

BLOG のトラックバック構造可視化システムの開発

鈴木 佑介[†] 石川 祥^{††} 木村 昌臣[†]
 芝浦工業大学[†] 芝浦工業大学大学院^{††}

1. 研究背景と目的

近年，BLOG のトラックバック構造を利用して BLOG 間の関連性を可視化することで，どのような話題がどのように分布しているのかを分析する研究が行われている。

この分野の先行研究に[1]がある．この研究では可視化方法に Kamada-Kawai モデル[2]を使用している．また，内田らの研究 [3]では LGL モデル[4]を使用して可視化している．これらの可視化方法では，ノード配置場所の決定にバネモデルを使用している．しかし，バネモデルを使用すると収束するまでの計算量が多いため，可視化するまでに時間が掛かってしまう．

そこで本研究では，数万件や数十万件，あるいは，それ以上の大規模なデータに対しても高速に可視化するシステムの開発を目的としている．

2. ノード配置アルゴリズムの考案

高速なノード配置アルゴリズムの考案については，バネモデルを使用しない新しいモデルを提案する．

具体的には以下の 4 つのステップで順にノードを配置する（図 1）．

始点となる根ノードを配置する．

根ノードとその各子ノードとの成す角度が均等になるように周囲に各子ノードを配置する．配置した子ノードを self ノードとし，self ノードとその親ノードを結ぶエッジを self ノード側に伸ばし，それを軸とした ± 45 度の範囲内に，self ノードの各子ノードを均等に配置する．

全ノードが葉ノードになるまで を繰り返す．

で子ノードを配置する範囲を ± 45 度にした理由は，この角度があまり大きすぎると，ノードが横に広がりすぎてしまうために，トラックバックのステップ数が増えると基点としたノードの方に戻ってきてしまうからである．

また，このモデルではバネモデルを使用していないため，親ノードと子ノードを繋ぐエッジ

の長さは均一である．

このようにバネモデルを使用せずに幾何学的にノードを配置することで，高速な可視化を実現する．

また，一つのノードを基点として，トラックバックで繋がれたノードが次々に広がっていく様子が水面の波紋のように見えることから，以下このモデルを「水紋モデル」と呼ぶことにする．

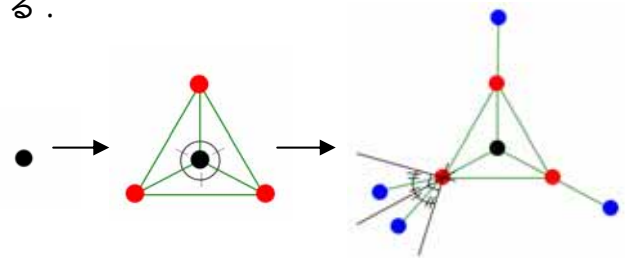


図 1 水紋モデル

水紋モデルで可視化した際には，まず BLOG 記事群のトラックバック構造全体を表示するようにしているが，可視化した一部分を拡大する機能を付けることで，ノードが密集している部分の詳細を見ることを可能にした．さらに，拡大後はエッジに色を付けることで有向グラフとして可視化できるようにした．

3. 実験方法

本研究では，[1]で取得したデータを使用した．映画「ダ・ヴィンチ・コード」（以下 DVC）について書かれた BLOG 記事を出発点とした約 4 千件のデータと，2006 年ドイツワールドカップ（以下 W 杯）について書かれた BLOG 記事を出発点とした約 5 千件のデータである．

この二つのデータに対して水紋モデルと Kamada-Kawai モデルを用いて可視化を行い，可視化結果の比較と可視化速度の検証を行った．

4. 結果と考察

DVC の 4 千件のデータを可視化したものを図 2 と図 3 に，W 杯の 5 千件のデータを可視化したものを図 4 と図 5 に示す．

この可視化手法では，Kamada-Kawai モデルでは見えない出発点となる記事と，そこからのトラックバックのステップ数を確認することが出来るだけでなく，図 2 及び図 4 の A のようにある BLOG 記事にトラックバックが集中したために，トラックバックされている BLOG 記事を

