

自己組織化を利用したネットワーク可視化の評価と比較

岩田泰士[†] 鈴木育男[†] 山本雅人[†] 古川正志[†]

[†]北海道大学 大学院情報科学研究科

1 序論

複雑ネットワークの可視化はネットワーク特徴量だけでは把握できない構造的特徴を実際にみることができるとして期待されている。近年では、大規模データを扱う機会が増加し、可視化対象となるネットワーク構造もまた大規模化している。本研究では大規模ネットワークの可視化に対して高速な可視化を実現する自己組織化マップの学習機構を取り入れたグラフィック手法を提案する。ネットワーク可視化研究では主に力学的手法が取り扱われることが多く、自己組織化マップを利用した可視化に対する検証や改良が十分ではない。そのため、本研究では新たな手法の有用性を検証し、問題点を改善することを目的とする。

2 関連研究

ネットワーク可視化には多くの場合、力学的手法が利用されている。力学的手法の中でも有名な手法として KK バネモデルがあげられる。KK バネモデルは可視化空間上に配置されたノード間のユークリッド距離をグラフ的理想距離に近くなるように配置する手法である。ノード間にバネを想定し、系全体でエネルギーを最小化することで可視化問題を解く。しかし、一般的に力学的手法は計算コストが高いことが知られている。

力学的手法とは異なるアプローチから大規模なネットワーク構造を可視化する目的で考案された手法も存在する。ISOM(inverted self-organizing map) は Bernd Meyer[1] によって提案された可視化手法で、自己組織化マップ (SOM:self organizing map) を利用している。これは教師なし学習を利用し、高速な可視化を実現する。一方で、可視化したネットワークの形状が可視化領域に依存して変形してしまう問題点をもつ。

Comparison and evaluation of Network Visualization based on Self-Organization
Yasushi IWATA[†], Ikuo SUZUKI[†], Masahito YAMAMOTO[†], Masashi FURUKAWA[†]
[†]Graduate School of Information Science and Technology, Hokkaido University, 060-0814, Sapporo, Japan
iwata@complex.eng.hokudai.ac.jp

3 提案手法

ISOM の問題点を解決するために、予め可視化領域を設定する必要のない新たなアプローチとして DSSOM(Dynamically-Signaling Self-Organizing Map) を提案する。DSSOM の特徴はその可視化領域、つまり、信号領域が動的に変化することにある。信号領域は各ノードの周囲に円形に発生させる。この信号領域はノードの座標が更新される度にその位置が変化するため学習ステップ毎に異なる信号領域が設定される。

3.1 可視化アルゴリズム

DSSOM のアルゴリズムを以下に述べる。

1. 全ノードにランダムに座標ベクトルを与える
2. ノードの周囲に半径 r の円形の信号領域を定義する
3. 信号領域の範囲からランダムに座標ベクトルを選択し、入力信号ベクトル \mathbf{x}_s とする
4. 入力信号ベクトルとのユークリッド距離が最も近い座標 \mathbf{x}_c を持つ勝者ノード n_c を式 (1) で定める

$$n_c = \arg \min_i |\mathbf{x}_s - \mathbf{x}_i| \quad (1)$$
5. 勝者ノード n_c の周りに近傍領域 N_s を定義する ($n_c \in N_s$)
6. $n_i \in N_s$ のノードの座標ベクトルを式 (2) で更新する。このとき近傍 N_s 以外のノードの座標は変化させない。

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + h_{ci}(t)[\mathbf{x}_s(t) - \mathbf{x}_i(t)] \quad (2)$$
7. 指定したループ回数なら終了、そうでなければ 2. に戻る

ここで、近傍 N_s は、勝者ノード n_c からの最短経路長 $d(n_c, n_i)$ が近傍半径 $\sigma(t)$ 以内に存在するノード群を指す。初期の近傍半径 $\sigma(0)$ はネットワークの直径に応じて調整する必要がある。また、 $\sigma = 0$ の状態は勝者ノードの座標ベクトルのみを更新することになり、グラフィックとは関係なくノードがランダムに移動してしまう。この状態で学習を続けることは望ましくないため、 σ の最小値は 1 とする。

ノード座標を入力信号ベクトルに近づけるアニーリングを意味する近傍関数 $h_{ci}(t)$ は式 (3) で表される。

$$h_{ci}(t) = \alpha(t) \cdot \exp\left(-\frac{d(n_c, n_i)^2}{2\sigma^2(t)}\right) \quad (3)$$

