

Web上からさまざまなエンティティ同士の社会ネットワーク抽出

Extracting Social Networks among various Entities from World Wide Web

金英子*¹ 松尾豊*² 石塚満*¹
Yingzi Jin Yutaka Matsuo Mitsuru Ishizuka

*¹ 東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technology, University of Tokyo

*² 産業技術総合研究所 情報技術研究部門
National Institute of Advanced Industrial Science and Technology

Social networks have recently attracted much attention for their importance to the AI and other research area. Several methods exist to extract social networks for people (particularly researchers) from the web using a search engine. In this paper we expand existing techniques to obtain social networks among various entities: firms and artists (of contemporary), which represent complex and inhomogeneous communities. This paper shows each of these algorithms and systems respectively, and furthermore proposes general extraction of a social network model that combines several ideas of social network extraction methods for various domain entities. Our approach contributes to existing studies by cultivating the applicability of social networks from various domains.

1. はじめに

社会ネットワーク分析は、人や組織のもつ関係性に注目して、つながりとその強度を測定してネットワーク化することにより、社会構造や現象を説明する社会学におけるひとつの方法論であり [金光 03, 安田 97, Wasserman 94], 企業の戦略を策定したり, クチコミマーケティングの分析など経済学や経営学への適応が注目されている。人間関係, 企業間関係, 産業間関係, 国家間関係など, あらゆるレベルのネットワークが対象となり, どの行為者が中心的なのか, どういった派閥が存在するか, そしてその背後にあるものは何なのかなどといった研究が行われている。さらに, 近年, 社会ネットワークは, セマンティック Web における情報の信頼性の計算 [Golbeck 06], クチコミマーケティングの分析 [Leskovec 05], 情報の共有・推薦 [Mori 05], コミュニティ抽出 [Newman 04], オントロジー抽出 [Mika 05] など, 多くの研究で着目されている。

社会ネットワークを抽出するために, 社会学では従来, 観察やインタビュー調査による事例研究を行うことで関係のデータを集めていた。1920年代のホーソン実験を初め, 対象の属性ではなく関係性に着目して現象を分析する社会ネットワーク分析の流れにつながっている。また, ネットワーク・クエスチョンにという, 人のネットワークを抽出するために社会学でよく用いられた方法がある。例えば, アメリカの GSS (General Social Survey) 調査では「あなたが過去半年のあいだに, あなたにとって重要なことを話しあった人々は誰でしたか?」というネットワーク質問を行う。このような質問により, 個人のもつネットワークや関係性の分析を行うことができる [安田 97]。しかし, 多くの人に対して定期的にこれらのアンケートや観察を行うことは難しい。

電子データや情報のデータベース化が進むにつれて, 構造化, あるいは半構造化されたデータから関係性の情報を得てネットワーク化する研究が多くなっている。例えば, 電子図書館 (electronic library) や DBLP (Digital Bibliography & Library Project) から, 論文引用関係のネットワークや共著関係のネットワークを抽出したり [Chen 99, Aleman-Meza 06], E-mail アーカイブや電話記録やスケジュールリングデータから連絡関係の情報を

連絡先: 金英子, 東京大学大学院 情報理工学系研究科

〒113-8656 文京区本郷 7-3-1 工学部新 2 号館 111C1 室

TEL: 03-5841-6774, FAX: 03-5841-8570

Email: eiko-kin@mi.ci.i.u-tokyo.ac.jp

抽出したり [McCarthy 04, Tyler 03], 企業の場合は, 金融機関の振替や取引などの電子データの記録から企業同士の資金や株所有関係のネットワークを抽出したり [福岡 03, 相馬 05], さらに, 社会ネットワーキングサービス (SNS) や Friend-Of-A-Friend (FOAF) ドキュメントから友人関係を抽出したり [Finin 05], Web リンクや Blog のトラックバックの関係情報を抽出する [Adamic 03, Miki 05, 古川 05] などがある。これらの構造化された電子データから抽出されたネットワークは質が高く, 分析結果の信用性が高い。ただ, ネットワークの範囲が限定されていることと, プライバシー問題やデータが高価でだれでも用意に入手することはできないという問題もある。

そして, 最近には検索エンジンを利用して Web 上からエンティティ同士の関係を推定して社会ネットワークを構築・分析する研究が多く行われている。研究者ネットワークの抽出 [Kautz 97, Mika 05, Matsuo 06] を初め, 抽出手法をさまざまなエンティティに拡張した研究として, アーティスト同士の弱い社会的関係抽出 [Jin 07] や企業間関係の抽出 [金 07] などがある。検索エンジンを利用することで, Web から多様なエンティティと, エンティティ同士の多様な関係を抽出することができる。また, 構造化されていない関係でも抽出することができるので有効である。

