リンク数に達するまで、ランダムに除去(付加)するものである。

最適化の対象とする指標は以下のものとする。

ネットワーク単位の指標:

- 次数相関
- K-コアネス
- リッチクラブ

ノードレベルの指標をネットワーク単位にした指標:

- 最短経路長の平均と分散
- クラスタ係数の平均と分散
- ノードコアネスの平均
- 次数の分散
- 媒介中心性の分散

ここで、K-コアネス [Borgatti 03] は、サブネットワークの中でそれを構成するノードの次数がすべて  $K$  以上であるようなサブネットワークの最大サイズ、ノードコアネスは、各ノードが属する最大の K-コアネスの  $K$  値をノードの指標値として与えたときの全ノード平均を表す。また、リッチクラブ [Colizza 06] はノードの次数が上位  $n\%$  のノード間の結合確率を表す。

表 1: 最適化ネットワークの各指標の値

	最短経路 A	クラスター A	次数相関	次数 V	ノードコア A	K-コア	リッチ	最短経路 V	クラスター V	媒介中心性 V
最短経路 A+	1.0	0.20	0.91	0.063	0.86	0.29	0.67	0.36	0.35	1.0
最短経路 A-	0.0	0.68	0.28	0.55	0.42	0.37	0.32	0.000	0.38	0.17
クラスター A+	0.072	0.88	0.45	0.076	0.96	0.54	0.39	0.010	0.28	0.17
クラスター A-	0.068	0.0	0.49	0.015	0.72	0.73	0.12	0.004	0.0	0.008
次数相関+	0.125	0.063	1.0	0.018	0.69	0.30	0.45	0.024	0.052	0.017
次数相関-	0.072	0.013	0.014	0.029	0.64	0.63	0.0	0.002	0.013	0.013
次数 V+	0.0	1.0	0.0	1.0	0.33	0.04	0.4	0.000	0.075	0.098
次数 V-	0.069	0.002	0.50	0.0	1.0	1.0	0.075	0.000	0.001	0.000
ノードコア A+	0.071	0.023	0.53	0.004	1.0	1.0	0.10	0.001	0.015	0.004
ノードコア A-	0.075	0.059	0.57	0.054	0.0	0.52	0.37	0.013	0.036	0.022
K-コア+	0.071	0.35	0.53	0.004	1.0	1.0	0.12	0.001	0.024	0.003
K-コア-	0.069	0.048	0.56	0.017	0.33	0.0	0.26	0.003	0.074	0.011
リッチ+	0.083	0.094	0.81	0.043	0.67	0.17	1.0	0.007	0.21	0.017
リッチ-	0.065	0.031	0.38	0.021	0.67	0.65	0.0	0.003	0.036	0.011
最短経路 V+	0.72	0.17	0.36	0.19	0.18	0.26	0.66	1.0	0.35	0.80
最短経路 V-	0.069	0.0078	0.33	0.002	1.0	1.0	0.017	0.000	0.004	0.000
クラスター V+	0.043	0.50	0.28	0.12	0.56	0.30	0.38	0.003	1.0	0.053
クラスター V-	0.068	0.0	0.49	0.013	0.68	0.65	0.075	0.004	0.0	0.008
媒介中心性 V+	0.30	0.40	0.40	0.18	0.62	0.15	0.85	0.065	0.77	0.63
媒介中心性 V-	0.067	0.0	0.50	0.000	1.0	1.0	0.013	0.0	0.0	0.0

## 2.3 ネットワーク構造空間の可視化

自己組織化マップとは、トポロジーを保持しつつ多次元空間を低次元空間（本研究においては  $X \times Y$  の二次元空間）に投影することが可能なニューラルネットワークの一種である。マップ上の各ニューロンは入力データと同じ次元  $D$  のベクトル重みを持ち、最も入力データに近かった勝者ニューロンとその周辺が優先的に入力に近い値に更新するという学習を  $T$  回だけ行う。学習においては、各ニューロンのベクトル重みの学習率を、勝者ニューロンからの距離とマップが学習した回数に対して指数関数的に減少させる。本研究では、ベクトルの次元としては最適化したネットワークに関して求めた各指標の値を用い、可視化の対象とする構造空間を表し、多次元空間上に配置される各最適化ネットワークを二次元平面に可視化することにより、構造間の関係性や指標の持つ意味を調べる。

