

サイバーコミュニティ形成支援システムのネットワーク分析による評価

藤田邦彦[†] 亀井剛次[†] Eva Jettmar[‡] 吉田仙[†] 桑原和宏[†]

[†]NTT コミュニケーション科学基礎研究所 [‡]スタンフォード大学

[†]{fujita|kamei|yoshida|kuwabara}@cslab.kecl.ntt.co.jp [‡]true@stanford.edu

概要：ネットワーク分析は、集団の特性を定量的に表現する社会科学的分析手法である。本稿では、まずネットワーク分析の概要を紹介し、その中で、ネットワーク分析の適用範囲を拡大する提案を行う。次に、我々が以前に構築した、サイバーコミュニティの発生・形成過程における人々の活動を支援するシステム **Community Organizer** の有効性を確認するために実施された社内実験の結果に、ネットワーク分析を適用した結果を報告する。

Network Analysis of a System Supporting the Formation of Cyber-communities

Kunihiko FUJITA[†], Koji KAMEI[†], Eva JETTMAR[‡], Sen YOSHIDA[†] and Kazuhiro KUWABARA[†]

[†]NTT Communication Science Laboratories [‡]Stanford University

Abstract : Network analysis is a method for discovering the structure of the communications flow between group members. In this paper, we introduce network analysis first. Next, we introduce **Community Organizer** a system designed to support network communities, and report results of an experiment applying network analysis.

1 はじめに

コンピュータネットワークの普及にともない、個人や企業の活動の場は、ネットワーク上にまで広がっている。ネットワーク上では、電子メールやWWWをはじめとして、掲示板やチャットなどさまざまなアプリケーションを媒介してコミュニケーションが行われている。そこに発生する人々のつながりは「サイバー社会」を形成しつつあり、その存在は無視できないものとなっている。

サイバー社会の特徴である、コンピュータが介在するコミュニケーションでは、ネットワークは単なる情報伝送路ではなく、積極的にコミュニケーションをサポートすることが可能である。ここには、人間とコンピュータが混在する新しいコミュニケーション形態が考えられ、サイバー社会には、現実社会を補完する機能を持ち、また現実社会と同様に魅力的なものとして機能することが求められる。そのような人々のネットワーク上での社会活動を支援するソフトウェア基盤として、コミュニティウェアやソーシャルウェアの研究が進められている [1, 2].

一方で、コミュニティウェアやソーシャルウェアは、対象を人間またはその集合体である社会としているため、システムとして評価を行う場合、単にパフォーマンスなどを測定するのみならず、なんらかの社会科学的知見を援用した評価手法が求められてきた [3, 4].

そこで我々は、集団の特性を定量的に表現する社会科学的分析手法であるネットワーク分析 [5] を評価手法として着目している。ここではまず、我々が以前に構築した、サイバーコミュニティの発生・形成過程における人々の活動を支援するシステム **Community Organizer** [6, 7] の、社内での実験結果の評価にネットワーク分析を適用した。分析の結果は、我々の仮説を立証するものであった。

2 ネットワーク分析

ネットワーク分析とは、集団に属するメンバー間のコミュニケーションの構造を見出す分析手法である [5]. 分析単位として対人関係を使い、コミュニケーションの流れやパターンについてのネットワークデータを分析する。ネットワークデータとは、だれとだれ

がどのくらいコミュニケーションをとったのかを数値的に表現した行列データのことである。行列データでは、関係のある2者間には1を、関係のない2者間には0を割り当てて表現する。頻度を考慮する場合は1ではなくその度数を割り当てて表現する [8]。

2.1 ネットワーク分析の指標

ネットワーク分析の指標の主なものとして、中心性・密度・直接結合度の3つが挙げられる。

1. 中心性 (Centrality)