本論文では, 検索エンジンを利用して Web 上から研究者, アーティスト, 企業などのエンティティの社会ネットワークの抽出する手法をそれぞれの示すとともに, さまざまなエンティティの社会ネットワークを抽出する汎用モデルを提案する。

2. Web上からさまざまなエンティティ同士の社会ネットワーク抽出

2.1 研究者ネットワークの抽出

検索エンジンを用いて Web 上から社会ネットワークを抽出する既存の研究では, 研究者を対象とすることが多い [Kautz 97, Mika 05, Matsuo 06]。これらの研究では, Web における名前の共起の強さは, その 2 人の実世界における関係の強さと高い相関があるという仮説に基づいて関係の有無を判断している。ここで, Web における名前の共起とは, 同一の Web ページ上に名前が同時に出現することを指す。例えば, 学会や研究会のプログラム, 研究室のメンバーのページ, 大学内の教官のメンバーリストなどのページに名前が多く共起するほど, 2 人の研究者の間には何らかの社会的関係が強い可能性が高いと推測で

Algorithm 2.1: NETWORKEXTRACTION (T)

Input: 名前のリスト L , 閾値 T
Output: ネットワーク G

```

for each  $x \in L$ 
  do set a node in  $G$ 
for each  $x \in L$  and  $y \in L$ 
  do  $r_{x,y} \leftarrow \text{GoogleCooc}(x, y)$ 

/* 絶対的ルールによりエッジを張る */
for each  $x \in L$  and  $y \in L$ 
  do if  $r_{x,y} > T$ 
    then set an edge between  $x$  and  $y$  in  $G$ .
return ( $G$ )

```

* ここで, GoogleCooc とは, 検索エンジン (Google) の共起ヒット件数 (“ x AND y ” で検索したもの) を返す関数である。

図 1: 絶対的ルールによるネットワーク抽出

きる。

Web における人物 x と y の共起の強さを計算する関連度の尺度として, 名前の共起頻度 ($|x \cap y|$) を直接用いる以外に, ダイス係数 ($2 \frac{|x \cap y|}{|x| + |y|}$), 相互情報量 ($\log \frac{N|x \cap y|}{|x||y|}$), コサイン類似度 ($\frac{|x \cap y|}{\sqrt{|x||y|}}$),

Jaccard 係数 ($\frac{|x \cap y|}{|x \cup y|}$), Overlap 係数 ($\frac{|x \cap y|}{\min(|x|, |y|)}$) などさまざまなものがある*1 [Manning 02]。ただし, $|x|$, $|y|$, $|x \cap y|$, $|x \cup y|$ はそれぞれ, 名前 x と名前 y の単独でのヒット件数, “ x AND y ” と “ x OR y ” でのヒット件数を表す。例えば, Referral Web [Kautz 97] や Flink [Mika 05] では, 国際会議などに参加している有名な研究者を対象に, Jaccard 係数を用いて研究者同士の関係の強さを計算している。Polyphonet [松尾 05, Matsuo 06] では, さまざまな指標を用いて人間関係の共起の強さを計算した場合について評価・考察を行い, Overlap 係数が人の協働関係の強さを表すのに最も適しているという知見を示している*2。原田らの NEXAS [原田 03] では, 与えられたトピックとそれに関するキーパーソンとの関連度を G スコアという指標を用いて計算している。

いずれの場合も, 図 1 に記述したように, ネットワーク全体で一貫した閾値 (絶対的な指標 T とする) を設定し, 共起指標がそれ以上であればエッジを張るという関連性の判断基準に基づいてネットワークを構成する。本論文では, このような対象全体に一貫した閾値を定めてエッジのあり/なしを決める方法を絶対的ルールによる関係抽出と呼ぶ。ノードの文脈に依存せず, 共起指標の値と閾値によって絶対的な基準でエッジの有無が決定されるからである。 T は, コミュニティの凝集度を反映し, オフライン状態で対象となるエンティティの訓練データから学習しておく必要がある。共通性の高い均一なコミュニティであるほど T の値が高くなる。

