

異なる通信サービスのデータを相補的に利用した 社会ネットワークの構造分析

会田 雅樹^{†a)} 高野 知佐^{††} 小頭 秀行^{†††} 中村 元^{†††}

Structures and Dynamics of Social Networks Revealed by Complementary Use of
Data Analyses of Different Communication Services

Masaki AIDA^{†a)}, Chisa TAKANO^{††}, Hideyuki KOTO^{†††}, and Hajime NAKAMURA^{†††}

あらまし 通信ネットワークに関する各種データ(トラフィック量, ユーザ数等)には, 何らかの形で人間社会の活動の特性が反映されていると考えることができる. 我々はこれまで, 情報通信サービスのデータに現れるべき乗則を利用して, 人間の情報交換関係に関する社会ネットワークの構造を調べてきた. 結果として得られた社会ネットワークの構造は, 分析のもとになった特定の情報通信サービスのデータに関する表面的なサービス種別に左右されるものではなく, 背後にある普遍的な社会ネットワーク構造を記述していることが望ましい. これを確認するためには, 特定のデータ分析から得られる社会ネットワーク構造の分析結果を, 他の通信サービスのデータによって検証することが有効であると考えられる. 本論文では, 複数の情報通信サービスのデータを相補的に用いて, データの自己無撞着性から許される社会ネットワークの構造やユーザ行動特性の分析を行う. また, 得られた次数分布特性が現実のデータとよく一致することを確認する.

キーワード トラフィック, スケールフリーネットワーク, ソーシャルネットワーク, 携帯電話

1. ま え が き

通信ネットワークに関する各種データ(トラフィック量, ユーザ数等)は, これまで主に通信ネットワークの需要予測や設備設計に利用されてきた. これはデータの「量的」な側面に注目した利用法であるといえる. 一方, 通信ネットワークに関する各種データには人間社会の活動の特性が何らかの形で反映されていると考えることができる. 本論文は, データからそのような社会的情報を抜き出し, 通信サービスの背後にある社会ネットワークの構造を知ること, 通信ネットワー

クデータの新たな「質的」な利用法を開発するための研究である.

近年, 社会ネットワークを含む様々な複雑ネットワークに対して, ネットワークの性質を調査する研究が行われてきた [1] ~ [8]. 多くの複雑ネットワークは, 人為的に設計されたわけではなく, ネットワーク構成要素の自律的な活動の結果として生じたものであるにもかかわらず, スケールフリーネットワークなどの共通の性質が見出されている. ここでスケールフリーネットワークとは, ネットワークの次数分布 $p(k)$ が以下のべき乗則に従うことをいう [1], [2].

$$p(k) = O(k^{-\gamma}) \quad (1)$$

ここで k はノード次数, $p(k)$ はノード次数 k をもつノードの存在確率である. また $\gamma > 0$ は定数である. 多くの複雑ネットワークでは, $2.0 \leq \gamma \leq 3.4$ となることが報告されている [2]. 特に, 通信ネットワークに関するデータを用いた分析としては, 大学内の電子メールの分析 [3], 電話トラフィックの分析 [4], [5], 携帯電話トラフィックの分析 [6] ~ [8] が挙げられる. これらの研

[†] 首都大学東京大学院システムデザイン研究科, 日野市
Graduate School of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino-shi, 191-0065 Japan

^{††} 広島市立大学大学院情報科学研究科, 広島市
Graduate School of Information Sciences, Hiroshima City University, 3-4-1 Ozuka-higashi, Asa-minami, Hiroshima-shi, 731-3194 Japan

^{†††} (株) KDDI 研究所, ぶじみ野市
KDDI R&D Laboratories Inc., 2-1-15 Ohara, Fujimino-shi, 356-8502 Japan

a) E-mail: maida@sd.tmu.ac.jp

究は、特定の通信サービスのデータを分析することで、そのサービス利用者のネットワーク構造を調べたものである。本論文においても通信サービスのデータを利用した分析を行うが、その目的は特定の通信サービス利用者のネットワークを調べるのではなく、その背後にある普遍的なネットワーク構造を調べることを目指すとしている。

情報通信サービスのデータに見られる何らかの特性が、特定のサービスやユーザに依存する固有の性質によるものではなく、ユーザの背後にある普遍的な社会ネットワーク構造に由来する性質に基づくものであれば、特定サービスのデータから普遍的な社会ネットワーク構造を調べることができる可能性がある。

一般に、社会ネットワーク構造を調べる直接的な方法としては、アンケートによる調査が考えられる。しかし、大規模な社会ネットワーク構造を調べるにはコストの面から困難である。本研究は、通信ネットワークに関する各種データを利用して、そのデータの背後にある大規模で普遍的な社会ネットワークを探ることを目指している。しかし、通信履歴等のデータ自体の中身を詳細に分析するだけでは、例えば組織内の人間関係や法人間の取引といった「ネットワークの中で起こっている個別の状況」の理解には有効であっても、背後にある普遍的な社会ネットワークの全体像を理解することは難しい。

本研究は、通信ネットワークに現れる各種のべき乗則をもとにして、それらの出現理由を背後にある社会ネットワークの特性に求めることで、社会ネットワークの全体像の理解を目指すものである。ここで扱うべき乗則は、細部の構造が分からない粗視化されたデータであるため、個々のユーザの詳細な振舞い（例えばだれとだれが通信しているかなど）を見ることはできないが、細部の構造によらないより普遍的で一般的な構造が潜んでいることが期待できる。分析の結果として得られる社会ネットワーク構造は、具体的な個人同士の関係などの「個別の状況」を記述するものではないが、そのような個別状況によらない普遍的で一般的な性質を把握することが期待できる。そうして得られた特定のサービス種別によらない普遍的な社会ネットワーク構造は、既に存在する通信サービスに関する普及特性を説明するだけでなく、いまだ見ぬ新規サービスに対するマーケティング戦略の策定などへの工学的なアプローチを与える基礎になるものと期待できる。更に、特定のサービス種別によらない社会ネットワー

ク構造に関する理解は、情報通信サービスだけでなく、より一般的な事業の商品やサービスに関するマーケティング戦略に応用できる可能性がある。

ここで、本研究においてデータのべき乗則に着目する意図を補足しておく。べき乗則には、指数分布や正規分布のような「分布の出現要因が何らかのランダム性に起因するもの」とは異なり、確定的な出現要因の存在が期待できる。このため、べき乗則の出現要因を探索することで、探索の過程がランダムな要因によって乱されることなく、何らかの別の性質に対応することが期待できる。

我々はこれまで、ある携帯電話サービスの普及期におけるトラフィック量の増加特性にべき乗則を見出し、べき乗則を分析することによって通信サービスの背後にある人間の情報交換関係に関する社会ネットワークの構造を調べてきた [9]。しかし、単一の通信サービスのデータでは、社会ネットワークの情報を十分に特定することはできなかった。

本論文では、上記分析に加え、ソーシャルネットワーキングサービス (SNS) 普及期のユーザ数増加特性に見られるべき乗則を分析し、背後にある社会ネットワーク構造を調べる。このとき、携帯電話サービスと SNS のユーザの母集団が共通 (日本在住者) であることを考慮し、双方の分析が同じ対象 (同じ社会ネットワーク) を記述していることを利用することにより、社会ネットワーク構造の性質を詳細化することができる。

結果として得られた社会ネットワークの構造が適切であることを確かめるには、何らかの検証が必要である。普遍的な社会ネットワーク構造を直接観測することができれば分析結果と比較可能だが、実際に大規模な調査を行うことは困難である。このため、分析で利用した通信サービスとは異なる他の通信サービスのデータによって、社会ネットワーク構造の主要な分析結果である通信の次数分布を検証する。

本論文の構成は以下のとおりである。まず 2. において、社会ネットワークの分析方法に関する概念的な全体像について説明する。3. では、文献 [9], [10] に従って、携帯電話サービスのデータを分析することにより得られる社会ネットワーク構造の部分情報を明らかにする。この部分情報からでは、社会ネットワーク構造を記述するモデルに任意パラメータが残り、このままでは検証可能な社会ネットワークモデルとはなっていないことが明らかになる。4. では、3. での分析に SNS のデータ分析を加えて相補的に利用することによ

り、任意パラメータを含まない検証可能な社会ネットワークモデルを導出する。5. では、分析で用いた通信サービスとは独立の新たな携帯電話の通信ログデータを用いて、4. で導出した社会ネットワーク構造の特性を検証する。最後に 6. でまとめを述べる。