中心性は、あるメンバーが他のメンバーとどれだけ関係を持っているかを示す指標である。この値が最大のメンバーは、他のメンバーと最も多くの関係を持っているといえ、その集団の中心人物である可能性が高い。中心性は下式によって算出される。中心性を計測する基準には様々なものがあるが、一例として各メンバーの全発言ノードに対する入次数と出次数を挙げる。式中の x_{ij} は、入次数を基準とする場合はメンバー j からメンバー i へ向かうリンクの本数であり、出次数を基準とする場合はメンバー i からメンバー j へ向かうリンクの本数である。 N はメンバー数である。

$$Centrality_i = \frac{\sum_{j=1}^N x_{ij}}{\sum_{i=1}^N \sum_{j=1}^N x_{ij}}$$

2. 密度 (Density)

密度は、集団内のメンバー間の関係がどのくらい緊密であるのかを示す指標である。密度が高い集団はメンバー間の関係が緊密であると言え、各メンバーの行動は均一的かつ類似的となる傾向がある。他方、密度が低い集団はメンバー間の関係が疎遠であると言え、各メンバーの行動は多様性に富んだものになる傾向がある。密度は下式によって算出される。式中の z_{ij} は、方向性を考慮してメンバー i からメンバー j の関係が一度でも存在していれば1を、存在していなければ0をとる。

$$Density = \frac{\sum_{i=1}^N \sum_{j=1}^N z_{ij}}{N(N-1)}$$

3. 直接結合度 (Cohesion)

直接結合度は、メンバー間の双方向的な関係がどの程度存在しているかを示す指標である。密度の算出式に似ているが、メンバー間の双方向的な関係があるときのみ分子がカウントされる。直接結合度は下式によって算出される。式中の $z_{ij} \oplus z_{ji}$ は、 $z_{ij} = 1$ は、かつ $z_{ji} = 1$ のときのみ1をとり、それ以外では0をとる。

$$Cohesion = \frac{\sum_{i=1}^N \sum_{j=1}^N (z_{ij} \oplus z_{ji})}{\frac{N(N-1)}{2}}$$

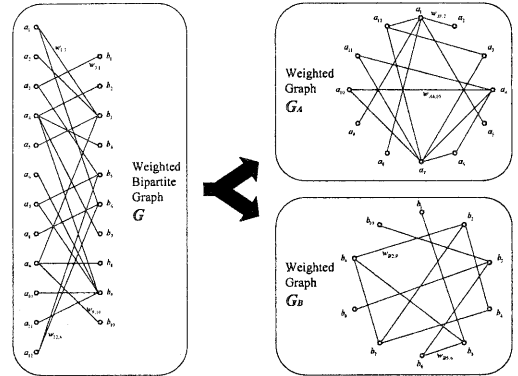


図 1: 2部グラフの分割

2.2 異種要素混在構造への適用

これまでに説明してきたネットワーク分析の対象は、人間という同種の要素のみから構成された、集団というネットワーク構造である。本節では、複数種類の要素から構成されたネットワーク構造に対してのネットワーク分析の手法を提案する。説明のため、分析対象を2種類の要素からなるネットワーク構造とし、そのような構造をとる例としてE-Commerceサイトでの販売実績を挙げる。

昨今では、amazon.comに代表されるインターネットを利用して商品を販売する、いわゆるE-Commerceサイトが増えつつあり、その市場も拡大傾向にある。E-Commerceサイトと、店頭における商品販売の違いとして、1) 購入者を特定できる、2) 購入商品履歴が存在する、3) 購入者の購入商品に対する評価が与えられている、などが挙げられる。これは、図1のような重み付き2部グラフで表現できる。

2部グラフ G とは、 $V(G) = V(A) \cup V(B)$, $V(A) \cap V(B) = \emptyset$ と分け、 G の全ての辺が $a \in V(A)$ と $b \in V(B)$ を結ぶようにできるグラフで、 $V(A)$ と $V(B)$ を部集合という。 $V(A) = \{a_1, a_2, \dots, a_m\}$, $V(B) = \{b_1, b_2, \dots, b_m\}$ とし、 a_i と b_j を結ぶ辺を e_{ij} と表し、 e_{ij} に付けられた重みを w_{ij} とする。簡単のため $w_{ij} > 0$ と仮定する。

