信学技報 TECHNICAL REPORT OF IEICE.

残存次数のエントロピーを用いたトポロジー構造の多様性の評価

諶 璐[†] 荒川 伸一[†] 村田 正幸[†]

† 大阪大学 大学院情報科学研究科 〒 565–0871 大阪府吹田市山田丘 1–5 E-mail: †{l-chen,arakawa,murata}@ist.osaka-u.ac.jp

あらまし インターネットの社会インフラ化が進み利用形態が多様化するにつれ、トラヒック需要の変化やネット ワーク機器故障に対する適応性や拡張性の高いネットワーク設計が重要になりつつある。しかし、需要の変化や機器 故障の事象は予測困難であるため、事前に発生頻度や発生規模を想定してネットワークを構築するのではなく、トポ ロジーの構造に多様性を持たせることで環境変化に対する適応性や拡張性を高めることが望ましい。本稿では、トポ ロジーが有する構造の多様性を測る指標として残存次数の相互情報量に着目し、その有用性を評価する。評価の結果、 ルーターレベルトポロジーの残存次数の相互情報量は約 1.0 となり、トポロジー構造の多様性が低いことがわかった。 また、相互情報量が小さいとトポロジー構造が多様となり、相互情報量が大きいとトポロジー構造に規則性が生じる ことを示した。

キーワード べき則、ルーターレベルトポロジー、トポロジー構造、相互情報量、構造の多様性、次数分布

Analysis of Network Heterogeneity by Using Entropy of the Remaining Degree Distribution

Lu CHEN^{\dagger}, Shin'ichi ARAKAWA^{\dagger}, and Masayuki MURATA^{\dagger}

[†] Graduate School of Information Science and Technology, Osaka University
 1–5 Yamadaoka, Suita, Osaka 565–0871, Japan
 E-mail: [†]{1-chen,arakawa,murata}@ist.osaka-u.ac.jp

Abstract As the Internet becomes the social infrastructure, a network design method that has the adaptability against the failure of network equipment and has the sustainability against changes of traffic demand is becoming important. Since we do not know when to occur the environmental changes and how large the changes in advance, it is preferable to have a heterogeneity of topological structures so that the network can evolve more easily. In this paper, we investigate the mutual information of remaining degree as a measure of the heterogeneity of topological structures. Our results show that the mutual information is high at the most of router-level topologies, which indicate that the route-level topologies are highly designed by, e.g., the network operators. We then discuss and show that the mutual information represents the heterogeneity of topological structure through illustrative examples.

Key words power-law network, router-level topology, topological structure, mutual information, network heterogeneity, degree distribution

1. はじめに

インターネットの社会インフラ化が進み利用形態が多様化す るにつれ、多種多様なプロトコルに基づくネットワーク制御 が複雑に干渉するため、ネットワーク設計が複雑・困難になり つつある。例えばオーバレイルーティングにより、アプリケー ション層で経路制御を行い性能向上を測る技術が検討されてい るが、ネットワーク層の管理者によるトラヒックエンジニアリ ングと干渉し、性能が著しく劣化することが指摘されている。 そのため、トラヒック需要の変化やネットワーク機器故障など のネットワーク環境の変化に対する適応性や拡張性を保ちつつ ネットワークにリンク・ノードの追加や処理能力の増強などを 行う、新たなネットワークの設計手法が重要である。

インターネットトポロジーを計測した結果、次数分布がべき 則に従うことが明らかとなっている [1]。次数分布がべき則に従 うトポロジーでは、ノードの次数 x の出現確率 P_x が $P_x \propto x^{-\gamma}$ で近似される (γ は定数)。次数分布がべき則に従うトポロジー 生成手法はこれまでも広く検討されており、その1つに BA モ デルがある [2]。BA モデルは (1) ノードを段階的に追加してい く Incremental Growth、(2) 追加の際に既存のトポロジーのノー ドの次数分布に応じて確率的に接続する Preferential Attachment の規則に基づいてノードおよびリンクを追加するトポロジー生 成手法である. BA モデルで生成したトポロジーでは、多数の ノードの次数は小さい一方で、次数の大きいノードが少数存在 する。BA モデルを用いて生成した次数分布がべき則に従うト ポロジーを対象とし、ネットワーク性能の様々な評価がなされ ている [3],[4]

