

研究会情報を用いた人間関係の抽出

神崎 浩貴^{†1} 蒔田 圭佑^{†1} 佐藤 充^{†2} 大塚 真吾^{†2}

自分の分野外のことについて知りたい場合，その分野の有名人を探すことができれば，その人の論文を見ることで，効率的に知識を集めることができる．また，行政や企業において，今後の指針や経営方針などについて，ある特定の分野の人物と情報交換をしたいと考えたとき，その分野のキーパーソンを容易に見つけられれば便利である．そこで，本稿では研究会や会議などの開催プログラムを用いることで，ある特定の分野における人間関係の可視化を行い，そこから，キーパーソンの特定や人間関係を俯瞰的に見ることができるツールを構築した．

Extraction of Human Relations using the Conference Program

HIROKI KANZAKI^{†1} KEISUKE MAKITA^{†1} MAKOTO SATO^{†2}
SHINGO OTSUKA^{†2}

1. はじめに

自分の分野外のことを調べる場合において，その分野における人間関係がわかれば，その分野で影響力のある人物や，強いつながりを持つ人々を探し出すことができれば，その分野についての情報を効率的に収集することが可能となる．そこで，本稿では研究会の開催プログラムから特定の分野における人間関係の可視化を行い，強いつながりを持つ人々を特定することが可能なツールの構築を行う．

研究会の開催プログラムは，一般的に同じ分野の研究をひとまとめでしたセッションから構成されている．セッションには座長と数件の発表があり，発表される論文には一人ないし複数の著者が記載されている．著者の所属が異なる複数人で構成されていれば共同研究を行なっている可能性が高く，強い人間関係が構築されていると推測できる．また，研究分野が近いものをまとめて一つのセッションとしていることから，同じセッションで発表をした著者達も何らかの関連はあると推測でき，座長も同様に考えることができる．多くの開催プログラムからそれらの関係を抽出することで，人間関係を抽出できるのではないかと考え，実際に電子情報通信学会の開催スケジュールにある，開催プログラムや独自に入手した研究会情報から，座長や著者の情報を抽出し，AT&T Research が提供する Graphviz[1]や Java を用いて人間関係のグラフ化を行った．

2. 関連研究

Web 上で人間関係を調べるツールとしては「あのひと検索スパイシー[2]」が有名である．スパイシーでは検索サイトのように人名を入力し，検索することでその人やその人と関係がある人をインターネット上から検索し，その人の情報と一緒に人とのつながりをグラフ化したものを閲覧することができる．しかし，このサイトは通常の検索サイトと同様に Web 上から該当する情報を抽出しているため，表示される情報の信憑性は高いものとは言えない．例えば，同姓同名の場合，その人物の情報が混じっており，また，誰かの主観が入っている情報や噂などの確実性の高い情報も提示されていることもある．そのためスパイシーでは確実性の低い情報をなくすためにユーザが複数の情報の中から正しい情報を選び，結びつける機能を用意するなど，信憑性の向上に努めている段階にあるといえる．

一方，論文の書誌情報から研究者の関係抽出や研究分野の俯瞰，および，論文のクラスタリングに関する研究も行われている[3,4,5,6]．これらの研究では，論文の参考文献や著者情報を用いることで，関連情報の抽出を行なっているため，我々の研究と関連が深い．参考文献は論文を書いている本人が関連ある研究だと認識したものであるため，主観的な判断になる．一方，我々が利用する研究会などの開催プログラムでは，主催者が関連のある研究をまとめてセッション毎に分けているため，客観的な判断で分類されている点異なる．

†1 神奈川工科大学大学院 工学研究科
†2 神奈川工科大学 情報学部 情報工学科

研究会 開催プログラム - 2011-04-IN - Opera

Web: www.ieice.org/jken/programs/index.php

情報ネットワーク研究会 (IN) [schedule] [select]

