1606 MPS を用いた複雑ネットワークの可視化法

Visualization Technique for a Complex Network Using MPS

正 白山 晋 (東大人工物)

Susumu SHIRAYAMA, RACE, the University of Tokyo

1. はじめに

現実世界の様々なネットワークにおいて、スモールワールド、スケールフリーといった性質が見いだされ、複雑ネットワーク分析の応用が盛んになっている¹⁾。このようなネットワークの構造を調べる上で可視化は重要ではあるが、確立された方法はない。距離空間の定義が難しいことと、データの大規模性が問題になるからである。

ネットワークは、データを示すノードとデータの接続関係を示すリンクによって表され、無向グラフ、あるいは有向グラフを用いて可視化されることが多い。具体的には、ノードを示すオブジェクト(頂点、ノード)と、リンクを示すオブジェクト(辺、エッジ、有向グラフの場合はアークと呼ばれる)が表示される。Fig.1で示すA,B,C,Dがノード、AB, AC, BD, CDがエッジである。このようなネットワークの性質はグラフ理論によって明らかにされてきた。また、社会科学の分野では、社会ネットワーク分析の手法が数多く提案されている²⁾。

Fig.1 ネットワークの表現例

複雑ネットワークとは、従来のグラフ理論や社会ネットワーク分析の手法では扱うことのできない、複雑で、大規模性をともなう現実のネットワークを指す(実際には、グラフ理論や社会ネットワーク分析の手法を援用して分析されている)。一般に、複雑ネットワークも、無向グラフ、有向グラフで表現される。ノードやリンクには、人、場所、流量などの属性が与えられている。属性によっては空間座標や距離に関連するものもあるが、そのような属性は排除され、位相構造として扱われることが多い。ただし、ネットワークとしての距離は定義されている(グラフ距離と呼ぶ)。2ノード間のグラフ距離は、その間の最も少ないエッジの数である。Fig.1の場合は、AとB, AとC, BとD, CとDの距離が1, AとD, BとCの距離が2である。

Fig.1の3つのネットワークは同じものを示している。このように、ノードとエッジを表示させることは単純ではあるが、物理座標に対応する表示空間の中での位置関係が、必ずしもノードの配置を適切に表現しえないという難しさがある。ただし、例えば、道路網のように、属性によって位置関係が必然的に決定するものや、補正されるものもある。

以上のように、複雑ネットワークの可視化の難しさは、空間座標の概念を直接用いられないことに起因して生じる。また、表示空間において、ネットワーク構造が視覚化されると表示空間での位置関係や距離という概念が生じてしまい、誤情報となる場合もある。したがって、位置関係や距離をどのように扱うかが視覚情報を分析する上で重要になる。このため、グラフドロー(グラフドローイング)^{3,4)}という分野で様々な提案がなされ

ている(グラフレイアウト, ネットワークドローイングと呼ばれることもある)。なお、それらは、Fig.1に示すようなネットワークの可視化だけを指すわけではない。例えば、グラフレイアウトの場合は、折れ線グラフ、棒グラフの配置のようなものも扱っている。本稿では、断りのない限りグラフレイアウトという言葉を用い、それを、Fig.1のようなノードとエッジの表現を指すものとする。

理想的なグラフレイアウトは表示空間におけるオブジェクトの配置の意味づけを行うことである。ネットワークに対しては、いくつかの特性量、あるいは特徴量が定義されており²⁾、そのような量に基づく方法がある。例えば、固有空間を形成し、固有空間を表示空間に対応づける方法である。しかしながら、ほとんども、オブジェクトの数が増えると有効性を失う。結局、大まかな配置を示し、インタラクティブ性を中核として、分析者の解釈を支援する操作を提供するという方法がとられることが多い。本稿でも、起点を、そのような見やすさやインタラクティブ性とする。代表的な方法が次節で述べる力学モデルによるものである。しかしながら、力学モデルでは、ノード分布の制御は難しい。そこで、MPS法における粒子数密度の制御法を、複雑ネットワークの効果的な表示法に応用することを試みる。

2. 力学モデルによるグラフレイアウト

Eadesは、Fig.2に示すようなバネモデルによるグラフレイアウトの方法を提案した⁵⁾。力学モデルにもとづくその後のグラフレイアウトの方法は、Eadesの方法の拡張であるものが多い。

Fig.2 4つのノードに加わる力の例

この方法は、エッジを仮想的なバネとみなし、ノード i に働く力を、直接リンクのあるノード j (グラフ距離1のもの)からの引力、あるいは斥力で求めることを基本としている。はじめに、表示空間を座標空間とみなす。初期状態として、ノードを表示空間上に適当に配置する。例えば、Fig.3に示すような楕円上での分布である。点がノード、線がエッジを示している。ノード i の表示空間上での座標を、 r_i とする。

力を表す関数には様々なものが提案されているが、仮想的なバネによる力 f^s は、

$$f_i^s = \sum_j k \log \left(\frac{r_{i,j}}{d_0} \right) \quad (1)$$

と定義されることが多い。ここで、 $r_{i,j}$ は、ノード i と j の表示

空間での距離 (以降, 単に距離と呼ぶ), j はグラフ距離 1 のノード, d_0 は, 表示空間上でのバネの自然長, k はバネ定数に相当する適当な定数である. このモデルでは, Fig.2 の C と D で示すようなグラフ距離が 1 以上のノードの配置を制御できない. そこで, 表示空間上で, ある程度近い距離にあるノードを次の斥力によって移動させる. 力 f^p を,

$${}^3r_{i,j} \leq D, \quad f_i^p = \sum_j \frac{\alpha}{\left(\frac{r_{i,j}}{d_0}\right)^2} \quad (2)$$

で与える. ここで, D は表示空間での距離とし, この距離よりも近いグラフ距離 1 以外のノードが影響を与えるものとする. α は適当な定数である.

