Semantic Webのための人の社会ネットワーク抽出と利用

友部博教[†1] 松尾豊[†2] 武田英明[†3,†4]
安田雪[†5] 橋田浩一[†2] 石塚満[†6]

Web上では，あらゆる人があらゆるトピックについて何でも書くことができる．セマンティックWebでは，情報の信頼性をWeb of trustにより計算することが重要な技術である．我々はこれまで，研究者の協働関係ネットワークをWeb上の情報から抽出する研究を行ってきた．本論文では，Webから抽出した情報を用い，人の関係性をセマンティックWebの枠組みの中で記述できること，得られたネットワークの中心性の解析をすることで，コミュニティにおける著名な人物が分かること，それが信頼性につながる別の尺度と相関していることを示す．セマンティックWebにおいて，人のネットワークを情報の信頼性の計算に用いる1つの新しい方法を提案していると考えている．

Extraction and Utilization of Human Social Network toward the Semantic Web

Hironori Tomobe,[†1] Yutaka Matsuo,[†2] Hideaki Takeda,[†3,†4]
Yuki Yasuda,[†5] Koiti Hashida[†2] and Mitsuru Ishizuka[†6]

Anyone can say anything about anything on the Web. In the context of Semantic Web, it is important to caluculate the trust of information based on "Web of Trust." We have developed an automatic extraction system of a social network among researchers from the Web. In this paper, we describe that the extracted relation is useful for human relation description in the context of Semantic Web. Moreover, we can extract the central persons in a community by analizing the network centality. Centrality measures are used as a surrogate of trust degree. Our approach shows an overall architecture to calculate the trust of information based on a social network.

1. はじめに

Web上では，あらゆる人があらゆるトピックについて何でも書くことができる．論文のような比較的客観的な立場に立ったコンテンツから，個人の日記，ゴシップ記事，宣伝，噂，フィクション，個人的な意見や思い込みなど，情報の信頼性は千差万別である．我々は，Web上には信頼できるものもあればそうでないものもあると理解したうえで，Web上の情報にアクセスしている．社会的な評価が高い組織や企業のサイトであるか，他のサイトでも同じようなことを言っているか，また有名な人のページであるか，自分の知り合いのページであるかといった手がかりをたよりに，コンテンツの信頼性を推測している．つまり，社会的な評判や関係性といったコンテンツを取り巻く周辺情報によって，コンテンツの信頼性を判断している．

これは，セマンティックWebの文脈でも同じである．セマンティックWebでは，計算機に理解できるRDFなどの形式的な記述の方法を用い，他との関係性を明示的に記述していくことで推論が可能になり，高度な情報検索や問題解決を期待することができる．ただし，誰でも何についてでも情報を発信できるが，推論による情報探索や問題解決を行うためには，どの情報が信頼できるかを計算機によって推測する仕組み

†1 名古屋大学21世紀COEプログラム「社会情報基盤のための音声・映像の知的統合」
21st Century COE Program "Intelligent Media Integration for Social Infrastructure", Nagoya University
†2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)
†3 総合研究大学院大学
The Graduate University for Advanced Studies
†4 国立情報学研究所
National Institute of Informatics
†5 東京大学大学院経済学研究科
Graduate School of Economics, The University of Tokyo
†6 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

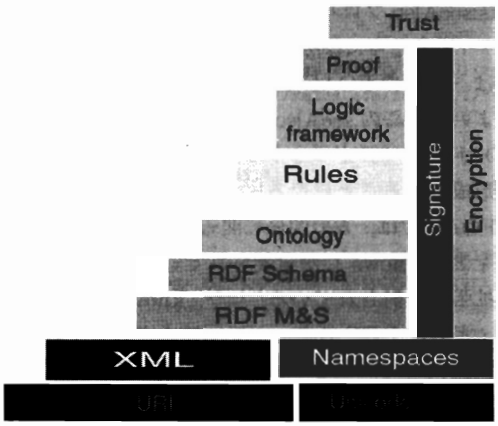

図 1 Layer cake
Fig. 1 Layer cake.

が必要である．Tim Berners-Lee が技術的な階層を示すために用いている Layer cake（図 1）☆1には，一番上に Trust という層があるが，これは最終的にどの情報を用いればよいかは情報の信頼性に帰着することが意図されている．

信頼性を決めるのは，信頼の輪（Web of trust）というネットワーク構造である[5),18)]．多くの人が信頼している人は信頼できるし，その人が信頼している人もまた信頼できる可能性が高い．現実社会でも，人を紹介するときに他の人との関係を説明したり，自社を説明するときに取引先企業を示したりする．これは，他の信頼性によって自分の信頼性を証明していることになる．PageRank も同じような考え方をページの評価に用いているが，セマンティック Web では，これをコンテンツの著者に対して計算する必要がある．

セマンティック Web 上で信頼の輪を実現するには，人が誰とどのような関係にあるかというネットワーク構造が必要である．最近，Orkut[17)] や Gree☆2，mixi☆3などのソーシャルネットワーキングサービス☆4（以下 SNS）の人気が高まっているが，これらのサイトでは，ユーザが趣味や興味などのほかに，自分の友人や知人をサーバに登録することで，知人関係のつながりによるコミュニティを形成している．また，昨年から日本でも Blog と呼ばれる Web サイトの形態が流行している[21)]．トラックバックやコメントといったコミュニケーション機能を備え，更新が頻繁で他サイトへのリンクも張られやすい Blog は，ユーザ間の関係性を把握する有用な情報源を提供していると考えられる．SNS や Blog は，信頼性の計算につながる可能性のある，「人のネットワーク」を得ることができるという点で，セマンティック Web に貢献する可能性を秘めている．