## 2.4 指標間の相関の分析

ベクトルの次元として用いた最適化したネットワークにおける各指標の値を用いて、相関係数も求める。具体的には、全てのネットワークのデータから指標の値の平均と分散を計算し、ピアソンの相関係数の式に代入することによる。

## 3. 結果

### 3.1 ネットワーク構造の各種指標による最適化

遺伝的アルゴリズムのパラメータは、個体数 100、規定ノード数  $N=100$ 、規定リンク数  $L=198$ 、遺伝子長 4950、交叉回数 20、交叉率 1.0、突然変異率 0.0012 と設定した。世代交代のモデルには MGG を用いており、次のようになる。まず、集団中から 2 個体をランダムに親個体として選ぶ。親個体に交叉・突然変異を行って生成した子個体を 40 個体作成する。そのうち、適応度最大の 1 個体とランダムに選んだ 1 個体を、親 2 個体と入れ替えた集団を次世代とする。K-コアネスに関しては  $K=3$ 、リッチクラブに関しては  $n=10$  を用いる。そして、各指標に対して最大化と最小化を 5 試行ずつ行った。なお、簡易記述法として、以降では、(略された) 指標名、A (平均) または V (分散)、+ (最大化) または - (最小化) の組み合わせで表現する。例えば、最短経路長の平均を最大化したネットワークは、「最短経路 A+」と表される。

結果を表 1 に示す。同表は、5 試行の平均に関して、各指標の最大値を 1、最小値を 0 として、規格化した値を示したものである。対角線の欄のペアは、それぞれのネットワークを最大化あるいは最小化した指標に関する値を示している。こ

れがほぼ 1 と 0 にそれぞれなっているので、基本的には最適化が機能していることが推測される。同表では、最適化の対象でない指標も 0 や 1 に近い値をとる場合があり、そのようなケースを編掛けによって示している。これより、最短経路長平均においては、最短経路長平均最大化だけでなく、最短経路長分散最大化、媒介中心性分散最大化がも同時に極めて大きな値をとっていることが見て取れる。また、クラスター係数平均、次数分散、最短経路長分散、媒介中心性分散においても、いくつかの指標が極めて大きな値を取っている。また、極めて小さな値を取る指標もありノードコアネス平均の欄を見ると、ノードコアネス平均最小化だけでなく最短経路長分散最大化もその傾向があることがわかる。これら特徴的な指標の値は、何かしらの特徴的な構造を反映した結果だと予想できる。このことに関して、次に、最適化によって得られたデータを元に構造空間を可視化し、特徴的な構造を明確化することを試みる。

### 3.2 ネットワーク構造空間の可視化

自己組織化マップのパラメータ設定は、 $X \times Y=50 \times 50$  の正方格子、 $T=10000$ 、学習範囲の重み 2.0、 $D=10$ 、入力データの選択方法はランダム選択とする。

次に、構造空間の可視化の結果を図 1 に示す。randomnet と表記した部分はランダムネットワークのデータである。このマップに関しては、基本的には同じ最適化の結果得られたネットワークは近くに位置され、ランダムネットワークの周辺はクラスター係数平均最小化やリッチクラブ最小化といったあまり極端な特徴を持たないものが位置づけられ、そこから離れるにつれ、リッチクラブ最大化、最短経路長分散最大化、次数分散最大化といった極端な特徴を持つものが位置づけられることがわかる。また、指標の最適化で特徴的な構造（ハブ、核、ひも、波状）の現れるものに関しては、フリーで配布されている描画ソフト Pajek<sup>\*1</sup> によって描画したネットワークを載せた。以降、ハブ構造とは次数の高い中心的なノードを持つ構造、核構造とは次数の高いノード同士が密に連結している構造、小クラスター構造とは次数の低いノードがクラスターを持っている割合が高い構造、ひも状構造とは次数の低いノードによる鎖状構造、波状構造とは次数の高いノードと低いノードが交互に現れる構造、一様構造とはランダムに近かったりノードの次数がほぼ一定構造とする。