2.2 アーティストのネットワーク抽出

均一なコミュニティに対しては, 一貫した閾値により関係の有無を判断することができるが, 不均一なコミュニティに対し

*1 このような検索エンジンによる共起の指標が, どのような状況でどのような人間関係に対して有効であるかを明確にした研究は現在のところない。本手法で対象となる関係性は, 多様な関係性の一部であることに注意いただきたい。

*2 Overlap 係数は, 分母に関して \min をとっており, ヒット件数の小さい方から見た距離感を表している。例えば, 学生からみた先生との関係, 有名でない人からみた有名な人との関係など, 両方からみて強い方の関係をとる。

Algorithm 2.2: NETWORKEXTRACTION (T, M)

Input: 名前のリスト L , 閾値 T
Output: ネットワーク G

```

for each  $x \in L$ 
  do set a node in  $G$ 
for each  $x \in L$  and  $y \in L$ 
  do  $r_{x,y} \leftarrow \text{GoogleCooc}(x, y)$ 

/* まず, 絶対的ルールによりエッジを張る */
for each  $x \in L$  and  $y \in L$ 
  do if  $r_{x,y} > T$ 
    then set an edge between  $x$  and  $y$  in  $G$ .

/* 次に, 相対的ルールによりエッジを張る */
for each  $x \in L$ 
   $Y_x \leftarrow \text{ConnectedNodes}(x)$ 
   $\bar{Y}_x \leftarrow L \setminus Y_x$ 
  while  $|Y_x| < M$  and  $\bar{Y}_x \neq \phi$ 
    do  $y = \underset{y_j \in \bar{Y}_x}{\text{argmax}} r_{x,y_j}$ 
      do set an edge between  $x$  and  $y$  in  $G$ 
       $\bar{Y}_x \leftarrow \bar{Y}_x \setminus \{y\}$ 
       $Y_x \leftarrow Y_x \cup \{y\}$ 
return ( $G$ )

```

* ConnectedNodes とは, あるノード (x) とつながっているノードの集合を返す関数である。 $|X|$ は, 集合 (X) の要素数を返す。

図 2: 絶対的・相対的ルールによるネットワーク抽出

ては一貫した閾値を決めることができない。例えば, 研究者ネットワークの抽出に用いた絶対的なルールによる関係抽出手法を, 国際展に参加しているアーティスト (例えば, 横浜トリエンナーレ 2005 に参加したアーティスト) に適応した場合, 多くの実際に存在する関係が抽出されなくなる。ここで, 不均一というのは, 関係の出力が不安定であることを指す。例えば, 異なる国や分野の人同士の関係は, 同じ国や分野より Web 上に名前の共起が少ないので, 関係が低く計算される。また, 新しく結成された関係なども抽出されなくなる。

提案手法では, 社会学のネットワーク・クエスチョンの考えに基づいて, 個々のノードからみた相対的な重要性によってネットワークを抽出する。ネットワーク全体では共起指標の値が低くても, 各ノードから見て値が高い場合には (M 人まで) エッジを張るという相対的ルールによる関係抽出を行う。これを, 本論文では相対的ルールによる関係抽出と呼ぶことにする。しかし, 相対的ルールを用いる手法では, ネットワークを構成するすべてのアクターを同等に扱うが, これは適切でない場合もある。例えば, 多くの人と関係を取り持つ人 (コネクター) の場合, 相対的ルールではうまく取り出せないこともある。つまり, ノードの活動量そのものが大きく異なるときにも同様に扱ってしまうのが相対的ルールの欠点である。したがって, 本論文では相対的ルールと絶対的ルールと相対的ルールを組み合わせた絶対的・相対的ルールによる関係抽出というネットワーク抽出のアルゴリズムを構築する。その処理は以下の通りである (図 2)。まず, 絶対的ルールによる関係抽出により, 関係の強さが一定の閾値 T 以上になるアクター同士をエッジでつなぐ。つぎに, 相対的ルールによる関係抽出により, エッジの数が少ないアクターに対して, そのアクターのもつ相対的に強い関係のエッジを, エッジの数が M になるまで追加して

Algorithm 2.3: NETWORKEXTRACTION ($score_{th}$)