2. 社会ネットワーク構造分析のための戦略

本論文では、日本の情報通信サービスの潜在的な利用者からなる社会ネットワークをグラフ $G(V, E)$ で表現することとし、人間をノード、2 人の人間の間の情報交換関係をリンクで表す。

我々の興味の対象は $G(V, E)$ の構造を明らかにすることであるが、残念ながら $G(V, E)$ の全体構造を直接観測することはできない。そこで、携帯電話サービスや SNS といった特定の情報通信サービスのデータを分析することで、間接的に $G(V, E)$ の構造を明らかにするアプローチを考える。表面上は特定の情報通信サービスのデータを分析することになるが、本論文の目的は、特定サービスの利用状況や事例を調べるのではなく、あくまでもこれらのサービスの背後にある人間の社会ネットワーク $G(V, E)$ の構造を調べることにある。このため、特定の情報通信サービス固有のデータ特性に惑わされずに、背後の普遍的な社会ネットワーク $G(V, E)$ の構造を抜き出す分析法が必要になる。

それでは、情報通信サービスのデータから、それら特定のサービスに依存しない普遍的な社会ネットワーク構造を導くにはどうすればよいであろうか。我々の研究アプローチのコンセプトを図示すると図 1 のように表すことができる。図 1 は、中心にある「立体的な」社会ネットワークを、三つの異なる通信サービスのデータによって異なる側面から「平面的な」構造

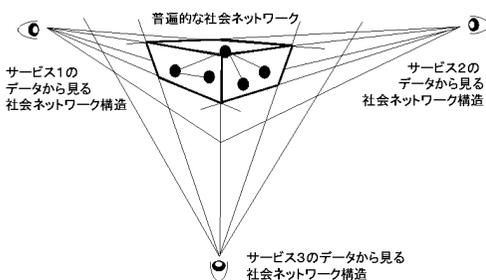


図 1 特定サービスのデータから見る社会ネットワークと普遍的な社会ネットワーク構造の関係の模式図

Fig. 1 The relationship between the real social network and observable images obtained through specific communication services.

をとらえたことを表す概念図である。図の中心にある特定の通信サービスの特性によらない「普遍的な社会ネットワーク」は、人間が直接的に全体像を知ることにはできないが、特定の情報通信サービスのデータを通して「普遍的な社会ネットワーク」の構造に由来する「社会ネットワークの部分情報」を得ることができるとする。複数の異なる情報通信サービスによって複数の「部分情報」を得たとき、それらがある一つの「普遍的な社会ネットワーク」に由来するものであれば、逆にそれらの「部分情報」を相補的に組み合わせることで「普遍的な社会ネットワーク」の全体像を間接的に明らかにすることができるかと期待できる。本研究のデータ分析では、いくつかの通信サービスのデータによって得られる社会ネットワークの断片的な情報から、社会ネットワーク構造の全体像を浮かび上がらせることをねらいとしている。

本分析の特徴は以下のようにまとめることができる。

- 実際の通信サービスのデータを分析するが、そのサービス自体に関する特性を知ることが目的ではなく、その背後にある普遍的な社会ネットワークの構造の抽出を行うこと。
- 単一の通信サービスのデータからでは、普遍的な社会ネットワークに関する限られた情報しか得られないが、共通の社会ネットワークを背後にもつ異なる通信サービスを対象としてそれらのデータ分析結果を相補的に利用することで、社会ネットワークの構造の詳細化を行うこと。
- データに現れるべき乗則を利用することで、個々のユーザの動作が記録されていないような大まかなデータから、次数分布やユーザの行動規則などの性質を導くこと。
- 大規模な社会ネットワークを対象としているので、分析結果を実験で検証することができないが、共通の社会ネットワークを背後にもち、かつ分析に利用したものは独立な情報通信サービスのデータを用いて結果の妥当性を検証すること。

3. 携帯電話によるデータ通信サービスのトラフィックデータを用いた社会ネットワーク分析

本章では、文献 [9], [10] の考え方に従って、携帯電話サービスのデータを分析することにより得られる社

会ネットワーク構造の部分情報を明らかにする。

3.1 分析データ

分析対象の情報通信サービスは、携帯電話端末を用いたインターネットアクセスサービス (i モード [11]) であり、携帯電話端末から Web アクセスと電子メールの利用を可能にしたものである。以降は本サービスをサービス I と呼ぶ。分析に用いるデータはサービス I のユーザ数とトラフィック量の関係である。サービス I は 1999 年 2 月の開始後から爆発的にユーザ数を増やし、1 年半後にはユーザ数が 1,000 万人を超えている。分析対象の期間は本サービスの普及初期にあたる 1999 年 8 月～2000 年 1 月末までの 6 か月間である。この期間にサービス I のユーザ数は約 120 万人から 360 万人程度に増加している。

分析に用いるデータは、社会ネットワーク構造を分析するのに相応しい以下のような特性を備えている。

- サービス I は (固定電話等に比べ) 短期間に爆発的に普及したため、ライフスタイルの変化等の外部要因がデータに与える影響は小さい。
- ほとんどの携帯電話端末はユーザ個人が独占的に使用するため、端末間のトラフィックは端末ユーザ間の情報交換が支配的である。
- サービス I のほとんどの電子メールは 1 対 1 の通信であるので、電子メールトラフィック量と情報交換を行う関係にあるサービス I のユーザペアの数には密接な関連がある。
- 携帯電話の音声通話に比べてサービス I の電子メールの送信コストが極めて低いため、ユーザの所得等にかかわらず通信が行われる。
- サービス I 開始から間もない普及初期は、携帯電話ユーザをターゲットにした広告等の電子メールが問題となる以前の期間であり、既存の社会ネットワークとは無関係なトラフィック量は少ない。

図 2 はサービス I の普及初期において、ユーザ数に対する Web トラフィック量 (アクセス数) 及び電子メールトラフィック量 (メッセージ数) の変化を明示したものである。縦軸及び横軸は、1999 年 8 月 1 日のトラフィック量とサービス I のユーザ数をそれぞれ 1 と正規化し、サービス I のユーザ数とトラフィック量の間を対数 (底は 10) で表示している。また、比較のため傾き 1, 2 の直線を併せて表示している。これらの図から、サービス I の普及初期では、ユーザ数 n ($n \ll N$) に対して

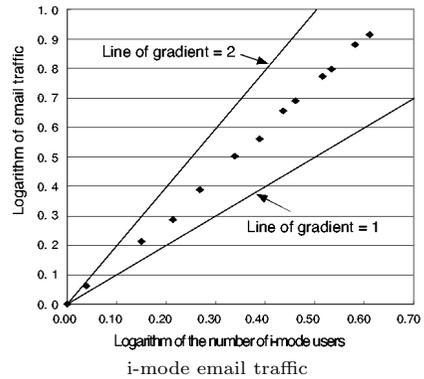
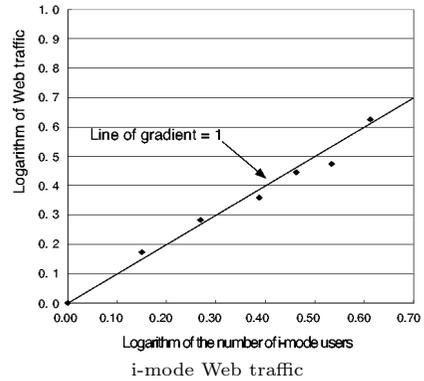


図 2 両対数表示によるサービス I ユーザ数とトラフィック量の関係

Fig. 2 The relationship between the number of users and the volume of traffic for Web and email.

$$(\text{Web traffic}) \propto n, \tag{2}$$

$$(\text{Email traffic}) \propto n^{5/3}, \tag{3}$$

となっていることが分かる。ただし、 N は潜在的ユーザを含んだ母集団の人間の数で、 $n \ll N$ の条件はサービスの普及初期を意味するものである。Web トラフィック量がユーザ数 n に比例することは、もし 1 人当りの Web アクセス回数が一定であれば自然な結果である。一方、電子メールトラフィック量は $n^{1+\alpha}$ ($\alpha \simeq 2/3$) に比例している。この α は、もしユーザ 1 人当りの電子メールトラフィック量がユーザ数 n に無関係であれば (Web トラフィックのように) $\alpha = 0$ となるべきである。つまり、 $\alpha \simeq 2/3$ は 1 人当りの電子メールトラフィック量が n の増加とともに大きくなることを示している。この原因は、ユーザ数 n の増加とともに、サービス I のユーザの中に通信相手が増加していくためであると考えられる。したがって、 $\alpha \simeq 2/3$ は社会ネットワークにおける人間同士の接続の強さを反映し