この構造空間の可視化マップより次のことがわかる。基本的に同じ最適化を用いたものはマップ上では近くに集まってい

\*1 <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

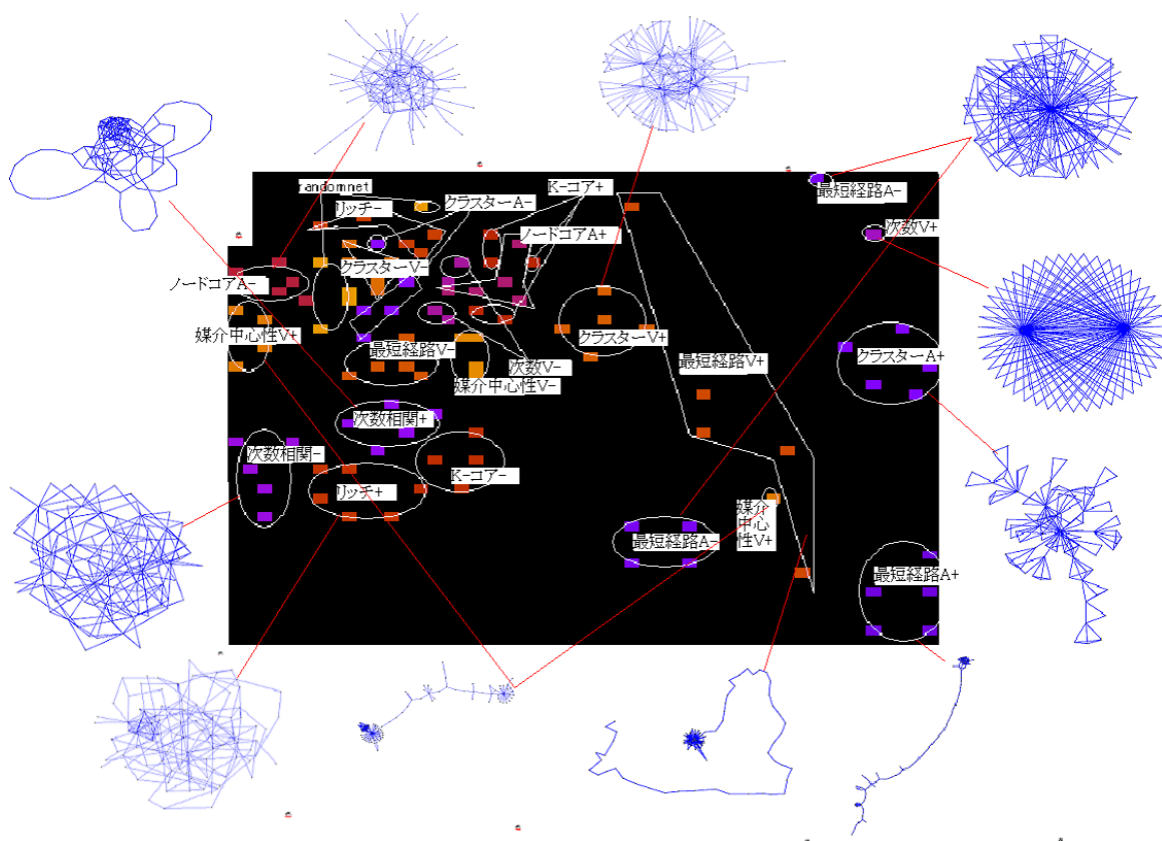


図 1: 自己組織化マップによって可視化された構造空間

表 2: 指標が特徴的に表す構造

	リッチ	ノードコア A	媒介中心性 V	最短経路 V	最短経路 A	次数 V	クラスター A	クラスター V	K-コア	次数相関
ハブ構造										
核構造	+	-	+	+	-	+		+		+
小クラスター構造					-	+	+	+		
ひも状構造			+	+	+					+
波状構造										-
一様構造	-	+	-	-		-	-	-	+, -	