しかし、次数分布のみがトポロジーの性質を決定することは ない。文献 [5] では、次数分布が同一で構造の異なる複数のト ポロジーを列挙し、それぞれのトポロジーがノード処理能力の 制限下で収容できるトラヒック量を評価している。その結果、 ノードの次数分布が同一であったとしてもトポロジーが有する 構造に依存して収容可能なトラヒック量が異なることが示され いてる。すなわち、次数分布よりも高次の性質が収容可能なト ラヒック量を決定付けることが示されている。ISP のルーター レベルトポロジーでは、回線の物理距離やルーターに接続可能 なポート数などの技術的・物理的な制約に起因してトポロジー の構造が定まり、それは、BA モデルなどの確率的なトポロジー 生成モデルでは説明することはできない。

我々の研究グループでは、通信ネットワークを対象としたト ポロジーのモデル化手法の検討を行っている[6]~[8]。文献[6] では、ISPのルーターレベルトポロジーは、地域ごとのモジュー ル内のルーター同士が局所的に密に連結されるモジュール性の 高い構造を有していることを示している。また、文献[7]では、 ISPのルーターレベルトポロジーの持つ高いモジュール性は、 回線の物理距離の最小化と回線容量制約下における収容可能な トラヒック量が減少しないよう設計した結果として出現するこ とを明らかにしている。文献[8]では、国内 ISP ネットワーク の回線容量分布が Zipf 則に従うことを指摘している。回線容 量分布が Zipf 則に従うとは、k 番目に大きな回線容量が、全回 線の回線容量に占める割合が 1/k に比例することである。文 献[6] では、回線容量分布がべき則に従うことで収容可能なト ラヒック量が増大することを明らかにしている。しかし、ネッ トワーク環境の変化に対する適応性や拡張性を保持するネット ワーク設計を行うためには、トポロジーのモデル化手法の検 討のみでは不十分である。特にトラヒック需要の変化やネット ワーク機器故障は予測困難な事象であり、事前に変動規模や故 障規模を想定してネットワークを設計するのではなく、トポロ ジーの構造やリンクの回線容量を含めて適応性や拡張性の高い ネットワークを設計する必要がある。

本稿では、適応性や拡張性の高いネットワーク設計への適用 を想定し、トポロジーが有する構造の多様性を測る指標を評価 する。ここで言う構造の多様性とは、トポロジーの一部のノー ド集合(以降、subgraph と呼ぶ)を取り出した際に、その集合 の接続形態が多様であることである。これまでにも subgraph の 接続形態をパターンとして定義し、そのパターンの出現頻度に よりトポロジーの構造的特徴を分析するネットワークモチーフ の概念が提案されている [9]。文献 [10] では、4 ノード subgraph の出現頻度から ISP のルーターレベルトポロジーの構造的特徴 の分析を行っている。そこでは、ISP のルーターレベルトポロ ジーでは、4 ノードが完全連結された subgraph から1 リンクを 除去した"sector"の構造を多く有することを示している。しか し、ネットワークモチーフは、パターンの出現頻度を抽出するも のであり、構造の多様性を測るものではない。ただし subgraph のノード数を増やし(最大で N、ただしトポロジーのノード数 を N とする)、すべてのパターンの出現頻度を抽出することで、 構造の多様性を議論することは可能となるが、そのためには膨 大な計算が必要となる。

そこで、本稿では、トポロジーが有する構造の多様性を測る 指標として、相互情報量に着目する。相互情報量は、情報理論 における解釈では、確率変数 X、Y があった時、Y を知るこ とにより得られる X の情報量となる。相互情報量を、トポロ ジーの構造の一部 (Y に相当)を知ることで得られる残りのト ポロジーの構造 (X に相当)の情報量と見立てることで、トポ ロジーが有する構造の多様性を測る。文献[11]では、残存次数 の相互情報量を用いたトポロジー分析がなされている。ソフト ウェアプログラムや電子回路等の人工システムと生物システム のトポロジーにおける残存次数の相互情報量を算出し、ランダ ムに接続されたトポロジーと比較して相互情報量が大きくなる 結果が示されている。そこで、本稿では残存次数の相互情報量 に着目し、トポロジーが有する構造の多様性を評価する。