た量であると考えることができる。

3.2 記号の定義と問題設定

サービス I のユーザ以外の人間を含む一般の人間 (通信サービスの潜在的ユーザ) の集合を V , V に含まれる人間の数を $|V| = N$, 何らかの形で情報交換がある人間のペア $(x, y) \in V \times V$ (リンク) の集合を E とし, 人間の情報交換関係 (社会ネットワーク) を表すグラフを $G(V, E)$ とする。

次に, サービス I のユーザ (ノード) の集合を $V_i(n)$ ($V_i(n) \subset V$), サービス I のユーザ数を $|V_i(n)| = n$, サービス I のユーザの集合 $V_i(n)$ によって $G(V, E)$ から誘導される誘導部分グラフを $G_i(V_i(n), E_i(n))$ とし, $E_i(n)$ は $G_i(V_i(n), E_i(n))$ のリンクの集合とする。ここで, 誘導部分グラフとは, ノードの部分集合からなる部分グラフにおいて, ノード間のリンクの有無を「元のグラフにおいてそのノード間にリンクがあるときに限ってリンクを張る」ことによって決めることで生成される部分グラフをいう。

電子メールトラフィックについて, リンク当りの平均トラフィック量は n によらず一定であると仮定する。当然のことながら, それぞれのリンクにはトラフィックの多い少ないの違いがあるが, 多くのリンクを選んで平均すると, 平均値はリンクの選び方によらず一定になることを仮定していることになる。この仮定の妥当性については, 携帯電話の音声通信トラフィックのデータを用いて確認済みである [12]^(注1)。この仮定のもとでは, 誘導部分グラフのリンク数 $|E_i(n)|$ は, 電子メールトラフィック量に比例することになる。したがって, 電子メールトラフィック量の振舞い (3) から要請される条件として, $n \ll N$ に対して

$$|E_i(n)| \propto n^{1+\alpha}, \quad (\alpha \simeq 2/3) \quad (4)$$

の関係を得る。

本論文での課題はサービス I ユーザ間の社会ネットワーク $G_i(V_i(n), E_i(n))$ を調べるのではなく, サービス I のトラフィックデータを利用してサービス I ユーザ以外の人間を含む普遍的な社会ネットワーク $G(V, E)$ を調べることである。図 3 は $G(V, E)$ と $G_i(V_i(n), E_i(n))$ の関係を示したものである。一番上のグラフは普遍的な社会ネットワーク $G(V, E)$ を表し, そこからサービス I ユーザのみを抜き出して作ったサービス I ユーザ間の社会ネットワーク $G_i(V_i(n), E_i(n))$ が一番下のグラフである。サービス I のユーザ数と電子メールトラフィック量から得られる情報は, サービス I ユーザ間の

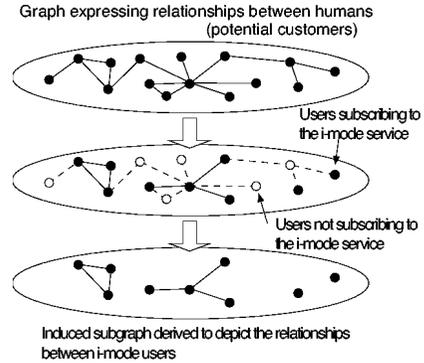


図 3 普遍的な社会ネットワーク $G(V, E)$ と, そこからサービス I ユーザによって誘導される誘導部分グラフ $G_i(V_i(n), E_i(n))$ の例

Fig. 3 Example of $G(V, E)$, a graph showing the structure of the social networks, and $G_i(V_i(n), E_i(n))$, the induced subgraph induced by considering i-mode users.

社会ネットワーク $G_i(V_i(n), E_i(n))$ におけるノード数とリンク数の関係 (4) である。この関係を実現するための条件として, 普遍的な社会ネットワーク $G(V, E)$ の構造や新規にサービス I のユーザとなるときに規則を考察する。

3.3 新規ユーザの選択規則と社会ネットワークの構造

分析に先立ち, 集合 V の要素に二つの異なる番号付けを導入する。

まず, 集合 V の要素を社会ネットワーク $G(V, E)$ に関して次数の大きい順に並べることを考える。このとき ℓ 番目に次数の大きいノードの次数を $D(\ell)$ とする。つまり $D(\ell)$ とは, ℓ 番目に知り合いの多い人の知り合いの数である。

次に, 集合 V の要素をサービス I のユーザになる時刻の順番に並べることを考える。もちろん, 競合他社の同様なサービスも存在するので, すべての人間がサービス I のユーザになることは現実にはあり得ない。しかし, 仮に競合他社のサービスがなく, サービス I の普及初期のような市場支配力の強いサービス普及がその後も持続したとして, 仮にいつかは全員がユーザになるとして, サービス I のユーザになる順番を考える。集合 V の要素をサービス I のユーザとなる時刻が早い順に並べ, ℓ 番目にサービス I のユーザになる人の $G(V, E)$ に関する次数を $D_i(\ell)$ とする。つまり $D_i(\ell)$

(注1): 詳しくは付録 1. 参照。

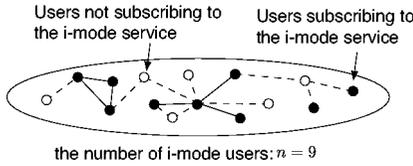


図 4 $c_i(n)$ の例
Fig. 4 Example of $c_i(n)$.

とは、 ℓ 番目にサービス I のユーザになる人の知り合いの数である。同様に、 ℓ 番目にサービス I のユーザになる人の $G_i(V_i(n), E_i(n))$ に関する次数を $d_i(n, \ell)$ とする。つまり $d_i(n, \ell)$ とは、 ℓ 番目にサービス I のユーザになる人のサービス I ユーザ $V_i(n)$ の中での知り合いの数である。

式 (4) に現れる $|E_i(n)|$ は、

$$|E_i(n)| = \frac{1}{2} \sum_{\ell=1}^n d_i(n, \ell) \quad (5)$$

である。これを社会ネットワーク $G(V, E)$ の性質に結び付けるためには、 $D_i(\ell)$ に関する条件を導く必要がある。そのため、以下の関係を満たすような関数 $c_i(n)$ を考える。

$$\sum_{\ell=1}^n d_i(n, \ell) = c_i(n) \sum_{\ell=1}^n D_i(\ell), \quad (6)$$

ここで、 $c_i(n)$ は、サービス I のユーザのすべての知り合いのうち、サービス I のユーザとなっている人の割合を示す関数

$$c_i(n) = \frac{2 \times (\text{サービス I ユーザ同士を結ぶリンク数})}{\text{サービス I ユーザの } G(V, E) \text{ に関する全次数}}$$

である。また、 $c_i(1) = 0$ 、 $c_i(N) = 1$ である。図 4 は $c_i(n)$ の例を示して、この場合 $N = 15$ 、 $n = 9$ であり以下ようになる。

$$\sum_{\ell=1}^n D_i(\ell) = 22, \quad \sum_{\ell=1}^n d_i(n, \ell) = 12, \quad c_i(n) = \frac{6}{11}.$$