4月21日(木) 午後 モバイルネットワーク 座長: 村山 純一 12:30 - 13:45		
(1) 12:30-12:55	TE-Advancedテクノロジーニアスネットワークにおける干渉を考慮した優先接続方式の検討	○松尾英範・湯田泰明・青山高久 (パナソニック)
(2) 12:55-13:20	無線通信ネットワークフレームワークの設計	○大塚義浩・角 武喜・横山陽介・清水直樹・渋谷昭宏 (三菱電機)
(3) 13:20-13:45	無線通信システムにおけるクロスレイヤ設計によるQoSマッピング法	○森 慎太郎・石井光治・生越重章 (香川大)
13:45-14:00 休憩 (15分)		
4月21日(木) 午後 一般 座長: 磯山 和彦 14:00 - 15:15		
(4) 14:00-14:25	VPN接続環境におけるTwice NAT方式の一提案	○小山高明・山田英樹・岸 寿春・深見公彦・北爪秀雄 (NTT)
(5) 14:25-14:50	通信ネットワークの安全性に関する検討 ～ノリノリ・トリック計算と分析精度の向上～	○船越裕介・松川達哉・中西靖人 (NTT)
(6) 14:50-15:15	冗長経路を考慮した分散省電力ルーティングプロトコルの提案	○荒井大輔・茂本信二・土岐 卓・吉原貴仁 (KDDI研)

図1 開催プログラム

person_list(ORG).txt - TeraPad

ファイル(F) 編集(E) 検索(S) 表示(V) ウィンドウ(W) ツール(T) ヘルプ(H)

50563	S: IN:				
50564	Z: 村山純一				
50565	松尾英範	湯田泰明	青山高久		
50566	大塚義浩	角武喜	横山陽介	清水直樹	渋谷昭宏
50567	森慎太郎	石井光治	生越重章		
50568	Z: 磯山和彦				
50569	小山高明	山田英樹	岸寿春	深見公彦	北爪秀雄
50570	船越裕介	松川達哉	中西靖人		
50571	荒井大輔	茂本信二	土岐卓	吉原貴仁	
50572	Z: 鈴木光				
50573	内田大誠				
50574	喜田弘司	中村暢達			
50575	Z: 鈴木光				
50576	丁農				
50577	Z: 石橋孝一				
50578	車谷駿介	外山将司	Eric Y. Chen		
50579	見上紗和子	佐藤正	落合敏功	喜田弘司	
50580	眞鍋孝次	舟阪淳一	石田賢治		
50581	池上大介	本多泰理	土屋利明		
50582	Z: 今井尚樹				
50583	池川美波	森慎太郎	生越重章		
50584	金崎佑子	森慎太郎	生越重章		
50585	神林隆	馬越健治	吉田学	武本充治	
50586	高橋洋介	石橋圭介			

図2 著者リスト

3. 研究会データの取得

研究会のページなどに掲載されている開催プログラムから取得した著者リスト、および、Graphviz で利用する dot ファイルの作成手順について説明を行う。

3.1 開催プログラムより著者リストを作成

研究者同士のつながりを抽出するため、研究会の開催プログラムページから、座長や著者などの情報を抽出し、以下の法則に従い著者のリスト化を行った。

- (1) 研究会名は S:研究会名と記述し、その後改行する。
- (2) 氏名の間の空白は詰める。
- (3) 座長は Z:人名と記述し、Tab による空白の後改行する。
- (4) 座長がない場合 Z:の後改行する。
- (5) どちらでもない場合発表者と判断する。
- (6) 発表者、共同研究者は行で表し、人名は Tab により区切る。

