

大規模なネットワーク構造の可視化

Webをはじめとする大規模で複雑なネットワークについて、内在するデータの構造的特徴が浮き彫りになるように効率よく可視化する方法を紹介します。本技術は3DWebブラウザなどに応用できます。

やまだ たけし
山田 武士

NTTコミュニケーション科学基礎研究所

ネットワーク構造の可視化とは

大量のデータを分析してそのデータ固有の構造的な特徴を明らかにしたいということはしばしばあります。その際、これらのデータの中に複雑な関係や相互作用が存在する場合には、これを分かりやすく表現する手段としてネットワークがよく用いられます。例えばインターネットのWebページを対象とする場合、各Webページをノード、ページ間のハイパーリンクをノード間のリンクとするハイパーリンクネットワークが用いられます。生物学においても、遺伝子やたんぱく質、代謝物

質などの相互作用は遺伝子制御ネットワークとして表現されます。また人間どうし、あるいは企業などの社会的存在どうしの関係の解析にはソーシャルネットワークが用いられます。さらに単語間の関係も意味ネットワークで表現されます。

ネットワーク表現を用いる利点は、グラフ理論などを駆使した数学的な解析が可能となることが挙げられます。しかし同時に重要なことは、ネットワークを低次元空間に埋め込んで配置し、実際に表示させることです。これにより、その配置上でノードからノードへのリンクを実際にたどったり、ノード

ごとの接続関係を相互に比較したりするなど、低次元の配置を直接調べることができます（ブラウジング）。その結果、内在するデータの構造的特徴が浮き彫りとなり、新たな科学的発見のための重要な手がかりが得られる可能性があります。

従来から、ネットワークを可視化する方法はいくつか提案されています。代表的なものとして線形手法^{*1}に基づくもの、後述のばねモデルに基づくものが挙げられます。図1、2にそれ