ここで、関数 $c(n)$ の関数形を以下のべき関数で試してみる。

$$c_i(n) \propto n^{1-\delta} \quad (7)$$

ここで δ は定数である。この時点で式 (7) はあくまで仮定にすぎないが、以下ではこの仮定の下で何が導かれるか考察を進める。仮定の妥当性は次章で改めて考

察する。

式 (4)、(5) から $n \ll N$ に対して

$$\sum_{\ell=1}^n D_i(\ell) \propto n^{\alpha+\delta} \quad (8)$$

が成り立つ。これが $n \ll N$ なるすべての n で成立するとすれば、 $\ell \ll N$ に対して

$$D_i(\ell) \propto \ell^{\alpha+\delta-1} \quad (9)$$

となる。

ここで、 $\alpha + \delta - 1$ の値によって三つに分けて考察する。まず $\alpha + \delta - 1 < 0$ の場合、 $D_i(\ell)$ は ℓ の増加に対して減少傾向を示す。したがって、サービス I のユーザになる順番で番号を付けた $D_i(\ell)$ は、実は次数の大きい順に並んでいることが分かる。この並び方はあくまでも、データの両対数表示でのべき乗則を観測する程度のオーダの意味で成り立つもので、厳密な順序関係を示しているわけではない。結局 $\alpha + \delta - 1 < 0$ の場合、 $\ell \ll N$ に対して

$$D_i(\ell) \simeq D(\ell) \quad (\text{オーダの意味で}). \quad (10)$$

となる。これはサービス I の加入順に人間を並べたものが、知り合いの多い順に並べたものにおおむね一致することを示している。このことから以下の結論を導く。

- 社会ネットワーク $G(V, E)$ の次数は、 $\ell \ll N$ に関してべき指数 $-(1 - \alpha - \delta)$ の Zipf 則 [13] に従う。

$$D(\ell) = O(\ell^{-(1-\alpha-\delta)}). \quad (11)$$

- サービス I のユーザは $G(V, E)$ に関する次数の高いノードから順にユーザになる傾向がある。言い換えると、知り合いの多い人ほど早い段階でサービス I のユーザになる傾向がある。

最後の結果は、マーケティングでいわれる「情報に敏感な人から流行に飛びつく」という傾向と関連する可能性がある。

次に $\alpha + \delta - 1 = 0$ の場合、 $D_i(\ell)$ は ℓ とは無関係になる。 $G(V, E)$ から次数とは無関係にノードを選んで誘導部分グラフを作ると、 $G(V, E)$ の構造には無関係に、誘導部分グラフのリンク数はノード数 n に対して n^2 となることが知られている [9]。これは $\alpha = 1$ と

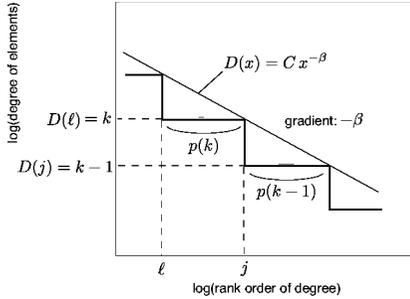


図5 Zipf 則に従うデータから選んだ 2 点の満たす性質
Fig. 5 Two points are extracted from data that satisfies Zipf's law.

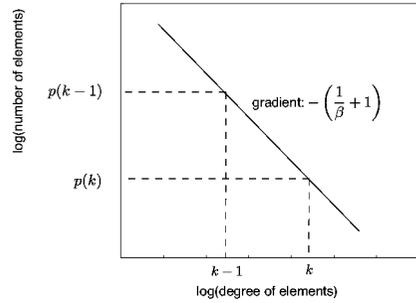


図6 Zipf 則に従うデータから選んだ 2 点を度数分布として表示
Fig. 6 The two extracted Zipf's law points are plotted as the distribution of degree.

なることを意味する．今回の場合，式 (3) からリンク数は $n^{1+\alpha}$ ($\alpha \simeq 2/3$) となるため， $\alpha + \delta - 1 = 0$ では実験事実を説明できない．

最後に $\alpha + \delta - 1 > 0$ の場合，最初の場合とは逆に，サービス I のユーザになるのは次数の低い人（知り合いの少ない人）から順に行われることになる．これは経験的に受け入れられない結果であると判断できる．

以上の考察から， δ は $\alpha + \delta - 1 < 0$ を満たすとして議論を進める．上記結果で得られたように，社会ネットワークの次数が Zipf 則に従うとき，社会ネットワークはスケールフリーネットワークになる．ここでスケールフリーネットワークとは，次数分布 $p(k)$ が式 (1) のべき乗則に従うことをいう [1], [2]．

$D(\ell)$ がべき指数 $-\beta$ の Zipf 則に従うとして， C を比例定数として

$$D(\ell) = C \ell^{-\beta} \tag{12}$$

と表すことにする (図 5 参照)．ここで， $\beta = 1 - \alpha - \delta$ である． $D(\ell) = k$ ， $D(j) = k - 1$ を満たす ℓ と j に対して

$$\ell = C^{1/\beta} k^{-1/\beta}, \quad j = C^{1/\beta} (k - 1)^{-1/\beta} \tag{13}$$

である．ここで， $D(\ell) = k$ のとき $p(k) = j - \ell$ であるので，

$$\begin{aligned} p(k) &= C^{1/\beta} \{ (k - 1)^{-1/\beta} - k^{-1/\beta} \} \\ &\simeq C^{1/\beta} k^{-1/\beta} \frac{1}{\beta k} \\ &= O(k^{-(1/\beta+1)}) \end{aligned} \tag{14}$$

となる．したがって，スケールフリーネットワークの次数分布 (1) に現れる γ は

Assumptions

- 付録 1. の図 A-1 より，電子メールトラフィック量は誘導部分グラフのリンク数 $|E_i(n)|$ に比例する．
- 図 2 より， $|E_i(n)| \propto n^{1+\alpha}$ ，($\alpha \simeq 2/3$)．
- $c_i(n)$ は n のべき関数で $c_i(n) \propto n^{1-\delta}$ ，($\delta < 1 - \alpha$)．

Results

- サービス I のユーザになる順序は $G(V, E)$ に関する次数の大きい順となる傾向がある．
 $D_i(\ell) \simeq D(\ell)$ (オーダの意味で)．
- $G(V, E)$ の次数は以下の Zipf 則に従う．
 $D(\ell) \propto \ell^{-(1-\alpha-\delta)}$
- $G(V, E)$ は以下の次数分布をもつスケールフリーネットワークである．
 $p(k) = O(k^{-\gamma})$ ， $\left(\gamma = \frac{1}{1-\alpha-\delta} + 1 \right)$

図7 サービス I のデータ分析における仮定と分析結果
Fig. 7 The assumptions of the analysis of i-mode data and the results.

$$\gamma = \frac{1}{\beta} + 1 = \frac{1}{1 - \alpha - \delta} + 1 \tag{15}$$

となる (図 6 参照)．

上記の議論の仮定と結論をまとめると 図 7 のようになる．この議論で未解決の問題点は， $c_i(n)$ がべき関数であることと，べき指数に関連するパラメータ δ の値が決定していないため，結論に現れるスケールフリーネットワークの妥当性の確認や次数分布を特徴づける γ の値の決定ができないことである．

4. SNS サービスのユーザ数データを用いた社会ネットワーク分析

本章では、SNS サービスのデータを用いた分析により、社会ネットワーク $G(V, E)$ の性質を調べる。また、前節の結果と組み合わせることにより、サービス I のデータ分析からだけでは未解決な課題として残っていた 2 課題： $c_i(n)$ の関数形に関する仮定の妥当性を確認すること、及び $G(V, E)$ のスケールフリーネットワーク性を特徴づける γ の値を決定すること、を解決する [12], [14]。

4.1 分析データ

分析対象の情報通信サービスは (株) ミクシィによって提供される SNS サービス (mixi [15]) であり、日本最大の SNS サービスである。以降は本サービスをサービス II と呼ぶ。サービス II は 2004 年 2 月にサービスを開始し、ユーザ数 100 万人の到達は 2005 年 8 月、200 万人の到達は 2005 年 12 月であり、最初の 100 万ユーザの到達が 17.5 か月かかっているのに対し、その時点から 200 万ユーザへの到達はわずか 4 か月で実現している。このように、サービス II もサービス I と同様に、急速にユーザ数を増やしたサービスであることが分かる。このサービスの特徴は、新たにサービス II のユーザになるためには、既にサービス II のユーザになっている人からの招待が必要であるという点である。

我々は、サービス II がサービスを開始した以降の、ユーザ数の経時変化を分析する。分析に用いるデータは、社会ネットワーク構造を分析するのにふさわしい以下のような特性を備えている。

- サービス II は短期間に爆発的に普及したため、ライフスタイルの変化等の外部要因がデータに与える影響は小さい。
- 新たにサービス II のユーザになるには既存ユーザからの紹介が必要であるため、サービス II の普及過程は社会ネットワークのリンクと密接に関係している。

図 8 の上図は、サービス II 導入初期における、サービス開始後の経過時間 (経過日数) に対するサービス II ユーザ数の変化を示したものである。また、下図は上記グラフを両対数表示したものであり、比較のために傾き 3 の直線を併せて表示している。サービス II ユーザ数の経時変化が、経過時間に対してべき乗則を

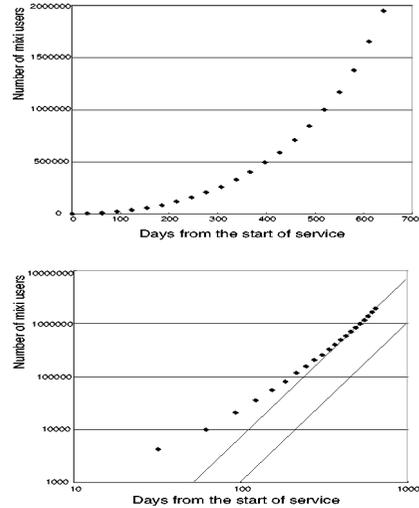


図 8 サービス II ユーザ数の経時変化：上図は線形スケールでの表示、下図は両対数表示で比較のため傾き 3 の直線を併せて表示

Fig. 8 Growth in the number of mixi users.