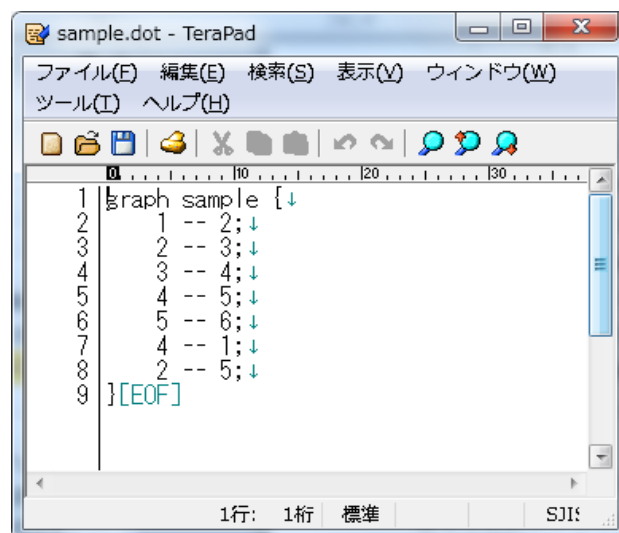


図3 dot 言語の例

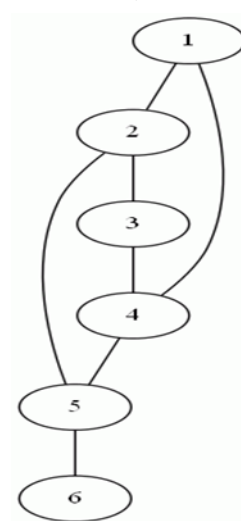


図4 Graphvizの例

実験は2つのデータセットを用いており、1つ目は著者らと関連が深いデータベースに関連する研究会のホームページにある開催プログラムから手作業で著者リストの作成を行った。もう一方は、電子情報通信学会[7]のサイトにある開催プログラムから著者リストを自動的に作成したものである。このサイトには研究会発表申込システムがあり、これを利用することで、図1に示すような開催プログラムページを自動的に作成することできる。自動生成されたページであるため、タグ情報の解析により、著者情報の自動抽出が可能となり、著者リストの自動抽出が可能となる。著者リストの一部を図2に示す。

3.2 著者リストから DOT 言語への変換

前節で作成した著者リストを DOT 言語に変換するプログラムを作成した。DOT 言語とはテキストでグラフを表現するためのもので図3のように記述する。コンピュータで処理しやすい形でありながら目で見てわかりやすくなっている。Graphviz を用いて描画した例を図4に示す。「1」

から「6」はつながっており、かつ「1」と「4」、および、「2」と「5」もつながりがあることがわかる。今回の実験では、無向グラフを用いているが、パラメータを変更することで、有向グラフにすることも可能であり、ノードやエッジの属性も指定することで、色や大きさを変更することも可能である。

3.3 Graphviz を用いた画像変換

作成された dot ファイルは Graphviz を用いて画像ファイルに変換を行う。今回の実験では、以下のように関連する人間を1対1で表現し、結びつきの強さを len で表現した。共著、同一セッション、座長関係になる回数が多いほど len の値は大きくなるようにした。len の値が大きいと近くに配置されるレイアウトとなる。また、表示される人数が多いため、強い関係は赤線で表現した。

```
graph "g" {
    "神崎 浩貴" -- "蒔田 圭佑" [len = 42.00];
    "蒔田 圭佑" -- "佐藤 充" [len = 41.00];
}
```

4. 実験結果

4.1 データセットの詳細

データセット A は DBWeb, DBWeb, DBWS, DEWS, DEIM などデータベースに関する 41 回分のワークショップやフォーラムの開催プログラムデータである。一方、データセット B は電子情報通信学会の研究会開催案内ページにある 2,457 リンクから取得できた 2,306 回分の開催プログラムデータにデータセット A を追加したものである。

この情報を元に、まず、データベースに関連する研究分野の人間関係の視覚化を行った。視覚化は Graphviz を用いた手法と、Java で作成した自作のグラフツールを用いた手法の2つで行った。

4.2 可視化例

まず、データセット A について Graphviz を用いて可視化した例を図 5 に示す。提案手法により作成された研究者相関図は中心部に非常に沢山の人と結びついている研究者が存在した。実際にこの研究者はこの分野での中心的人物であり、数多くの弟子も活躍していることが分かった。扇形になっている部分が数多く存在するが、これは、師弟関係を表しており、大きな研究室ほど扇型が大きいことが分かる。

次に、自作ツールを用いて可視化を行った例を図 6 に示す。図中の中央下の頻度を変化させるスライドバーがあり、表示させる関連語の数を調節することができる。また、ノード(人名)間で関連度が高い場合にはエッジが表示され、ノード間の関連度が高いほどエッジは短くなるように設定

した。各ノードの位置に意味は無く、関連語を表示させた直後はランダムに表示されるが、時間が経つと関連度が高いノード同士が近くなる。また、あるノードを左クリックするとの一どの固定ができ、右クリックするとその人物と関連のある人物の色が変化する。この例からは、データベース分野の著名な先生を中心として円状にノードが集まっていることがわかる。また、右クリックした人物が有名な先生のため、その先生と関係がある人物が多数いることがわかる。

最後に、データセット B について自作ツールで可視化を行った例を図 7 に示す。ノード間が密になっている部分があるものの、全体的にバラけた形となっていることがわかる。データセット B の分野は電子情報通信学会全体であり、全体で 6 万人弱の研究者を表示可能だが、実験でを使用した 27 インチ画面の大きさでは 1,000 ノード程度の表示が限界のため頻度をかなり高い値にしている。そのため、ここに表示されている人物のほとんどは、その分野で有名な研究者と思われる。