Development of visible making trackback structure system of BLOG

[†] Yuusuke Suzuki, Masaomi Kimura · Shibaura Institute of Technology

^{††} Sho Ishikawa · Shibaura Institute of Technology graduate school

中心にして弧を描いている様子も見る事ができる。

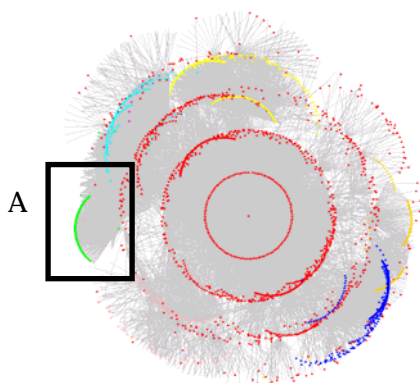


図2 水紋モデルによる DVC の可視化

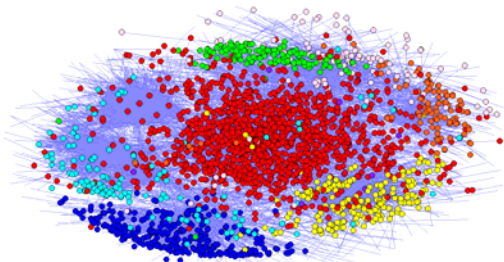


図3 Kamada-Kawai モデルによる DVC の可視化

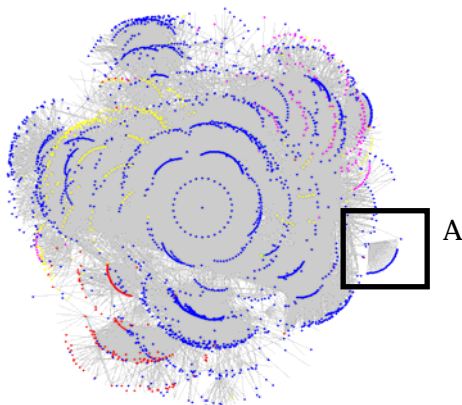


図4 水紋モデルによる W 杯の可視化

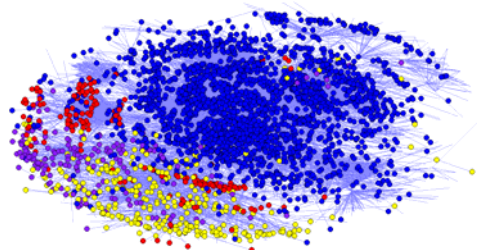


図5 Kamada-Kawai モデルによる W 杯の可視化

同じ話題を持つノードのクラスは、水紋モデルと Kamada-Kawai モデルの両方ともノードの色で判別することができる。しかし水紋モデルでは、他の映画や時事問題など、他の話題が

どのノードで切り替わるかを容易に確認することができた。

次に、可視化時間の比較については DVC と W 杯のデータに対してそれぞれのモデルで可視化を 3 回ずつ行った後、平均可視化時間を求めた(表 1)。

表 1 可視化時間の比較

	DVC	W 杯
Kamada-Kawai モデル	3 分 10 秒 7	4 分 43 秒 7
水紋モデル	57 秒 63	61 秒 818

この結果からもわかるように、可視化時間に関しては水紋モデルの方が短いのは明らかである。パネモデルを使用せず、単純なアルゴリズムを用いることで計算量が少なくなったことが主な要因の一つとして考えられる。

また、新しく DVC と W 杯の両方の約 5 万件のデータを取得して、それぞれ水紋モデルで可視化を行った。可視化時間は DVC のデータで約 1 時間 10 分、W 杯のデータで約 56 分であった。

5. まとめ

本稿では高速性を重視した水紋モデルを提案し、[1]で収集した DVC の 4 千件のデータと W 杯の 5 千件のデータに対して、水紋モデルと Kamada-Kawai モデルを使用して可視化結果の比較と可視化速度の検証を行った。可視化結果の比較は、出発点となった BLOG 記事と話題の推移も確認することができ、同時にトラックバックのステップ数を確認することができた。また、水紋モデルの可視化の高速性も確認できた。

参考文献

- [1]石川祥, 木村昌臣: BLOG のトラックバック構造における話題の可視化, 電子情報通信学会総合大会 D-4-14, 2006
- [2]西尾元宏, 増田宏: 板ばねモデルを用いたインタラクティブな曲線グラフ描画手法とその応用に関する研究, 第 2 章グラフィックアウト, 東京大学大学院工学系研究科修士論文, 2003
- [3]内田誠, 柴田尚樹: ブログ記事ネットワークからの emerging topic の抽出と可視化, 第 20 回人工知能学会 3D2-03, 2006
- [4]Alex T. Adai, Shailesh V. Date, Shannon Wieland and Edward M. Marcotte: LGL: Creating a Map of Protein Function with an Algorithm for Visualizing Very Large Biological Networks, J Mol Biol, 340(1), 179-90, 2004 Jun 25