一方，我々はこれまで，Web 上の情報から人間関係ネットワーク，特に研究者の協働関係を抽出する研究を行ってきた．検索エンジンと機械学習の手法を用いることで，高い精度で研究者の関係の強さおよび種類（共著関係や同研究室関係など）を判別することができる．研究者という特殊な対象ではあるが，数百人から数万人規模の人のネットワークを自動的に獲得することができる．

本論文では，Web から抽出した情報を用い，人の関係性をセマンティック Web の枠組みの中で記述できること，得られたネットワークの中心性の解析をすることで，コミュニティにおける著名な人物が分かること，それが信頼性につながる別の尺度と相関していることを示す．本研究は，セマンティック Web において，人のネットワークを信頼性の計算に用いる枠組みの 1 つの方向性を提示していると考えている．

以下，2 章では人のネットワークの取得について関連研究を述べる．3 章では Web から人のネットワークを自動抽出する手法について概要を述べる．そして Web から自動抽出した人の関係性を FOAF によって記述することで，セマンティック Web の枠組みで記述できることを示す．4 章では，信頼性の計算法について先行研究を紹介し，Web から抽出したネットワークを用いて，中心性の計算を行い，信頼性を推測する．ネットワークの中心性の解析からコミュニティにおける著名な人物を抽出し，この中心性と各研究者のパフォーマンスの尺度との相関から，信頼性となる適切な指標を示す．そして 5 章でまとめる．

2. 関連研究

セマンティック Web では，1 章で述べたように，情報の信頼性を測るために，信頼の輪（Web of trust）をいかにして作るかが鍵となる☆5．そのためには，人の関係を表すネットワークを獲得することが必要であ

☆1 Tim Berners-Lee http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/
☆2 http://www.gree.jp/
☆3 http://mixi.jp/
☆4 もしくはソーシャルネットワーキングサイト

☆5 信頼性（trust）という言葉はさまざまな意味で用いられることがある．たとえばアドレス認証に関する研究では，電子署名を trust としてアクセスコントロールを行っている．本論文で意図しているのは，情報が当人によって書かれたものかどうかという点での信頼性（authentification）ではなく，社会的な信頼性（reputation）に近いものである．

る．そのためのアプローチとしては，次のようなものが考えられる[15),16)]．

- ユーザが人の関係について記述するコストを下げる．メーラやブラウザ，スケジューラといった日常的なツールの中に組み込んだり，FOAF-a-matic☆のようなユーザフレンドリなインタフェースによって，記述しやすくしたりする．
- 既存の情報源を再利用する．メールアーカイブ[6)]やWebページの情報から人の関係を抽出する．

前者に関して，最近着目を集めているのがSNSであり，さまざまなツールが作られている．また，後者のアプローチの1つが本論文の方法である．

2.1 ソーシャルネットワーキングサービスとFOAF

SNSの先駆け的な存在として有名なのがFriendster☆☆である．2003年のサイト開設以来，全世界でユーザ数は数百万人に達している．また，Googleの社員が作ったことで有名なOrkutは日本でもユーザ数を延ばしている．日本初のSNSは2004年9月の時点で少なくとも10以上あり，なかでもGreeとmixiが数万人から10万人規模のユーザ数である．これらのサイトでは，ユーザが自分の自己紹介を書き込んだり，友人とのつながりを登録したりすることができる．友達の友達を見たり，周りの人のお薦めの本やCDの情報が手に入る，Blogを読めるなど面白いサービスではあるが，特定の目的やビジネス的な位置づけがはっきりしているわけでない．SNSでは，人のネットワークが手に入るため，セマンティックWebとの関連が着目されている．

一方，W3Cの活動として，BrickleyとMillerが中心となって進めているFOAF（Friend Of A Friend）プロジェクトでは，個人の情報を記述するFOAFと呼ばれるRDFの語彙の構築を進めている[3),20)]．FOAFを使うことで，個人のメールアドレスを識別子として，ホームページや所属，職場や学校のURI，Blogの URI，所属するグループやプロジェクト，興味などさまざまな情報をRDFで記述することができる．例を**図2**に示す．FOAFは，SNSの共通データ形式として期待されているが，プライバシやデータの管理と検索など，さまざまな問題を解決する必要がある．

FOAFでは，特に，自分の知り合いをfoaf:knowsプロパティにより記述することができる．図2の下半分にrdfs:seeAlsoで始まる行があるが，これは，rdfs:seeAlsoプロパティによって，知り合いのFOAFファイルをさらに参照していることを表している．ユーザは，自分のFOAFファイルをどこに置いておいてもよいが，このseeAlsoプロパティによって，次々とFOAFファイルがつながっていき，人のネットワークが形成されることになる．

FOAFでは，人の関係に関して，知り合いであることを表すknowsしか定義されていないが，さまざまな拡張が検討されている[14),16)]．knowsのサブプロパティとしてRDFスキーマで関係を定義しておくことで，拡張した関係をFOAFファイルの中で記述できるようになる．