4.3 考察

実験結果から、Graphviz を用いることで、その研究分野での研究グループを俯瞰的に捉えることができることがわかり、ノードが動的に動く自作ツールでは、研究者間のつながりをより詳細に把握することができることがわかった。

現時点での問題点として、Graphviz で表現できるノード数に限界があるため、表示するノード数を予め間引かなければならない点が挙げられる。また、画像への変換を行っているため、ユーザの動的な要求に答えることが難しい。一方、自作ツールにおいても、ユーザが把握できるノードの数が画面サイズに依存しているため、この点に関して改良の余地がある。

5. おわりに

おわりに 本稿では、研究会や会議のプログラムから人間関係を可視化することを試み、実際に学会のページからデータを抽出し、可視化を行った。今後は、大きなモニタでの表示方法に関して検討を行なっていきたい。

参考文献

- 1) Graphviz, <http://www.graphviz.org/Home.php>
- 2) あのひと検索スパイシー, <http://spysee.jp/>
- 3) 難波 英嗣, 神門 典子, 奥村 学, 論文間の参照情報を考慮した関連論文の組織化, 情報処理学会論文誌, Vol.42, No.11, pp.2640-2649(2001)
- 4) 吉田 誠, 小林 隆志, 横田 治夫, 公開されている論文 DB からのマクロ情報抽出に対するリサーチマイニング手法と他手法の比較, 情報処理学会論文誌. データベース, Vol.45, No.SIG_7(TOD_22), pp.24-32(2004)
- 5) 江藤 正己, 論文の構成単位に基づいた共引用関係の尺度, 情報処理学会論文誌. データベース, Vol.49, No.SIG_7(TOD_37),

pp.1-15(2008)

6) 松尾 豊, 友部 博教, 橋田 浩一, 中島 秀之, 石塚 満: Web 上の情報から人間関係ネットワークの抽出, 人工知能学会論文誌, Vol.AI 20, pp.46-56(2005)