示すことは既に報告されている [16]。このグラフから、サービス開始直後を除き、そのべき指数は 3 に近いことが確認できる。

4.2 SNS ユーザの増加特性から見た社会ネットワーク構造

時刻 t におけるサービス II のユーザ数を $m(t)$ とする。図 8 の観測結果から、サービス普及初期において

$$m(t) \propto t^3 \tag{16}$$

となるとする。また、サービス II のユーザ数の時間変化 dm/dt を考えると、式 (16) を微分して

$$\frac{dm}{dt} \propto t^2 \tag{17}$$

となる。ここで、ユーザ数の時間変化 dm/dt が時刻に依存することは本質的な関係ではなく、ユーザ数 $m(t)$ に依存することが本質的であると考えられる。つまり、サービス II のユーザ数の増加は、既にユーザとなっている人からの招待によってのみ行われるので、ユーザ数の増加に関しては、時刻ではなくユーザ数自体が第一義的な要因であると考えられるからである。そのため、式 (16) と (17) から t を消去すると、

$$\frac{dm}{dt} \propto m^{2/3} \tag{18}$$

を得る。

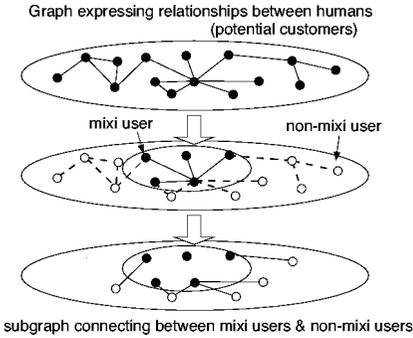


図 9 社会ネットワーク $G(V, E)$ とサービス II ユーザのみから構成される部分ネットワークの例
 Fig.9 Example of $G(V, E)$, a graph showing the structure of the social networks, and the subgraph consisting of mixi users alone.

次にサービス II ユーザの次数を考察していく．サービス II の潜在的ユーザの集合は，サービス I の潜在的ユーザの集合と共通の V とする．つまり，両サービスとも日本の情報通信サービスの潜在的な利用者を対象としている．集合 V の要素をサービス II のユーザになる時刻の順番に並べることを考える．もちろん，すべての人間がサービス II のユーザになることは現実にはあり得ないが，サービス I の場合と同様に，仮にいつかは全員がユーザになるとして，サービス II のユーザになる順番を考える． ℓ 番目にサービス II のユーザになる人の， $G(V, E)$ での次数を $D_x(\ell)$ とする．また， ℓ 番目にサービス II のユーザになる人の，サービス II のユーザからなる誘導部分グラフ（ユーザ数 m ）に関する次数を $d_x(m, \ell)$ とする．

サービス I の場合と同様に， $d_x(m, \ell)$ と $D_x(\ell)$ を関係づけるために，以下の性質を満たす関数 $c_x(m)$ を導入する．

$$\sum_{\ell=1}^m d_x(m, \ell) = c_x(m) \sum_{\ell=1}^m D_x(\ell). \quad (19)$$

ここで， $c_x(m)$ は，サービス II のユーザ数が m のとき，サービス II の知り合いの総数のうち，サービス II のユーザになっている人の割合

$$c_x(m) = \frac{2 \times (\text{サービス II ユーザ同士を結ぶリンク数})}{\text{サービス II ユーザの } G(V, E) \text{ に関する全次数}}$$

を示す．また， $c_x(m)$ は m に関して単調増加で $c_x(1) = 0$ ， $c_x(N) = 1$ であるとする．

図 9 の上段と中段の図は，社会ネットワーク $G(V, E)$ と，サービス II ユーザによって $G(V, E)$ から誘導さ

れた誘導部分グラフとの関係を示すものである．ここで，サービス II ユーザとそれ以外の人を結ぶリンクを外線と呼ぶことにする．図 9 の下段の図は，外線とそれに関するノードのみを表示したものである．この図の例では 6 本の外線がある．一般に外線の本数は

$$\sum_{\ell=1}^m D_x(\ell) - \sum_{\ell=1}^m d_x(m, \ell) = (1 - c_x(m)) \sum_{\ell=1}^m D_x(\ell).$$

で与えられる．サービス II の新規ユーザになるためには，既存ユーザからの招待が必要である．そのため，サービス II ユーザ数の増加速度は，外線の本数に比例すると考えるのが自然である．この根拠は，付録 1. で示されているように，リンク数とトラフィック量の比例関係があり，交換するトラフィック量に比例した頻度で新規ユーザの招待が行われると考えるためである．このとき，

$$\frac{dm}{dt} \propto (1 - c_x(m)) \sum_{\ell=1}^m D_x(\ell) \quad (20)$$

となる．式 (18) より

$$(1 - c_x(m)) \sum_{\ell=1}^m D_x(\ell) \propto m^{2/3} \quad (21)$$

を満たす必要がある．

このように，サービス I とサービス II のデータ分析の大きな違いは，サービスユーザ同士の関係を考えるか，サービスユーザとそれ以外のノードとの関係を考えるかの違いにある．サービス I の分析ではサービスユーザ同士の関係を考えていたため，関数 $c_i(n)$ の形が決定できなかった．サービス II の分析では，サービスユーザとそれ以外のノードとの関係を考えるため，関数 $c_x(m)$ が $(1 - c_x(m))$ の形で条件式に現れる．このため，サービス I の分析とは異なる観点から分析することが可能である．

今， $c_x(m) \ll 1$ となるような m の領域を考えてみる．このとき $(1 - c_x(m))$ の大きさは 1 が支配的なので，対数のスケールでは

$$(1 - c_x(m)) \simeq \text{constant} \quad (22)$$

とみなすことができる^(注2)．これを利用して，式 (21) から $c_x(m)$ に無関係な次数のみの性質を抜き出すと，

(注2)：式 (22) の妥当性は付録 2. 参照．

$$\sum_{\ell=1}^m D_x(\ell) \propto m^{2/3} \quad (23)$$

が得られる． $c_x(m)$ は単調増加関数でとしているので， $c_x(m) \ll 1$ となるのは $m \ll N$ に対応する．式 (23) は $m \ll N$ となるすべての m について成立するとすれば， $\ell \ll N$ において

$$D_x(\ell) \propto \ell^{-1/3} \quad (24)$$

となる．この結果は， ℓ の増加に対して $D_x(\ell)$ が減少することを示している．つまり，サービス I の場合と同様に，これはサービス II のユーザになる順番が知り合いの多い順番におおむね等しいことを示しており， $\ell \ll N$ において

$$D_x(\ell) \simeq D(\ell) \quad (\text{オーダの意味で}). \quad (25)$$

であることを示している．また，これらのことから $\ell \ll N$ で $G(V, E)$ の度数に関して

$$D(\ell) \propto \ell^{-1/3} \quad (26)$$

が成り立つことが分かる．

サービス I とサービス II の分析によって得られた度数 $D(\ell)$ の特性を表す式 (11) と (26) は，特定の通信サービスの特性を表すものではなく，その背後にある普遍的な社会ネットワーク $G(V, E)$ の構造に関する情報を記述したものである．2. で図 1 を用いて説明したように，本論文が前提としている社会ネットワーク分析の考え方に従えば，両者は同じものを表しており，一致することが求められる．式 (11) と (26) が一致するための第一のポイントは，式 (11) が ℓ のべき関数となるべき，ということである．式 (11) が得られた理由は，式 (7) で $c_i(n)$ を n のべき関数と仮定したからであり， $c_i(n)$ として他の関数を考えると式 (11) が ℓ のべき関数とはならない．したがって，この時点で式 (7) の仮定の妥当性が確認された．式 (11) と (26) が一致するための第二のポイントは，べき関数のべき指数が一致することである．そこで，式 (26) と (11) を比較し， $\alpha \simeq 2/3$ であることを使うと， $\ell \ll N$ において，サービス I の分析に関する以下の条件を確認することができる．

