

ネットワーク構造の局所エネルギー最小化による可視化の高速化

茂尾 亮太[†] 鈴木 育男[†]
山本 雅人[†] 古川 正志[†]

データ関係をノードとエッジで表すネットワーク構造は多くの分野で共通に利用されている。ネットワークの可視化技術は単なるノード間のつながりからは発見できないネットワークの構造や特徴を見出すのに有用な技術である。従来の研究で様々な可視化手法が提案されてきたが、本研究では力学的手法に焦点を当てる。力学的手法は実装と拡張が容易な最も一般的な可視化手法であるが、可視化のための計算量が大きく大規模なネットワークには適用することができない。しかし、比較的小規模なネットワークに対してはネットワークの特徴を捉え高速に可視化が可能である。提案手法では可視化対象のネットワークのある範囲内のノード群に対し、局所的エネルギー最小化によるノード配置をランダムに繰り返し行うことにより、ネットワーク全体の大域的なノード配置を導出する。これにより、計算量を削減し高速化に可視化を行う。また、従来手法との比較を行い提案手法の有用性を検証した。

Fast visualization of the complex network by use of local elastic energy minimization

Ryota SHIGEO[†] Ikuo SUZUKI[†] Masahito YAMAMOTO[†]
and Masashi FURUKAWA[†]

It becomes to make use of the network structure for representing many kind of relationship in many fields, where a set of entities and relationship among entities are described by a set of nodes and a set of links. Visualizing the network is a useful method to find and comprehend the network structure and its characteristic intuitively. this study focuses on a force-directed method, which is one of the most general methods. However, the computation time for visualization using this method becomes extremely large, as a size of networks grows larger. We propose a fast visualization method, which based on repeating local elastic energy minimization for randomly selected sub-networks. Numerical experiments verifies that the proposed method is capable of visualizing network fast with low computation time in comparing to conventional force-directed methods.

1. 序論

データ関係をノードとエッジの単純な構成要素で表すネットワーク構造は多くの分野に共通に利用されている。電力網や航空網などといった身近なものから、神経細胞、WWW のようなり

ンクで結ばれた仮想的なオブジェクト関係まで、1つのネットワークとして表現できるからである。これらは、ネットワークの構成要素1つ1つは単純な振る舞いしか行わないにも関わらず、ネットワーク全体としては様々な振る舞いを示すことから、近年では複雑ネットワークと呼ばれる。従来の複雑ネットワークの研究により、社会、経済、生物、情報などの分野の異なるネットワークがスモールワールド性やスケールフリー性などの共通の性質を有することが報告されてきた。現在では計算機の性能向上により扱うデータも年々規模が拡大しているため、ネットワーク構造も大規模化している。大規模化に伴い、ネットワーク構造の理解は特徴量という数値データを導出することで行われてきた。しかし、数値データのみでネットワークの特徴の全てを理解することは不可能であることも事実である。従って、ネットワークの可視化はネットワークの構造から人間の高い認識能力を利用し、新たな知的発見を導く重要な技術として考えられている。このネットワークの可視化もネットワークの規模の拡大により、高速に大規模なネットワークを可視化することが要求されてきている。

本研究では、力学的手法 (Force-directed Method) を用いた可視化の高速化を目的とする。力学的手法は広く一般的に用いられている可視化技法で実装と拡張が容易な手法であるが、規模が大きくなるにつれ計算量が極端に増大する問題点がある。しかし、力学的手法はネットワークの規模が小さいと比較的速く可視化が可能である。提案手法では可視化対象のネットワークのある範囲内のノード群の局所的最適配置を導出することで計算量を減少させ高速化を図る。更に、繰り返し局所最適化を行うことでネットワーク全体の大域的な最適配置を導出する。また、提案手法と既存の可視化手法の比較を行い、提案手法の有用性を検証する。

以下に本論文の構成を示す。第2章では従来の可視化における研究から現在の大規模ネットワークの可視化技術について述べる。第3章では可視化問題の問題提起と本研究の目的について述べる。第4章では提案手法における可視化アルゴリズムを説明する。第5章では各手法の評価を行うための評価式について述べる。第6章では数値実験を行った結果を示し、その考察を行う。最後に第7章でこれらのまとめを行う。

