

# Yahoo!掲示板が盛り上がるダイナミズム

## The Dynamism of Yahoo!Japan Message Boards

松村 真宏\*1

Naohiro matsumura

\*1大阪大学大学院経済学研究科

Graduate School of Economics, Osaka University

In this paper, I first present an approach to extract directed and weighted social networks from message boards on the Internet. Then, based on the structural features named communication gaps, I cluster 3,000 social networks extracted from 3,000 message boards in Yahoo!Japan Message Boards to prove the communication types. I also identify the roles of participants from the social networks, and verify the same communication types. These results imply the existence of dominant communication types underneath Yahoo!Japan Message Boards.

### 1. はじめに

Yahoo!Japan 掲示板（以下では Y!掲示板と略す）は日本最大のポータルサイトである Yahoo!Japan 内に設けられた電子掲示板サービスである。膨大な数の掲示板が設置され、様々な話題について活発な意見交換・議論が行われている。Y!掲示板では投稿されたメッセージは一般に公開されており誰でも読むことができるが、メッセージを投稿する際にはあらかじめ取得した ID を使ってログインする必要がある。しかし、投稿したメッセージには投稿者の ID が付与されるので、誰が投稿したのかが ID レベルで分かるようになっており、この仕組みによって相手を意識したコミュニケーションが実現されている。コミュニケーションは高度に社会的な行動であり、そこには常に他者との関わりが反映されている。したがって、投稿者を識別できるかどうかはコミュニケーション行動に大きな影響をもたらす。2ちゃんねる等に代表される匿名性の高い電子掲示板では、参加者が投稿者を識別することがほぼできないので、他者との関係性よりも議論傾向がコミュニケーションの大きな要因となっている [松村 04] が、同様の傾向が Y!掲示板にも当てはまる保障はない。

一般にコミュニケーション様式が異なればそこで繰り広げられる人々の振る舞いも異なることが様々な研究によって指摘されている [Wallace 1999]。そこで本稿では、Y!掲示板に参加する人たちが他の人たちとどのような関わりを持ちながらコミュニケーションを行っているのかを明らかにすることによって、識別性のある電子掲示板上で繰り広げられるコミュニケーションの性質の一端を明らかにすることを目指す。ここでは特に、会員間のコミュニケーションの距離的な隔たりと会員の役割の 2 つの視点に着目する。

### 2. コミュニケーションにおける影響力

コミュニケーションにおける対人距離は、物理的な距離ではなく心理的な距離で表される。また、対人関係は活動内容や社会的地位などさまざまな要因によって対等ではない。そこで、そのような非対称な対人関係を正確に反映するために、Y!掲示板におけるコミュニケーションを、ノードを人、リンクの向きをコミュニケーションの方向、リンクの重みを対人距離とする重み付き有向ネットワークで表現することを考える。

本稿では、コミュニケーションにおいて会員間に流れる発言の影響力の大きさによって対人距離を定義する。影響力はメッセージ間を伝播する語に基づいて計測される。基本的なアイデアは IDM [松村 02] と同じであるが、メッセージの影響力、会員の影響力、語の影響力、会員のプロフィール、メッセージ間の影響力、会員間の影響力・被影響力、語間の影響力を求めようにアルゴリズムを改良・拡張している。以下では、本稿で用いるメッセージ間の影響力、会員の影響力・被影響力を求めアルゴリズムの部分だけを示す。

まず、メッセージ  $x$  がメッセージ  $y$  に及ぼす影響力  $i_{x \rightarrow y}$  を

$$i_{x \rightarrow y} = |w_x \cap \dots \cap w_y|, \quad (1)$$

と定義する。ここで  $w_x$  と  $w_y$  はそれぞれ  $x, y$  に含まれる語であり、 $|w_x \cap \dots \cap w_y|$  は  $x$  から他のメッセージを介して  $y$  に伝播する語の数である。つまり、 $x$  から  $y$  に伝播した語の数によって影響力を求める。

$i_{x \rightarrow y}$  が求まれば、会員間の影響力を求めることができる。会員  $p$  が会員  $q$  に及ぼす影響力  $j_{p \rightarrow q}$  を  $p$  のメッセージが  $q$  のメッセージを介して他の人のメッセージに及ぼした影響力の総和と定義すると、 $j_{p \rightarrow q}$  は

$$j_{p \rightarrow q} = \sum_{x \in \alpha} \sum_{z \in \beta} \sum_{y \in \xi_z} i_{x \rightarrow y}. \quad (2)$$