*1 線形手法：行列演算などの線形演算（一次式で表される演算）で表現できる計算手法のこと。

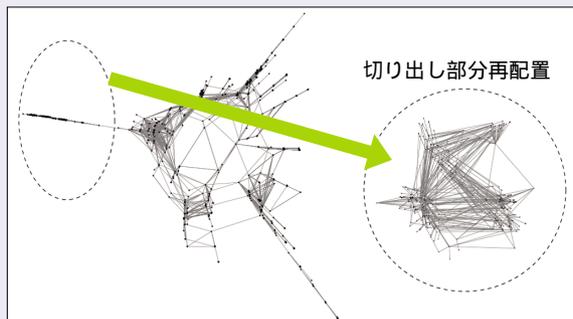


図1 従来法（古典MDS）によるNTTドメインに属するWebページの可視化結果

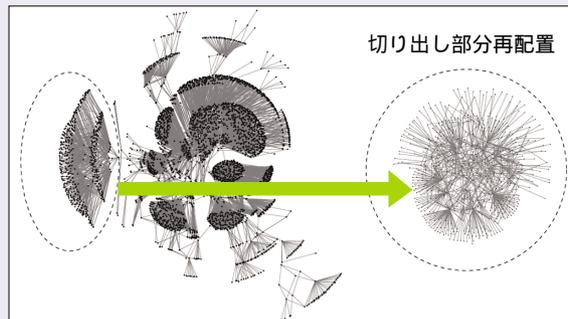


図2 ばねモデルによるNTTドメインに属するWebページの可視化結果

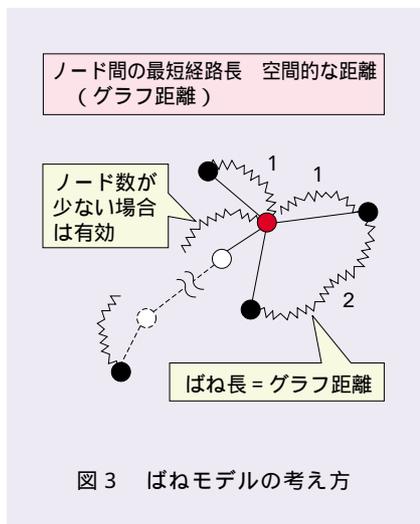


図3 ばねモデルの考え方

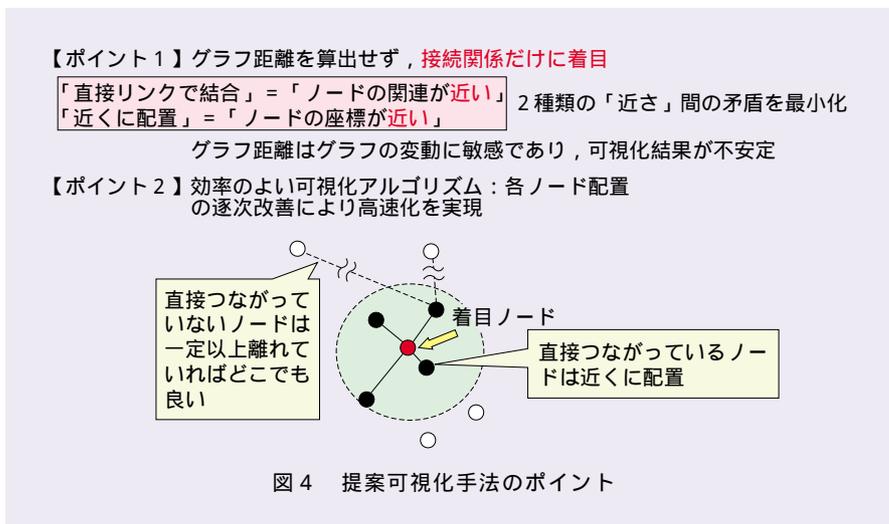


図4 提案可視化手法のポイント

それぞれの手法を用いてNTTのWebサイトであるwww.ntt.co.jpドメインに属するすべてのWebページを可視化した結果を示します。図1は「古典的MDS (Multi-Dimensional Scaling: 多次元尺度法)」と呼ばれる線形手法に基づく可視化結果ですが、多くのノードが縮退し、分かりにくい結果になっています。図2は同じデータをばねモデルと呼ばれる手法を用いて可視化した結果です。こちらのほうが見やすいですが、それでも多くのノードが半円状にかたまって配置されてしまい、かつ、場所によって粗密の差が大きくなっています。これは「たんぼぼ効果」として知られる現象で、特にブラウジングの観点では問題があると考えられます。

またいずれの手法においても、図中の矢印のようにネットワークの一部を切り出して再度可視化すると全体の構造が変化してしまうことが分かります。ブラウジングにおいては、注目すべきネットワークの一部を取り出して拡大し、より詳細な構造を観察したい、というケースが想定されますが、その際、拡大の前後で全体の構造が大きく変化してしまえばブラウジングの目的を果たすことができなくなります。

提案可視化手法

NTTコミュニケーション科学基礎研究所では、大規模なネットワークの構造を分かりやすく可視化する方法を提案しました。従来のばねモデルなどではネットワークをグラフとみなし、任意の2ノード間の最短経路長(グラフ距離)を計算します。さらに図3に示すように、この長さを「ばね長」とするばねを考え、これらのノードが空間的にほぼ「ばね長」に等しい距離となるような配置を考えます。この方法はノード数が少なく配置に余裕がある場合には有効ですが、ノード数が大きい大規模なネットワークの場合は、先に述べたような問題が生じます。

一方、提案可視化手法ではグラフの局所的な接続関係だけに着目しています。図4に提案可視化手法の考え方を示します。提案可視化手法は「直接つながっているノード(隣接ノード)は近くに配置し、直接つながっていないノード(非隣接ノード)は一定以上離れていれば、どこに配置されてもよい」というより柔軟な考え方に基づく方法です。

まず、隣接ノード間の「近さ」を1、非隣接ノード間の「近さ」を0と

する離散的な類似度を考えます。次に配置されたノード間の座標の「近さ」を指数関数を用いて0と1の間に正規化した連続的な類似度を考えます。さらにこの2種類の類似度間のクロスエントロピーという量を考え、これを最小化することによって2種類の類似度の間でできるだけ矛盾が生じないように配置を計算します。この最小化を逐次改善法に基づく効率のよいアルゴリズムによって高速に実行することができます。

図5は図1、2と同じデータに対する提案可視化手法による可視化結果です。図1、2に比べ、図5のほうが、全体として一様に放射状に広がり、空間を効率的に活用したバランスの取れた可視化結果が得られていることが分かります。またネットワークの一部を切り出して再度可視化しても構造の変化が少なく、その意味で安定な手法であることが分かります。

同様に図6は図5と同じデータを用い、提案可視化手法を用いて3次元で可視化した例です。表示にはVRML (Virtual Reality Modeling Language) を用い、自由に視点を変えたり、拡大・縮小などの操作が可能です。またWeb以外の例として、図7に共著

者ネットワークの可視化例を示します。NIPS (Neural Information Processing Systems) という毎年開催されている神経科学・機械学習に関する国際会議において、2人の研究者が、過去10回の会議のうち、1回でも共著で論文発表をしている場合はその著者間にリンクがあるとみなしてネットワークを作成し、それを可視化したものです。図は可視化結果の1部を表示しています。