2. 関連研究

ネットワークの可視化はグラフィケアウト問題と共に発展してきた。グラフィケアウトの目的は、閲覧者に誤解を与えず適切なオブジェクト (ノード) の配置を自動的に行うことである。一般的にグラフ構造を描画する際に使われる審美的基準としては、交差するエッジの最小化、エッジ長の均一化、ノード分布の均一化、グラフ全体の対称性の有無などがあげられる。これらの制約の中には互いに反するものもあり、1つの単純な制約を満たす配置を導き出すにもほとんどの場合 NP 困難であることが知られている。従って、実際のネットワークの可視化においては可視

[†]北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University

化の目的や可視化対象のネットワーク構造などに合わせて審美的基準の優先度を定めレイアウトを行っていく。既存の多くの可視化手法はエッジを直線として描画しているが、直線以外の描画手法も提案されている。例えば、曲線描画はノード間を結ぶエッジを曲線として表現する。エッジを曲線的に表現することで閲覧者にやわらかい印象を与え、可視化におけるノードの配置の自由度を増やす手法である。また、ネットワークの規模が大きくなっても可視化空間を有効的に活用できるので、エッジのクロスや、ノード同士の重なりを防ぐことができる。

現在、様々なグラフィックレイアウト手法が提案されている[1,2]が、無向グラフを描画対象とし、その自動配置を目的として開発された力学的手法が有名である。力学的手法は P.Eades[3]により提案されたばねモデルが基礎となっている。この手法はノード間を結ぶエッジを仮想的なばねとみなし、グラフの各ノードがばねから受ける力が系全体で最小となるようなノード配置(系の安定状態)を求める。しかし、この手法は各ノードの計算式が非常に簡略されているため(1)綺麗なレイアウトが得られない、(2)解が収束しない可能性がある、などの問題点が存在する。ばねモデルを改良した手法として Kamada & Kawai Model (KK 法) [4]がある。この手法は力学的手法の中で最も一般的に使用される手法である。ばねモデルとの違いは、全てのノードが隣接・非隣接に関わらず決められた自然長のばねにより接続され、1回の計算においてグラフを構成するノードの中で、移動した際に最も系全体のエネルギーが減少するノード1つのみを移動する点である。このモデルは、各ノードに掛かる力をより厳密に定義しているため、ばねモデルの問題点を解決しているが、計算量が大きいため大規模なグラフには適用することが難しい。他にも、Fruchterman, Reingold[5]らは、ばねモデルにおける引力・斥力の定義を単純なものに改変し、最適化の過程にアニーリングを導入した FR 法を提案した。この手法は KK 法の計算量の大きさをかなり改善しており、アルゴリズムも比較的簡単なため、一般的なグラフのノードの最適配置を導出するには有用である。T. Matsubayashi[6]らは FR 法を改良し、ノード毎に固有の更新頻度を持たせポテンシャルエネルギーが高いものを優先して計算する手法を提案している。力学的手法以外のグラフィックレイアウト手法も提案されており、代表的な可視化手法としては自己組織化マップ (SOM: Self-Organizing Map) を利用した手法や、多次元尺度法などがあげられる。SOM に基づく手法は、力学的手法に比べ高速であり、隣接ノード同士が可視化空間において近い位置に配置されるためある程度の意味を持つ結果が出力される。しかし、パラメータの設定により結果が大きく変わってくるため、可視化対象に合わせたパラメータの設定が非常に難しい。SOM に基づく可視化手法としては、Bernd Meyer[7,8]により提案された ISOM (Inverted Self-Organizing Map) がある。多次元尺度法 (MDS: Multi-Dimensional Scaling) は、ノード間のパス長や隣接行列を用いて類似度を導き出し、高次元のデータ群を 2 次元あるいは 3 次元の可視化空間に投影することで可視化を行う。近年では代数的なアプローチによるレイアウト手法も提案されている。ACE (Algebraic multigrid Computation of Eigenvectors)[12]は代数的マルチグリッド法により高速にラプラス行列の固有値を導出しレイアウトを行う。HDE (High-Dimensional Embedding)[13]はグラフィックレイアウトにおいて高次元空間への埋め込みが比較的容易に行えることに着目した手法である。高次元空間へのノードレイアウトを導出し、主成分分析を利用して任意の低次元空間に