7) 電子情報通信学会, <http://www.ieice.org/jpn/index.html>

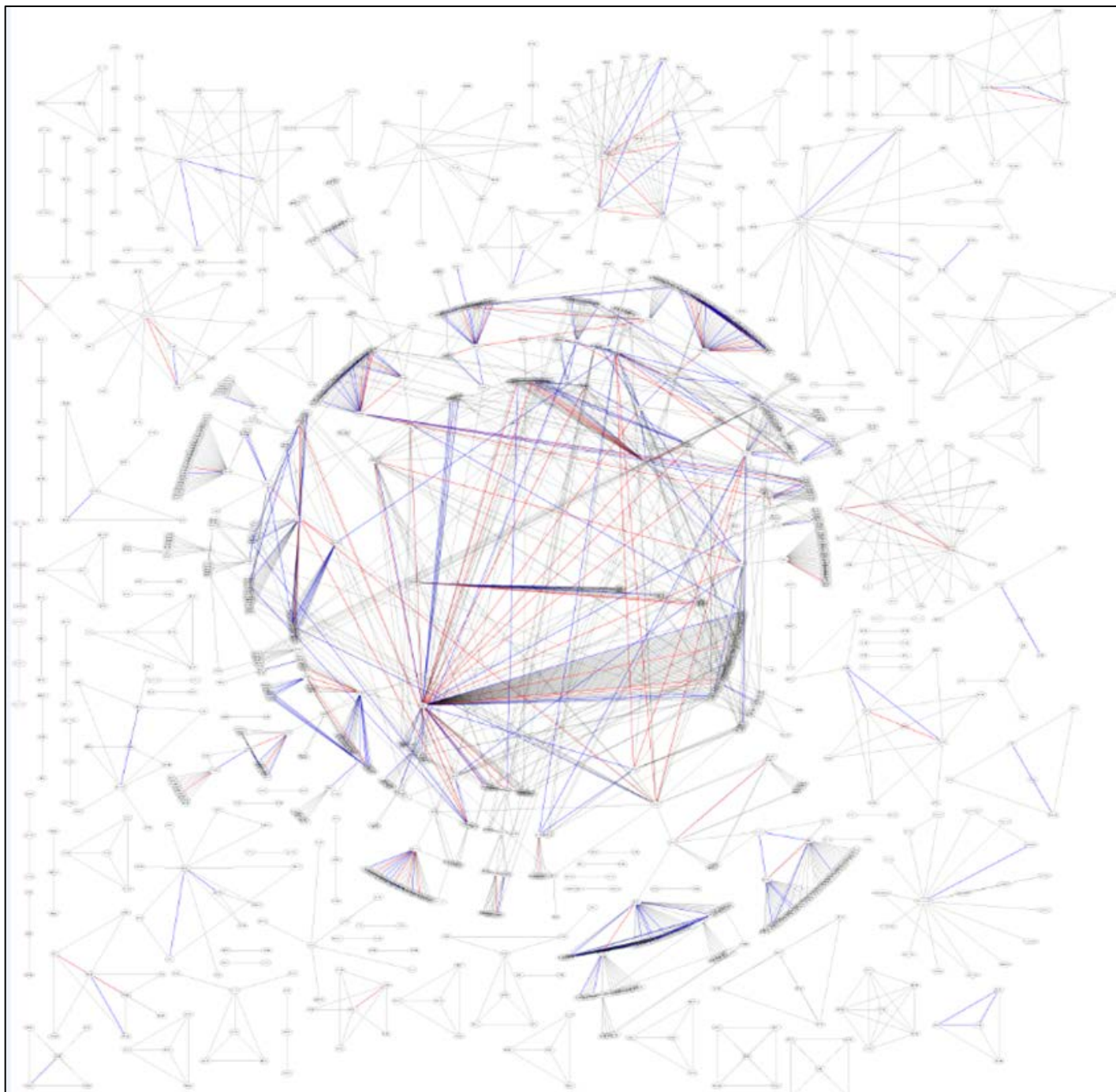


図5 Graphviz の描画例 (データセット A)