る。各構造別に見ると可視化マップにおいて各特長な構造の配置は、比較的ランダムに近い構造はランダムネットワークの部分と K-コアネスの部分の間の領域、核構造はその周辺、ひも状構造は真ん中から比較的右側にかけて、ハブ構造はもっとも右側に集まっている。この様な位置関係になったのは、核構造はあまり極端な構造を取らないものが多いので、同じく極端な構造をとらないランダムに近い構造の周辺に、ひも状構造は核構造を持つので、さらにその隣に、ハブ構造はマップの学習の結果最終的に端に配置されたものと思われる。ここで、媒介中心性分散最大化が二箇所に分かれているのは右側の一点はひも状構造がかなり長く、左側の四点は短い結果となったが、右側の周辺にはひも状構造が、左側の四点の周辺には核構造を持つノードコアネス最小化、リッチクラブ最大化といったものがあることからそれらの影響が大きいことがわかる。また、最短経路長平均も二箇所に分かれているが、右上の一点はハブが二つ、真ん中下の四点はハブが一つだけの構造となっていて、後者の構造の時は小クラスター構造はあまり顕著には見られない。このことから、特に右側のこの三指標はハブ構造と小クラスター構造が原因で集まった可能性がある。

次に構造空間の位置関係を加味して指標が特徴的に表す構造間の関係を考察する。

表 2 では、構造的に特に似たものを集め、太線で区切って表示している。また、表において、+-は最適化の種類を示し、指標に対して最大化、ないし、最小化を行ったときのような構造が現れるかを示している。左から順に核構造、核構造+小クラスター構造、小クラスター構造、一様構造、波状構造である。リッチクラブ最大化において、ノードコアネス最小化と位置が離れているが、これは核構造の大きさによるものと思われるので、リッチクラブの核構造を大きくすればおそらくは、より近い位置にすることが予想される。また、この表よりひも状構造は核構造を、ハブ構造は小クラスター構造を伴っていることがわかるが、これらの関係はリンク数の制限より比較的自明なものである。

以上の結果より、指標が表す特徴的な構造間には関係があることがわかる。特に、ひも状構造は核構造と、ハブ構造は小クラスター構造と、リンク数制限という根本的な原因をもって繋がっているため、これらの構造的特長を持つ指標を考察するときには注意が必要である。また、この結果を用いることによって、ネットワーク設計問題に対して最適手法を用いてアプローチするときには、おおよどのような構造が現れて、どのような構造が現れないか、ということに関して予測を立てることが可能となる。

表 3: 最適化ネットワークに基づく指標間の相関

	最短経路 A	クラスター A	次数相関	次数 V	ノードコア A	K-コア	リッチ	最短経路 V	クラスター V	媒介中心性 V
最短経路 A	1									
クラスター A	-0.043	1								
次数相関	0.33	-0.40	1							
次数 V	-0.072	0.78	-0.51	1						
ノードコア A	-0.029	-0.14	0.17	-0.34	1					
K-コア	-0.26	-0.47	-0.037	-0.46	0.63	1				
リッチ	0.48	0.33	0.33	0.23	-0.22	-0.68	1			
最短経路 V	0.78	0.0091	0.050	0.050	-0.25	-0.24	0.39	1		
クラスター V	0.28	0.47	-0.091	0.18	-0.12	-0.46	0.56	0.22	1	
媒介中心性 V	0.93	0.21	0.16	0.13	-0.091	-0.39	0.59	0.74	0.48	1

### 3.3 指標間の相関の分析

ここでは指標間の相関を各最適化のデータを基に導出した結果を考察する。Jamakovic らは実世界のネットワークのデータを用いて同様の計算を行うことにより実世界のネットワークに対して有用な指標のサブセットを導くことを行った。本研究においても同様の計算を行うことにより最適化ネットワークに対して有用な指標のサブセットを導くことを行う。

