

博士論文

Friendsterにおけるユーザーネットワーク崩壊の
メカニズム

関 和 紀

目次

| | | |
|-------|---|----|
| 第 1 章 | はじめに | 5 |
| 第 2 章 | 用語の定義 | 7 |
| 2.1 | SNS とオンラインコミュニティ | 7 |
| 2.2 | Friendster | 7 |
| 2.2.1 | Friendster の歴史 | 7 |
| 2.2.2 | Friendster の機能 | 10 |
| 2.3 | ネットワーク科学の用語 | 10 |
| 第 3 章 | 先行研究 | 14 |
| 3.1 | ネットワーク構造に着目した, SNS の成長・衰退過程の分析 | 14 |
| 3.2 | Friendster の衰退過程の分析 | 15 |
| 3.2.1 | Garcia のモデル | 15 |
| 3.2.2 | Yu のモデル | 16 |
| 3.3 | 影響伝播モデルを用いたオンラインコミュニティの分析 | 16 |
| 3.3.1 | SIR モデル | 17 |
| 3.3.2 | irSIR モデル | 18 |
| 3.3.3 | Ribeiro のモデル | 19 |
| 第 4 章 | Friendster のコア構造の分析 | 20 |
| 4.1 | データセット | 20 |
| 4.2 | 分析方法 | 22 |
| 4.2.1 | 各時刻におけるアクティブネットワークの推定 | 22 |
| 4.2.2 | 有向 k -core decomposition 法 | 23 |
| 4.3 | 分析結果 | 24 |
| 4.3.1 | コア構造の時間変化の可視化 | 24 |
| 4.3.2 | 指標に基づいた評価 | 25 |
| 4.4 | 分析の仮定の妥当性の検討 | 27 |
| 4.4.1 | ID を時刻とみなしたことの影響 | 27 |
| 4.4.2 | δ の値による影響 | 29 |
| 4.4.3 | 時刻 T から $T + \delta$ に登録したユーザーは時刻 T にアクティブであるとみなすか否かによる影響 | 31 |
| 4.4.4 | データセットを有向ネットワークとみなすか無向ネットワークとみなすかの影響 | 32 |

| | | |
|--------------|---|-----------|
| 4.4.5 | k -core を用いるか k -辺連結成分を用いるかの影響 | 33 |
| 4.5 | 本章のまとめ | 33 |
| 第 5 章 | Friendster のユーザーネットワークがコア構造の内側から縮小したメカニズム | 35 |
| 5.1 | 本章で解明したい問題 | 35 |
| 5.2 | ネットワーク irSIR モデル | 35 |
| 5.3 | Friendster のユーザーネットワークの時間変化はネットワーク irSIR モデルの遷移確率と同じ式を満たすかの検証 | 36 |
| 5.3.1 | 状態 $S \rightarrow I$ の遷移確率 | 37 |
| 5.3.2 | 状態 $I \rightarrow R$ の遷移確率 | 41 |
| 5.4 | ネットワーク irSIR モデルにしたがって時間変化するネットワークではコアの中心部が全体より先に縮小するかの検証 | 44 |
| 5.4.1 | ネットワーク生成アルゴリズム | 44 |
| 5.4.2 | ネットワーク irSIR モデルのアルゴリズム | 45 |
| 5.4.3 | パラメータがコア中心部の縮小に与える影響 | 47 |
| 5.5 | パラメータが状態 I の頂点数の時間変化に与える影響 | 49 |
| 5.6 | 本章のまとめ | 49 |
| 第 6 章 | ネットワーク irSIR モデルの拡張 | 51 |
| 6.1 | 状態 I への遷移確率が隣接する状態 I の頂点数に比例しない場合 | 51 |
| 6.2 | 状態 R への遷移確率に次数に反比例する項を加えた場合 | 53 |
| 6.3 | ネットワーク irSIRI モデル | 54 |
| 6.3.1 | ネットワーク irSIRI モデルのアルゴリズム | 55 |
| 6.3.2 | ネットワーク irSIRI モデルのシミュレーションの結果 | 55 |
| 6.3.3 | 実験結果の解釈 | 58 |
| 6.4 | 直接遷移ネットワーク irSIRI モデル | 58 |
| 6.4.1 | 直接遷移ネットワーク irSIRI モデルのアルゴリズム | 58 |
| 6.4.2 | 直接遷移ネットワーク irSIRI モデルのシミュレーションの結果 | 59 |
| 6.4.3 | 実験結果の解釈 | 60 |
| 6.5 | 復帰の連鎖反応モデル | 60 |
| 6.5.1 | 復帰の連鎖反応モデルのアルゴリズム | 62 |
| 6.5.2 | 復帰の連鎖反応モデルのシミュレーションの結果 | 63 |
| 6.5.3 | 実験結果の解釈 | 65 |
| 6.6 | 本章のまとめ | 65 |
| 第 7 章 | おわりに | 67 |
| 7.1 | 結論 | 67 |
| 7.2 | 今後の展望 | 68 |