- サービス I の分析では決定できなかった δ の値が以下のように決定される．

$$\delta \simeq 0 \quad (27)$$

- 関数 $c_i(n)$ は式 (7) で表せるべき関数となる，特に $\delta \simeq 0$ なので n に関する一次関数である．
- $\delta \simeq 0$ なので $\alpha + \delta - 1 < 0$ は満たされる．

また， $c_i(n)$ と $c_x(m)$ は，どちらも同じ意味の関数を表していて，社会ネットワークから次数の大きい順にノードを n 個 (または m 個) 選んで誘導部分グラフを作ったとき，どちらも選ばれたノードの総次数に対する誘導部分グラフの総次数の比を意味する．したがって $c_x(m) \propto m$ から， $m \ll N$ において式 (22) の性質が妥当であることが分かる．このように，サービス I 及び II のデータから導いた社会ネットワークの構造は，互いのデータに対して自己無撞着なモデルとなっている．

関数 $c_i(n)$ が n に比例するということは，次数の大きい順にユーザを選んで互いのつながりやすさを見たときに，ノードの総次数に対する誘導部分グラフの総次数の比は，選択したユーザ数 (具体的なサービスの例ではサービス普及率) に比例することになる．これは，ノード同士の接続の仕方について，次数が高いノード同士がつながりやすい，などといった次数による偏りが見られないことを示している．

δ の値が決定されたことで， $G(V, E)$ のスケールフリー性を特徴づける次数分布のべき指数が，次数の高いノードについて (つまり普及初期のユーザに関して)

$$\gamma = \frac{1}{1 - \alpha - \delta} + 1 \simeq 4 \quad (28)$$

と決定される．ここで，次数分布

$$p(k) = O(k^{-4}), \quad (k \gg 1) \quad (29)$$

が成り立つ範囲は大きな次数 $k \gg 1$ に対応する．つまり，ここで得られた $G(V, E)$ のスケールフリー性に関する特性 (28), (29) は，次数分布 $p(k)$ の裾に関する性質であり，スケールフリーネットワークの性質として重要なハブの存在確率を適切に与えることが期待できる．

サービス II の分析の仮定と分析結果，またサービス I の分析との相補的な組合せによる分析結果は，図 10 のようにまとめることができる．サービス II データの分析の仮定は，実験事実やサービス II の特性に起因するものである．それに基づいて結論される中間的な結

Assumptions

- 図 8 より $m(t) \propto t^3$.
- サービス II ユーザ数の増加速度は外線数に比例する .

$$\frac{dm}{dt} \propto (1 - c_x(m)) \sum_{h=1}^m D_x(\ell)$$

- $m \ll N$ の領域で $1 - c_x(m) \simeq \text{const.}$

Intermediate results

- サービス II ユーザとなる順序は $G(V, E)$ に関する次数の大きい順となる傾向がある .

$$D_x(\ell) \simeq D(\ell) \quad (\text{オーダの意味で}).$$

- サービス I で用いた仮定の検証 .

$$c_i(n) \propto n^{1-\delta}, \quad \delta < 1 - \alpha$$

Results

- 社会ネットワーク $G(V, E)$ の次数分布に関する特性として以下のスケールフリー性を得る .

$$p(k) = O(k^{-4}), \quad (\text{for large } k)$$

- 社会ネットワークのリンクの接続に関する特性として、次数による接続の偏りがないことが分かる .

$$c_i(n) \propto n^{1-\delta} \simeq n$$

図 10 サービス II データを用いた分析における仮定と分析結果

Fig. 10 The assumptions made for analysis of mixi and the results.

果によって、サービス I の仮定の妥当性が検証される。ここで妥当性とは、複数のデータの自己無撞着性を意味する。最終的な結論は、仮定の妥当性が検証されたサービス I の分析結果と、サービス II の分析を組み合わせることで、社会ネットワーク $G(V, E)$ の構造を導いた。

5. 通信サービスの次数分布

前章で導いた社会ネットワークの特性が、サービス I やサービス II といった特定のサービスの詳細によらない普遍的な結果であれば、別のデータを用いた分析によって分析結果が支持されるはずである。本章では、サービス I とサービス II のデータから導かれた社会ネットワーク構造のうち、スケールフリー性 (29) の性質を、サービス I やサービス II とは独立な情報通信サービスのデータにより検証する。

検証に利用する情報通信サービスは携帯電話の音声

通信サービス (au by KDDI [17]) であり、以下ではサービス III とする。分析の方法は以下のとおりである。まず、サービス III の通話ログデータをもとにして、携帯電話ユーザの間の関係を表すグラフを生成する。グラフの生成は、ユーザをノードとし、通話ログデータの期間中に通話の行われたユーザ間をリンクで結ぶ方法で行う。リンクは発信信を区別し、有向グラフを作成する。このとき、期間中に同一ユーザペア間の同一方向の発信が複数回行われたとしても、リンクは 1 本である。また、通話時間の長短にもよらない。このような方法で生成した携帯電話ユーザの間の関係を表すグラフについて、その次数分布を調べ、前節の分析結果と比較する。

実際のデータ分析では、普遍的な社会ネットワーク $G(V, E)$ を直接調べることができないので、ユーザはサービス III のユーザに限定し、また、データの取得地域が限定された通話履歴データを用いることになる。このように、対象とするユーザや対象地域を限定して作ったグラフの次数分布が普遍的な社会ネットワーク $G(V, E)$ の次数分布と比較できるものかどうかを考察する必要がある。一般に、対象ユーザの限定がノード次数とは無関係に行われていれば、社会ネットワークの次数分布特性は保存されることが分かっている^(注3)。この分析の場合、サービス III のユーザの次数とは無関係に選ばれていると考えてよい。また、データ収集地域の分割もユーザの次数とは無関係に行われていると考えてよい。したがって、普遍的な社会ネットワーク $G(V, E)$ の次数分布が $p(k)$ であれば、サービス III のユーザに限定したある地域のユーザから作成したグラフの次数分布も $p(k)$ となる。

分析するデータは、サービス III の 6 箇所の異なる携帯電話交換機の通話ログデータであり、ユーザ ID ごとに通話履歴を分析して、発信信別に次数分布を調べた。分析期間はどのエリアについても連続した 14 時間分である。発呼に関しては、あるユーザが発呼したあて先ユーザ ID 数を次数とし、着呼に関しては、あるユーザが着呼した発信元ユーザ ID 数を次数とした。

図 11、図 12 は、それぞれ発呼と着呼に関する次数分布 (PDF) をエリアごとに表示したものである。横軸は次数 k 、縦軸は次数 k をもつユーザ ID 数の PDF $p(k)$ を表し、両対数表示を行っている。比較のために、次数分布と平均値の等しい指数分布の曲線 (点線) と

(注3): 付録 3. 参照。

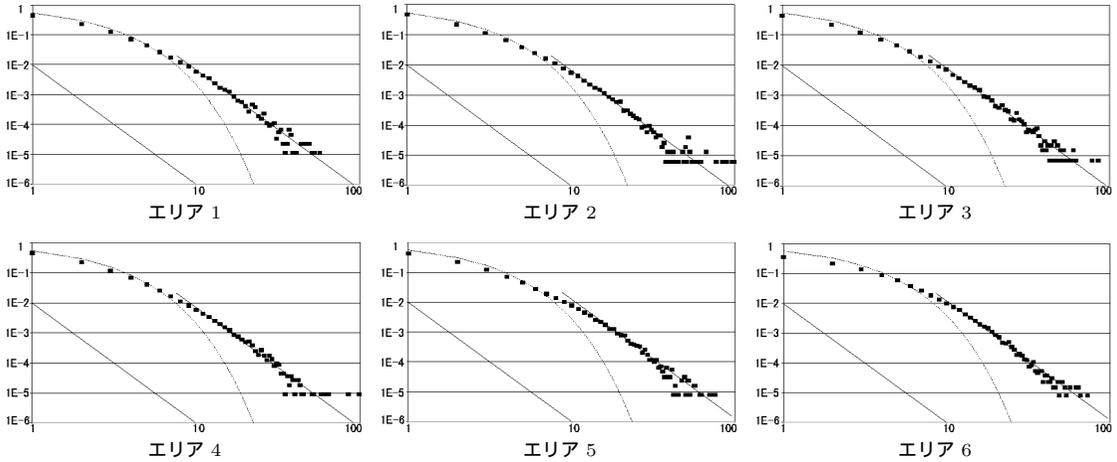


図 11 発呼の次数分布

Fig. 11 The degree distributions of outgoing calls.

