

線分交叉を伴う系図の自動配置に関する基礎的研究 — 効率的な探索のための新たなデータ構造の検討 —[§]

杉山 正治*, 松浦 亨**
大谷大学*, 北海道大学病院**

1. はじめに

系図を可視化する際、結合関係を木構造のみに限定すれば、自動配置は容易に実現できる。実際、多くの系図表示ソフトウェアが木構造をベースに実装されている。しかし、一般的な系図では図1のように、再婚[†]や血縁同士による親族婚などで結合関係に閉路が発生し、木構造から逸脱するため、表示が破綻する例が見られる。

一方、系図情報格納形式のデファクトスタンダードであるGEDCOM[1]は、個性データ列と家族データ列の2系列に分離されたデータ構造であり、複雑な婚姻関係を完全に定義できる。ところが、既存ソフトウェアでは、この2系列から単一の木構造を再構築し、実装を簡略化していることが多いため、自動配置は実現しているが、木構造以外の系図には対応しきれない。

図1のような複雑な系図を紙媒体に書く場合、一般に、1つの個性を1箇所にのみ配置し、垂直・水平線分のみを用いて個性同士を結び、線分が交叉する位置に半円弧などを配置する。このようなスタイルを用いれば、個性と個性の関係性を狭い領域に視認性良く配置できる。また、古来より多くの人がこのスタイルに馴染みがあるだけでなく、PC上でもこの紙媒体と同じようなスタイルで可視化して欲しいという要求がある。

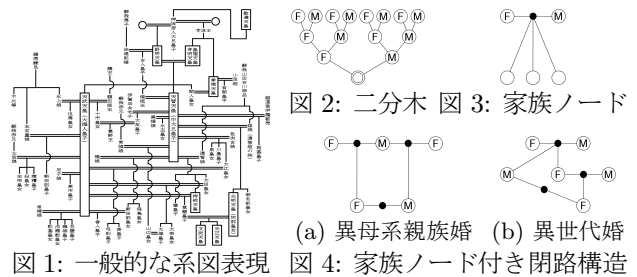
我々はこれまでに、線分交叉を伴う複雑な系図を表示出来るソフトウェアWHiteBasEを開発し、この問題を解決してきた[2]-[8]。WHiteBasEでは、個性データ列と家族データ列をそのままの形で利用しながら、親子を結合する線分の交点に不可視結節点を配置し、線分交叉の探索と表示を行っている。その結果、家族同士の配置の重なりがある場合のみの少ない探索回数で交叉位置を決定できる。しかし、このような系図を全自動で配置する問題は極めて難しく、現状ではユーザ入力を補助する半自動的な実装しか実現していない。

そこで、本研究では系図のデータ構造を抜本的に見直し、効率的な個性配置を探索するための新たな手法「父→母→子」型構造を提案し、その有効性を検討する。

2. 既存の系図データ構造の問題点

2.1. 二分木を用いた場合

ある個性をルートとした直系の先祖（父母のみ）に限定すれば、データ構造は図2に示すような二分木となり、実装は極めて単純である。ここで、Mは男性（父



親)を、Fは女性(母親)を、◎はルートとなる個性を、それぞれ表す。この方法で可視化できる系図表示ソフトウェアはAncestry[9]をはじめ、多数存在している。このスタイルは、血統などを辿るだけの用途に限定すれば有用である。しかし、以下に示す問題がある。

- 全ての個性の兄弟姉妹を追加できない
- 親族婚の関係になっている親を追加できない

2.2. 家族ノード付き木構造を用いた場合

1組の親子関係を家族ノードによって結合出来るようにすれば、データ構造は図3に示すような木構造となる。ここで、○は子を、●は家族ノードを、それぞれ表す。この関係はGEDCOMで採用されている2系列のデータ構造から直ちに再構築可能である。この方法で可視化出来る系図表示ソフトウェアはMyHeritage[10]をはじめ、多数存在している。このスタイルは、2.1節で示した二分木と異なり、複数の兄弟姉妹を追加できるため、一般的な親子関係を連ねる用途に限定すれば有用である。しかし、以下に示す問題がある。

- 親族婚の関係になっている親を追加できない。無理矢理追加すると同一個性が複数箇所に表れる。
- 本来必要のない家族ノードが増える分だけ、個性配置の探索に時間がかかる。