E-Commerceサイトの販売実績の例でいうなら、一方の部集合の要素が購入者、他方の部集合の要素が購入商品であり、辺が張られているのは、購入者が当該商品を購入したことを示す。購入者の購入商品に対する評価は、辺に付与された重みとして表される。

ネットワーク分析では、購入者と購入商品という全く性質の異なるものを同時に分析対象にすることはできないので、この2部グラフを、片方の部集合のみか

ら構成される2つのグラフに分割する。分割手法は以下の通りである。

ここで、2部グラフを、片方の部集合のみから構成される2つのグラフに分割するというのは、 G から、 $V(A)$ のみから構成されるグラフ G_A と、 $V(B)$ のみから構成される一般のグラフ G_B に分割する、ということである。定義より $V(G_A) = V(A)$ 、 $V(G_B) = V(B)$ である。 $a_k \in V(G_A)$ 、 $a_l \in V(G_A)$ 、 $a_k \neq a_l$ とするとき、 a_k と a_l の間に張られる辺に付与される重み w_{Akl} は、下式で求められる。

$$w_{Akl} = \sum_{p=1}^m w_{kp} \cdot w_{lp}$$

また、 $w_{Akl} = 0$ のときは a_k と a_l の間に辺を張らない。このようにして G から G_A を分割できる。 G_B についても同様の手順で分割できる。

この手順を踏むことによって得られた、購入者のみから構成された購入者グラフは、同じ商品を購入した者同士の間には辺が張られることになる。これは、同じ商品を購入したもの同士は、興味や関心を共有する度合いが高いだろうという仮定を表現したものになる。よって、全体の購入者グラフから、グラフの分割を行うことにより、いくつかの、COIを抽出することが可能となる。また、購入者グラフにネットワーク分析を施すことにより、中心性の高い者や、孤立している者など、個々の購入者の性格付けが行え、よりきめ細かなマーケティングを行うことが可能となる。

同様に、購入商品のみから構成された購入商品グラフからは、購入の因果関係（例えば、ある商品 α を購入した人は別の商品 β も購入することが多い、という関係）を発見できたり、商品群のカテゴリズが行えるため、商品の販売促進活動に寄与するであろう。

3 Community Organizer 評価実験結果分析

3.1 Community Organizer の概要

Community Organizer の目的は、ネットワーク上のコミュニティの形成を支援することである。ここでいうコミュニティとは COI (community of interest) と呼ばれるもので、興味や関心の共通する人々が集い、対話が行われている状況を想定している。ネットワーク上のコミュニティを支える最も重要な基盤は、人々の間に発生するコミュニケーションである。

パソコン通信における電子会議室や、WWW における掲示板・チャットルームなどのいわゆるオンラインコミュニティサービスと、Community Organizer の最大の違いは、コミュニティの「境界」の有無である。既存のオンラインコミュニティサービス

では、コミュニティに境界が存在するため、コミュニティライフサイクルの初期である形成段階においてユーザは、境界を作らなければならないという心理的なハードルと、新規参入者が境界をまたがなければならないというハードルが存在する。これに対し Community Organizer においては、境界が存在しないため、ユーザの心理的なハードルが低くなり、コミュニティの形成がスムーズに行われる。

3.2 評価実験

本評価実験は対照実験である [6, 7]。Community Organizer のもっとも大きな特徴である、コミュニティの形成の度合いを可視化する部分（ビューワ）だけを、Community Organizer のものを用いた場合（Full version（以下“Full”と称す）：図2）と、一般的な WWW 検索システムで用いられているような、興味・関心の一致度に基づいて列挙するものを用いた場合（Listing version（以下“Listing”と称す）：図3）とで、どのようにユーザ間のコミュニケーション活動に変化が生じるかを観察した。

被験者は、著者らが属する研究所内で募り、20才台から50才台までの42名が集まった。この42名の被験者を2つに分け、最初の5日間は21名の被験者が Full を使い、次の5日間は残りの21名の被験者が Listing を使った。被験者を割り当てる際には、性別や年齢の分布が均等になるようにし、プレテストを実施してメールや WWW の使用頻度なども均等になるようにした。実験中は、各被験者の匿名性を保つために、自身でつけたニックネームを使用し、実験が終了するまでは実験について話すことを禁じた。