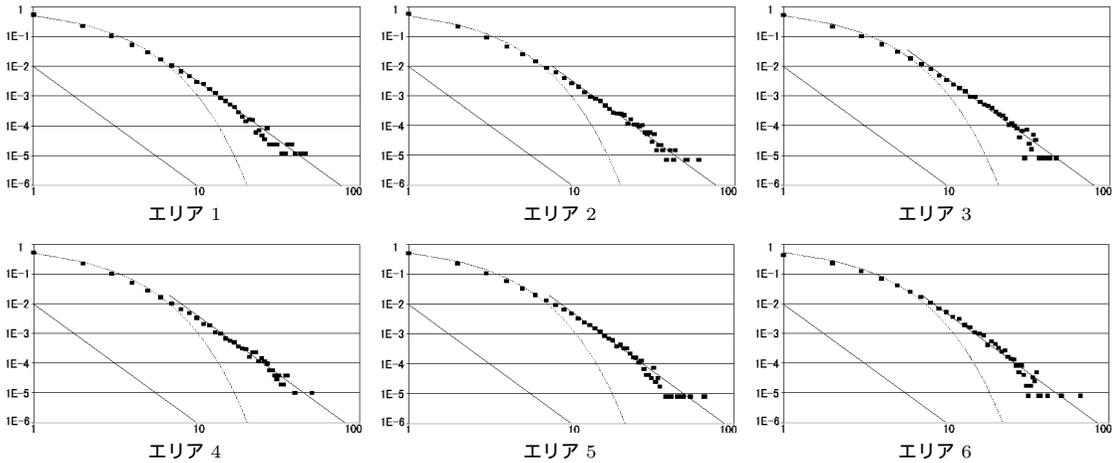


図 12 着呼の次数分布

Fig. 12 The degree distributions of incoming calls.

傾き -4 の直線を併せて表示した。これらの図から、分布の裾の振舞いが k^{-4} に比例していることが分かる。この結果から、社会ネットワーク $G(V, E)$ の分析結果として得られたスケールフリー性 (29) が妥当であることが確認できる。

6. む す び

本論文は、複数の情報通信サービスに関するデータを利用して、それらのサービスの背後にある社会ネットワークの構造を分析し、その結果を検証したものである。社会ネットワーク構造の分析利用したデータは、携帯電話サービスにおけるユーザ数とトラフィック量の

関係、及び SNS のユーザ数経時変化といった大まかなもので、個別のユーザの通信ログが分かるような詳細なデータではない。そのような大まかなデータをもとにしているが、データに現れるべき乗則を利用すること、複数サービスのデータを相補的に利用することにより、社会ネットワークの次数分布特性やノードの接続に関する規則、またユーザの行動特性に関する規則を導くことに成功した。導出された次数分布特性に関しては、分析に利用した情報通信サービスとは独立な他の携帯電話サービスの音声通話ログデータを用いて検証を行った。

分析によって得られた社会ネットワークの特性は、

特定の通信サービスに限られた特性ではなく、普遍的なものであると期待できる。事実、本論文で登場する異なる通信サービスに対して、その背後にある社会ネットワークの構造として矛盾しない特性を備えている。そのため、社会ネットワーク構造に関する知見は、まだ見ぬ新しい情報通信サービスや情報マーケティングの手法などを考案する上で工学的なアプローチを可能にすると期待できる。具体的にいうと、分析で得られた社会ネットワークの特性（度数分布、ノードの繋がり方の規則）を用いて社会ネットワークのモデルを生成し、その上で情報伝搬特性などのシミュレーション実験を行うことができるようになることを期待できる。例えば、新しい情報通信サービスを導入する際に、普及の速度を高めるためには、初期ユーザとしてどのようなユーザを選べばよいか（ユーザモニタの選択方法）の検討などが行われつつある [18]。

謝辞 付録 1. にあるリンク数とトラフィック量の比例関係は、文献 [12] から概要を引用したもので、導出には NTT ドコモの佐々木純氏の協力によるところが大きい、ここに謝意を記す。また本研究の一部は（財）国際コミュニケーション基金 (ICF) の平成 17 年度調査研究助成、及び文部科学省科学研究費補助金基盤研究 (S)18100001 より研究費の援助を受けて実施した。

文 献

- [1] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol.286, pp.509–512, 1999.
- [2] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol.74, pp.47–97, 2002.
- [3] H. Ebel, L.I. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," *Phys. Rev.*, vol.E66, 035103(R), 2002.
- [4] J. Abello, P.M. Pardalos, and M.G.C. Resende, "On maximum clique problems in very large graphs," *DI-MACS Series in Discrete Mathematics and Theoretical Computer Science*, vol.50, pp.119–130, 1999.
- [5] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," *32nd ACM Symposium on the Theory of Computing*, pp.171–180, 2000.
- [6] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks," *Proc. National Academy of Science (PNAS)*, vol.104, no.18, pp.7332–7336, 2007.
- [7] C.A. Hidalgo and C. Rodriguez-Sickert, "The dynamics of a mobile phone network," *Physica A*, vol.387, pp.3017–3024, 2008.
- [8] J. Candia, M.C. Gonzalez, P. Wang, T. Schoenharl,

G. Madey, and A.-L. Barabási, "Uncovering individual and collective human dynamics from mobile phone records," *J. Phys. A: Math. Theor.*, vol.41, no.22, 224015, 2008.

- [9] M. Aida, K. Ishibashi, H. Miwa, C. Takano, and S. Kuribayashi, "Structure of human relations and user-dynamics revealed by traffic data," *IEICE Trans. Inf. & Syst.*, vol.E87-D, no.6, pp.1454–1460, June 2004.
- [10] M. Aida, K. Ishibashi, C. Takano, H. Miwa, K. Muranaka, and A. Miura, "Cluster structures in topology of large-scale social networks revealed by traffic data," *IEEE GLOBECOM 2005*, St. Louis, 2005.
- [11] NTT ドコモホーム, <http://www.nttdocomo.co.jp/>
- [12] 会田雅樹, 佐々木純, "通信サービスの普及プロセスを利用した社会ネットワークの構造分析," *信学技報*, IN2006-41, 2006.
- [13] R. Rousseau, "George Kingsley Zipf: Life, idea, his law and informetrics," *Glottometrics*, vol.3 (To Honor G.K. Zipf), pp.11–18, 2002.
- [14] 会田雅樹, "物理の現象論に学ぶ: 通信ネットワークに現れるべき乗則を利用した社会ネットワーク構造の解明," *信学誌*, vol.91, no.10, pp.891–896, Oct. 2008.
- [15] (株)ミクシィ, <http://mixi.co.jp/>
- [16] 湯田聡夫, 小野直亮, 藤原義久, "ソーシャル・ネットワーキング・サービスにおける人的ネットワークの構造," *情報学論*, vol.47, no.3, pp.865–874, 2006.
- [17] au by KDDI, <http://www.au.kddi.com/>
- [18] 平野達矩, 上嶋聖人, 高野知佐, 会田雅樹, "社会ネットワークの特性を利用した通信サービス普及戦略の検討," *信学技報*, IN2008-135, 2009.