埋め込みを行いレイアウトを行う。これらの代数的なアプローチは、力学的手法や SOM などの手法に比べ非常に高速であり木構造や格子構造に近い特定のグラフ構造を持つネットワークに対しては比較の意味のある可視化結果が得られるが、一般的なグラフに対しては必ずしも良い結果が得られない。

高速化とは異なるが大規模なネットワーク構造の可視化のための技術として粗視化 (粗雑化: coarsening) [14]がある。粗視化はコミュニティ分割手法などを用いて、共通の性質をもつノード群を 1 つのノードとみなす。その結果、可視化対象のノード数を擬似的に減らし可視化を行える。グラフィックレイアウトにおいて可視化空間は有限空間であるため、ネットワークの規模が大規模になるほどレイアウトが困難になる。また、ネットワークの全てのノードを表示するよりも特徴的な構造のみを表示したほうが理解しやすい場合もあるので、大規模なネットワークの可視化には有益な手法といえる。

力学的手法は SOM や MDS を用いた手法に比べ可視化に時間を要する。これは各ノードに働く力を求める際の計算量が大きいためである。したがって、可視化に要する時間はネットワークの規模が大きくなると急激に増加する。本研究ではネットワークの一部に対して力学的手法を適用し局所的な最適配置を導出し、これを繰り返し行うことで最終的にネットワーク全体を最適な配置へと導く方法を提案する。力学的手法を適用する範囲を限定することで問題点である計算量の大きさが改善可能である。

3. ネットワークの可視化問題

ネットワークの可視化問題には明確な解が存在しない。したがって、可視化の目的やネットワーク構造に合わせて擬似的に最適解を設定し、その最適化問題を解くことにより最適なグラフィックレイアウトを導く。

本研究では以下を可視化目的とする。

- (1) 力学的手法の特徴である可視化した際の美しさを、極力損なわずに高速化を図る
- (2) 非隣接ノードは隣接ノードよりも可視化空間上において相対的に離れた位置に配置する
- (3) エッジ長の均一化を図る

可視化目的(1)の数値的比較はスカラーポテンシャルエネルギーにて行う。力学的手法はグラフィックレイアウトを対象の系におけるスカラーポテンシャルエネルギーの最小化問題と捉える。スカラーポテンシャルエネルギーは可視化の過程が進行するほど減少していき、最終的にはある一定の値に収束する。したがって、この値が低い方がより可視化が進行していると捉えることができる。可視化目的(2)の数値的比較は山田[8]らが提案した接続 F 尺度を用いる。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も近い位置に配置された場合に最大値を取る値である。したがって、可視化の精度を測定するには有用な手法である。なお、可視化目的(3)はエッジ長の分布により比較を行う。

4. 可視化アルゴリズム

4.1 提案手法における可視化アルゴリズム

ネットワークのノードの集合 $V = \{i\}$ とそのエッジの集合 (結合関係) $E = \{e_{ij}\}$ が与えられたときの提案手法のアルゴリズムを以下に示す.

- (1) 各ノードに初期値としてランダムな座標を与える
- (2) ランダムにノード n_0 を選択する
- (3) n_0 の m 次近傍まで各ノードに対して働く力を導出
- (4) n_0 の m 次近傍までの $x_i(t+1)$ における座標値の計算
- (5) $t = t+1$ とし時刻を更新
- (6) $t = T_{END}$ ならば終了, そうでなければ(2)へ戻る

座標の更新式およびノードに働く力の定義式に関しては次節以降で述べる.

4.2 力の定義

グラフィケアウトにおいて隣接ノードは非隣接ノードよりも空間的に近い位置に存在するほうが良い. 力学的手法では, 各ノードに働く力によってレイアウトを行う. 力学的手法の基本的な考え方は, 隣接ノード同士は互いに引き寄せあう力 (引力) を設定し, 互いの位置を近づける. 逆に非隣接ノードとは互いに反発しあう力 (斥力) を設定し, 互いの位置を遠ざける. 提案手法ではこれらの力に加え, 対象のサブグラフに属するノードに対し, ある力を加えることにより高速化を図る.