と表される。ここで、 $\alpha$  は会員  $p$  が投稿したメッセージ集合、 $\beta$  は会員  $q$  が投稿した  $p$  への返信メッセージ集合、 $\xi_z$  はメッセージ  $z$  を基点とするメッセージ・チェーンであり、 $j_{p \rightarrow q}$  は  $q$  が  $p$  のメッセージの普及に貢献している程度を表している。直接の返信関係のない会員間の影響力を求めることもできるが、ここでは対人関係における影響の流れを正確に捉えるために、直接の返信関係における影響力を求めている。

会員の影響力、被影響力は  $j_{p \rightarrow q}$  によって求めることができる。 $p$  の影響力を  $I_p^1$ 、被影響力を  $I_p^2$  とし、それぞれを他の会員に及ぼす影響力の総和、他の会員から受ける影響力の総和と定義すると、 $I_p^1, I_p^2$  は

$$I_p^1 = \sum_{q \in \gamma} j_{p \rightarrow q} \quad (3)$$

$$I_p^2 = \sum_{q \in \gamma} j_{q \rightarrow p} \quad (4)$$

と表される。ここで  $\gamma$  は  $p$  以外の会員集合である。

連絡先: 松村真宏, 大阪大学大学院経済学研究科, 〒560-0043 大阪府豊中市待兼山町 1-7, matumura@econ.osaka-u.ac.jp

メッセージ・チェーンは関連する話題の連鎖を表している。提案方法はメッセージ・チェーンを介して伝播する語を影響力として計測するので、同じ語でも異なる文脈に登場した語に対しては影響力を計上しない。つまり、影響力は文脈によって形成される関係を取り出していると言える。

なお、語が伝播する経路(メッセージ・チェーン)には、メッセージの返信関係だけを利用する方法 [松村 02] と、メッセージ間の類似度を計算する方法 [松村 03] の2通りある。Y!掲示板はメッセージの正確な返信関係が得られるので、本稿はメッセージの返信関係のみを考慮する。

### 3. 対人関係の重み付き有向グラフ

成員間の影響力は共有する興味の程度を反映している。そこで、興味が近いほど心理的距離も近くなると考えると、成員間の心理的距離  $d$  は影響力  $i$  の関数  $f(i)$  として表すことができる。影響力が大きいほど心理的距離は近くなると考えられるので、本稿では心理的距離は影響力に反比例するというシンプルなモデル  $f(i) = 1/i$  (但し、 $i \neq 0$ ) を採用する\*1。このモデルにより、影響力の大きさに応じて 0 から 1 の範囲の心理的距離が得られる。影響力が 0 のときの心理的距離は求まらないので、 $i = 0$  のときは  $n$  人の成員を一行に並べたときの列の最大長である  $n - 1$  を与えることにする。まとめると、本稿では影響力と心理的距離の関係を以下のように定義する。

$$d = \begin{cases} f(i) & (i \neq 0) \\ n - 1 & (i = 0) \end{cases} \quad (5)$$

ここで、成員をノード、成員間の心理的距離をリンクとすると、成員間の関係を重み付き有向グラフで表すことができる。

### 4. コミュニケーション・ギャップ指標

最小の対人関係である 2 人の中の心理的距離に注目すると、コミュニケーションは基本的な 3 パターンに分類することができる。

両方向型コミュニケーション：互いに相手への心理的距離が近い場合であり、互いが等しく意見交換・議論をする。

一方型コミュニケーション：一方からの心理的距離は近いが他方からの心理的距離は遠い場合であり、一方が話し手になり他方が聞き手に回る。

疎コミュニケーション：どちらからも心理的距離が遠い場合であり、互いのコミュニケーションが疎で、コミュニケーションとして成立していない。

$n$  人の成員が互いに連絡を取り合うと  $nP_2$  通りのコミュニケーション・チャンネルが必要になり、すべての成員が密にコミュニケーションを取ることは現実的な方法ではない。一方、対人関係における互いの心理的距離の差はその間のコミュニケーションの隔たり、つまりコミュニケーション・ギャップを表しており、ここは改善の余地がある。そこで本稿では、対人関係の重み付き有向グラフから成員間の心理的距離の差をコミュニケーション・ギャップ指標として算出することにより、成員間のコミュニケーションの状態を把握することを考える。つまり、

