

局所エネルギー最小化法におけるパラメータの効果

北海道大学大学院 ○辻本陽平, 北見工大 鈴木育男, 北大 山本雅人, 北大 古川正志

要旨

WWW や電力網, 神経網, 通信網, 知人ネットワークなどの様々なデータをノードとエッジで表現したグラフ理論において, 複雑ネットワークの可視化は, 大規模データを扱う上で重要な技術である. 本研究では, 力学的手法を用いて高速な可視化手法である「局所エネルギー最小化法 (LEM)」のノード間に働く引力・斥力にかかる重みによる影響と効果を検証する. LEM の引力・斥力の重みをパラメータ化しレイアウトを行い, エネルギーの比較を行う.

1. はじめに

様々なデータをノードとエッジ (リンク) の単純な組み合わせで表現するネットワークは, 多くの分野において幅広く共通に用いられている. WWW, 電力網, 神経網, 通信網, 知人ネットワーク, 生態系などの関係はネットワークで容易に表現できるが, 単純なルールしかないにもかかわらず, 大域的に見ると複雑な振る舞いをしていることから, 複雑ネットワークと呼ばれる分野が注目を集めている.

こうしたネットワークの可視化は, 人間の高い認識能力を利用したネットワークを直感的に理解する方法として, 高速な方法が求められている.

本研究では力学的手法を基にした, 高速にレイアウトするネットワークの可視化手法である LEM アルゴリズムにおいて, ノード間に働く引力・斥力にかかる重みによる影響と効果を検証する. この目的を達成するため, LEM の引力・斥力の重みを調整しレイアウトを行い, そのときのエネルギーの比較を行う.

2. グラフレイアウト問題

ネットワークの可視化はグラフレイアウト問題として扱うことができるが, その定量的な評価はいまいであり, 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の構造に合わせて評価関数を設定し, その最適化問題を解き, 準最適なグラフレイアウトを求めるのが一般的である.

多くの提案されたレイアウトアルゴリズムの中では, 無向グラフのレイアウトを目的とした力学的手法がよく知られている. これは, P.Eades[1] によって提案されたバネモデルが基礎としている. バネモデルは, ノードをスチールリング, エッジをバネと見なし, そこに働く引力と斥力を力学系の平衡状態へと導くようにレイアウトを行うが, 計算コストが高く, ネットワークのノード集合を N とすると $O(N^2)$ の計算量を必要とする.

現在では, ばねモデルを改良した KK 法[2], FR 法[3] が, 可視化レイアウトに, 幅広く用いられている. これらの力学的手法は複雑な構造を持つネットワークデータをレイアウトできるが, 自己組織化マップを用いた手法などの他の手法に比べレイアウト時間がかかる問題点がある. その理由は, 力学計算における引力と斥力の計算コストが高いためである.

2.1 LEM (局所エネルギー最小化法)

LEM は, 茂尾ら[4] が提案した力学的手法である. LEM は, 計算コストを抑え, 高速にレイアウトが行える特徴をもつ. このレイアウトアルゴリズムは, 力学的な計算範囲ランダムに限定し, その範囲の局所的エネルギーの最小化により計算量を削減するとともに, これを繰り返すことで大域的なレイアウト最適化を実現する.

LEM は, ネットワークのノード集合とエッジ集合の関係が与えられたとき, 以下のアルゴリズムで行う.

1. 各ノードをランダムに配置する
 2. ランダムにノード i を選択する
 3. 近傍定数 M (パス長) から $\pm M/2$ の範囲でランダムに近傍数 m を定める
 4. ノード i の m 次近傍までのノード集合 Γ_i^m を決定する
 5. Γ_i^m に対して働く力を計算する
 6. 座標値 x_i を更新する
 7. 終了条件を満たせば終了, そうでなければ(2)へ戻る
- LEM では, Γ_i^m に対して働く力は3つ存在する. 第1はノード i に働く引力 F_i^a (式 (1)) である.

$$F_i^a = \frac{1}{\alpha} \sum_{i \neq j, j \in E_i} \|x_{ij}\| x_{ij} \quad (1)$$

ここで, α は引力の任意定数, $x_{ij} = x_j - x_i$, E_i はノード i の隣接ノード集合である.

第2はノード i を中心とした半径 R 内のノードに働く斥力 F_i^r (式 (2)) である.

$$F_i^r = -\beta \sum_{j=1}^N \begin{cases} \frac{x_{ij}}{\|x_{ij} + \varepsilon\|^2} & : \|x_{ij}\| \leq R \\ 0 & : \text{otherwise} \end{cases} \quad (2)$$

ここで, β は斥力の任意定数, N は全ノード集合である.