本稿の構成は以下の通りである。まず、2. 章で、相互情報量 の定義について述べる。次に、3. 章で、相互情報量を用いて ルータレベルトポロジを評価し、構造の多様性について述べる。 4. 章では、相互情報量がトポロジー構造の多様性を示す指標で あることについて述べる。最後に、まとめと今後の課題につい て述べる。

2. 残存次数の相互情報量

本章では文献[11]で述べられている残存次数の相互情報量を



図1 残存次数



 \boxtimes 2 Abilene ($H = 3.27, H_c = 2.25$)

説明する。

残存次数 k とはリンクを一本取り除いた時に、その片方に接 続されていたノードの残り次数である。残存次数分布 q(k) は 次数分布 P(P₁,..., P_x,..., P_K)を用いて以下の式で表される。 ただし K はトポロジーの最大次数である。

$$q(k) = \frac{(k+1)P_{k+1}}{\Sigma_k k P_k} \tag{1}$$

残存次数の相互情報量 *I*(*q*) は残存次数分布 q=(*q*(1), ..., *q*(*i*), ..., *q*(*N*)) を用いて式 (2) で表される。

$$I(\mathbf{q}) = H(\mathbf{q}) - H_c(\mathbf{q}|\mathbf{q}') \tag{2}$$

式 (2) における第一項 *H*(q) は残存次数のエントロピーを表 しており、残存次数分布 *q*(*k*) を用いて式 (3) で定義される。

$$H(q) = -\sum_{k=1}^{N} q(k) \log(q(k))$$
(3)

これはリンクの接続先のノード次数の多様性を測っており、ト ポロジーの次数の不均質性を表す指標であることが文献[11] で 述べられている。残存次数のエントロピー H が 0 となるのは、 例えばリングトポロジーのような次数が均質な正則グラフであ る。また、リンクの接続先のノード次数が多様になればなるほ ど H は大きくなる。例えば、文献[5] で示されている Abilene トポロジー(図 2) では、リンクの接続先のノード次数は不均 質となり H は大きい。

式 (2) における第二項 $H_c(\mathbf{q}|\mathbf{q'})$ は残存次数の条件付きエント ロピーを表している。残存次数 k'を持つノードと接続されて いるノードの残存次数が k である条件付き確率 $\pi(k|k')$ と、残 存次数分布 \mathbf{q} を用いて以下の式で定義される。

$$H_c(\mathbf{q}|\mathbf{q}') = -\sum_{k=1}^N \sum_{k'=1}^N q(k')\pi(k|k')\log\pi(k|k')$$
(4)

Abilene トポロジーでは、あるノードと接続されているノード の残存次数に偏りがあるため、*H*_cが大きくなる。

表1 ルータレベルトポロジーの相互情報量

トポロジー	ノード数	リンク数	H(G)	$H_c(G)$	I(G)
Level3	623	5298	6.04	5.42	0.61
Verio	839	1885	4.65	4.32	0.33
ATT	523	1304	4.46	3.58	0.88
Sprint	467	1280	4.74	3.84	0.90
Telstra	329	615	4.24	3.11	1.13
BA	523	1304	4.24	3.98	0.26
Random	523	1304	3.22	3.15	0.07



図3 次数分布を維持するリンク張り替え

表 2 BA トポロジーと生成トポロジー

トポロジー	ノード数	リンク数	H(G)	$H_c(G)$	I(G)
BA	523	1304	4.24	3.98	0.26
T_{Imin}	523	1304	4.24	4.13	0.12
T_{Imax}	523	1304	4.24	1.54	2.70