Input: 名前のリスト L , 関係語リスト W , 閾値 $score_{th}$
 Output: ネットワーク G

```

for each  $x \in L$ 
  do set a node in  $G$ 
for each  $x \in L$  and  $y \in L$ 
  do  $D_{x,y} \leftarrow GoogleTop(x, y, W)$ 

for each  $x \in L$  and  $y \in L$ 
   $score_{x,y} \leftarrow 0$ 
   $S \leftarrow GetSentences(D_{x,y})$ 
  for each  $s \in S$ 
  do
    if  $s$  contains " $x$ " and  $s$  contains " $y$ "
    do
       $score_s \leftarrow \sum_{w_i \in W} \text{contained in } s \cdot t_{w_i}$ 
      if  $score_s > score_{x,y}$ 
      then  $score_{x,y} \leftarrow score_s$ 
    if  $score_{x,y} > score_{th}$ 
    then set an edge between  $x$  and  $y$  in  $G$ 
return ( $G$ )
    
```

* GoogleTop とは、検索エンジンの上位のページ (“ x AND y AND w ” で検索したもの) を返す関数である。そして、GetSentences は、文書集合から文の集合を返す関数である。

図 3: 関係の識別によるネットワーク抽出

いく。したがって、提案手法は共起指標の絶対的な大きさと相対的な大きさの両方を考慮し、活動量の多い人は多くのエッジをもつし、そうでない人でも少なくとも M 本のエッジは取り出されることになる。

ネットワークの抽出は、まず、オフライン状態で訓練データから各パラメータの値を学習する(例えば、訓練データを F 尺度最大になるように分類できるパラメータの値を取得する)。そして、オンライン状態で、名前ペア x AND y で検索クエリを生成し、Google によりそれぞれのヒット件数を求める。ネットワークを構築する段階では提案の手法(閾値調整)を用いる。実際のシステム (<http://release.nikkei.co.jp/>) [Jin 07] の中では、複数の指標(Overlap 係数と共起頻度)を用いて、各ノードの閾値を徐々に下げていくルールで、検索エンジンのばらつきに対処し、関係をよりロバストに抽出している。

2.3 企業間関係のネットワーク抽出

Web 上のメディア効果などにより、名前が頻出するエンティティに対して、ヒット件数は関係の強さを適切に表すことができない。例えば、企業名や組織名、または国の名前や有名人の名前は、ポータルサイトやニュースサイトなどで頻繁に共起するが、必ずしも関係があるとは言えない。

ネットワークの紐帯として、連絡の頻度表すものと関係の種類で表すものがある。Web 上に名前が頻出するエンティティ、例えば、企業の場合には、企業間にどのような関係をもっているかによって、関係の強さが決まる。我々は、エンティティの名前 x と y を含む文脈を分析することで関係の有無を判断している。まず、文脈の信頼度を計算するために、文脈に含まれる語の重要度を計算しておく。また、そういった文脈を Web 上から集めるために、クエリに関係を特定する「関係語」という語を加える。我々は、関係の識別という手法を提案して、企業間の社会ネットワークを抽出している [金 07]。すなわち、図 3) のように、Web 上からエンティティ x と y のある関係に

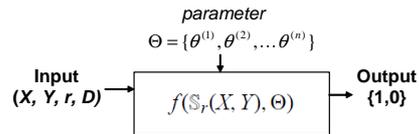


図 4: Web 上で社会ネットワークを抽出する汎用モデル

ついて記述している文脈を集めて、その文脈から関係を表すスコアを計算して、ネットワークを抽出している。文脈のスコアは、その文脈に含まれる語(関係語)のスコアを足し合わせて計算する。

3. Web 上から社会ネットワークを抽出する汎用モデル

本章では、検索エンジンを利用して Web 上から社会ネットワークを抽出する汎用モデルを提案する。Web 上からエンティティ同士の関係を同定するという事は、エンティティ同士の実世界における関係の強さを、エンティティ同士の Web 空間における関係の強さから推定することである。ここでは、Web 空間において関係があると判断されたエンティティ同士は、実世界においても関係があるだろうという仮説に基づく。

検索エンジンを利用して実世界の関係を推定するためには、一つは、Web 空間における関係の強さをどのように測るかの「関連性の尺度」を定める必要がある。また、測定された Web 空間の強さがどのような条件を満たせば、実世界にも関係があると推定するかという「関連性の判断基準」が必要になる。この 2 つの指標を用いて、本研究で、ネットワークの抽出モデル(図 4)を、下記の数式に一般化する。

$$f(\mathbb{S}_r(x, y), \Theta) \rightarrow \{0, 1\} \quad (1)$$

ここで、 $\mathbb{S}_r(x, y)$ は、関連性の尺度であり、エンティティ x と y の関係 r における関係の強さのさまざまな指標を、 m 次元のベクトル空間 ($S_r^{(1)}(x, y), S_r^{(2)}(x, y), \dots, S_r^{(m)}(x, y)$) で表す。例えば、 $S_r^{(i)}(x, y)$ は、共起頻度や、Jaccard 係数、或いは Overlap 係数などの指標で計算することができる。さらに、複数の指標を組み合わせたスコア関数であることも、関係を記述する文脈 D のスコアであることもある。そして、 Θ は関連性の判断基準を表し、 n 次元のベクトル空間 ($\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}$) のパラメータを用いて関連性を判別するルールを表すことができる。例えば、 Θ は、 T, M との組み合わせで不均一なコミュニティにおける関係抽出を行う。