ノード i に働く引力は式(1)に従う. ノード間に働く引力は隣接ノード同士を可視化空間において近い位置に配置させる働きを持つ.

$$F_{i,j}(t) = \frac{1}{\alpha} \sum_{j \neq i} \frac{E_{ij}}{|x_{ij}|} \cdot x_{ij} \quad (1)$$

ここで, α は任意定数, $x_{ij} = |x_i - x_j|$ である.

ノード i に働く斥力を式(2)~(3)に示す. 斥力はノード同士の重なりを防ぎ, 非隣接ノード同士を可視化空間において遠ざける働きを持つ. 従来の力学的手法では斥力の計算に必要な計算量が大きい. 提案手法ではノード i を中心とする半径 R の球を仮定し, この球内に存在するノードとのみ斥力計算を行うことで計算量を削減する.

$$F_{i,r}(t) = -\alpha^2 \sum_{j=1}^N F_i(i, j) \quad (2)$$

$$F_i(i, j) = \begin{cases} x_{ij} / |x_{ij} + \varepsilon|^2 & (|x_{ij}| \leq R) \\ 0 & (|x_{ij}| > R) \end{cases} \quad (3)$$

ここで, ε は $|x_{ij}| = 0$ のとき $F_i(i, j) = \infty$ になることを防ぐための制御定数である.

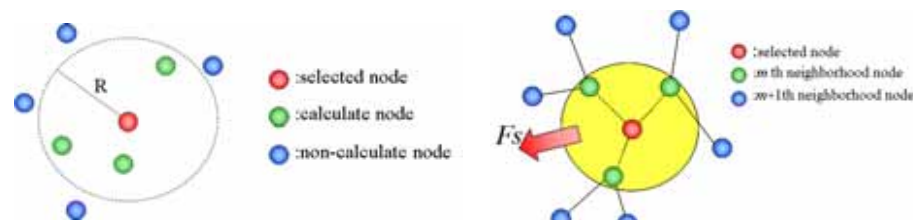


図1 斥力 F_r
Fig.1 Definition of F_r

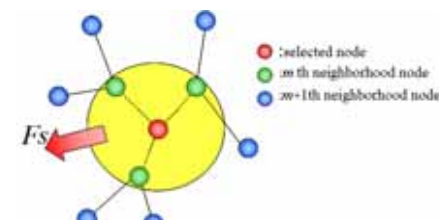


図2 サブネットワークにかかる力 F_s
Fig.2 Definition of F_s

ノード i の移動はその近傍ノードの座標に従う. 従って, ノード i の配置は自身の隣接ノードである1次近傍の配置の影響を受けることになる. また, 1次近傍ノードはその近傍ノードの影響を受けるため, ノード i は2次近傍の影響も1次近傍よりは少ない範囲で受けることになる. したがって, 提案手法ではノード i に近傍の回数に応じた力を与えることにより, エネルギー最小化問題における解の収束を高速化する.

ノード i の m 次近傍までに属するノードに働く力 F_s は以下の式で定義する.

$$F_{i,s}(t) = \sum_{j=1}^{m-n} \frac{m-n-j}{(m-n) \cdot (m-1-n)} \cdot F_i(i, j) \quad (4)$$

$$F_i(i, j) = \frac{1}{\#\Gamma_i^{j+1}} \sum_{k \in \Gamma_i^j} |x_{ik}| \cdot x_{ik} \quad (k | x_k \in \Gamma_i^j, l | x_l \in \Gamma_i^{j+1}) \quad (5)$$

n はノード i の n 次近傍目を表し, Γ_i^m はノード i の m 次近傍ノードの集合, $\#\Gamma_i^m$ はそのノード数である.

4.3 座標の更新式

力学的手法において座標の移動はノードに働く力の合力によって決まる.

式(6)~(7)に座標の更新式を示す.

$$x_i(t+1) = x_i(t) + F_i(t) \cdot \Delta t \quad (6)$$

$$F_i(t) = F_{i,r}(t) + F_{i,s}(t) + F_s(t) \quad (7)$$

Δt は時間の変化量で本研究では $\Delta t = 1$ を使用する.