近傍関数は勝者ノード n_c からのグラフ的距離が離れるにつれ更新の度合いを減少させる役割を果たす。 t は計算ステップであり、 $\alpha(t)$ と、 $\sigma(t)$ は共に計算ステップ t に応じた減少関数である。

4 数値実験

4.1 実験条件

DSSOM による可視化を評価するために、従来手法である ISOM と、力学的可視化アプローチとして知られる KK バネモデルとの比較を行う。可視化対象とするネットワークデータはネットワーク分野の論文共著ネットワーク (NetScience) を利用する。論文共著ネットワークはノード数 367, エッジ数 2742 である。本実験では目視による評価に加え、数値的に評価するため指標を導入し結果を観察する。

4.2 評価手法

4.2.1 接続 F 尺度

接続 F 尺度は文献 [2] にて提案されている評価法である。この評価法はネットワークを低次元の可視化空間に埋め込んだ際にノードの接続関係を忠実に再現した結果が得られているかを定量的な値 (F 値) として評価することができる。接続 F 尺度は全ノードで、隣接ノードが非隣接ノードより相対的に近い位置に配置された場合に最大値となる。

4.2.2 エッジ長分布

一般的な審美的基準に於いて、ノード間を結ぶエッジの長さは均一なものが良いとされている。そこで、各可視化手法についてそのエッジ長分布を観察し、分散をみることでエッジ長の均一性を評価する。エッジ長は描画されたネットワークのスケールに依存する。そのため、描画結果からバウンディングボックスを計算し、その対角長 L をもとにエッジ長を正規化する。

4.2.3 結果・考察

論文共著ネットワーク (NetScience) の可視化結果を図 2~図 3 に示す。ISOM では正方形の可視化領域にあわせて可視化結果の形状も歪んでいる (図 1)、一方で DSSOM では結果が可視化領域に影響を受けていない (図 2)。F 値の時間変化から DSSOM, ISOM は、共に速い段階で値が収束している。つまり、KK バネモデルと比較して高速に可視化結果を出力している。また、F 値自体も高い水準に達している。エッジ長の分散は KK, DSSOM, ISOM の順に小さい (表 1)。つま

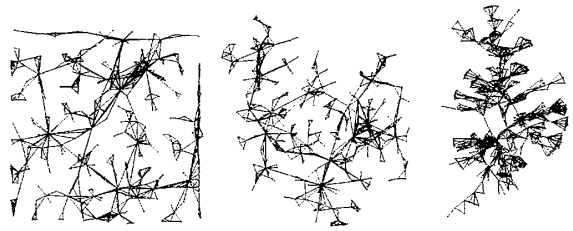


図 1: ISOM



図 2: DSSOM



図 3: KK

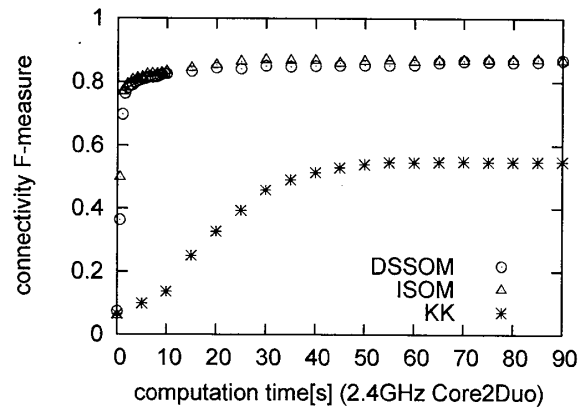


図 4: F 値の時間変化

表 1: エッジ長の分散

ISOM	DSSOM	KK
1.820×10^{-3}	0.482×10^{-3}	0.271×10^{-3}

り、エッジ長の均一性では KK バネモデルが最も良いことになる。ISOM は可視化結果が可視化領域全体に引き延ばされているため、エッジ長が他に比べ明らかに長いものが存在する。ISOM と比べると DSSOM では極端に長いエッジは削減されているため、分散は小さくなっている。

5 結論

本研究では、自己組織化の学習機構を利用した可視化手法として、新たに DSSOM を提案した。DSSOM は、F 値が高い評価を出していることから、隣接ノードを非隣接ノードより相対的に近い位置に配置する可視化を実現していることがわかった。結果出力までの計算時間も高速である。また、エッジ長の均一性では KK バネモデルに及ばないまでも、ISOM の問題点を改善し極端に長いエッジを削減することができた。

参考文献

- [1] Bernd Meyer. Competitive learning of network diagram layout. IEEE Symposium on Visual Languages, 56-63, 1998.
- [2] 山田武士, 齊藤和己, 上田修功. 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. 情報処理学会論文誌, Vol.44, No.9, 2401-2408, 2003.