コミュニケーション・ギャップ指標では、コミュニケーションが一方になっている程度を求める指標となる。

なお、実際に行われるコミュニケーションを考えると、直接の関係よりも知人を介した関係のほうがコミュニケーションを取り易いことも多い。例えば、面識のあまりない新メンバーが旧メンバーと意見を交換する場合には、直接メッセージをやり取りするよりも 2 人のどちらとも仲の良い人を介したほうがコミュニケーションはうまくいくだろう。そこで、成員間の距離を求める際には、その 2 人を結ぶ最短距離を最短心理的距離として定義する。

ここで、成員  $p$  から成員  $q$  への最短心理的距離を  $d_{p \rightarrow q}$  ( $p = q$  のときは  $d_{p \rightarrow q} = 0$ )、成員の集合を  $\gamma$  とし、3 種のコミュニケーション・ギャップ指標を提案する。

- $G_{diff}$  は成員間の最短心理的距離の差に基づくコミュニケーション・ギャップ指標であり、以下の式で定義される。

$$G_{diff} = \left( \frac{1}{2} \sum_{p \in \gamma} \sum_{q \in \gamma} |d_{p \rightarrow q} - d_{q \rightarrow p}| \right) \left( \frac{1}{nC_2(n-1)} \right) \quad (6)$$

ここで  $|d_{p \rightarrow q} - d_{q \rightarrow p}|$  は  $(d_{p \rightarrow q} - d_{q \rightarrow p})$  の絶対値である。また、 $nC_2(n-1)$  は右辺第 1 項の理論的最大値であり、 $G_{diff}$  の正規化に用いられる。

- $G_{max}$  は成員間の最短心理的距離の大きいほうをコミュニケーションを妨げるボトルネックと考えるコミュニケーション・ギャップ指標であり、以下の式で定義される。

$$G_{max} = \left( \frac{1}{2} \sum_{p \in \gamma} \sum_{q \in \gamma} \max(d_{p \rightarrow q}, d_{q \rightarrow p}) \right) \left( \frac{1}{nC_2(n-1)} \right) \quad (7)$$

ここで  $\max(d_{p \rightarrow q}, d_{q \rightarrow p})$  は  $\{d_{p \rightarrow q}, d_{q \rightarrow p}\}$  の大きいほうを返す。また  $nC_2(n-1)$  は右辺第 1 項の理論的最大値であり、 $G_{max}$  の正規化に用いられる。

- $G_{min}$  は成員間の最短心理的距離の小さいほうがコミュニケーションを妨げるボトルネックになると考えるコミュニケーション・ギャップ指標であり、以下の式で定義される。

$$G_{min} = \left( \frac{1}{2} \sum_{p \in \gamma} \sum_{q \in \gamma} \min(d_{p \rightarrow q}, d_{q \rightarrow p}) \right) \left( \frac{1}{nC_2(n-1)} \right) \quad (8)$$

ここで  $\min(d_{p \rightarrow q}, d_{q \rightarrow p})$  は  $\{d_{p \rightarrow q}, d_{q \rightarrow p}\}$  の小さいほうを返す。また  $nC_2(n-1)$  は右辺第 1 項の理論的最大値であり、 $G_{min}$  の正規化に用いられる。

上記では成員の組ごとに最短心理的距離を求めてコミュニケーション・ギャップ指標を定義したが、ある成員から他の全成員への距離を表す中心性指標を使うこともできる [Freeman 1978]。そこで、中心性指標を利用した 2 種のコミュニケーション・ギャップ指標も提案する。

- $C_{diff}$  は内向きの中心性指標と外向きの中心性指標との差に基づくコミュニケーション・ギャップ指標であり、以下の式で定義される。

$$C_{diff} = \left( \sum_{p \in \gamma} |c_p^{in} - c_p^{out}| \right) \left( \frac{1}{n(n-1)^2} \right) \quad (9)$$

\*1 もちろん、 $f(i)$  には他のモデルも考えられるが、本稿では影響力を心理的距離に変換する妥当な求め方が本題ではないので、これ以上深入りはしないでおく。

表 1: Y!掲示板におけるコミュニケーション・ギャップ指標 .