5. 評価式

5.1 スカラーポテンシャルエネルギー

力学的手法において可視化の速度はスカラーポテンシャルエネルギーを用いて比較可能である. 力学的手法はグラフィケアウトを対象の系におけるスカラーポテンシャルエネルギーの最小化問題と捉え可視化を行う. 可視化の過程が進むと系全体のスカラーポテンシャルエネルギーは



図3 FRによるCollege footballの可視化結果
Fig.3 FR (College football)



図4 FRによるCollege footballの可視化結果
Fig.4 ISOM (College football)



図5 提案手法によるCollege footballの可視化結果
Fig.5 Proposed method (College football)

減少し、最後には一定値に収束する。従って、系のスカラーポテンシャルエネルギーが小さいほど可視化が進行していることになる。スカラーポテンシャルエネルギーは以下の式で与えられる。

$$\Psi(\mathbf{x}_{ij}) = \frac{1}{3\alpha} \sum_{i,j} |\mathbf{x}_{ij}|^3 - \frac{\alpha^2}{2} \sum_i \sum_{j \neq i} \log[|\mathbf{x}_{ij}|^2 + \epsilon^2] \quad (8)$$

また、ノードに働く力 F とスカラーポテンシャルエネルギー Ψ の関係を以下に示す。

$$F(\mathbf{x}_{ij}) = \nabla \Psi(\mathbf{x}_{ij}) \quad (9)$$

5.2 接続F尺度

接続F尺度は山田[9]らにより提案された評価法で、隣接ノードは非隣接ノードよりも可視化空間において近い位置に配置されるべきという考えに基づいている。この評価法はある N 次元の可視化空間にネットワークを埋め込んだ際に、ノードの接続関係をどれだけ忠実に再現できたかを F 値を定義して表現できる。 F 値はノード毎に異なる半径 r_i の超球 $B_i(r_i)$ を仮定する。この F 値は対象ノード i の超球 $B_i(r_i)$ 内にノード i との全隣接ノードが存在し、かつ非隣接ノードが全く存在しないとき最大値 1 を得る。しかし、実際の可視化において可視化空間は有限であるため、 F 値が最大値となるネットワークは稀である。また、可視化空間の次元が小さくなるほどノード配置が限定されるため、同じネットワークを可視化しても F 値は低くなる。

ノード i における F 値は適合率 $P_i(r_i)$ と再現率 $R_i(r_i)$ にて定義される。

$$P_i(r_i) = \frac{\#\{j | x_j \in B_i(r_i), a_{ij} = 1, j \neq i\}}{\#\{j | x_j \in B_i(r_i), j \neq i\}} \quad (10)$$

$$R_i(r_i) = \frac{\#\{j | x_j \in B_i(r_i), a_{ij} = 1, j \neq i\}}{\#\{j | a_{ij} = 1, j \neq i\}} \quad (11)$$

ここで、 $\#X$ は集合 X の要素数を表し、 a_{ij} は隣接行列を表す。各式より $P_i(r_i)$ は超球 $B_i(r_i)$ に含まれるノードにおける隣接ノード割合を示し、 $R_i(r_i)$ は隣接ノードのうち超球 $B_i(r_i)$ に含まれる隣接ノードの割合を示す。ノード i の F 値は適合率と再現率の調和平均にて得られる。本研究で

表2 エッジ長分布

Table 2 Variance of the edge length

	College football	Power grid
Proposed method	0.0132	1.23×10^{-4}
FR	0.0188	2.37×10^{-4}
ISOM	0.0194	1.97×10^{-4}

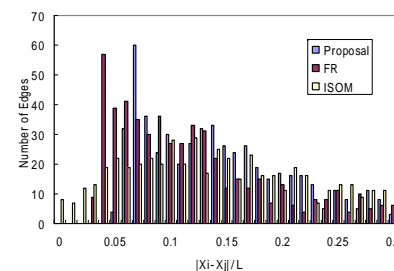


図7 エッジ長の度数分布(College football)
Fig.7 Histogram of edge length (College football)