2.2 既存の情報源からの関係抽出

本研究では，Web上の情報から人のネットワークを抽出するが，それに関連する研究としては次のようなものがある．Kautzらによる Referral Webは，Web上での対象人物へのつながりを順次発見し，人間関係を可視化する研究として有名である[10)]．Referral Webでは，まずシードとして与えられた人名の出現するWebページを検索エンジンを用いて収集し，検索されたページに出現する人名を人名抽出の技術を用いて抽出する．次に，人名どうしの共起関係の強さを，検索エンジンのヒット件数を用いたJaccard係数により計算する．また，原田らは，ある単語で検索したWebページ集合（最大1,000件）から固有表現抽出により名前を抽出する[23)]．そして，独自に定義した共起度を用い，共起関係から人物の関係を表すネットワークを抽出する方法を提案している．基本的なアイディアは本研究の手法と同じであるが，本手法では人名のリストを最初から与える点，また，共起の尺度やテキスト解析による関係の種類の判別など，より精密なネットワークの抽出を行うことができる点で異なる．

人以外にも，Webページの関係の抽出[22)]，EUのプロジェクトの関係の抽出[7)]なども行われている．また，人の関係を，Web上の情報ではなく，社内のメールのやりとりから抽出する研究もある[8),19)]．

野村らはWeb上のリンク関係や論文の共引用関係によるネットワークに対して，中心性の分析やクラスタ分析を行っている[25)]．本手法とはネットワークの構築法は異なるものの，海外の著名な情報系研究者が中心性の上位としてあげられている．また，Web上の情報から人のネットワークを抽出し，ネットワーク分析を行うという点で，Mikaの研究は本研究と非常に方向性が近い[16)]．共起関係に基づいて計算した関係性や，収集されたFOAFのプロファイルを用いて，研究者のネットワークを構築し，さまざまな尺度での中

☆ http://www.ldodds.com/foaf/foaf-a-matic.html
☆☆ http://www.friendster.com/

```
<rdf:RDF
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
      xmlns:foaf="http://xmlns.com/foaf/0.1/"
      xmlns:admin="http://webns.net/mvcb/">
 <foaf:name>Yutaka Matsuo</foaf:name>
 <foaf:title>Dr</foaf:title>
 <foaf:givenname>Yutaka</foaf:givenname>
 <foaf:family_name>Matsuo</foaf:family_name>
 <foaf:mbox rdf:resource="mailto:y.matsuo@aist.go.jp"/>
 <foaf:homepage rdf:resource="http://www.carc.aist.go.jp/~y.matsuo/"/>
 <foaf:depiction rdf:resource="http://www.carc.aist.go.jp/~y.matsuo/homepage/matsuo-narita.jpg"/>
 <foaf:phone rdf:resource="tel:+81-3-3599-8214"/>
 <foaf:workplaceHomepage rdf:resource="http://www.carc.aist.go.jp/"/>
 <foaf:workInfoHomepage rdf:resource="http://www.carc.aist.go.jp/~y.matsuo/homepage/research.htm"/>
 <foaf:schoolHomepage rdf:resource="http://www.u-tokyo.ac.jp/"/>
 <foaf:knows>
  <foaf:Person>
   <foaf:name>Masahiro Hamasaki</foaf:name>
   <rdfs:seeAlso rdf:resource="http://www-kasm.nii.ac.jp/~hamasaki/foaf.rdf"/>
   <foaf:mbox rdf:resource="mailto:hamasaki@grad.nii.ac.jp"/>
  </foaf:Person>
 </foaf:knows>
 <foaf:knows>
  <foaf:Person>
   <foaf:nick>i2k</foaf:nick>
   <rdfs:seeAlso rdf:resource="http://www-kasm.nii.ac.jp/~i2k/foaf.rdf"/>
   <foaf:mbox_sha1sum>031e4846ef8666f0bd04d0a7a4fa5985cee7e6b9</foaf:mbox_sha1sum>
  </foaf:Person>
 </foaf:knows>
 <foaf:knows>
  <foaf:Person>
   <foaf:name>Hideaki Takeda</foaf:name>
   <foaf:mbox rdf:resource="mailto:takeda@nii.ac.jp"/>
  </foaf:Person>
 </foaf:knows>
</rdf:RDF>
```

図 2 FOAF の例
Fig. 2 Sample code of FOAF.

心性の解析を行っている．しかし，本手法は，関係性の強さの尺度や関係の種類の抽出法を評価実験に基づいて設計しているため，精度が高く多様なネットワークを得ることができる点，研究者のパフォーマンスと比較しながら中心性の尺度を検討している点で，より進展したものである．

3. Web 上の情報からの人のネットワークの抽出

本章では，人のネットワークを Web 上の情報から抽出する手法[12),13)] を簡単に述べ，それがセマンティック Web の枠組みで記述できることを示す．抽出されるネットワークは，ノードが人であり，関係の強さがエッジの有無，関係の種類がエッジのラベルで表現される．