付 録

1. リンク数とトラフィック量の関係

トラフィック量に関するデータを用いて社会ネットワークのリンク数を議論するためには、トラフィック量とリンク数が比例関係にあることを確認する必要がある。このためには、個々のユーザの通信ログを分析することが必要であり、図 2 のような個々のユーザの振舞いが観測できないようなデータからは判断することができない。

ここでは、文献 [12] での分析結果を引用し、トラフィック量とリンク数の比例関係を確認する。この文献では、某社の提供する携帯電話サービスにおいて、2004 年 9 月のある一日（平日）のサービスエリア内での発着信のログデータをもとにした分析を行っている。

まず、発着信のログデータによってユーザ ID 間の関係を示すグラフを作成する。ユーザ ID 間にリンクを張る基準は以下のようなものである。リンクの有無は通話の有無によって決まり、通話が行われなかったユーザ ID 間はリンクを張らない。また、電話勧誘な

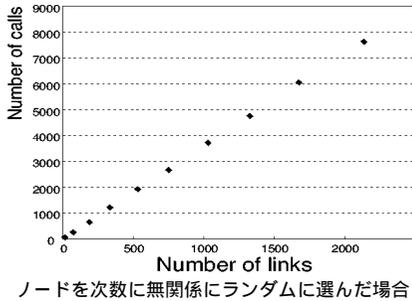
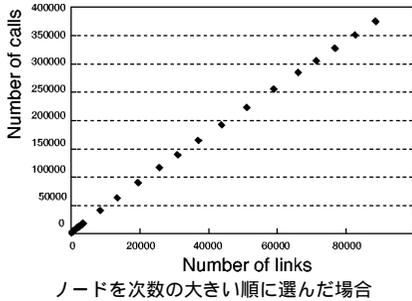


図 A.1 誘導部分グラフのリンク数とそれらのリンク上のトラフィック量の関係

Fig. A.1 Relationship between the actual number of calls and the number of links in induced subgraphs.

どの知人関係と無関係な通話の影響を極力排除する目的で、双方向の発信が少なくとも一度あった場合のみにリンクを張り、片方向のみの発信にはリンクを張らない。リンクが張られる場合には、通話回数の多寡によらずに1本のリンクとする。また、各リンクの実際のトラフィック量（通話回数）を別途記録しておく。

このように作成した大規模なグラフを社会ネットワーク $G(V, E)$ とみなして以下の実験を行った。まず、 $G(V, E)$ のノードを次数の大きい順に選択し、選択されたノードによって $G(V, E)$ から誘導される誘導部分グラフを作っていく。このとき、誘導部分グラフのリンク数と、それらのリンクに対応するトラフィック量を比較したものが図 A.1 の上図である。また、 $G(V, E)$ のノードを次数とは無関係にランダムに選択し、選択されたノードによって $G(V, E)$ から誘導される誘導部分グラフを作ったとき、誘導部分グラフのリンク数と、それらのリンクに対応するトラフィック量を比較したものが図 A.1 の下図である。どちらも線形のスケールで表示している。これらの結果から、誘導部分グラフのリンク数の和とトラフィック量には比例関係があるとよいことが分かる。

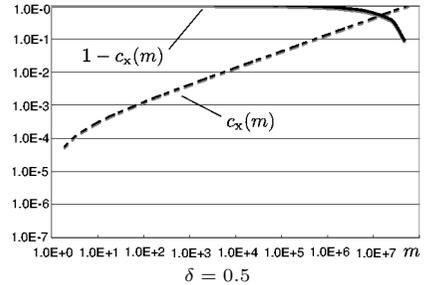
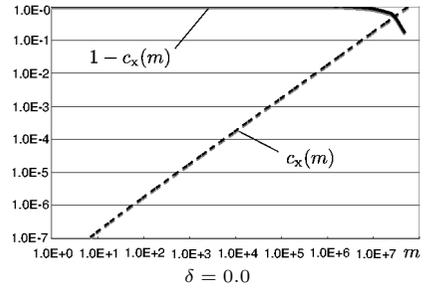


図 A.2 $1 - c_x(m)$ の振舞いの例

Fig. A.2 Examples of behavior of $1 - c_x(m)$.

2. $1 - c_x(m)$ の振舞い

関数 $c_x(m)$ は $c_x(1) = 0$ で $c_x(N) = 1$ となるような単調増加関数であるが、 $m \simeq 0$ において急激に1に近づくような極端な振舞いはもたず、 $m \ll N$ で $c_x(m) \ll 1$ となるとする。このとき、 $m \ll N$ では $1 - c_x(m) \simeq 1$ となる。

これを簡単な例で確認してみる。上記の性質を満たすような $c_x(m)$ の例として

$$c_x(m) = \left(\frac{m-1}{N-1} \right)^{1-\delta}. \quad (\text{A.1})$$

となる関数を考える。図 A.2 は、横軸を m として $c_x(m)$ と $1 - c_x(m)$ を表示したものである。ここで潜在的なユーザ数 $|V| = N$ の値は 60,000,000 とし、 $\delta = 0.0$ と 0.5 の2種類の場合を例示している。これらのグラフより、 m が比較的小さい範囲で $1 - c_x(m)$ を定数とみなしてよいことが分かる。

3. 通信サービスの次数分布と社会ネットワークの次数分布の関係

社会ネットワークの次数分布を調べる際に、すべての人間を対象とした調査を行うことは事実上不可能である。ここでは、地域分割などにより、次数分布とは独立の方法で特定のユーザを選択し、それらの選択されたユーザに対して次数分布を調べる状況を想定する。その上で、もしすべての人間を対象とした社会ネット

ワークの次数分布がべき分布であった場合、選択されたユーザの次数分布も同じべき分布となることを示す。

まず、次数分布とは独立の方法で選択したユーザの全次数の分布を考える。社会ネットワーク全体の次数分布を $p(k)$ とし、 $p(k) \propto k^{-\gamma}$ ($\gamma > 0$ は定数) とする。次数 k をもつノードが存在する割合は $p(k)$ なので、ある確率 ξ でランダムにノードを選択したとき、次数 k のノードが選択される割合は $\xi \cdot p(k)$ である。したがって、選択されたノードの次数分布 $a(k)$ は

$$a(k) \propto \xi \cdot p(k) \propto k^{-\gamma}$$

となる。ここでの次数分布 $a(k)$ は、リンク先が選択されたノードに限定されておらず、選択されたノードがもつすべてのリンクをカウントする分布である。次に、同様な状況において、選択したユーザの同士を結ぶリンクのみを対象とした次数分布 $b(k)$ を考える。ユーザが本来もっている次数 k のうち、選択されたノード同士を結ぶもののリンク数(次数)は平均で $\xi \cdot k$ となる。したがって、選択されたノード同士を結ぶリンクの次数分布 $b(k)$ は平均的には

$$b(\xi k) \propto \xi \cdot p(k) \propto \xi \cdot k^{-\gamma}$$

となる。ここで $K := \xi k$ とすると、

$$b(K) = \xi \cdot p(K/\xi) \propto \xi \cdot \left(\frac{K}{\xi}\right)^{-\gamma} \propto K^{-\gamma}$$

である。以上から、選択されたユーザに関する次数分布 $a(k)$ と $b(k)$ はともに社会ネットワークの次数分布 $p(k)$ と一致する。

(平成 22 年 2 月 25 日受付, 5 月 12 日再受付)



会田 雅樹 (正員)

昭 62 立教大・理・物理卒。平元同大大学院博士課程前期課程原子物理学専攻了。同年日本電信電話(株)研究所入所。以来、ネットワーク制御技術、通信トラヒック設計技術等の研究に従事。平 17 首都大学東京システムデザイン学部准教授, 平 19 同大大学院システムデザイン研究科教授。博士(工学)。IEEE, 日本オペレーションズ・リサーチ学会各会員。本会平 7 年度学術奨励賞受賞。



高野 知佐 (正員)

平 12 阪大・工・通信卒。同年 NTT アドバンステクノロジ(株)トラヒックリサーチセンタ入社。以来、通信システムの性能評価技術、自律分散制御技術等の研究開発に従事。平 20 首都大学東京大学院システムデザイン研究科博士課程後期課程了。同年、広島市立大学大学院情報科学研究科准教授。博士(工学)。本会平 14 年度学術奨励賞受賞。



小頭 秀行 (正員)

平 14 早大・理工・電子情報通信卒。平 16 同大大学院修士課程了。同年 KDDI(株)入社。以来、研究所にて、ネットワーク設計及びトラヒック制御技術の研究に従事。現在(株)KDDI 研究所ネットワーク設計グループ研究員。



中村 元 (正員)

昭 63 早大・理工・電子通信卒。平 2 同大大学院修士課程了。同年国際電信電話(株)研究所入所。以来、ネットワーク設計技術、通信トラヒック、ネットワークアーキテクチャ等の研究に従事。現在(株)KDDI 研究所ネットワーク設計グループリーダー。工博。IEEE 会員。本会平 7 年度学術奨励賞, 平 9 年度及び平 15 年度論文賞等各受賞。