Categories	$C'_{dist}$	$C'_{diff}$	$G'_{diff}$	$G'_{max}$	$G'_{min}$
家庭と住まい	0.032	0.028	0.032	0.048	0.017
健康と医学	0.065	0.065	0.073	0.100	0.029
芸術と人文	0.068	0.070	0.081	0.108	0.029
科学	0.072	0.068	0.078	0.110	0.034
生活と文化	0.085	0.061	0.071	0.120	0.051
出会い	0.081	0.079	0.089	0.125	0.038
ホビー	0.095	0.092	0.106	0.146	0.043
地域情報	0.161	0.120	0.142	0.230	0.093
エンターテイメント	0.151	0.129	0.157	0.228	0.075
政治	0.217	0.167	0.197	0.313	0.120
ビジネスと経済	0.241	0.160	0.184	0.331	0.150
学校と教育	0.253	0.161	0.195	0.349	0.158
スポーツ・レジャー	0.239	0.208	0.253	0.362	0.116
コンピュータとインターネット	0.447	0.221	0.272	0.579	0.315
ニュース	0.455	0.220	0.271	0.588	0.322

\*\*\*\*\*HIERARCHICAL CLUSTER ANALYSIS\*\*\*\*\*

Dendrogram using Average Linkage (Between Groups)

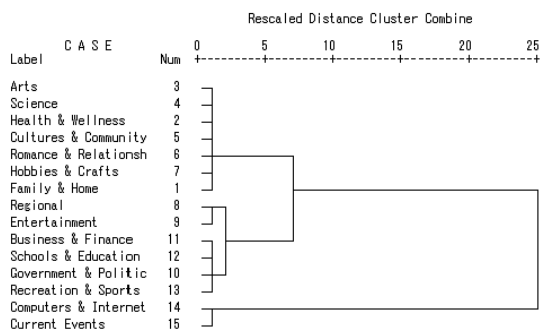


図 1: 表 1 のデータに対する階層的クラスタリング結果 .

ここで  $c_p^{in}$  は成員  $p$  の内向きの中心性指標であり、他の成員から  $p$  への最短心理的距離の和によって表される。また  $c_p^{out}$  は外向きの中心性指標であり、 $p$  から他の成員への最短心理的距離の和によって表される。また、 $n(n-1)^2$  は右辺第 1 項の理論的的最大値であり、 $C_{diff}$  の正規化に用いられる。

- $C_{dist}$  は外向きの中心性に基づくコミュニケーション・ギャップ指標であり、以下の式で定義される。

$$C_{dist} = \left( \sum_{p \in \gamma} c_p^{out} \right) \left( \frac{1}{n(n-1)^2} \right) \quad (10)$$

なお、外向きの中心性の代わりに内向きの中心性を用いても  $C_{dist}$  は同じ値になる。

### 5. コミュニケーションの 3 分類

前章で定義したコミュニケーション・ギャップ指標を用いて、Y!掲示板から獲得した 3,000 個の重み付き有向グラフの分析を行った。分析の手順を以下に示す。

Step 1. Y!掲示板の 15 カテゴリから各 200、合計 3,000 の電子掲示板を対象として、各掲示板から 300 メッセージをダウンロードした。前処理としてメッセージ本文を JUMAN [黒橋 03] を用いて形態素解析し、名詞、形容詞、動詞以外の語の削除、および人手で作成した不要語リストを用いて話題と関係のない語の削除を行った。

Step 2. Step 1 で得られた各掲示板に 2. の手法を適用し、合計 3,000 の重み付き有向グラフを取り出した。なお、グラフのノード数を揃えるために、各掲示板ごとに影響力(式(3))の上位 10 名からなるグラフを作成した。

Step 3. Step 2. で得られた 3,000 の重み付き有向グラフから 5 種のコミュニケーション・ギャップ指標  $G_{diff}$ ,  $G_{max}$ ,  $G_{min}$ ,  $C_{dist}$ ,  $C_{diff}$  を算出した。各指標のカテゴリごとの平均を表 1 に示す。

ここで、カテゴリとコミュニケーション・ギャップ指標との関係を探るために、表 1 のデータに対して平方ユークリッド距離に基づく階層的クラスタリング [Everitt 01] による分析を行った。得られたデンドログラムを図 1 に示す。図 1 より、特徴的なコミュニケーション傾向を持つ 3 つの主要なクラスタの存在が明らかになった。それぞれのクラスタを以下に示す。