3.1 エッジの抽出

ここでは，ネットワークの抽出法を人工知能学会（JSAI）の研究者を例にとって説明する．まず，ネットワークを構成するのは，2003 年度および過去 4 年間の人工知能学会の全国大会の著者・共著者とし，ネットワークのノードとする．ネットワークに含める研究者は，あらかじめ目的とする研究コミュニティの研究者リストを何らかの方法で入手しておけばよい．なお，本手法では，個人に関する情報として用いるのは，氏名と所属だけである．

次に，ノード間にエッジを付与する．基本的なアルゴリズムは非常にシンプルである．たとえば，「松尾豊」と「石塚満」の関係を調べるときには，検索エンジンに

“松尾豊 AND 石塚満”

と入力する．「松尾豊 AND 石塚満」の場合には，156

件のヒットがあるのに対し☆,「松尾豊 AND 溝口理一郎」の場合には 7 件のヒットしかない.「石塚満」単独では 1,120 件のヒット件数,「溝口理一郎」単独では 1,130 件のヒット件数であり,ほぼ同数であるから,「松尾豊」と AND をとったときの件数の違いは,氏名の共起関係の強さの違いを表していると考えることができる.

氏名が共起するページというのは,研究室のメンバのページ,業績リストのページ,論文データベース,学会や研究会のプログラム,大学内の教官メンバリストなどさまざまである.そして,このようなページが多くあるほど,両者が何らかの社会的関係にあり,またその関係が強い可能性が高いというヒューリスティックを本研究では用いている.

共起の強さを測る指標として,次の Simpson 係数(もしくは Overlap 係数)を用いる.

$$R(X,Y) = \begin{cases} \dfrac{|X \cap Y|}{min(|X|,|Y|)} & \text{if } |X| > k \text{ and } |Y| > k, \\ 0 & \text{otherwise} \end{cases}$$

$R(X,Y)$ は,「X」と「Y」の関係の強さを表す関数であり,k は閾値である.JSAI2003 の場合,$k = 50$ とした.つまり単独でのヒット件数が 50 件以下の人はエッジが張られない.評価実験により,この指標は Jaccard 係数や共起頻度などの他の指標よりも関係の強さの適切な推測値となることが分かっている.

また,同姓同名の問題に対処するために,氏名とともに所属もクエリとして用いた.たとえば,「松尾豊」の場合には,"松尾豊産業技術総合研究所" というクエリを用い検索する.なお,複数の所属機関にまたがっている場合や所属が変わった場合は,それらを OR でつなげたものを用いる.また,東大と東京大学など,代表的な機関の略称や別名については,同義語辞書を作り,同義語拡張を行ったうえで検索を行う.

3.2 エッジラベルの抽出

次に,検索にヒットしたページから関係の種類を判別する.研究者の関係の種類として,本手法では次のようなクラスを定めた.

共著関係 共著の論文がある関係.

同研究室関係 同じ研究室や研究所のメンバなど所属が同じである(あった)関係.

同プロジェクト関係 同じプロジェクトや委員会など,組織をまたがる同グループに所属している(いた)関係.

同発表関係 同じ研究会で発表する(した)関係.

1 つのエッジは複数のラベルを持つことができる.

このような関係を抽出するために,まず検索エンジンに「X and Y」をクエリとして入力し,上位 5 ページを取得する.次に,それぞれのページから属性の値を抽出する.ここでいう属性とは,たとえば,X と Y が同行内で共起したか,X および Y の出現回数,タイトルや最初の 5 行に別に定義した語群に含まれる語が出現するかなどである.この属性を用い,判別ルールによって共著や同研究室などどのクラスにあたる関係かを判断する.この判別ルールは,あらかじめ人手で付与した訓練例を用い,C4.5 を用いて生成する.

人手による正解の評価および対象となっている研究者への実際のアンケート調査によって,8 割から 9 割程度の精度で関係を判別することができることが示されている[13].

3.3 抽出したネットワークの FOAF による表現

Web から抽出した情報を用い,人の関係性をセマンティック Web の枠組みの中で記述するには,FOAF で表現すればよい.その例を**図 3** に示す.ACRE(ACademic RElationship)という独自に定義した RDF スキーマにより,Coauthor,Lab,Proj,Conf というプロパティを,foaf:knows のサブプロパティとして定義している.この例では,松尾豊氏が石塚満氏と共著関係にある(さらに知り合いでもある)ことを示している.

ユーザが自分で記述する knows 関係と,他の情報源から自動的に取得される人同士の関係は異質なものであるが,うまく融合すれば効果的である.たとえば,

```
<rdf:RDF
  xmlns:rdf=
  "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:acre=
  "http://www.carc.aist.go.jp/~y.matsuo/ACRE/0.1/">

<foaf:Person>
  <foaf:name>Yutaka Matsuo</foaf:name>
  <foaf:mbox>y.matsuo@aist.go.jp</foaf:name>
  <foaf:workplacehomepage
    rdf:resource="http://www.carc.aist.go.jp/">
  <acre:Coauthor>
    <foaf:Person>
      <foaf:name>Mitsuru Ishizuka</foaf:name>
    </foaf:Person>
  </acre:Coauthor>
</foaf:Person>
```

図 3 Web から抽出した人間関係を表す FOAF の例
Fig. 3 Sample code of FOAF made from a mined relation from the Web.

☆ 2004 年 1 月 8 日時点での Google による検索結果.以下の例でも同様.Google では姓と名の間をつめて正確な氏名の検索が可能である.

