

固有楕円ポテンシャルを利用したラベル付きグラフ可視化の座標計算

松林 達史[†] 山田 武士[†] 藤村 滋[‡] 藤村 考[‡]
日本電信電話株式会社 [†]NTT コミュニケーション科学基礎研究所
[‡]NTT サイバーソリューション研究所

概要: 本研究では、ラベル付きグラフ可視化のための、ラベル同士が重ならない効率的な可視化座標計算の手法を提案する。従来のラベル付きグラフ可視化の座標計算アルゴリズムは Force-directed 法やバネモデルといった、ノードを“点”として扱う座標計算を行うために、文字列や、異なるサイズのラベルを扱う時にはラベル同士が重なってしまうという問題が生じる。ラベルの重なりを回避する手法はいくつか提案されて来たが、いずれの手法も可視化計算を行った後に、再び座標計算処理を必要とする。これ等の手法では大規模なグラフ可視化では計算量が莫大になり、さらに元の可視化結果を破壊してしまうという問題も生じる。本提案手法では、各ノードにラベルサイズに依存した固有楕円ポテンシャルを与え、ラベルの大きさを考慮する事によって、ラベル同士の重なりを回避する座標計算を可能とした。さらに、斥力項に対し点対称ポテンシャルと固有楕円ポテンシャルの関数の重ね合わせにより、従来手法の可視化結果と比べ、大局的な構造を保ちつつ局所的にはラベルの重なりを回避する手法を提案する。また、提案手法を GPU を用いて計算させることにより、座標計算速度の数百倍の高速化を実現した。

Labeled Graph Drawing based on the Individual Ellipsoidal Potentials

Tatsushi Matsubayashi[†] Takeshi Yamada[†] Shigeru Fujimura[‡] and Ko Fujimura[‡]

[†]NTT Communication Science Laboratories

[‡]NTT Cyber Communications Laboratory

Abstract: We present a new graph layout algorithm for a graph with node labels. The proposed algorithm can generate a graph layout efficiently avoiding overlapping of node labels. Most of the conventional algorithms such as force-directed and spring methods compute node positions by treating nodes as “points”, and thus may cause node overlapping when strings or labels with various sizes are added to the nodes. Most of the conventional approaches to solve this problem require an initial graph layout generation phase without considering label sizes and a separate overlap removal phase as its post processing. These methods are not suitable for a large-scale graph layout, because their computational costs are high and they may even destroy the initial graph layout. The proposed algorithm gives an individually different ellipsoidal potential to each node depending on its label size as well as a common point-symmetric potential, and the combination of these two potentials defines the repulsion force. This enables an efficient graph layout that can avoid local node overlaps while maintaining the global structure. By implementing it on the GPU, the proposed algorithm can compute high quality layouts of large-scale graphs in some hundredths fraction of time required by the conventional algorithms.

1 はじめに

複雑な関係データの表現方法として、ネットワークもしくはグラフを用い、これ等を可視化するという手法は様々な研究分野で広く用いられている。

一般的なグラフ可視化においてはネットワークの関連性を元に、2次元もしくは3次元のユークリッド空間にノードを「点」として、エッジを点と点を繋ぐ「辺」として投影し描画する。しかしながら、各ノードが固有情報、例えばノードの名称や画像などを持つ場合、ノード上に直接文字列を埋め込んだり、縮小版画像を埋め込みたい場合もある。これらいずれの場合も、ノードは「点」ではなく大きさを持つ「ラベル」として描画する必要がある。

近年の研究では、周囲にラベルが重ならないように表示する手法としてボロノイ分割を利用した手法や、Slide model という方法も提案されている。これ等の手法は、与えられた座標値を基準にラベル同士が重ならないように再度座標計算を行う手法である。しかしながらこれ等の手法では、ノードの増加に伴い計算量が莫大に増加し、かつ元のネットワーク構造が保存されない等の問題がある。算量が莫大となるため、用途が限定されている。

本研究では、無向かつラベル付きグラフデータを対象に、ラベル同士が重ならない可視化手法を提案する。

2 従来研究

グラフ可視化において最もよく知られているのは“force-directed method”と呼ばれる手法であり、この手法は、ノード間のグラフ距離や接続関係などに基づき、系にポテンシャルを与え、加速度方向に座標を更新してゆくことによって系の最小エネルギー状態を求める手法である。Kamada and Kawai (1991)[1]は(以下KK法と略す)全てのノード間に仮想的なバネを仮定し、エネルギー準位が高いものから逐次的に1ノードずつ座標更新を行う。しかしながら、この手法ではノード数 N に対して $O(N^2)$ のメモリ領域が必要になる。それ故、この手法は大規模な可視化計算には適していない。