対話型コミュニケーション：このクラスタは 7 つのカテゴリ（芸術と人文、科学、健康と医学、生活と文化、出会い、ホビー、家庭と住まい）からなり、コミュニケーション・ギャップ指標はいずれも小さい。これらのカテゴリでは日常的に経験する身近な話題を扱っている。成員は誰でも自然に議論に参加することができ、活発なコミュニケーションが繰り返されている。

分散型コミュニケーション：このクラスタは 6 つのカテゴリ（地域情報、エンターテイメント、ビジネスと経済、学校と教育、政治、スポーツ・レジャー）からなり、コミュニケーション・ギャップ指標は対話型コミュニケーションより大きい。これらのカテゴリで扱われている話題は比較的専門的なため、知識をもつ成員が中心となってコミュニケーションが形成されている。

一方向型コミュニケーション：このクラスタは 2 つのカテゴリ（コンピュータとインターネット、ニュース）からなり、コミュニケーション・ギャップ指標は 3 つのクラスタの中で最も大きい。ここで対象となる話題は最新情報であり、一部の人たちが流す情報にその他大勢が群がっている。情報を得ることが主目的であり、それについて議論することはあまりしない。

以上の分析より、Y!掲示板は大きく 3 タイプのコミュニケーションがあることが明らかになった。なお、ここで用いた 5 種のコミュニケーション・ギャップ指標間のピアソンの積率相関係数は 0.9 以上であり、どの指標も同じ振る舞いを示している。したがって、どれか 1 つの指標を用いるだけで Y!掲示板の分類は可能である。

### 6. 成員の役割

次に、コミュニケーションにおける成員の役割に注目する。一般に効率的かつスムーズなコミュニケーションは成員の相互依存構造によって実現され、成員間の階層的関係や成員の役割が重要な役割を果たす。組織の場合だとそのような構造や役割は明示的に与えられるが、同様の現象は Y!掲示板においても自然発生している可能性がある。そこで本章では、そのような構造が Y!掲示板においても観察されるのか、また存在するとしたらどのような形態を成しているのかを調べる。

そこでまず、影響力、被影響力によって、成員の 4 つの役割（リーダー、専門家、調整役、追従者）を定義する。

表 2: Y!掲示板における成員の役割ごとのメッセージ数の割合 .

Categories	Leader	Maven	Coordinator	Follower
家庭と住まい	0.5263	0.0289	0.0426	0.4022
健康と医学	0.5137	0.0453	0.0605	0.3804
芸術と人文	0.5331	0.0312	0.0438	0.3920
科学	0.5337	0.0372	0.0505	0.3785
生活と文化	0.5752	0.0415	0.0428	0.3405
出会い	0.5277	0.0450	0.0533	0.3741
ホビー	0.5158	0.0293	0.0496	0.4054
地域情報	0.5088	0.0407	0.0595	0.3910
エンターテイメント	0.4844	0.0275	0.0540	0.4341
政治	0.4291	0.0224	0.0558	0.4927
ビジネスと経済	0.4506	0.0365	0.0596	0.4534
学校と教育	0.4212	0.0348	0.0591	0.4849
スポーツ・レジャー	0.4101	0.0420	0.0626	0.4852
コンピュータとインターネット	0.3611	0.0289	0.0691	0.5409
ニュース	0.2931	0.0303	0.1010	0.5756

\*\*\*\*\*HIERARCHICAL CLUSTER ANALYSIS\*\*\*\*\*

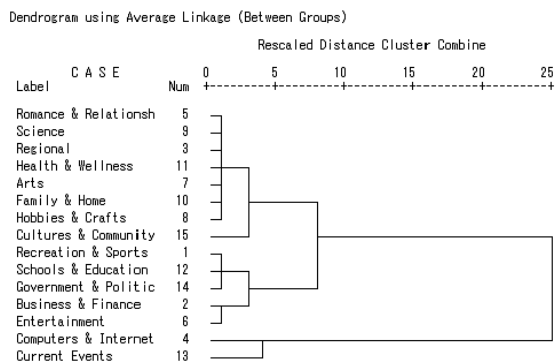


図 2: 表 2 のデータに対する階層的クラスタリング結果 .