JSAI2004 では学会支援の 1 つとしてユーザが聴講する発表のスケジューリングを管理するスケジューリング支援システムが運用された．聴講する発表の推薦にいかすために，自分の知り合い情報（knows 関係）を登録できるシステムであるが，その中で我々の手法によって自動的に抽出された関係の強い人を「知り合いではありませんか」と提示することによって，より洩れが少なく知り合いを登録できるようになっている．

4. 信頼性と中心性の計算

4.1 信頼性の計算に関する先行研究

複数の FOAF ファイルや，Web 上の情報からの抽出によって人のネットワークが得られたとして，そこからどのように人の信頼性を計算するかに関しては，これまでにいくつか研究が行われている．

Richardson らは，trust と belief を分け，前者を人の人に対する信頼性，後者を人の自分自身の言明に対する信頼性と定義している[18]．たとえば，ノード A がある言明 s に対して 0.8 の belief があり，ノード B からノード A に対して 0.5 の trust，ノード C からノード B に対して 0.6 の trust があったとすると，ノード C がノード A の言明 s に対する信頼度は $0.6 \times 0.5 \times 0.8 = 0.24$ と計算される．確率的な視点からの解釈も行っている．

Golbeck らは，ノード A からノード C に直接エッジがない場合に，その中間にあるノードを介して信頼性を計算する手法を提案している[5),6)]．信頼度を 0-1 値とし，ノード A が信頼するノードのノード C に対する信頼度の平均値が 0.5 以上ならノード A は C を信頼するというモデルでは，信頼度の評価に誤りが多い場合でもロバストに信頼性を推測できることをシミュレーションによって検証している．また，信頼度を 1 から 9 までの整数値で定義するアプリケーションでは，ノード A の B に対する信頼度，B の C に対する信頼度のかけ算によりノード A のノード C に対する信頼度の計算を行っている．

信頼性をどのように計算するかに関しては，以前から研究が行われている[11)]が，多くは信頼度を何らかの方法でネットワーク上を伝搬させるものである．Guha らは，この伝搬の方法を，直接的な信頼関係の場合，共参照的な信頼関係にある場合，被信頼関係の場合，共参照的な被信頼関係の場合の 4 つにわけ，信頼度の伝搬を特徴づけている．

さて，個人の他の人に対する信頼性の計算と，コミュニティ全体における一般的な信頼性，いわば reputation とはどのような関係にあるのだろうか．Kamvar らは，P2P を対象としたネットワークにおける個人の信頼性の値から，一般的な信頼性の値（global trust value）をどのように計算するかについて考察している[9)]．個人 i の他の人 j に対する信頼性 s_{ij} を，全体で 1 になるように正規化したうえで，他の人の global trust value に s_{ij} をかけあわせることで自分の global trust value を決定する，というアルゴリズムで，EigenTrust と呼ぶ一般的な信頼性が出るとしている．この EigenTrust は，ネットワークの接続行列の固有ベクトルにあたるものであり，PageRank におけるランダムサーファーモデル[2)]や社会学のネットワーク分析における固有ベクトル（もしくはボナチッチ）の中心性[24)]とも同様の手法である．

4.2 中心性の計算

次に抽出した人のネットワークを用いて，ネットワークの中心性を算出することで人の信頼性を推測する．本節では，4.1 節で述べた固有ベクトルの中心性を使って，一般的な信頼性の計算を試みる．論文などを通じて互いに間接的に関係しあうことが多いと考えられる研究者を対象としたネットワークであるため，他者との間接的結合関係の多寡を考慮する固有ベクトルによる中心性の計算が妥当だと考えられる．

まず，それぞれのノード v は反復 n において信頼度 $A_n(v)$☆を持つとする．信頼度は近接ノードの信頼度を用いて，次のように反復的に計算される．

$$A_{n+1}(v') = c \sum_{v' \in N(v)} \frac{rel(v, v')}{rel\_sum(v')} A_n(v') + cE(v) \tag{1}$$

$$rel\_sum(v) = \sum_{v'' \in N(v)} rel(v, v'') \tag{2}$$

ここで $N(v)$ はノード v とつながっているノード集合を，c は定数を表す．$rel(v, v')$ はノード v と v' の関係の強さであり，ここでは Web 上の共起頻度を用いている．

次に，JSAI の全国大会で過去に発表した 1,509 人に対して，$A(v)$ を計算し，その上位を示したものが**表 1** である☆☆．15 人のうち 11 人が JSAI の理事や理事を勤めたことがある人である☆☆☆．また，若手で

☆ 社会学では中心性の値，PageRank アルゴリズムでは PageRank という値に対応する．

☆☆ 本論文では人工知能学会コミュニティにおけるキーパーソンの人数として，1,509 人の 1%にあたる 15 人が妥当と考えた．

☆☆☆ 2005 年 1 月 31 日時点で，石塚氏は人工知能学会会長，溝口氏は副会長，奥野氏，武田氏，山口氏は理事である．また，田中氏は前学会長，西田氏，石田氏，中島氏，橋田氏，古川氏は理事経験者である．

表 1 JSAI における固有ベクトル中心性上位の人物
Table 1 Result of authority propagation.