可視化結果の定量的評価

以上、具体的な可視化の方法と、定性的な可視化結果の良さについて説明しましたが、「見た目」は主観に依

存しますので、定量的に客観的な評価を行うことも必要です。また人間が実際に目で見る「可視化」の場合3次元が限界ですが、定量的な分析を行うのであればもう少し次元の高い空間への埋め込みも考えることができます。

そこで、種類の異なる複数のネットワークを古典的MDS、ばねモデル、提案法の3種類の方法で低次元のユークリッド空間^{*2}に埋め込んで配置し、隣接ノードが非隣接ノードより近くに配置されている割合を尺度として用いて定量的に評価した結果を図8に示します。テストに用いたネットワークは左からそれぞれ大腸菌の遺伝子制御ネットワーク(ノード数328, リンク

数456), NIPS掲載論文の共著者ネットワーク(ノード数1061, リンク数2080), NTTドメインに属するWebページのハイパーリンクネットワーク(ノード数2870, リンク数9337)であり、それぞれのグラフは縦軸が矛盾のない埋め込み結果である度合(隣接ノードが非隣接ノードより近くに配置されている割合)を、横軸は埋め込みの次元を示します。結果から容易に考察できるように、提案法が矛盾の少ない優れた可視化法であることが分かります。特にもっともネットワークのサイズが大きくて複雑なネットワークであるハイパーリンクネットワークの場合について従来法を顕著に上回っています。

3DWebブラウザなどへの応用

今後はさまざまな分野のネットワーク可視化ツールに本技術を応用していきたいと考えていますが、その中でも次世代の3DWebブラウザなどは興味深いアプリケーションであると考えています。図9に本技術を応用した

*2 ユークリッド空間: n 次元の2点 (x_1, x_2, \dots, x_n) と (y_1, y_2, \dots, y_n) との距離が $(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2$ で定義される空間。この距離をユークリッド距離といいます。例えば平面は2次元ユークリッド空間。