Fruchterman and Reingold (1991)[2]の提案した手法(以下FR法と略す)では、連結するノード間(エッジ)に働く力を引力を与え、更にノード間に斥力を与え、二つの力の和を与えることによりノードの座標更新を行う。ここで系の総エネルギーは以下のように表わされる、

$$\Psi_{FR} = \frac{1}{3k} \sum_{i,j}^E |\mathbf{x}_{ij}|^3 - \frac{k^2}{2} \sum_i \sum_{i \neq j}^N \ln |\mathbf{x}_{ij}| \quad (1)$$

FR法では、ポテンシャル勾配である加速度の方向に、全ノードの座標を同時更新する。更新幅についてはこの際、あらかじめ設定した繰り返し演算数に対し、最大更新幅が0に収束するように冷却関数を与え座標更新を行う。

3 提案手法

従来の手法において、グラフ可視化の座標値計算では、ノード間の力学方程式は位相空間上での座標間距離に比例した関数系で与える。つまりKK法やFR法では、 i 番目と j 番目のノードに関わる加速度公式はユークリッド距離 $|\mathbf{x}_{ij}|$ に比例した関数で与えられ、その加速度の方向ベクトルは \mathbf{x}_{ij} と平行である。ここで本研究では、固有楕円ポテンシャルを用いて各ノードに対して異なる形状の斥力を与える事によって、個々のラベルの大きさを考慮した最適配置を求める手法を提案する。しかしながら、全てのノードが文字列のような場合、全てのノードのポテンシャルが横方向に扁平になり、その結果可視化結果が全体的に扁平になるという問題が生じる。従って、単純にFR法に固有楕円ポテンシャルを導入するだけでは、可視化結果に影響を及ぼす可能性がある。

そこで本研究では、楕円ポテンシャルを斥力項のみに与え、さらにポテンシャル関数を二重化させる事によりこの問題を解決する。

この手法による効果は主に三点あり、第一に、可視化結果が全体的に潰れることを防ぐ効果がある。第二に、遠方で点対称ポテンシャルによる斥力を残しつつ、近傍ではラベルサイズに依存した斥力を強くすることが出来る、という利点がある。つまり、

大局的な構造を保ったままラベル同士が重ならないという制限をより強くすることが出来る。第三に、遠方でのラベルサイズに依存した楕円ポテンシャルの影響を弱くする事により、Tree-codeのような既存の近似手法を用いる事が可能である。

具体的には i 番目のノードに対して x 軸方向、 y 軸方向に A_i , B_i という大きさを持つラベル付きのグラフに対して、ラベルの大きさに依存した斥力のポテンシャルを用いる。ここで、 $|\tilde{\mathbf{x}}_{ij}|$ を以下のように定義する、

$$|\tilde{\mathbf{x}}_{ij}| = \left\{ \frac{(x_j - x_i)^2}{A_i^2} + \frac{(y_j - y_i)^2}{B_i^2} \right\}^{1/2} \quad (2)$$

この時 i 番目のノードの作るポテンシャルの斥力成分を

$$\Phi_{r,i}(x, y) = f(|\tilde{\mathbf{x}}_{ij}|) + g(|\mathbf{x}_{ij}|) \quad (3)$$

と定義することにより、点対称ポテンシャル成分 $g(|\mathbf{x}_{ij}|)$ と楕円ポテンシャル成分 $f(|\tilde{\mathbf{x}}_{ij}|)$ とにわけて考える。この時注意すべき点は、関数 $f(|\tilde{\mathbf{x}}_{ij}|)$ が $g(|\mathbf{x}_{ij}|)$ よりも近傍では強く、遠方では早く減衰するように関数を定める必要がある。

一般的に大規模ネットワークデータは、一部に次数の非常に高いノードが存在し、そのようなノード付近では多くのノードが密集し、可視化結果が潰れてしまうと言う問題がある。そこで提案手法では、ノード間の斥力には力を与える側の次数のみを重みとして与え、大局的な可視化結果が大きく変わること無く、かつ次数の高いノードが集まる領域においてはラベルが重ならない手法を提案する。具体的には、式(3)の楕円ポテンシャル成分に対して以下のようにノードに重みを与える、

$$f(|\tilde{\mathbf{x}}_{ij}|) = \frac{k_\alpha}{\alpha} q_j |\tilde{\mathbf{x}}_{ij}|^{-\alpha} \quad (4)$$

座標の更新計算は、FR法をベースとし、可視化の座標計算手法は以下のように設定する、

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \eta \times \mathbf{a}_i \times \min(1, 1/|\mathbf{a}_i|) \quad (5)$$