は $\alpha = 1/2$ とする。

$$F_i(r_i) = \frac{1}{\alpha / P_i(r_i) + (1 - \alpha) / R_i(r_i)} \quad (12)$$

$$F = \frac{1}{N} \sum_{i=1}^N \max(F_i(r_i)) \quad (13)$$

なお、ノード i における超球の半径 r_i は $F_i(r_i)$ が最大値となる値を採用する。

5.3 エッジ長分布

一般的な審美的基準では、グラフィアウトにおいてエッジ長は均一であることが望ましいとされている。そこで各手法におけるエッジ長の分散を比較する。しかし、対象のネットワークの規模やスケールなどによってエッジ長は変化してしまい一概に比較することはできない。そのため、可視化結果から、ネットワークのバウンディングボックスを導出し、これを用いてエッジ長を正規化することで比較を行う。

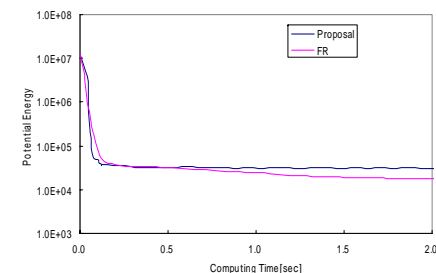


図6 ポテンシャルエネルギーの変化
Fig.6 Potential Energy (College football)

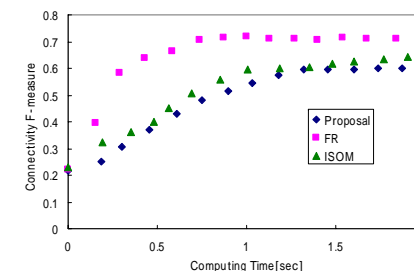


図8 College footballにおける接続F尺度の変化
Fig.8 F-measure (College football)



図9 FRによる Power grid の可視化結果
Fig.9 FR (Power grid)

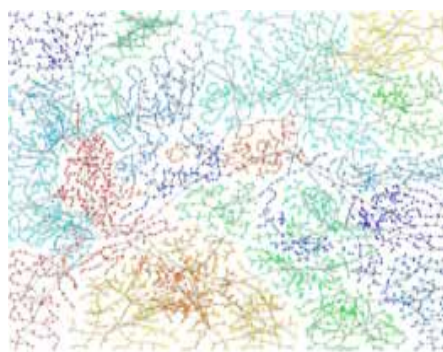


図10 ISOMによる Power grid の可視化結果
Fig.10 ISOM (Power grid)

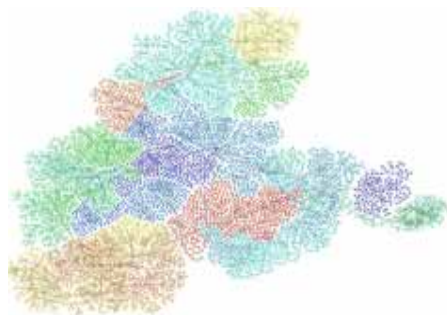


図11 提案手法による Power grid の可視化結果
Fig.11 Proposed method (Power grid)

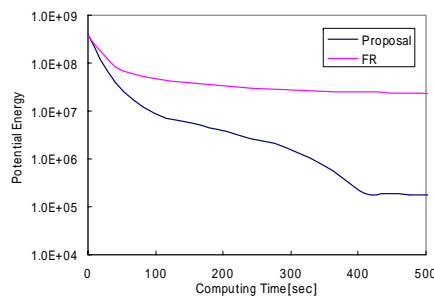


図12 power grid における
ポテンシャルエネルギーの変化
Fig.12 Potential Energy (Power grid)

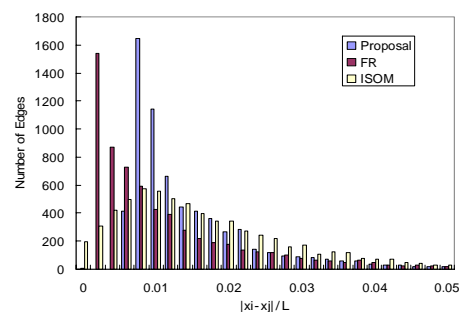


図13 エッジ長の度数分布(power grid)
Fig.13 Histogram of edge length (Power grid)