3.3 分析・考察

本稿では、評価実験の結果として得られたデータの内の、Community Organizer のコミュニケーションツールの1つである公開メッセージの発言ログをネットワーク分析の対象とする。理由は、Community Organizer にて用意されているコミュニケーションツールの内、最も多く使われたツールが公開メッセージだったこと、ネットワーク分析が行い易いこと、などが挙げられる。

公開メッセージは一般的な掲示板とほぼ同等の機能をもつ。ただし、一般的な掲示板は1つの掲示板内に複数のスレッドが存在することが多いが、公開メッセージは1つの公開メッセージ内には1つのスレッドのみが存在する。ユーザは誰でも、何らかの話題について他のユーザと対話したいとき、公開メッセージを作成することができる。作成された公開メッセージは、作成者のアイコンとともにビューワ（図2,3）の中にアイコンとして表示される。他のユーザは誰で

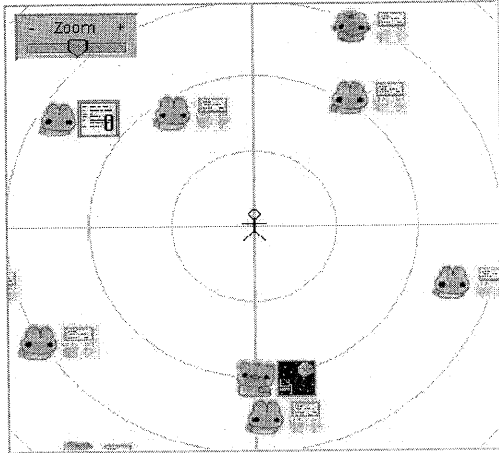


図 2: Full

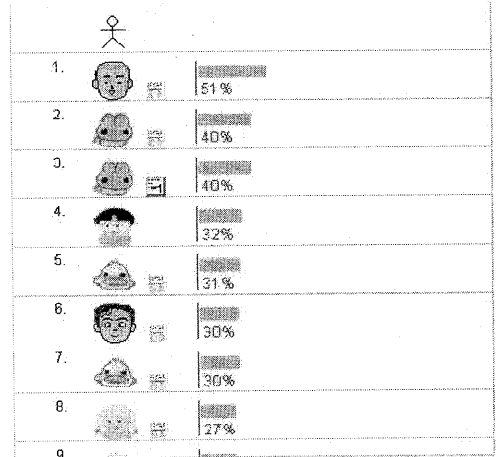


図 3: Listing

も、その公開メッセージ内で対話できる。

以下、ネットワーク分析の対象を、ユーザ間の関係とする場合と、公開メッセージ間の関係とする場合に分けて述べる。

3.3.1 ユーザ間の関係

実験ログからコメントチェーンを作成した結果、Fullは図4、Listingは図5のようになった。図の最上部に横に並んでいるのが各ユーザを識別するIDで、番号付けされた各ノードが発言を示しており、ノードに付与された番号は発言の順番である。破線は日の境界である。有向リンクにおいて、出次にあたるノードは、新規に公開メッセージを作成し、そこで発言したことを意味し、入次にあたるノードは、既存の公開メッセージ内で発言したことを意味している。

このコメントチェーンを元にネットワーク分析を行う。指標として、中心性・密度・直接結合度の3つを用いた。各々の指標について計算した結果が、表1・表2である。

入次数を基準とする中心性の標準偏差は、Fullが0.034に対し、Listingは0.051である。これは、Fullの方が中心性の偏在が少なかったことを意味し、特定の被験者に発言が集中することなく、多くの被験者が発言したことを示す。また、出次数を基準とする中心性の標準偏差は、Fullが0.067に対し、Listingは0.060であり、入次数の結果とは逆転しているが、入次数の結果ほどの差はない。表1によれば、Fullのユーザの内、新規に公開メッセージを作成していない者が6名いるのに対し、Listingでは3名であり、このことが結果に影響したものと考えられる。