ルータレベルトポロジーにおける構造のパ ターン化

文献[12]の手法を用いて抽出されたルータレベルトポロジー Level3、Verio、AT&T、Sprint、Telstra の相互情報量を求めたと ころ、表1の通りになった。また、比較のため、AT&Tトポロ ジーと同一のリンク数、ノード数の BA モデル [2] で生成した トポロジと ER モデル [13] で生成したトポロジーの相互情報量 を求めている。以降では、それぞれ BA トポロジー、ER トポロ ジーと呼ぶ。表1から、Verio 社を除くルータレベルトポロジー は残存次数の相互情報量が大きいことが分かる。確率的に生成 される BA トポロジーや ER トポロジーでは相互情報量が小さ いことが分かる。ISP トポロジーにおいて故障耐性向上のため の接続パターンが繰り返し出現しているためであり、ルータレ ベルトポロジーが意図的に設計されていることに起因している と考えられる。一方で、Verio 社のトポロジーの相互情報量は 他の ISP と比較して小さいことが分かる。これは、Verio 社の ネットワーク拡張の経緯によるものと考えられる。Verio 社は小 規模な地域 ISP の買収を繰り返して規模を拡大したため [14]、 さまざまな ISP の設計指針を内包し、結果として構造が多様に なっている。

4. 相互情報量とトポロジーの性質

4.1 残存次数のエントロピー H(q) の影響

まず、残存次数分布の違いによるトポロジーの性質を明らか にするために、式 (2) における *H*(q) の異なる複数のトポロジー を生成し、その平均ホップ長および次数分布を調査した。BA ト ポロジー(523 ノード、1304 リンク)を初期トポロジーとして 与え、ランダムにリンクを張り替えつつ、焼きなまし法を用い て下記のポテンシャル関数 U を最小化することにより求めた。

$$U(G) = \sqrt{(H - H(G))^2 + (H_c - H_c(G))^2}$$
(5)

ここで H、H_c はそれぞれ事前に指定した目標エントロピー H、目標条件付きエントロピー H_c であり、H(G)、H_c(G) は それぞれ生成過程のトポロジー G のエントロピー、条件付きエ ントロピーである。上記の手法で、H = H_c となる値を目標値 に設定して、異なる残存次数分布を持つトポロジーを生成し、 それらの平均ホップ長を求めた結果を図6に示す。図を見ると、 残余次数分布のエントロピー H が3を超えて増加すると、平 均ホップ長が減少することが分かる。これは、H の増大ととも にノード次数に偏りが生じ、H が4 付近においては次数分布が べき則となるためである。

4.2 残存次数の条件付きエントロピー H_c(q|q')の影響

次に、相互情報量とトポロジーの性質の関係を明らかにする。 本節では、ルーターレベルトポロジーの次数分布がべき則に従 うことから *H*(q) が大きいトポロジーを対象とする。

本節では、BAトポロジー(523 ノード、1304 リンク)を初期 トポロジーとして与え、生成トポロジー G の相互情報量 I(G) が目標相互情報量 I に近づくよう、焼きなまし法を用いて下記 のポテンシャル関数 U^I を最小化することにより求めた。

$$U^{I}(G) = |I - I(G)| \tag{6}$$

図4と図5を比較すると、相互情報量 I が大きくなるとトポ ロジー構造の多様性が失われ、構造に規則性が出現することが 見て取れる。次節では、T_{Imin} と T_{Imax} を例として、相互情報 量とトポロジー構造の多様性の関係を詳しく述べる。

4.3 相互情報量とトポロジー構造の多様性

相互情報量を最小化したトポロジー*T_{Imin}*(図4)と相互情 報量を最大化したトポロジー*T_{Imax}*(図5)を見比べると、相 互情報量*I*が大きくなるとともにトポロジー構造に規則性が出 現していることが分かる。本節では、*T_{Imin}*と*T_{Imax}*の2つの トポロジーを用いて相互情報量がトポロジー構造の多様性を示 す指標であることを説明する。