**リーダー:** 率先して情報や意見を出すことによってコミュニケーションの主導権を握るだけでなく、他の成員の意見にも耳を傾ける。高い  $I^1$  と  $I^2$  を持つ。

**専門家:** 他の成員の意見や情報に耳を傾けるより、積極的に専門的な情報や意見を提供する。高い  $I^1$  を持つ。

**調整役:** 他の成員の出した情報や意見を受けとめて反応を返すことによって、協調的な雰囲気を作り、コミュニケーションを円滑に行うための調整を行う。高い  $I^2$  を持つ。

**追従者:** 情報や意見を探するためにコミュニケーションに潜り込んでおり、自分から何かの活動を行うことはしない。 $I^1$ ,  $I^2$  はどちらも低い。

### 7. 再びコミュニケーションの3分類

5. で Y!掲示板でのコミュニケーションが3分類できることを示したが、これらは成員の役割とも関わっている可能性がある。そこで Y!掲示板のデータに対して  $I^1$ ,  $I^2$  を求め、掲示板ごとに成員の役割を同定した。ここでは  $I^1$  が平均値の2倍より高ければ“High”, それ以外を“Low”として分析を行った。合計 119,957 人の成員の役割を同定し、うち 17% が  $I^1$  もしくは  $I^2$  が“High”であった。また、彼らの投稿したメッセージは全体の 62% を占めた。役割ごとの投稿メッセージ数の割合をカテゴリごとにまとめた結果を表 2 に示す。

ここで、成員の役割とカテゴリにどのような関係があるのかを調べるために、表 2 のデータに対して 5. と同じく平方

ユークリッド距離に基づく階層的クラスタリングを行った。得られたデンドログラムを図 2 に示す。図 2 より、再び3つの主要なクラスタの存在が明らかになった。それぞれのクラスタを以下にまとめる。

**リーダー主導型コミュニケーション** このクラスタは8つのカテゴリ(家庭と住まい, 健康と医学, 芸術と人文, 科学, 生活と文化, 出会い, ホビー, 地域情報)からなり、リーダーが率先して情報発信していることが分かる。

**リーダー・追従者協働型コミュニケーション** このクラスは5つのカテゴリ(エンターテイメント, 政治, ビジネスと経済, 学校と教育, スポーツ・レジャー)からなり、リーダーと追従者がバランスよくメッセージを投稿している。

**追従者主導型コミュニケーション** このクラスは2つのカテゴリ(コンピュータとインターネット, ニュース)からなり、追従者によるメッセージが多い。

興味深いことに、上記の3分類に含まれるカテゴリは5. で得られた3分類とほとんど同じであった(地域情報だけが異なる)。つまり、対話型コミュニケーションとリーダー主導型コミュニケーション, 分散型コミュニケーションとリーダー・追従者協働型コミュニケーション, 一方向型コミュニケーションと追従者主導型コミュニケーションはそれぞれ関係があり、Y!掲示板の盛り上がる基本軸となっていることが明らかになった。

### 8. まとめ

本稿ではコミュニケーション・ギャップ指標および成員の役割という観点から Y!掲示板を分析した。全く異なる切り口からの分析にも関わらずほぼ同一のクラスタが得られたことは Y!掲示板の盛り上がりや根底から支配する要因の存在を示唆しており、大変興味深い。今後はさらに様々な角度からの分析を進め、Y!掲示板が盛り上がるメカニズムを解明していきたい。

### 参考文献

[松村 02] 松村真宏, 大澤幸生, 石塚満: テキストによるコミュニケーションにおける影響の普及モデル, 人工知能学会論文誌 第 17 巻 3 号, pp. 259-267, 2002.

[松村 03] 松村真宏, 加藤優, 大澤幸生, 石塚満: 議論構造の可視化による論点の発見と理解, 日本知能情報ファジィ学会, Vol. 15, No. 5, pp. 554-564, 2003.

[松村 04] 松村真宏, 三浦麻子, 柴内康文, 大澤幸生, 石塚満: 2ちゃんねるが盛り上がるダイナミズム, 情報処理学会 45 巻 3 号, pp. 1053-1061, 2004.

[黒橋 03] 黒橋禎夫, 河原大輔: 日本語形態素解析システム JUMAN version 4.0 マニュアル (2003)

[Everitt 01] Brain Everitt, Sabine Landau, and Morven Leese: Cluster Analysis, Hodder & Stoughton Educational, 2001.

[Freeman 1978] Freeman, L.C. (1978) Centrality in Social Networks, *Social Networks*, 1, 215-39.

[Wallace 1999] Wallace, Patricia (1999) The Psychology of the Internet, Cambridge University Press.