密度については、Fullの方がListingより高い値に

なっている(表2)。これは、5日間の実験期間中にFullの方がより多くの被験者間でコミュニケーションが行われたことを意味する。

直接結合度についても、Fullの方がListingより高い値になっており(表2)、被験者間のコミュニケーションが緊密であったことを意味する。

密度と直接結合度については、対象となる集団の形成から時間が経過するにつれ、その値は累積して増加していく。そして、ある一定の時間が経過すると、その値の増加がストップする。そのときの状態はいわば集団内コミュニケーションの飽和状態といえ、その集団にとって必ずしも好ましいものとは言えず、むしろ閉塞状況に陥っていると判断できる。しかし今回の評価実験結果分析では、集団の形成(実験開始)からの5日間を対象としているので、高い値を得たFullの方がより活発なコミュニケーションを行ったと肯定的に判断するのが妥当であろう。

実験後のアンケートの分析結果[6, 7]からは、Fullの方がListingよりも、コミュニティの存在を感じられ、多くの人との出会いを感じられたという結果が得られており、上で示した3つの指標の示すところと一致する。

3.3.2 公開メッセージ間の関係

ユーザ間の関係とは異なり、公開メッセージ間には直接的な関係性は表面化していない。しかし、2.2節で述べた手法を用いれば、その関係性が浮かびあがる。つまり、一方の部集合の要素はユーザであり、他方の部集合の要素は公開メッセージであるような2部グラフを構成し、その2部グラフから、公開メッセー

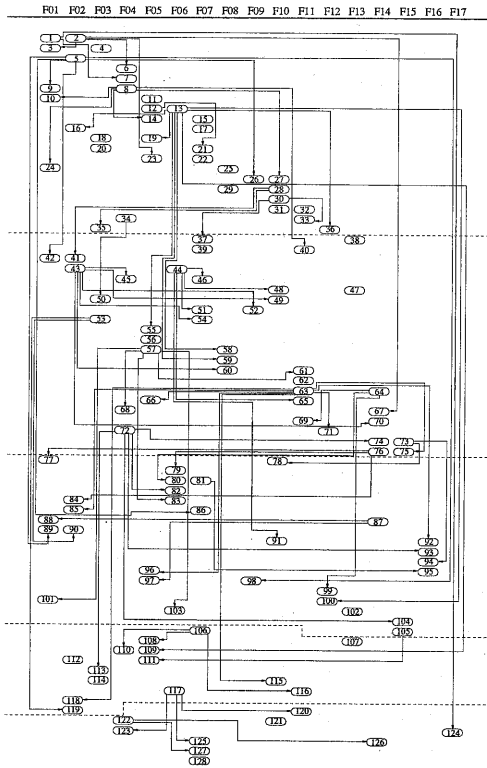


図 4: Full のコメントチェーン

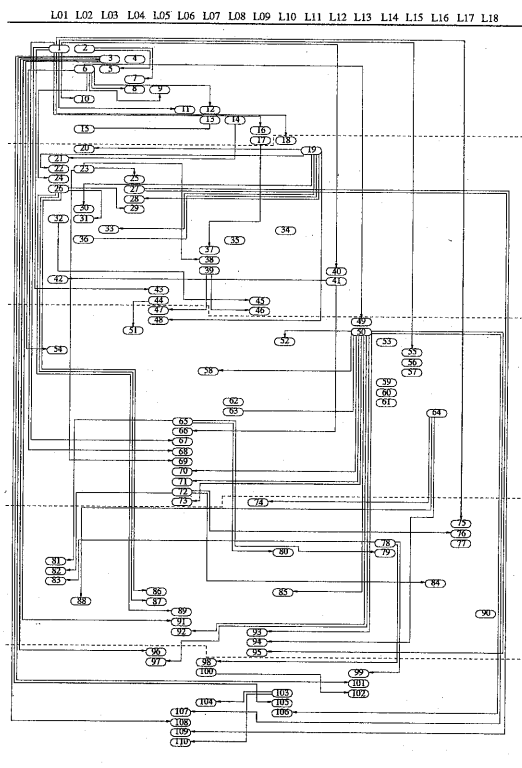


図 5: Listing のコメントチェーン

ジのみからなるグラフを分解することにより、公開メッセージ間の関係を抽出できる。この2部グラフにおいて、辺が張られているのは、ユーザが当該公開メッセージ内で発言したことを示し、辺に付与された重みは、ユーザの当該公開メッセージ内での発言回数であるとする。