相互情報量の定義式は式(2)であり、残存次数のエントロ



図4 相互情報量を最小化したトポロジー*T_{Imin}*:相互情報量 0.12。初 期トポロジーとして BA トポロジーを与えた



図 5 相互情報量を最大化したトポロー *T_{Imax}*:相互情報量 2.70。初 期トポロジーとして BA トポロジーを与えた



ピー H と条件付きエントロピー H_c により決定される。しかし T_{Imin} と T_{Imax} を生成するにあたり、次数分布を維持したリン ク張り替えを行っているため、式 (1) より残存次数 q(k) は変化 せず、従って残存次数のエントロピー H も同一となる。すなわ

ち、残存次数の条件付きエントロピー *H*。の変化が相互情報量 *I*を変化させている。また、残存次数の条件付きエントロピー は、式 (4)を変形すると式 (7) になる。

$$H_c(\mathbf{q}|\mathbf{q}') = \sum_{k'=1}^{N} q(k') \{-\sum_{k=1}^{N} \pi(k,k') \log \pi(k,k')\}$$
(7)

このとき、条件付き確率 $\pi(k|k')$ のエントロピー $H_{\pi}(k')$ を、

$$H_{\pi}(k') = -\sum_{k=1}^{N} \pi(k, k') \log \pi(k, k')$$
(8)

とおくと、

$$H_{c}(\mathbf{q}|\mathbf{q}') = \sum_{k'=1}^{N} q(k') H_{\pi}(k')$$
(9)

となる。次数分布を維持したリンク張り替えを行う場合、残存 次数の分布 q(k') は変化しないため、 H_c の変化は条件付き確率 $\pi(k|k')$ のエントロピー $H_{\pi}(k')$ の変化によってもたらされる。

相互情報量 *I* が小さい場合、つまり残存次数の条件付きエントロピー $H_c(k')$ が大きい場合のトポロジーの構造を説明する。 次数分布がべき則に従う場合、残存次数 k' が大きい時、その残 存次数 k' を持つノード数は少ない。また、残存次数 k' が小さ い時、その残存次数 k' を持つノード数は多く存在する。そこ で、式 (9) から、条件付き確率 $\pi(k|k')$ のエントロピー $H_{\pi}(k')$ の大小が表すトポロジーの構造を、残存次数 k' が最大のもの と残存次数 k' が最小のものに着目して説明する。

残存次数 k' が最大のノードに着目した場合、条件付き確率 $\pi(k|k')$ のエントロピー $H_{\pi}(k')$ が大きくなるのは、残存次数 k'を持つノードの接続先ノードの残存次数 k が多様であるときで ある。図 7 は BA モデルの相互情報量を最小化したトポロジー T_{Imin} の最大残存次数 k'の条件付き確率 $\pi(k|k')$ の分布を表し ており、最大残存次数を持つノードの接続先ノードの残存次数 がばらついていることが分かる。

残存次数 k' が最小のノード集合に着目した場合、条件付き 確率 $\pi(k|k')$ のエントロピー $H_{\pi}(k')$ が大きくなるのは、残存 次数 k' が最大のノードに着目した場合と同じく、残存次数 k'を持つノードの接続先ノードの残存次数 k が多様であるときで ある。図 8 は相互情報量を最小化したトポロジー T_{Imin} の、最 大残存次数 k' の条件付き確率 $\pi(k|k')$ の分布を表しており、最 大残存次数を持つノードの接続先ノードの残存次数がばらつい ていることが分かる。

次に、相互情報量 I が大きいとき、つまり残存次数の条件付 きエントロピー $H_c(k')$ が小さい場合のトポロジーの構造を説 明する。残存次数 k' が最大のノードに着目した場合、条件付 き確率 $\pi(k|k')$ のエントロピー $H_{\pi}(k')$ が小さくなるのは、残 存次数 k' を持つノードの接続先ノードの残存次数 k が均一と なるときである。図 9 は相互情報量を最大化したトポロジー T_{Imax} の、最大残存次数 k' の条件付き確率 $\pi(k|k')$ の分布を表 しており、最大残存次数を持つノード集合の接続先ノードの残 存次数が 1 である確率が高いことが分かる。すなわち、相互情 報量の大きなトポロジーの残存次数 k' が大きいノードは図 11