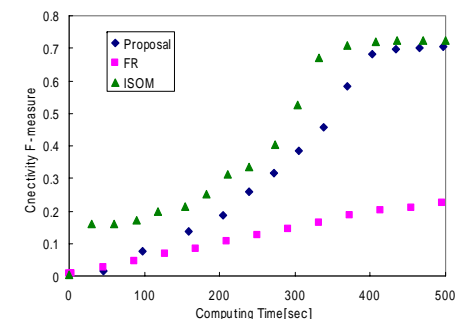


図14 power grid における接続 F 尺度の変化
Fig.14 F-measure (Power grid)

6. 数値計算実験

6.1 実験条件

ネットワークの可視化問題に対する提案手法の有効性を検証するために、提案手法と同じ力学的手法であるFRと自己組織化マップを可視化に応用したISOMを用いて比較を行う。使用するデータはアメリカの大学におけるフットボールチームのネットワーク[10]とアメリカ西部の電力網のデータ[11]を使用する。フットボールのネットワークはノード数115、エッジ数616の規模の小さなネットワークデータのため、力学的手法のような計算量が大い可視化手法でも短時間で可視化が可能である。一方、アメリカ西部の電力網ネットワークはノード数4941、エ

ッジ数6591の比較的大規模なネットワークデータである。このデータを使用する利点としては、地理的關係という2次元関係データのため2次元空間上に可視化を行った際に構造が理解しやすく評価を行い易いことがあげられる。

6.2 結果と考察

アメリカの大学におけるフットボールチームのネットワーク(College football)の可視化結果を図3~図5に示す。どの手法も即時に結果が出力され、College footballは規模が小さいネットワークデータであるため手法による可視化結果の違いはあまり見られなかった。提案手法とFRのポテンシャルエネルギーの変化を図6に示す。提案手法の方がFRよりも速くエネルギーが減少するが、最終的にFRの方が低いエネルギーまで低下している。力学的手法は可視化対象の系のエネルギー最小化問題を各ノードに従って移動させることにより解いている。提案手法は系全体のエネルギー最小化を直接解かず、系の一部のノード群のエネルギーを最小化させ、この処理を繰り返すことで系全体のエネルギー減少させる。したがって、系全体のエネルギーに着目したFRよりもエネルギー最小化問題の局所解に陥る可能性が高くなってしまいうため、エネルギーが下がらないと考えられる。また、系の一部に着目することで力学的手法の欠点である計算量の大きさを改善し可視化の高速化を図ったが、ネットワークの規模が小さく力学的手法でも十分に高速に対応できたため、速度面での提案手法の利点が見られなかった。エッジ長の分布では提案手法の分散がFRとISOMと比較して最も小さくなっている(表2, 図7)。提案手法は選択されたノード群が近傍のノードの動きに合わせるような力を加えたためだと考えられ、エッジ長の均一性では提案手法は優れた手法である。一方、F値は他の手法よりも悪い(図8)。これはノードの配置が他の手法と比べ均一になるよう配置されるため、非隣接ノードも可視化空間上において比較的近い位置に存在してしまう。従って、F値の採択基準を満たさず悪い評

価につながったと考える。ISOM が力学的手法に比べエッジ長の均一性が良くないのは ISOM の特性によるところが大きい。ISOM は決められた信号空間上にランダムな信号を入力する。この入力信号に最も近いノードを勝者とし、勝者ノードとその近傍回数で与えられる近傍ノードが入力信号に向かって移動させる。これを繰り返し行うことで可視化を行うため、与える信号の範囲の影響を受ける。本研究では正方形の 2 次元空間上でランダムな点を選択し入力信号として ISOM に与えている。ISOM は入力信号と同じ正方形上に引き伸ばしてしまうため、力学的手法に比べ極端に長いエッジが多くなり、ノードの均一性が悪くなった。