2.3. 家族ノード付き閉路構造を用いた場合

一方、2.2節で示した家族ノードを用いながら異母系親族婚や異世代婚など木構造からの逸脱を許容すれば、データ構造は図4(a),(b)に示すような閉路を含む形になる。この方法で可視化出来る系図表示ソフトウェアはMadeline[11]をはじめ、多数存在している。このスタイルは、再婚や親族婚のうち、線分交叉の無い比較的簡単な関係までは適切に追加できるため、医療系図方面などで有用である。しかし、以下に示す問題がある。

- 本来必要のない家族ノードが増える分だけ、個性配置の探索に時間がかかる。
- 複雑な婚姻関係の場合、データ構造自体に線分交叉が発生やすく、個性配置の探索は容易ではない。

[§]A Study of New Data Structure for Automatic Layout of Genealogy with Segment Intersections

*Seiji Sugiyama: Otani University

**Tohru Matsuura: Hokkaido University Hospital

[†]「再婚」(離婚後に別の人と婚姻)と「複数婚」(一夫多妻等)は図像的な配置が同じになるため、本稿では「再婚」で統一表記する。

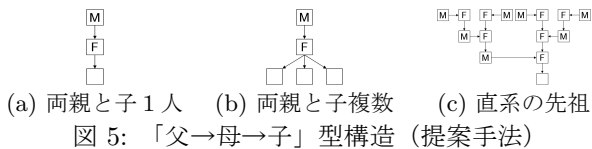


図 5: 「父→母→子」型構造 (提案手法)

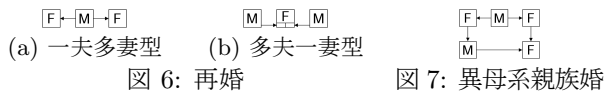


図 6: 再婚

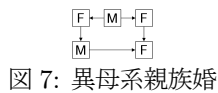


図 7: 異母系親族婚

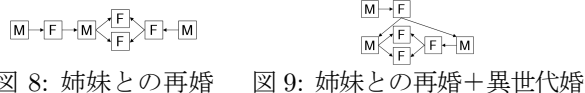


図 8: 姉妹との再婚

図 9: 姉妹との再婚+異世代婚



図 10: 異世代婚+異母系親族婚

図 11: カリエラ型

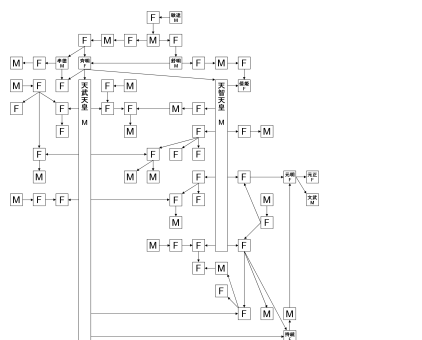


図 12: 図 1 の系図を「父→母→子」型構造で表現

3. 「父→母→子」型構造の提案

2.1 節~2.3 節に示した問題を解消し、紙媒体のような系図スタイルにするには、以下3点が必要になる。

- 複雑な婚姻関係があっても破綻無く親子関係を表現できるデータ構造
- 家族データ列や家族ノードのような余分な結節点を必要としないシンプルなデータ構造
- 可視化前のデータ構造の段階で、出来る限り線分交叉をしなくて済むようなデータ構造

これらを満たす新しいデータ構造を検討した。従来、家族ノードは親子関係を1つの家族とする社会的関係に依存したデータ構造として用いられてきた。一方、子の発生プロセスを時系列的に捉えれば、通常は、婚姻関係(男女の交合)の後、母親からのみ子が発生する。したがって、父→母→子のように順番にたどるデータ構造さえあれば必要十分であり、家族ノードや父→子の直接の結合が無くても関係性を表現できる。

以上より、親子関係は図5(a)-(c)のように書ける。ここで、**M**は男性を、**F**は女性を、**□**は子を、それぞれ表す。なお、**M**→**F**は婚姻関係(男女の交合)を、**F**→**M**および**F**→**F**は子の発生を、それぞれ表す。このデータ構造を「父→母→子」型構造と呼称する。