注意しなければならないのは、公開メッセージのみからなるグラフのリンクは無向である、ということである。これは、前節のユーザ間の関係は、公開メッセージを作成したユーザから、当該公開メッセージ内で発言したユーザへと有向リンクが張られたのに対し、公開メッセージ間の関係を求めるにあたっては、まず2部グラフで表現しなければならず、この際、作成したユーザと、発言だけをしたユーザの区別をつけることができず、一律に当該公開メッセージ内で発言したものとして扱うからである。このため、ネットワーク分析の指標である中心性・密度・直接結合度の内、直接結合度は密度と同じ値になる。

公開メッセージのうちのいくつかは、作成者だけが発言者がいないようなものもある。それらを除くと、Fullでは公開メッセージは26個存在し、Listing

では公開メッセージは22個存在する。これらの公開メッセージを対象に、中心性と密度を算出した結果が表3、表4である。

中心性の標準偏差は、Fullが0.021に対し、Listingは0.032であった。これは、Fullの方が、中心性の偏在が少なかったことを意味し、特定の公開メッセージに発言が集中することなく、多くの公開メッセージにおいて発言されたことを示す。

密度については、Listingの方がFullより高い値になっている(表4)。この結果の解釈は、中心性の偏在がListingの方がFullより高い値を示していることと合わせて考慮すると、Listingの方が公開メッセージで対話された内容が均一的かつ類似的であったととらえる方が自然であろう。換言すると、Fullの方がより多様な話題が展開されたということである。

4 おわりに

本稿では、サイバーコミュニティの発生・形成過程における人々の活動を支援するシステム **Community Organizer** の社内での実験結果の評価手法として、集団の特性を定量的に表現する社会科学的分析

表 1: 中心性 (ユーザ間の関係)

| Full | 入 | 出 | Listing | 入 | 出 |
|------|-------|-------|---------|-------|-------|
| F01 | 0.110 | 0.024 | L01 | 0.105 | 0.211 |
| F02 | 0.085 | 0.207 | L02 | 0.092 | 0.158 |
| F03 | 0.024 | 0.024 | L03 | 0.026 | 0.053 |
| F04 | 0.098 | 0.146 | L04 | 0.079 | 0.026 |
| F05 | 0.110 | 0.073 | L05 | 0.105 | 0.013 |
| F06 | 0.061 | 0.195 | L06 | 0.197 | 0.079 |
| F07 | 0.098 | 0.049 | L07 | 0.079 | 0.053 |
| F08 | 0.037 | 0.000 | L08 | 0.013 | 0.013 |
| F09 | 0.037 | 0.000 | L09 | 0.092 | 0.013 |
| F10 | 0.073 | 0.049 | L10 | 0.079 | 0.026 |
| F11 | 0.085 | 0.098 | L11 | 0.000 | 0.092 |
| F12 | 0.049 | 0.000 | L12 | 0.013 | 0.026 |
| F13 | 0.000 | 0.000 | L13 | 0.053 | 0.158 |
| F14 | 0.049 | 0.098 | L14 | 0.013 | 0.039 |
| F15 | 0.024 | 0.037 | L15 | 0.013 | 0.000 |
| F16 | 0.049 | 0.000 | L16 | 0.013 | 0.039 |
| F17 | 0.012 | 0.000 | L17 | 0.026 | 0.000 |
| | | | L18 | 0.000 | 0.000 |

表 2: 密度と直接結合度 (ユーザ間の関係)

| | Full | Listing |
|-------|-------|---------|
| 密度 | 0.238 | 0.195 |
| 直接結合度 | 0.092 | 0.052 |