本研究により得られた指標間の相関を表 3 に示す。また、網掛けの部分は指標間の相関が 0.6 以上の部分である。

指標間の相関が 0.6 以上の高い部分に関して次のグループを得る。(最短経路長平均、媒介中心性分散、最短経路長分散) (ノードコアネス平均、K-コアネス) (クラスター係数平均、次数分散) と残りの独立した指標という形になることがわかる。これらは、主に次の点が原因で高い相関を得ているものと思われる。第一、第二のグループに関しては、グループ内の指標の定義が似ているので、必然的に高い相関を得ている。第三のグループに関しては、ハブ-小クラスター間に強い関係、すなわち、ハブがあれば、小さいクラスターが大量に作りやすく、逆に、ハブがなければ小さいクラスターが作りにくいという関係による。これらのグループ分けにより最適化ネットワークを調べる際にはこれらを念頭に置くことにより、より容易にトポロジーが把握できるようになる。

Jamakovic らの研究と本研究において共通する指標 (最短経路長平均、クラスター係数平均、次数相関、ノードコアネス平均、リッチクラブ) の間の相関で彼らの研究において 0.6 以上の強い相関が確認されているのは、クラスター係数とノードコアネス平均の相関となるが、本研究においては確認されていない。他の関係に関してはどちらの研究においても強い相関は確認されていない。また、クラスター係数平均の定義、すなわち各ノードがどの程度の割合でクラスターを形成しているかとノードコアネスの定義、すなわちどのレベルのサブグラフに所属しているかという二つは関係がないことがわかる。よって、実世界のネットワークのデータにおいてはどのような理由でこの二つの指標間に強い相関を持っているかということ調べるのが実ネットワークの相関関係を考える上では重要であることがわかる。

## 4. おわりに

本論文では、様々な複雑ネットワークの指標が表す構造的特徴や、それらの普遍的な関係を明らかにすることを目的とし、各指標に関する最適化ネットワークを遺伝的アルゴリズムで探索し、構造空間を自己組織化マップで可視化した。その結果、ネットワークの構造間には複雑な関係があることが判明した。さらに、各指標間の相関に注目し、得られた最適化ネットワークにおける相関と、Jamakovic らによる実世界のネットワークにおける相関との比較を行い、その相違点を明らかにした。

本研究で得られた結果は、複雑ネットワークの設計や最適化

の際に有用な知見を与えうると言える。例えば、ある指標を重視したネットワークを設計するにあたり、それが取り得る構造や、他の指標に与える影響に関して予想を与えることができると考えられる。

本手法では、最適化の対象とする指標の種類を増やすほど、特徴的な構造にどんなものがあり、どう関係しているかといった構造空間の詳細がより明らかになり、一般的な知見が得られると考えられる。今後は指標の種類を増やした解析を含め、構造空間の全体像のより詳細な把握を進める予定である。

## 参考文献

- [Newman 02] Newman, M. E. J.: Assortative mixing in networks, *Physical Review Letter*, Vol. 89, 208701 (2002).
- [Jamakovic 08] Jamakovic, A. and Uhlig, S.: On the relationships between topological measures in real-world networks, *Networks and Heterogeneous Media*, Vol. 3, No. 2, pp. 345-359 (2008).
- [Kohonen 82] Kohonen, T.: Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, Vol. 43, pp. 59-69 (1982).
- [花田 07] 花田良子, 佐藤史隆, 廣安知之, 三木光範, 鈴木泰博: 遺伝的アルゴリズムによるネットワーク特性量に着目したネットワーク設計法, *コンピュータソフトウェア*, Vol. 24, No. 1, pp. 91-100 (2007).
- [佐藤 96] 佐藤浩, 小野功, 小林重信: 遺伝的アルゴリズムにおける世代交代モデルの提案と評価, *人工知能学会誌*, Vol. 12, No. 5, pp. 734-743 (1996).
- [Borgatti 03] Borgatti, S. P. and Everett, M. G.: Mathematical results on scale-free random graphs, *Handbook of Graphs and Networks: From the Genome to the Internet*, Wiley-VCH (2003).
- [Colizza 06] Colizza, V., Flammini, A., Serran M. A. and Vespignani, A.: Detecting rich-club ordering in complex networks, *Nature Physics*, Vol. 2, pp. 110-115 (2006).