4. 様々な関係性への適用例

「父→母→子」型構造を用いて、これまでに調査してきた線分交叉の最小単位の系図[2]などへの適用例、および、その閉路の個数を以下に示す。

再婚: 一夫多妻型の再婚は図6(a)のように書ける。多夫一妻型の再婚は、父親と子の関係を区別できるようにするための識別子を設けて、図6(b)のように書ける。再婚の結合だけであれば閉路が生じることはない。

異母系親族婚: 異母兄妹(姉弟)の婚姻は図7のように書ける。これは上段の2つの**F**(異母)から生まれた下段2人が婚姻した場合であり、閉路が1つ生じる。

姉妹との再婚: 男性が姉妹と再婚した場合は図8のように書ける。これは、上下に並んだ**F**(姉妹)と、その左側の**M**(男性)との再婚であり、閉路が1つ生じる。

姉妹との再婚+異世代婚: 兄弟の子(異世代)となる姉妹と再婚した場合は図9のように書ける。下段両側の**M**が兄弟になった場合であり、閉路が2つ生じる。

異世代婚+異母系親族婚: 異母系親族婚の親が異世代婚等を含む場合は図10のように書ける。これは図9の閉路の内側2人の婚姻であり、閉路が3つ生じる。

カリエラ型: 少数部族間で行われる交換型の婚姻(カリエラ型)は図11のように書ける。これは兄弟と姉妹がそれぞれ婚姻した場合であり、閉路が1つ生じる。

複雑な親族婚の例: 図1の系図を「父→母→子」型構造で表現した例を図12に示す。閉路は12箇所になる。

なお、全ての適用例で線分交叉は全く必要なかった。

5. おわりに

本研究では、系図のデータ構造を抜本的に見直し、効率的な自動配置を探索するための新たな手法を検討し、「父→母→子」型構造を提案した。このデータ構造の有効性は以下5点にまとめられる。

- 複雑な婚姻関係があっても破綻なく記述できる
- 家族ノードが不要である
- ノードの数が個性の数と一致し、必要最小になる
- 辺の数も必要最小になる
- データ構造自体の線分交叉が劇的に少なくなる

今後は、養子縁組関係、同性婚、生殖補助医療、モノザネへの対応と、自動配置手法を検討する予定である。

参考文献

- [1] GEDCOM LETTER, <http://en.wikipedia.org/wiki/GEDCOM>
- [2] 杉山正治, 生田敦司, 柴田みゆき, 松浦亨, “線分交叉を伴う系図表示の基礎的研究—不可視結節点を用いた線分交叉位置探索手法—”, 情報処理学会・じんもんこんシンポジウム, pp. 1-8, 2009 (In Japanese)
- [3] S. Sugiyama, et. al., “A Study of an Event Oriented Data Management Method for Displaying Genealogy: Widespread Hand to InTErconnect BASic Elements (WHItEBasE)”, IEEE Int. Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Vol. 3, pp. 280-289, 2011
- [4] S. Sugiyama, et. al., “Displaying Genealogy with Various Layouts by Using the WHItEBasE”, IEEE Int. Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Vol. 6, pp. 102-115, 2013
- [5] S. Sugiyama, et. al., “Displaying Genealogy with Adoptions and Multiple Remarriages Using the WHItEBasE”, Lecture Notes in Computer Science (LNCS) 8104, pp. 325-336, 2014
- [6] S. Sugiyama, et. al., “Simple Displaying Method for Genealogy with Assisted Reproductive Technologies”, Lecture Notes in Computer Science (LNCS) 9339, pp. 204-215, 2015
- [7] S. Sugiyama, et. al., “Displaying Genealogy with Mythological Relations by Using the WHItEBasE Method”, 情報処理学会・じんもんこんシンポジウム, pp. 75-82, 2017 (In English)
- [8] 系図・紋章画像化研究会じんもん!, <http://www.bartok.jp/msb0909/>
- [9] The Generations Network: Genealogy, Family Trees and Family History Records on line. <http://ancestry.com>
- [10] MyHeritage. <http://myheritage.jp>
- [11] Edward H. Trager, et.al., “Madeline 2.0 PDE: a new program for local and web-based pedigree drawing”, In Bioinformatics ‘Applications Note’, Vol. 23, No. 14, pp. 1854-1856, 2007