第3は, ノード集合 Γ_i^m で構成されるサブネットワークに働く力 F_i^s (式 (3)) であり, ノード i からパス長が離れれば離れるほど力は弱くなる. ここで, ノード i から近傍距離 $k, k+1$ に存在するノード集合を E_{ij}^k, E_{ij}^{k+1} とし, $[a_{uv}]$ はノード u とノード v の隣接行列, Γ_{k+1} を $E_{ij}^k - E_{ij}^{k+1}$ 間に存在するエッジ本数とする.

$$F_i^s = \sum_{k=2}^{m-1} \frac{m-k+1}{m-1} \frac{1}{\Gamma_{k+1}} \sum_{u \in E_{ij}^k, v \in E_{ij}^{k+1}} F_{uv}^s \quad (3)$$

$$F_{uv}^s = \begin{cases} \|x_{uv}\| x_{uv} & : a_{uv} = 1 \\ 0 & : \text{otherwise} \end{cases}$$

座標の更新は、上記3つの力の合力に重み Δt を乗じて加算することで行う。

3. 数値計算実験

KK法やFR法などの力学的手法によるグラフレイアウトでは、斥力の計算コストが特に高く、選択ノードとそれ以外のすべてのノードとの力を計算するからである。LEMでは、選択ノード i を中心とした半径 R 内のノードに斥力計算を行うことで計算コストを抑えている。斥力計算の半径を抑えることは、引力と斥力の均衡が一部崩れるため、半径 R に適した引力の重み α 、斥力の重み β を設定する。

本研究では、半径 R を固定し2種の重み α, β を調整する。また、LEMのポテンシャルエネルギー Ψ を式(4)に示す。

$$\Psi = \frac{1}{3\alpha} \sum_{j=i}^{E_i} \|x_{ij}\|^3 - \beta \sum_i^{N-1} \sum_{j=i+1}^N \log \|x_{ij} + \varepsilon\| \quad (4)$$

3.1 実験条件

本研究ではNewmanが作成したネットワーク科学の共著ネットワークデータを用いて、ネットワーク可視化をFR法、提案したLEMで行う。またそれらを、ポテンシャルエネルギー、エッジクロス数、山田らにより提案された接続F尺度[5]で比較する。接続F尺度($0 \leq F \leq 1$)は、任意のノードがその非隣接ノードより隣接ノードに近い位置にあれば高くなる。

LEMの α, β はそれぞれ、10,30,100とする。初期配置はランダム配置とし、一定時間を経過するまで可視化を行う。また、斥力の影響する半径を $R=30$ 、 $\Delta t=0.01$ とする。

3.2 結果・考察

可視化の結果を図1~3に示す。図1はFR法、図2~4はLEMの引力・斥力の重みを調整して可視化した結果である。図より、LEMはFR法よりノードが密集しないことがわかる。表1は、可視化終了時における各々 α, β の組合せのポテンシャルエネルギーを示す。表2、表3も同様に、可視化終了時におけるエッジクロス数、接続F尺度を示す。

表1より、ポテンシャルエネルギーは β に大きく影響されるが、これは式(4)右辺第2項によるものである。表2より、設定値の中では α, β が小さければエッジクロス数が減少することがわかる。表3より、 α, β が小さければ接続F値が増大することがわかる。

図および表より、本実験で用いたネットワークの斥力影響半径 $R=30$ においては、 α, β が大きいと引力が減少、斥力が増大し、エッジが長くなることがわかる。このため、エッジが長くなると、近くの別コミュニティにあるノード群と干渉しやすくなり、したがって、エッジクロス数の増大、接続F値の減少につながると考えられる。



Fig.1 FR-method

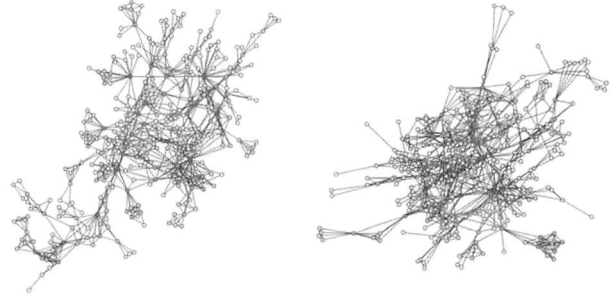


Fig.2 LEM($\alpha=10, \beta=10$)

Fig.3 LEM($\alpha=10, \beta=100$)

Table.1 Potential energy

$\alpha \setminus \beta$	10	30	100
10	-2.79E+6	-8.76E+6	-31.78E+6
30	-2.98E+6	-9.07E+6	-31.32E+6
100	-3.11E+6	-9.59E+6	-32.75E+6

Table.2 Number of edge cross

$\alpha \setminus \beta$	10	30	100
10	2437	3633	3083
30	2471	4402	5561
100	4798	6650	8102

Table.3 F-measure

$\alpha \setminus \beta$	10	30	100
10	0.514	0.450	0.456
30	0.411	0.374	0.364
100	0.394	0.347	0.300

4. まとめ

グラフレイアウトは大規模で複雑なネットワークの特徴を理解するのに重要な問題である。本研究は、LEMにおける引力・斥力の重みを調整し、レイアウトを行い、ポテンシャルエネルギー、エッジクロス数、接続F尺度への効果を調査した。今後は、引力・斥力の重みをより柔軟にし、その効果を検証する。

参考文献

- [1] P.Eades: A Heuristic for Graph Drawing, Congressus Numerantium, 42, pp.149-160, 1984.
- [2] T.Kamada and S.Kawai: An Algorithm for Drawing General Undirected Graphs, Information Processing Letters, 31, pp.7-15, 1989.
- [3] T.Fruchterman and E.Reingold: Graph Drawing by Forcedirected Placement, Software-Practice and Experience, 21, pp.1129-1164, 1991.
- [4] 茂尾亮太, 鈴木育男, 山本雅人, 古川正志: 局所エネルギー最小化による可視化の高速化 情報処理北海道シンポジウム 2009, pp.60-65 2009.
- [5] Saito K. Yamada T. and Ueda N. Embedding network data based on crossentropy minimization (knowledge processing). Transactions of Information Processing Society of Japan, Vol.44, No.9, pp.2401-2408, 2003.