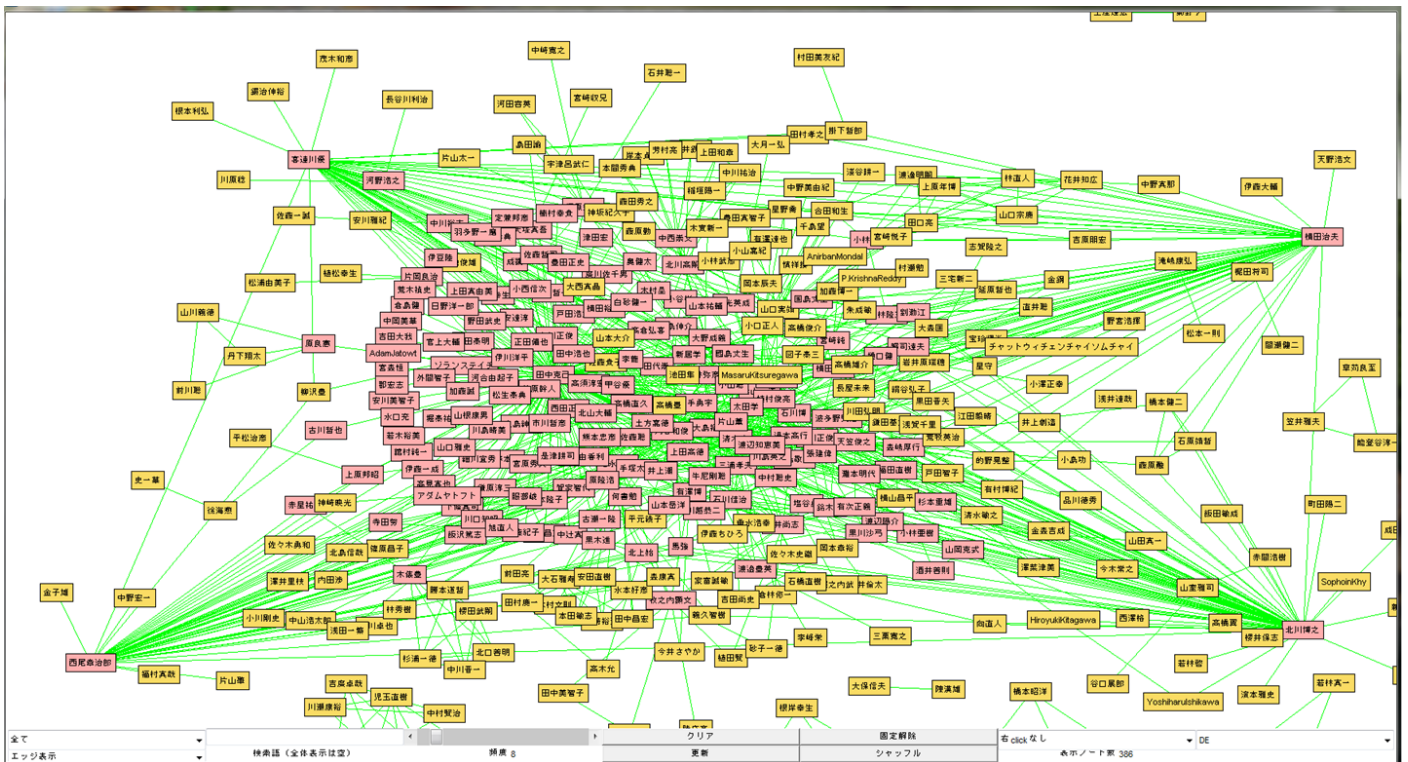


図6 自作ツールの描画例（データセット A）

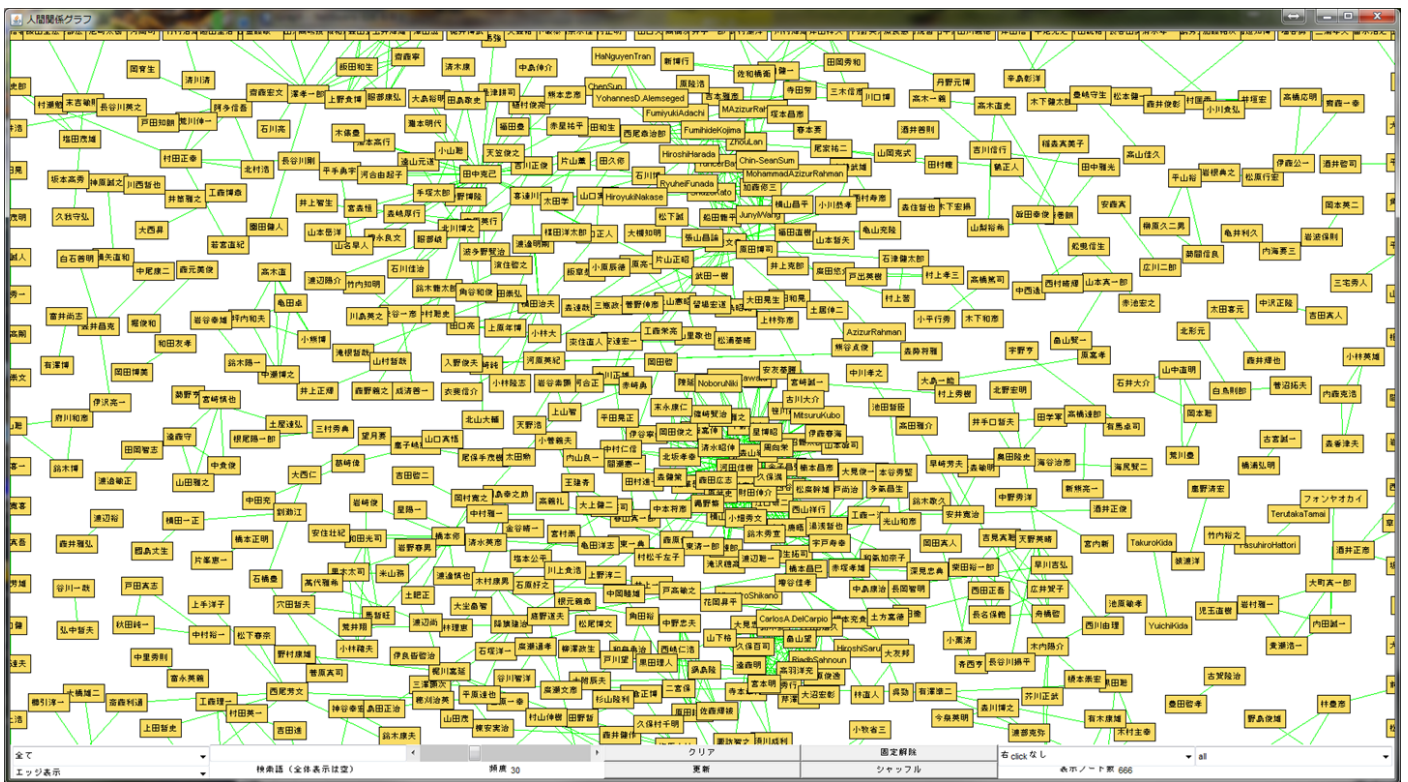


図7 自作ツールの描画例（データセット B）