図 11 相互情報量が大きいトポロジーに現れる構造



図 12 複数の k'、 $H_{\pi}(k')$ が大きなノードの存在により現れる構造



図 13 複数の k' が小さく、H_π(k') が大きなノードの存在により現れる構造

のトポロジー構造を有している。また、残存次数 k' が大きく かつ H_π(k') が大きいノードが複数存在すると、図 12 のトポロ ジー構造が出現する。

一方、残存次数 k' が最小のノード集合に着目した場合、条 件付き確率 $\pi(k|k')$ のエントロピー $H_{\pi}(k')$ が小さくなるのは、 残存次数 k' が最大のノード集合に着目した場合と同じく、残 存次数 k' を持つノードの接続先ノードの残存次数が均一のと きである。図 10 は相互情報量を最大化したトポロジー T_{Imax} の、最小残存次数 k' の条件付き確率 $\pi(k|k')$ の分布を表してお り、最小残存次数を持つノード集合の接続先ノードの残存次数 が図 8 より偏っていることが分かる。このような相互情報量が 大きなトポロジーの場合、残存次数 k' が小さいノード集合は、 図 13 に示すトポロジー構造が出現すると考えられる。

5. まとめと今後の課題

本稿では、トポロジーが有する構造の多様性を測る指標とし て、相互情報量に着目した。相互情報量を用いてルータレベル トポロジを評価しており、確率的に生成されるトポロジの相互 情報量が小さいのに対し、多くのルータレベルトポロジの相互 情報量が大きいことが分かった。また相互情報量の異なるトポ ロジを用いて、相互情報量 I が小さいとトポロジー構造が多様 となり、大きいとトポロジー構造に規則性が出現することを説 明した。

今後の課題として、構造の多様性を有するトポロジと有さな いトポロジでネットワークの性能評価を行い、適応性や拡張性 の高いネットワーク設計に適用することが挙げられる。



図9 T_{Imax} の残存次数が最大のノードの $\pi(k|k')$ の分布

文 献

- M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," ACM SIGCOMM Computer Communication Review, vol. 29, pp. 251–262, Oct. 1999.
- [2] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, Oct. 1999.
- [3] R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, 2000.
- [4] K. L. Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale–free networks," *Physical Review Letters*, vol. 87, Dec. 2001.
- [5] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first-principles approach to understanding the Internet's router-level topology," ACM SIGCOMM Computer Communication Review, vol. 34, pp. 3–14, Oct. 2004.
- [6] 荒川伸一, 滝根哲哉, 村田正幸, "通信ネットワークのトポロジー 構成のモデル化と性能評価への応用," 電子情報通信学会技術研 究報告 (NS2011-110), vol. 111, pp. 37–42, Nov. 2011.
- [7] 松井 佑記, "ISP トポロジーにおける局所連結性が回線容量制約 下のフロー量分布に与える影響," 大阪大学情報科学科特別研究 報告, pp. 1–33, Feb. 2012.
- [8] T. Hirayama, S. Arakawa, S. Hosoki, and M. Murata, "Models of link capacity distribution in ISP's router-level topology," *International Journal*, vol. 3, pp. 205–216, Sept. 2011.
- [9] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, pp. 824–827, Oct. 2002.
- [10] S. Arakawa, T. Takine, and M. Murata, "Analyzing and modeling router-level Internet topology and application to routing control," to appear in *Computer Communications*.
- [11] R. Solé and S. Valverde, "Information theory of complex networks: On evolution and architectural constraints," *Complex net-*





図 10 T_{Imax} の残存次数が最小のノード集合の $\pi(k|k')$ の分布

works, vol. 650, pp. 189-207, Aug. 2004.

- [12] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with rocketfuel," *IEEE/ACM Transactions on Networking*, vol. 12, pp. 2–16, Feb. 2004.
- [13] P. Erdős and A. Rényi, "On random graphs," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.
- [14] M. Pentz, "Verio grows big with small clients," *Business Journals*, Feb. 1999.
- [15] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic topology analysis and generation using degree correlations," in ACM SIGCOMM Computer Communication Review, vol. 36, pp. 135–146, Oct. 2006.