	氏名	中心性の値	頻度
1	西田豊明	5.53	624
2	石田亨	4.98	574
3	中島秀之	4.52	278
4	橋田浩一	4.499	345
5	石塚満	4.24	377
6	奥乃博	3.89	242
7	溝口理一郎	3.60	404
8	山田誠二	3.35	168
9	武田英明	3.22	435
10	山口高平	3.10	236
11	大澤幸生	2.98	185
12	田中穂積	2.90	465
13	徳永健伸	2.89	302
14	古川康一	2.77	141
15	河原達也	2.74	440

も人工知能学会を中心に活発に活動している研究者が上位に入っている☆1☆2.

表中の頻度は，氏名と所属による Web ページのヒット件数を示している．頻度が高い人が必ずしも信頼度で上位にくるわけではない．特に，本を出版している人やソフトウェアを開発している人は，Web ページのヒット数が非常に多くなる傾向がある．それに対して，ネットワークを抽出して信頼度を計算した場合には，たとえ頻度が低くてもネットワーク構造上で重要な位置をしめている人は高い値となる．

4.3 中心性と研究者のパフォーマンス

前節では，固有ベクトルによる中心性のアルゴリズムを使って信頼度を計算したが，社会学では古くから中心性を計算するさまざまな指標が提案されている．本節では，いくつかの中心性の指標を用い，さらに研究者の信頼性を測る別の指標としてパフォーマンスに関するデータとの比較を行う．

ネットワーク解析に一般的に用いられることの多い，次の 4 つの中心性の尺度を用いる[1),4)].

E（Eigenvector Centrality） 固有ベクトルによる中心性．

C（Closeness） 近接性．対象ノードと他のすべてのノードとのグラフ上の距離の平均

D（Degree） 次数．対象ノードの持つエッジの数．

B（Betweeness） 媒介性．他のノードペアの最短パスに，対象ノードが含まれる割合．

また，本研究のネットワーク抽出の手法では，Simpson 係数によるネットワークのほかに，関係の種類を考慮することができるため，共著関係によるネットワーク，研究室関係によるネットワーク，プロジェクト関係によるネットワークを含めた，計 4 種類のネットワークを対象とする☆3.

さらに，研究者のパフォーマンスの指標として，論文データベース☆4を用いて，以下を用意した．

総発表文献数 対象の研究者が発表した論文数

JSAI 発表文献数 JSAI で発表した論文数

総発表ページ数 発表した論文の総ページ数

JSAI 発表ページ数 JSAI で発表した論文の総ページ数

総被引用数 他の論文に引用された回数

JSAI 被引用数 JSAI の論文に引用された回数

表 2 に，中心性とパフォーマンスの相関係数を示す．人工知能学会の研究者 298 人☆5に対して計算した値である．Simpson 係数によるネットワークでは，媒介中心性（B）がどのパフォーマンスとも強い相関があることが分かる☆6．特に，総発表文献数の相関係数は 0.619 と最も高い．また，固有ベクトルによる中心性（E）は JSAI 発表文献数に対して相関係数が 0.546，JSAI 発表ページ数に対して相関係数が 0.515 と強い相関がある．

共著関係によるネットワークでは，媒介中心性（B）は強い相関があるが，固有ベクトルによる中心性（E）とは相関が弱い．特に，JSAI 発表文献数に対しては，共著関係の媒介中心性の相関係数が 0.670 と高く，最も良い指標となっていることが分かる．研究室関係も共著関係と同様の傾向を示している．また，プロジェクト関係は全体に相関が弱く，プロジェクト関係のネットワーク構造から研究者のパフォーマンスをとらえるのは難しいことが分かる．

以上から，研究者のパフォーマンスが論文数や論文ページ数によって測られるとすると，Simpson 係数や共著関係，研究室関係のネットワークにおける媒介中心性，もしくは Simpson 係数のネットワークにおける固有ベクトル中心性が良い指標になっていることが分かる．固有ベクトル中心性は他者との間接的結合の

☆1 人工知能学会で採録された論文数（共著を含む）を活動の指標とし，活発な活動をしていると判断している．

☆2 一方で，ネットワーク抽出の特性により，同姓同名の著名人がいる人，最近所属が変わった人などは，値が低くなる傾向がある．

☆3 なお，発表ネットワークは，他の 3 種類と性質の異なるネットワークであり，ここでは分析の対象としていない．

☆4 NII が運用する引用文献情報ナビゲータ CiNii (http://ci.nii.ac.jp/) のデータベース．実験時において，国内の計算機科学の論文が約 547,000 編集められている．

☆5 人工知能学会の研究者 1,509 人のうち，Web のヒット件数が閾値以上の人を選んでいる．

☆6 本論文では相関係数が 0.4 を超える場合には強い相関があると判断している．

表 2 各ネットワークに対する中心性とパフォーマンスの相関係数
Table 2 Correlation coefficient between centrality and performance.