Fig.3 ノードの初期分布

力学モデルによる表示空間(物理空間)でのノードの配置は, 以上の力を使い, すべてのノードに対して以下の運動方程式を数値的に積分することによって, 決められる.

$$m \frac{d^2 r_i}{dt^2} + c \frac{dr_i}{dt} = f_i^s + f_i^p \quad (3)$$

ここで, m は仮想的な質量, c は減衰係数である.

表示空間上のバネの自然長は, グラフ距離の 1 に相当する. グラフ距離が 1 のノード間は, (1)式によって自然長に戻ろうとする. 一方, Fig.2 の C と D はグラフ距離で 3 離れているが, (1)式や(2)式で, 表示空間上でグラフ距離に相当する長さに配置することは難しい. そこで, (2)式に相当するグラフ距離 1 以外のノードの制御を以下のバネモデルによって行うという方法が提案されている⁶⁾. 本稿でもこの方法を採用する.

$$f_i^p = \sum_j k' \log\left(\frac{r_{i,j}}{L_G d_0}\right) \quad (4)$$

ここで, L_G はグラフ距離, k' は適当な定数である.

グラフレイアウトに要求される条件を, 土井と伊藤⁷⁾に従って纏めると以下ようになる. グラフのレイアウトを,

- (a) グラフ距離に応じてエッジの長さを均等化する
- (b) グラフ距離の近いものを近くに配置する
- (c) ノード全体をできるだけ均質的に配置する
- (d) ノードやエッジの重なりを避ける

ように決めるというものである. (d)の重なり以外は, 力学モデルにもとづくグラフレイアウトによってほぼ解決しているといえる. しかしながら, ノードの数が数百以上になると, (a)~(c)を満たすことが困難になる. (3)式から, 全体の定常状態(均衡状態)を得ることが難しくなるためである. このため, (3)式の数値積分を強制的に終了させ, インタラクティブに位置を補正するといったプロセスが追加されることになる. 数千になる

と, このプロセスに, 初期分布の調整を加えて, ようやく(a)~(c)が実現できる. 数十万以上では画素の並びがノードを表すことになり, エッジが表現できないという本質的な問題が生じる. この場合は, 別種の表現方法が望ましいだろう. 本稿では, 数万までの規模で考える. この場合の問題は, (3)式の数値積分を終了させた後の補正プロセスの自動化にある.

3. MPSを利用したノード配置の制御

ノードを粒子とみなし, MPS法⁸⁾の連続の式を満たすアルゴリズムを応用してノード分布を調整する.

MPS法に従い, 重み関数を

$$w(r) = \begin{cases} \frac{h}{r} - 1 & (0 \leq r < h) \\ 0 & (h \leq r) \end{cases} \quad (5)$$

とし, 粒子数密度 n_i を,

$$n_i = \sum_{j \neq i} w(|r_j - r_i|) \quad (6)$$

とおく. n^0 を目標となる粒子密度とする (例えば, 一定値を与える). n^k を現時点での粒子数密度とする.

はじめに, 以下のポアソン方程式を数値的に解くことによって, あるスカラー場 ϕ を求める.

$$\nabla^2 \phi = -\frac{\rho_0}{\Delta t^2} \frac{n^k - n^0}{n^0} \quad (7)$$

この ϕ の勾配によって, 粒子を移動させることでノードの配置を調整する.

$$r^{k+1} = r^k + \varepsilon \Delta t \left(-\frac{\Delta t}{\rho_0} \nabla \phi \right) \quad (8)$$

ここで, ε は適当な正数である.

4. おわりに

グラフレイアウトには, (a) グラフ距離に応じてエッジの長さを均等にする, (b) グラフ距離の近いものを近くに配置する, (c) ノード全体をできるだけ均質に配置する, (d) ノードやエッジの重なりを避ける, という要件があるが, これらをすべて満たすことは難しい. 本稿では, 力学モデルによって, 大体の配置を行った後に, MPSの連続の式を満たすアルゴリズムによって, 所与の粒子数密度になるようにノードを移動させるという方法を提案した. 具体的な結果については講演時に示す.

参考文献

- 1) 増田直紀, 今野紀雄: 複雑ネットワークの科学, 産業図書, 2005
- 2) 金光 淳: 社会ネットワーク分析の基礎, 勁草書房, 2003
- 3) Tollis, L., Di Battista, G.: *Graph Drawing: Algorithms for Geometric Representations of Graphs*, Prentice Hall, 1998
- 4) 杉山公造: *グラフ自動描画法とその応用*, コロナ社, 1993
- 5) Eades, P.: A Heuristic for Graph Drawing, *Congressus Numerantium*, Vol.42, 1984, pp.149-160.
- 6) Kamada, T. and Kawai, S.: An Algorithm for Drawing General Undirected Graphs, *Information Processing. Letter*, Vol.31, No.1, 1989, pp.7-15
- 7) 土井 淳, 伊藤貴之: 力学モデルを用いた階層型グラフデータ画面配置手法の改良手法とウェブサイト視覚化への応用, 芸術科学会論文誌, Vol.3, No.4, 2004, pp.250-263
- 8) 越塚誠一: 粒子法, 丸善, 2005