関数 f は、これらの複数の指標とパラメータに基づいて、ネットワーク上に x と y の間にエッジでつながりかどうかを決める 2 値関数 $\{1, 0\}$ である。 f 関数の内部構造は、エンティティの種類や紐帯を示す関係、コミュニティの特徴により異なる。すなわち、エンティティの種類や抽出したネットワークの種類によって関連性の尺度と関連性の判断基準が異なったりする。そのため、オフラインモジュールが必要になり、訓練データを基に、適切な指標とパラメータの値を定めておく必要がある。

前章で説明した研究者ネットワーク、アーティストのネットワーク、企業関係のネットワークの 3 つの抽出手法を、汎用モデルにより一般化すると、それぞれ下記のようになる。

$$\begin{cases} \mathbb{S}_r(x, y) = GoogleCooc(x, y) \\ \Theta = T \end{cases} \quad (2)$$

$$\begin{cases} \mathbb{S}_r(x, y) = \text{GoogleCooc}(x, y) \\ \Theta = (T, M) \end{cases} \quad (3)$$

$$\begin{cases} \mathbb{S}_r(x, y) = \sum_{w \in (x, y)}^{W(w)}(x, y) \\ \Theta = \text{score}_{\text{thre}} \end{cases} \quad (4)$$

検索エンジンを利用して Web 上から社会ネットワークを抽出するシステムでは、いかに効率的に Web 上から社会ネットワークを抽出することが重要になる。基本的には、検索クエリの生成、Google 検索、ネットワークの構築という3つのプロセスにより Web 上から社会ネットワークを抽出する。まず、検索クエリの生成段階では、入力されたエンティティを用いて検索クエリ x AND y を生成する。特定の関係に絞るときには、その関係の関係語をクエリに加える。複数のクエリを用いて複数回の検索を行うことで、Web 記述の多様性に対応することができる。エンティティのあいまい性解消はクエリの生成段階において重要であり、[Aswani 06, Bekkerman 05] などの研究を参照することができる。そして、Google 検索段階では、生成されたクエリを検索エンジン（例えば、Google）に入力する。検索結果からヒット件数と上位の Web ページ（或いは、スニペット） D 取得する。ヒット件数の統計的な結果と、Web ページのコンテンツを用いることで情報の抽出がより効果的になり得る。最後に、定めた指標とパラメータにより関数 f の値を求めて（1 の場合エッジでつなぐ）ネットワークを構築する。パラメータの値を変えることで、動的にネットワークを抽出することができる。

4. まとめ

本稿では、検索エンジンを利用して Web 上から研究者、アーティスト、企業などのエンティティの社会ネットワークの抽出する手法をそれぞれの示すとともに、さまざまなエンティティの社会ネットワークを抽出する汎用モデルを提案した。今後は、より多様なエンティティ同士の社会ネットワークを Web 上から抽出する手法を検討し、汎用モデルを拡張していく。そして、ネットワークを抽出するルールを訓練データなしで自動的に学習する手法について検討する。