η は1stepあたりの最大更新幅で、初期配置は乱数で与え、 $\eta = 0.001$ として 10^5 回の繰り返し演算を行った。なお、本研究においては $k = k_\alpha = k_\beta = 1$ として計算を行った。

4 計算結果と考察

本研究ではモデルデータとして、Newmanのサイト¹から取得可能な3つのグラフデータを用いた。これらのデータは各ノードのラベル名、各ノード間のエッジの連結情報を持つ。ラベルサイズはサンプルデータとして、 i 番目のノードに対して文字の大きさ B_i をノードの次数に比例し与え、 A_i は文字列の長さに対しさらに文字の大きさに規格化させた値を用いた。

¹<http://www-personal.umich.edu/~mejn/netdata/>



図 1: FR 法による“Dolphins”の可視化

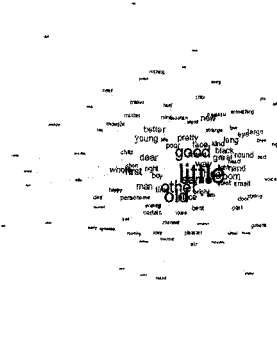


図 2: FR 法による“adjnoun”の可視化

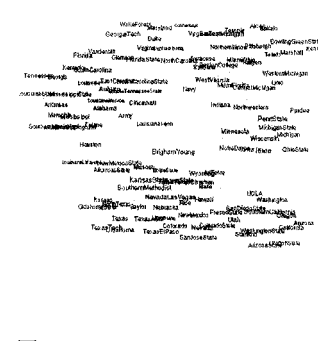


図 3: FR 法による“football”の可視化

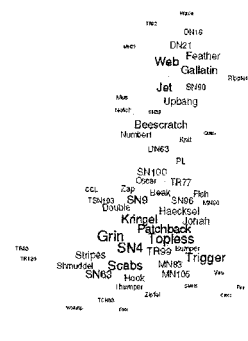


図 4: 提案法による“Dolphins”の可視化

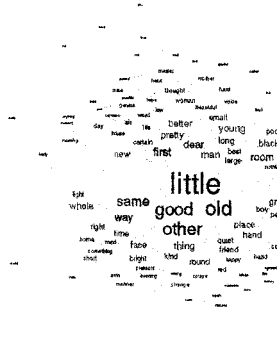


図 5: 提案法による“adjnoun”の可視化



図 6: 提案法による“football”の可視化

図 1 は FR 法による“Dolphins”の可視化結果。ノードが密集したな場所ではラベル同士が重なり合ってしまうことが分かる。一方図 4 は提案手法による可視化結果である。図 1 と比べ、ラベル同士の重なりも解消されていることが分かる。

“adjnoun”は名詞と形容詞について単語間の接続関係をネットワークとして表現したものであり、単語「little」「good」「old」「other」等が高い次数を持つ。FR 法による可視化結果である図 2 に対し、提案手法の図 5 では明らかに「little」を中心とした文字の関連性が見られるように配置されている。

“football”はアメリカの大学のアメリカンフットボールチームの対戦履歴をネットワークとして表現したものであり、次数分布のばらつきが非常に少ないグラフデータである。FR 法による可視化図 3 ではチーム名が重なってしまう一方、提案手法による可視化図 6 はチーム名の重なりが抑えられていることが分かる。

4.1 結論と考察

以上、これらモデルデータの可視化結果より、提案手法によるラベル付き可視化はラベル同士の重なり

りを回避可能であることが示された。提案手法では、斥力項に対し点対称ポテンシャルと固有楕円ポテンシャルの関数を二重に重ね合わせる事により、従来手法の可視化結果に対して大局的な構造を保ちつつ、局所的にはラベルの重なりを回避する事が可能であることが示された。また、本提案手法は点対称ポテンシャル項および固有楕円ポテンシャル項がそれぞれ独立に関数を与える事が出来るため、ラベル同士の排他制限をより強くすることが可能であることが示された。

5 応用例 -BLOGRANGER TG-

“BLOGRANGER TG”とは、NTTサイバースリユーション研究所とNTTコミュニケーション科学基礎研究所が共同で研究を進めているプロジェクトであり、2007年12月よりgooラボにて公開される、タグクラウド・ブログ地図である。このプロジェクトでは、従来のタグクラウドのように順番に羅列するのではなく、類似性を持ったタグ同士をノード間の類似度を基準に、関連性の高いもの近傍に配置するように2次元可視化計算を行い配置したものである。図 7 は現在開発中のタグクラウド地図であり、