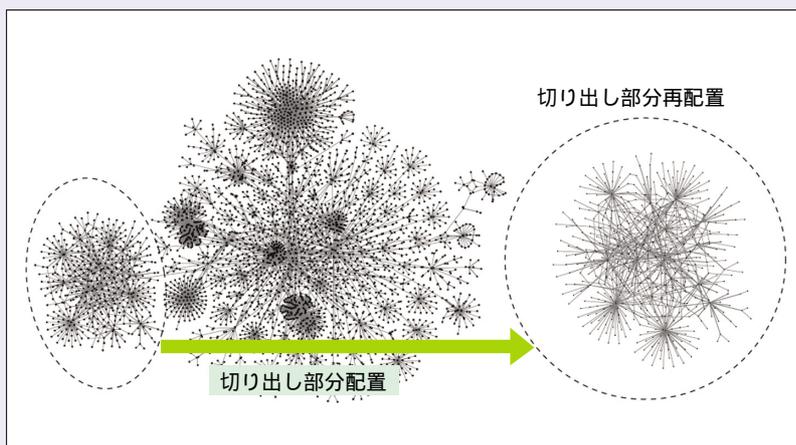


図5 提案法によるNTTドメインに属するWebページの可視化結果

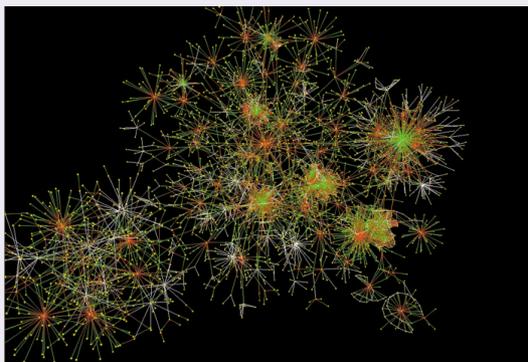


図6 提案法によるNTTドメインに属するWebページの3次元での可視化結果

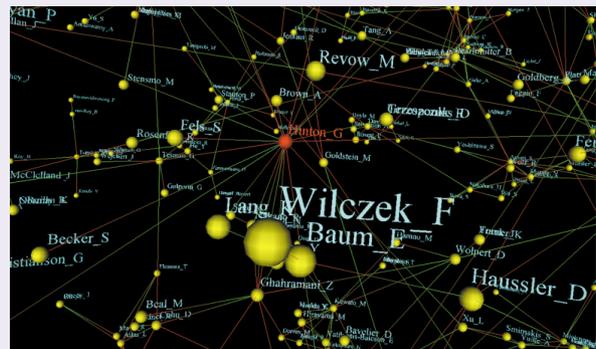


図7 提案可視化手法によるNIPS掲載論文の共著者ネットワークの可視化結果

d次元ユークリッド空間に埋め込んだ際の忠実度を比較

比較手法 古典MDS (線形手法)
ばねモデル (非線形手法)
提案法 (非線形手法) の計3手法

縦軸は、隣接ノードが非隣接ノードより近くに配置されている割合、横軸は埋め込みの次元

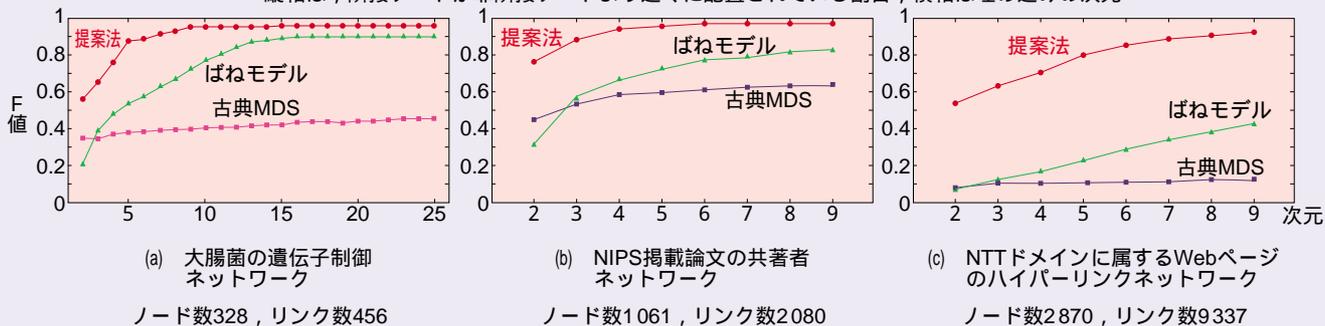


図8 可視化性能の定量的評価

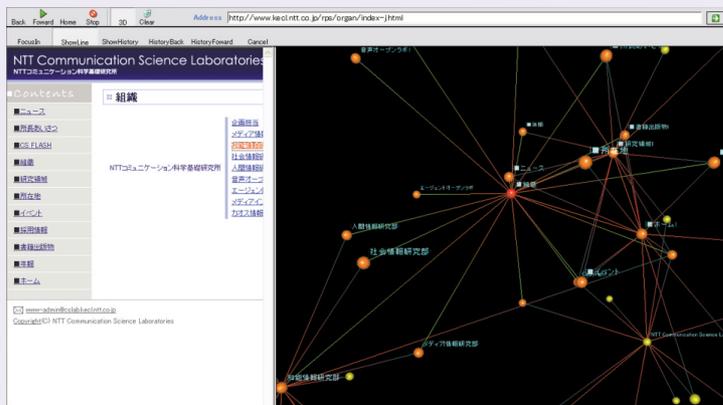


図9 3DWebブラウザの表示例

れています。

参考文献

- (1) 山田・斉藤・上田：“クロスエントロピー最小化に基づくネットワークデータの埋め込み,” 情報学論, 第44巻, 第9号, 2401-2408.
- (2) T.Yamada, K.Saito and N.Ueda: “Cross-Entropy Directed Embedding of Network Data,” in Proceedings of the Twentieth International Conference on Machine Learning, pp.832-839, 2003.



山田 武士

今後は、アルゴリズムをさらに高速化し、また重みや向きをついたネットワークへと適用範囲を広げていくとともに、真に分かりやすい可視化とは何かをさらに追求し、新しいインターフェースとしての可能性を探っていきます。

問い合わせ先

NTTコミュニケーション科学基礎研究所
企画担当
研究総括担当
TEL 0774-93-5040
FAX 0774-93-5045
E-mail yamada@cslab.kecl.ntt.co.jp

3DWebブラウザの実験システムの画面例を示します。画面左側は通常のWebブラウザと同じですが、右側に現在巡回中のネットワークの様子が逐次表示されています。通常のように左側のページにあるハイパーリンクをクリックしても、また右側のネットワークのノードをクリックしても、ページを移動することができます。ユーザがハイパーリンクをたどってページ間を移動すればするほど、得られた情報に基づいてネットワークは成長し、これまで得られたネットワークの全貌が明らかになります。

このように、ユーザは自分が現在巡回中のネットワークを確認し、実際のネットワークをたどりながらページ間を移動することができます。自分がこれまでたどってきたページの履歴を、ネットワーク上の軌跡として表示したり、軌跡上の移動を再現しながら確認したりできる機能なども備えており、よりリアルに、より快適に、インターネットを巡回することができます。

そのほかにも例えば、本可視化技術は前稿の『パラメトリック混合モデル(PMM)による多重トピック抽出』における可視化エンジンとしても活用さ