手法であるネットワーク分析を採用した。その結果、**Community Organizer** がコミュニティ内のコミュニケーションを促進する効果があることが確認された。今後は、ネットワーク分析の適用の範囲をさらに広げたい。

謝辞

Community Organizer の実装にあたって御協力頂いた、NTT コミュニケーションズ先端ビジネス開発センターのみなさまに感謝致します。また、実験に御参加頂いたNTT コミュニケーション科学基礎研究所のみなさまに感謝致します。また、2部グラフについて御教授頂いた佐々木淳氏に感謝致します。

参考文献

- [1] Hattori, F., Ohguro, T., Yokoo, M., Matsubara, S. and Yoshida, S.: Multiagent systems for supporting network communities, *CACM*, Vol. 42, No. 3, pp. 55 - 61 (1999).
- [2] Ishida, T.(ed.): *Community Computing*, John Wiley & Sons (1997).

表 3: 中心性 (公開メッセージ間の関係)

| Full | | Listing | |
|---------|-------|---------|-------|
| T_f01 | 0.033 | T_l01 | 0.041 |
| T_f02 | 0.024 | T_l02 | 0.051 |
| T_f03 | 0.013 | T_l03 | 0.017 |
| T_f04 | 0.017 | T_l04 | 0.034 |
| T_f05 | 0.068 | T_l05 | 0.092 |
| T_f06 | 0.017 | T_l06 | 0.048 |
| T_f07 | 0.050 | T_l07 | 0.022 |
| T_f08 | 0.042 | T_l08 | 0.026 |
| T_f09 | 0.024 | T_l09 | 0.015 |
| T_f10 | 0.040 | T_l10 | 0.028 |
| T_f11 | 0.027 | T_l11 | 0.065 |
| T_f12 | 0.012 | T_l12 | 0.034 |
| T_f13 | 0.050 | T_l13 | 0.124 |
| T_f14 | 0.004 | T_l14 | 0.068 |
| T_f15 | 0.004 | T_l15 | 0.021 |
| T_f16 | 0.054 | T_l16 | 0.029 |
| T_f17 | 0.043 | T_l17 | 0.016 |
| T_f18 | 0.044 | T_l18 | 0.015 |
| T_f19 | 0.030 | T_l19 | 0.041 |
| T_f20 | 0.065 | T_l20 | 0.034 |
| T_f21 | 0.077 | T_l21 | 0.055 |
| T_f22 | 0.064 | T_l22 | 0.125 |
| T_f23 | 0.076 | | |
| T_f24 | 0.063 | | |
| T_f25 | 0.031 | | |
| T_f26 | 0.026 | | |

表 4: 密度 (公開メッセージ間の関係)

| | Full | Listing |
|----|-------|---------|
| 密度 | 0.600 | 0.710 |

- [3] NIFTY ネットワークコミュニティ研究会, 金子郁容, 松岡正剛, 中村雄二郎, 岡田智雄: 電線交響主義—ネットワークコミュニティの出現, NTT 出版 (1997).
- [4] 池田謙一 (編): ネットワーキング・コミュニティ, 東京大学出版会 (1997).
- [5] 安田雪 (編): ネットワーク分析, 新曜社 (1997).
- [6] 亀井剛次, Jettmar, E., 藤田邦彦, 吉田仙, 桑原和宏: ネットワーク上のコミュニティ形成を支援するシステム “Community Organizer” の実装と評価実験, 情報処理学会研究報告, Vol. 2000, No. 26, pp. 43-48 (2000).
- [7] Kamei, K., Jettmar, E., Fujita, K., Yoshida, S. and Kuwabara, K.: Community Organizer: Supporting the Formation of Network Communities through Spatial Representation, *SAINT 2001*, San Diego, IEEE Computer Society, pp. 207-214 (2001).
- [8] 高橋正道, 北山聡, 金子郁容: ネットワーク・コミュニティにおける組織アウェアネスの計量と可視化, 情報処理学会論文誌, Vol. 40, No. 11, pp. 3988 - 3999 (1999).