図 9~ 図 11 にアメリカ西部の電力網ネットワークの可視化結果を示す。Power grid はコミュニティ分割手法の一手法である CNM 法にて 63 個のコミュニティに分割が可能であり、各ノードの色は所属するコミュニティを表す。Power grid は College football に比べエッジ数が 10 倍、ノード数が 40 倍の大規模なデータである。可視化結果の出力には提案手法と ISOM が数分程であるのに対して FR は数時間の時間を要した。データの規模が大きくなったことにより、各手法による可視化結果の特長が顕著に現れている。FR は同じコミュニティに属するノード同士が他の 2 つの手法よりも明確に表現されている (図 9)。ISOM は入力信号の信号領域への依存性がより強く現れている。その結果、可視化結果が入力信号領域と同じ正方形上に引き伸ばされてしまい、ネットワーク構造がわかりにくい結果となる (図 10)。提案手法も FR 同様に同じコミュニティに属するノード同士が空間的に近い位置となっている (図 11)。提案手法と FR のポテンシャルエネルギーの変化を比較すると、提案手法が FR に比べ急激にエネルギーが落ちていることがわかる (図 12)。提案手法は、可視化対象のネットワークに対して局所的にエネルギーが最小化になるように処理を行っている。従って、大規模な 1 つのネットワークを、無数の小規模なネットワークの集まりとして扱うことができる。力学的手法は College football の数値実験からもわかるように、小規模なネットワークに対しては比較的高速とされる ISOM と同等の速度で可視化が可能である。よって、データの規模を小さく抑えることができたため、提案手法は FR よりもかなり高速にエネルギーを減少させることができたと考えられる。F 値は増加率、収束値ともに ISOM には及ばないが非常に高い値に収束していることがわかる (図 14)。ISOM が可視化空間全体に均一にノードが配置されるのに対して、提案手法では同一のコミュニティに属するノード同士が密集する形状のため可視化空間におけるノード数の密度にばらつきが生じる。従って、ISOM より狭い空間内に同数のノードが存在するため提案手法の F 値の評価値が増加しにくかった。表 2 より、分散が小さいことからノードの均一性は提案手法が他の手法よりも優れているといえる。

7. 結論

本研究では力学的手法の高速化を目的とした手法を提案し、数値実験を通してその有効性を検証した。それらをまとめると以下ようになる。

- (1) ランダムな局所的エネルギーの最小化によりネットワーク可視化を実現出来ることを示した

- (2) 大規模なネットワークも無数の小さなネットワークとして扱うことが出来るため、従来の力学的手法に比べ計算コストを下げる事が可能
- (3) ネットワークの一部のノードを同時に移動させることにより可視化速度が向上することを示した
- (4) 提案手法はノードの均一性に優れた手法である

本研究では大規模なネットワークデータに局所エネルギー最適化を適用した。これは並列計算に適したアルゴリズムのため、並列計算への適用を行い更なる高速化が可能だと考えられる。一方で、同時に移動させる近傍数による形状への影響をより詳しく調査する必要がある。

参考文献

- [1] G.Battista, P.Eades, R.Tamassia, and I.Tollis, Algorithms for Drawing Graphs an Annotated Bibliography, Computational Geometry Theory and Applications,4, 235.282, 1994.
- [2] G.Battista, P.Eades, R.Tamassia, and I.Tollis, Graph Drawing - Algorithms for the Visualization of Graphs, Prentice-Hall, 1999.
- [3] Eades, A Heuristic for Graph Drawing, 1984
- [4] Kamada, Kawai, An algorithm for drawing general undirected graphs, 1989
- [5] Fruchterman, Reingold, Graph Drawing by Forcedirected Placement, Software - Practice and Experience, 1991
- [6] T. Matsubayashi, T. Yamada, A Force-directed Graph Drawing based on the Hierarchical Individual Timestep Method, 2007
- [7] Bernd Meyer, Competitive learning of network diagram layout, In Visual Languages, pp. 56.63, 1998
- [8] Bernd Meyer, Self-organizing graphs A neural network perspective of graphlayout Lecture Notes in Computer Science, Vol. 1547, pp. 246.262, 1998.
- [9] 山田武志, 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, pp.2401-2408, 2003
- [10] Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA 99, 7821-7826, 2002
- [11] D. J. Watts and S. H. Strogatz, Nature 393, 440-442, 1998
- [12] Y.Koren, L.Carmel, and D.Harel, ACE:A fast multiscale eigenvectors computation for drawing huge graphs, Proceeding of the IEEE Symposium on Information Visualization 2002, pp.137-144, 2002
- [13] D.Harel and Y.Koren, Graph drawing by high-dimensional embedding, Graph Drawing 2002, LNCS, vol2528, pp.207-219, 2002
- [14] D.Archambault, and T.Munzner, TopoLayout: Multi-Level Graph Layout by Topological Features, Graph drawing 2004, vol.3383, pp285-295, 2005