	Simpson 係数				共著関係			
	E	C	D	B	E	C	D	B
総発表文献数	0.390	−0.052	0.394	0.619	0.181	−0.126	0.391	0.487
JSAI 発表文献数	0.546	−0.041	0.505	0.533	0.297	−0.153	0.583	0.670
総発表ページ数	0.344	−0.046	0.349	0.555	0.153	−0.113	0.325	0.395
JSAI 発表ページ数	0.515	−0.039	0.478	0.498	0.276	−0.146	0.530	0.605
総被引用数	0.360	−0.038	0.364	0.518	0.139	−0.090	0.329	0.379
JSAI 被引用数	0.390	−0.027	0.377	0.424	0.198	−0.103	0.382	0.409
	研究室関係				プロジェクト関係			
	E	C	D	B	E	C	D	B
総発表文献数	0.087	−0.185	0.294	0.396	0.346	−0.276	0.352	0.276
JSAI 発表文献数	0.153	−0.181	0.464	0.647	0.337	−0.233	0.384	0.392
総発表ページ数	0.060	−0.163	0.249	0.338	0.309	−0.244	0.308	0.241
JSAI 発表ページ数	0.135	−0.174	0.439	0.621	0.305	−0.220	0.352	0.358
総被引用数	0.075	−0.135	0.291	0.402	0.252	−0.208	0.266	0.218
JSAI 被引用数	0.092	−0.127	0.347	0.474	0.176	−0.156	0.222	0.241

多寡を意味し，媒介中心性は他者との関係を連結させるために不可欠かどうかを意味する指標である．これらの指標は人工知能学会のような学術コミュニティにおいて，研究者のパフォーマンスが他者との協働関係と強く関係するため，強い相関を示したと考えられる．

本論文で対象としている信頼性と，ここで取り上げている研究者のパフォーマンスには少しギャップがあるが，少なくとも，人工知能の研究に関わる内容であれば，一般的に研究業績をあげている人の情報や意見は，そうでない人の情報や意見より信頼性が高いと考えるのが妥当であろう．その意味では，Web 上から抽出したネットワークの媒介中心性や固有ベクトル中心性を測ることで，情報の信頼性の尺度として用いることは可能であろう．

5. おわりに

信頼性を計算することは，セマンティック Web において重要な課題である．本論文では，Web から抽出した人のネットワークを抽出し，セマンティック Web の枠組みの中で記述する方法について述べた．そして抽出したネットワークの中心性を測ることで，業績の高い研究者を抽出することができる可能性があることを示した．少なくとも，人工知能学会のコミュニティが対象としている人工知能に関連する情報に関しては，この業績の高い研究者は信頼性が高いと考えられる．

今回は人工知能学会という特定の学術組織を対象にしており，現在この手法を学術組織以外へのコミュニティへの適用も試みている．「共著」や「同研究室」などのクラス分けを，それぞれのコミュニティごとに定義する必要があるだろう．そして信頼性の検証には，コミュニティに応じた尺度を利用しなければならない．今回の解析は研究者に限定したものであり，信頼性の尺度として研究者のパフォーマンスを用いたが，コミュニティが異なれば別の尺度によって検証しなければならない．

セマンティック Web において，コンテンツの信頼性を測るにはコンテンツの著者に対する信頼性の計算が必要となる．本論文は研究者という特殊な条件ではあるが，Web から抽出した人のネットワークを用いて，信頼性を測定するという 1 つの方向性を示している．

参考文献

1) Bonacich, P.: Factoring and weighting approaches to status scores and clique identification, *Journal of Mathematical Sociology*, Vol.2, pp.113–120 (1972).
2) Brin, S. and Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine, *Proc. 7th WWW Conf.* (1998).
3) FOAF: the 'friend of a friend' vocabulary. http://xmlns.com/foaf/0.1/
4) Freeman, L.C.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol.1, pp.215–239 (1979).
5) Golbeck, J. and Hendler, J.: Inferring reputation on the semantic web, *Proc. WWW2004* (2004).
6) Golbeck, J. and Hendler, J.: Reputation network analysis for email filtering, *Proc. 1st Conference on Email and Anti-Spam* (2004).
7) Grobelnik, M. and Mladenic, D.: Approaching Analysis of EU IST Projects Database, *Proc. International Conference on Information and Intelligent Systems (IIS-2002)* (2002).
8) Guimera, R., Danon, L., Diaz-Guilera, A.,