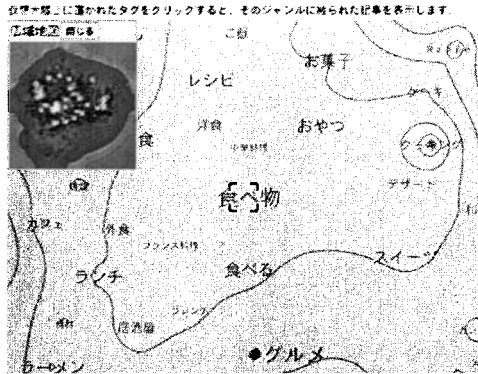


図 7: BLOGRANGER TG によるタグクラウド地図。ただし画像は 2007 年 11 月現在、開発中のものである。

左上に全体の広域図が表示され、中央に一部分の広域図が示されている。ただし画像は 2007 年 11 月現在、開発中のものである。

6 GPU による高速化

本提案手法は GPU への実装により、高速に可視化計算を行うことができる。

松林ら [3, 4] は、粒子間相互作用専用計算機 MD-GRAPES3 を用いて、Force-directed 法による可視化手法により百倍を超える計算の高速化を実現した。一方、GPGPU (General Purpose computation on Graphics Processing Units)² ではさらに安価に座標計算の高速化を行うことが可能である。

なお本研究では、Geforce8800 GTX を用いて、NVIDIA が提供する CUDA の API を用いた C プログラムによる開発を行った。提案手法ではさらに、GPU で斥力計算を行う間に、CPU 側でエッジ引力を計算させ計算の効率化を行った。

本研究では、の可視化計算に要した時間は 10 万回の繰り返し演算に対し、Intel Core2Quad Q6600 (2.4GHz) を用いて 10 秒以内であった。しかしながら、“BLOGRANGER-TG” のグラフデータは $N = 4,457$ と規模が大きく、さらに次数の高い領域が存在するために 2×10^7 回の更新計算を行った。この計算は上記汎用 CPU を用いた場合約 800 時間の計算時間を要する。一方、CUDA の API を利用したプログラムでは約 450 倍の高速化を実現し、GPU を用いた計算では約 2 時間で計算が終了した。

このように、GPU による並列処理を利用することにより、数万ノードを超える可視化計算が可能になり、提案手法のような複雑な関数も高速演算可能となる。

²<http://www.gpgpu.org>

7 おわりに

本研究では各ノードに固有の楕円ポテンシャルを与えることにより、ラベル同士の重なりを回避する座標計算手法を提案した。

従来手法では、ノードを“点”として扱うためにラベル付きグラフ可視化において限界がある。しかしながら提案手法では、ラベルサイズに依存した固有楕円ポテンシャルを斥力項として与え、さらに、斥力項に対し点対称ポテンシャルと固有楕円ポテンシャルの関数を二重に重ね合わせる事により、従来手法の可視化結果に対して大局的な構造を保ちつつ、局所的にはラベルの重なりを回避する事を可能とした。また、斥力項のポテンシャル関数を重ね合わせたことにより、Tree-code 等の近似手法も実装可能であり、さらに、GPU の利用により、大規模グラフデータに対し高速に計算を行うことを可能とした。

本研究では比較実験対象として FR 法を用いて行った。これは KK 法に二重のポテンシャルを与えることが困難であり、かつ KK 法ではノード間の最短経路のデータを保持する必要があり、ノード数 N に対して $O(N^2)$ のメモリ領域が必要になるという問題があるためである。それ故、KK 法は大規模なグラフ可視化計算には適しておらず、本研究においては提案手法の妥当性の評価を、FR 法を用いて行った。また、force-directed 法は並列処理の実装が可能であり、GPU 等を用いる事により大規模なグラフ可視化が可能であるという利点がある。

従来グラフ可視化においては引力、斥力のポテンシャル関数は自然科学的な (たとえば万有引力のような) 普遍的な関数が存在せず、ポテンシャルの関数の設定には自由度が生じる。それ故当該分野において、ポテンシャル関数による可視化結果の定量的評価は重要課題であり、今後はグラフデータの特徴に適したポテンシャル関数について研究を進めてゆく予定である。

参考文献

- [1] Kamada, T. and Kawai, S.: An algorithm for drawing general undirected graphs, *Information Processing Letters*, 12, 31, 7-15 (1989)
- [2] Fruchterman, T.M.J. and Reingold, E.M.: *Graph Drawing by Force-directed Placement*, Software - Practice and Experience, 11, 21, 1129-1164 (1991)
- [3] Matsubayashi, T. and Yamada, T.: A Force-directed Graph Drawing based on the Hierarchical Individual Timestep Method, *International Journal of Electronics, Circuits and Systems*, 1, 2, 116-121, (2007)
- [4] 松林達史 山田武士: 階層的独立固有時間刻み法によるグラフ可視化計算の高速化, *情報処理学会論文誌*, Vol. 48, SIG15 (2007)