参考文献

- [Adamic 03] Adamic, L. and Adar, E.: Friends and Neighbors on the Web, *Social Networks*, Vol. 25, No. 3, pp. 211–230 (2003)
- [Aleman-Meza 06] Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Sheth, A., Arpinar, I., Ding, L., Kolari, P., Joshi, A., and Finin, T.: Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection, in *Proc. WWW2006* (2006)
- [Aswani 06] Aswani, N., Bontcheva, K., and Cunningham, H.: Mining Information for Instance Unification, in *Proc. ISWC2006* (2006)
- [Bekkerman 05] Bekkerman, R. and McCallum, A.: Disambiguating Web Appearances of People in a Social Network, in *Proc. WWW2005* (2005)
- [Chen 99] Chen, C.: visualising semantic spaces and author co-citation networks in digital libraries, *Inf.Process.Manage.*, Vol. 35(3), pp. 401–420 (1999)
- [Finin 05] Finin, T., Ding, L., Zhou, L., and Joshi, A.: Social Networking on the Semantic Web, *The Learning Organization*, Vol. 12, No. 5, pp. 418–435 (2005)
- [Golbeck 06] Golbeck, J. and Parsia, B.: Trust network-based filtering of aggregated claims, *International Journal of Metadata, Semantics and Ontologies* (2006)
- [稲岡 03] 稲岡 創, 二宮 拓人, 清水 季子, 高安 秀樹: 金融機関の資金取引ネットワーク, Technical Report ワーキングペーパー 2003-J-2, 日本銀行金融市場局 (2003)
- [Jin 07] Jin, Y., Matsuo, Y., and Ishizuka, M.: Extracting Social Networks Among Various Entities on the Web, in *ESWC2007 (Proc. 4th European Semantic Web Conf., Innsbruck, Austria), LNCS 4519, pp.251–266, Springer* (2007)
- [Kautz 97] Kautz, H., Selman, B., and Shah, M.: The Hidden Web, *AI magazine*, Vol. 18, No. 2, pp. 27–35 (1997)
- [Leskovec 05] Leskovec, J., Adamic, L. A., and Huberman, B. A.: The Dynamics of Viral Marketing (2005), <http://www.hpl.hp.com/research/idl/papers/viral/viral.pdf>
- [Manning 02] Manning, C. D. and Schütze, H.: *Foundations of statistical natural language processing*, The MIT Press, London (2002)
- [松尾 05] 松尾 豊, 友部 博教, 橋田 浩一, 石塚 満: Web 上の情報からの人間関係ネットワークの抽出, *人工知能学会論文誌*, Vol. 20, No. 1E, pp. 46–56 (2005)
- [Matsuo 06] Matsuo, Y., Mori, J., Hamasaki, M., Takeda, H., Nishimura, T., Hasida, K., and Ishizuka, M.: POLYPHONET: An advanced social network extraction system, in *Proc. WWW2006* (2006)
- [McCarthy 04] McCarthy, J., McDonald, D., Soroczak, S., Nguyen, D., and Rashid, A.: Augmenting the Social Space of an Academic Conference, in *Proc. CSCW2004* (2004)
- [Mika 05] Mika, P.: Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks, *Journal of Web Semantics*, Vol. 3, No. 2 (2005)
- [Miki 05] Miki, T., Nomura, S., and Ishida, T.: Semantic Web Link Analysis to Discover social Relationship in academic communities, in *Proc. SAINT 2005* (2005)
- [Mori 05] Mori, J., Ishizuka, M., Sugiyama, T., and Matsuo, Y.: Real-world Oriented Information Sharing Using Social Networks, in *Proc. ACM GROUP2005* (2005)
- [Newman 04] Newman, M. E. J. and Girvan, M.: Finding and evaluating community structure in networks, *Physical Review E*, Vol. 69, p. 026113 (2004)
- [相馬 05] 相馬 亘: 経済における複雑系ネットワーク –日本の経済ネットワークは特殊か?–, *人工知能学会誌特集*, Vol. 20, No. 3, pp. 289–295 (2005)
- [Tyler 03] Tyler, J., Wikinson, D., and Huberman, B.: *Email as spectroscopy: automated discovery of community structure within organizations*, pp. 81–96, Kluwer, B.V. (2003)
- [Wasserman 94] Wasserman, S. and Faust, K.: *Social network analysis. Methods and Applications*, Cambridge University Press, Cambridge (1994)
- [安田 97] 安田 雪: 社会ネットワーク分析 –何が行為を決定するか–, 新曜社 (1997)
- [金 07] 金 英子, 松尾 豊, 石塚 満: Web 上の情報を用いた企業間関係の抽出, *人工知能学会論文誌*, Vol. 22, No. 1, pp. 48–57 (2007)
- [金光 03] 金光 淳: 社会ネットワーク分析の基礎 –社会的関係資本論にむけて–, 勁草書店 (2003)
- [原田 03] 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*, 第 DBS-130/FI-71 巻 (2003)
- [古川 05] 古川 忠延, 松澤 智史, 松尾 豊, 内山 幸樹, 武田 正之: Weblog におけるユーザの繋がりと閲覧行動の分析, *電子情報通信学会論文誌*, Vol. J88-B, No. 7, pp. 1258–1266 (2005)