Giralt, F. and Arenas, A.: *Self-similar Community Structure in Organizations* (2002).
9) Kamvar, S., Schlosser, M. and Garcia-Molina, H.: The EigenTrust Algorithm for Reputation Management in P2P networks, *Proc. WWW2003* (2003).
10) Kautz, H., Selman, B. and Shah, M.: Referral web: Combining social networks and collaborative filtering, *Comm. ACM*, Vol.40, No.3, pp.63–65 (1997).
11) Khare, R. and Rifkin, A.: Weaving a Web of Trust, *WWW Journal*, Vol.2, No.3, pp.77–112 (1997).
12) 松尾　豊，友部博教，橋田浩一，石塚　満：イベント空間支援における人間関係ネットワーク抽出技術の活用，人工知能学会全国大会，pp.3C1–04 (2004).
13) 松尾　豊，友部博教，橋田浩一，中島秀之，石塚満：Web 上の情報からの人間関係ネットワークの抽出，人工知能学会誌，Vol.20, No.1 (2005).
14) Matsuo, Y., Hamasaki, M., Mori, J., Takeda, H. and Hasida, K.: Ontological Consideration on Human Relationship Vocabulary for FOAF, *Proc. 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web* (2004).
15) Matsuo, Y., Tomobe, H., Hasida, K. and Ishizuka, M.: Finding Social Network for Trust Calculation, *Proc. 16th European Conference on Artificial Intelligence* (*ECAI2004*), pp.510–514 (2004).
16) Mika, P.: Bootstrapping the FOAF-Web: An experiment in social networking mining, *Proc. 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web* (2004).
17) Orkut. http://www.orkut.com/
18) Richardson, M., Agrawal, R. and Domingos, P.: Trust Management for the Semantic Web, *Proc. ISWC2003, LNCS2970*, pp.351–368 (2003).
19) Tyler, J.R., Wilkinson, D.M. and Huberman, B.A.: Email as spectroscopy: automated discovery of community structure within organizations, *International Conference on Communities and Technologies* (2003).
20) XML-Watch: Finding friends with XML and RDF (2002). http://www-106.ibm.com/developerworks/xml/library/x-foaf.html
21) 武田英明，大向一輝：Weblog の現在と展望 — セマンティック Web とソーシャルネットワーキングの基盤として，情報処理，Vol.45, No.6 (2004).
22) 村田剛志：参照の共起性に基づく Web コミュニティの発見，人工知能学会誌，Vol.16, No.3, pp.316–323 (2001).
23) 原田昌紀，佐藤進也，風間一洋：Web 上のキーパーソンの発見と関係の可視化，情報処理学会研究報告，Vol.DBS-130/FI-71 (2003).
24) 安田　雪：実践ネットワーク分析，新曜社 (2001).
25) 野村早恵子，三木　武，石田　亨：コミュニティマイニングにおける Web 引用解析と文献引用解析の比較，電子情報通信学会論文誌 D-I，Vol.J87-D-I, No.3, pp.382–389 (2004).

(平成 16 年 9 月 27 日受付)
(平成 17 年 4 月 1 日採録)

友部　博教（正会員）

1999 年東京大学工学部電子情報工学科卒業．2004 年同大学院情報理工学系研究科博士課程修了．博士（情報理工学）．同年より，名古屋大学 21 世紀 COE プログラム「社会情報基盤のための音声・映像の知的統合」ポスドク研究員．実世界における人間活動からの知識抽出，再利用に興味を持つ．人工知能学会会員．

松尾　　豊（正会員）

1997 年東京大学工学部電子情報工学科卒業．2002 年同大学院博士課程修了．博士（工学）．同年より，産業技術総合研究所サイバーアシスト研究センター勤務．2004 年より産業技術総合研究所情報技術研究部門勤務．GBRC 社会ネットワーク研究所研究員．(株) ホットリンク 技術アドバイザー．2003 年度人工知能学会論文賞受賞．高次 Web マイニングに興味がある．人工知能学会，AAAI 各会員．

武田　英明（正会員）

1991 年 3 月東京大学大学院工学系研究科博士課程修了．1993 年 4 月奈良先端科学技術大学院大学助手．1995 年 4 月同助教授．2000 年 4 月国立情報学研究所助教授．2004 年 5 月同教授．現在に至る．総合研究大学院大学教授（複合科学研究科情報学専攻）を併任．特に知識共有，ネットワークコミュニティ，実世界エージェント等の研究に従事．AAAI，人工知能学会等会員．工学博士．

安田　雪

東京大学大学院経済学研究科特任助教授.(有) 社会ネットワーク研究所代表取締役. 専門：社会学，ネットワーク分析，組織科学，雇用機会問題等. 著書：『ネットワーク分析』『実践ネットワーク分析』(ともに新曜社),『人脈づくりの科学』(日本経済新聞社),『働きたいでも働けない…高校生就職難の社会構造』(勁草書房),『大学生の就職活動』(中央公論新社) 等. 現在はコンサルティング業務のほか，東京大学ものづくり経営研究センターにて自動車サプライシステムの取引構造のネットワーク分析，ものづくりインストラクター養成のための中核人材育成プロジェクトに従事. Ph.D. (社会学).

橋田　浩一 (正会員)

1981 年東京大学理学部情報科学科卒業. 1986 年同大学院理学系研究科博士課程修了. 理学博士. 1986 年電子技術総合研究所入所. 1988 年から 1992 年まで (財) 新世代コンピュータ技術開発機構に出向. 2001 年より産業技術総合研究所サイバーアシスト研究センター副研究センター長，2004 年に研究センター長. 同年 7 月より産業技術総合研究所情報技術研究部門副部門長. 専門は自然言語処理，人工知能，認知科学. 最近の興味は，セマンティックコンピューティングおよびその応用としての文脈依存型情報サービス，知の社会的共創等.

石塚　満 (正会員)

1971 年東京大学工学部電子工学科卒業. 1976 年同大学院博士課程修了. 同年 NTT 入社，横須賀研究所勤務. 1978 年東京大学生産技術研究所助教授 (1980～1981 年 Purdue 大学客員準教授)，1992 年東京大学工学部電子情報工学科教授，2001 年情報理工学系研究科電子情報学専攻，2005 年同創造情報学専攻 (電子情報学専攻兼任). 研究分野は人工知能，生命的エージェントによるマルチモーダルメディア，Web インテリジェンス，次世代 Web 情報基盤. IEEE，AAAI，人工知能学会 (現在，会長)，電子情報通信学会，映像情報メディア学会，画像電子学会等の会員.