第1章 はじめに

近ごろは、世界中に多種多様なソーシャル・ネットワーキング・サービス (SNS) が存在し、人々の生活に大きな影響を与えている。SNS とはその名の通り、ユーザーが自由に社会ネットワークを構築することができるようなウェブサービスのことである。SNS のユーザーの多くは、「実社会においても繋がりを持つ人々」や、「実際に顔を合わせたことはないが同じ趣味を持っている人々」などと、SNS 上でのつながりを形成している。

SNS の歴史はまだ比較的浅いが、この短い歴史の中で、数多くのサイトが成長と衰退を繰り返してきた [8]。1990 年代末には SixDegrees.com¹ などいくつかの SNS が出現し、今日における SNS の繁栄の礎を築いた。2000 年代初めには Friendster² が急成長を見せたものの、すぐに MySpace³ に取って代わられた。2000 年代の後半からは Facebook⁴ が強い影響力を持ち続けている。むろん、以上で挙げたような（たとえ短期間であっても）大きな成功を収めた SNS の影には、ほとんど注目を集めることもないまま消滅した SNS も多数存在する。

こうした現象は、SNS や、SNS 上で形成されるオンラインコミュニティに関して、様々な疑問を提起する。SNS やオンラインコミュニティの成長や衰退を司っているものは一体何だろうか？ そこには普遍的な法則が存在するのだろうか？ もしそうであるならば、衰退を防ぐための効果的な方法もそこから導き出せるのだろうか？ 多種多様な SNS が人々の生活に浸透し、社会に大きな影響力を与えている現代において、これらの疑問に対する考察を深め、正しい見解を得ることは、大きな意義を持つ。

そのような知見を得るための先行研究として、Garcia[13] と Yu[37] によるものがある。Garcia は、ネットワーク上のユーザーがコスト・ベネフィットに基いて合理的に意思決定を行うモデルを用いて、Friendster という SNS の衰退メカニズムを説明することを試みた。Yu は Garcia のモデルに改良を加えた。Friendster とは、2002 年にアメリカで設立された SNS である。主にアメリカとアジアで広く利用されたものの、その後衰退し、2011 年にサービスが終了している。彼らは、提案したモデルがアクティブユーザー数の時間変化を良く説明できることを示したが、ユーザーネットワーク構造の時間変化を分析したわけではない。

そこで、本論文ではいくつかの仮定をおくことで Friendster のユーザーネットワークの時間変化を再構成し、その際構成されたデータに基づいて Friendster のネットワーク構造の時間変化について明らかにし、その時間変化を説明できるモデルを構築する。次に、そのモデルを拡張して、他の SNS のネットワークの時間変化について考える。

本論文の成果は以下の通りである。

- Friendster のサービス終了直前のある一つの時刻におけるネットワーク構造を示すデータセットから、

¹<http://sixdegrees.com/>

²<http://www.friendster.com/>

³<https://myspace.com/>

⁴<https://www.facebook.com/>

いくつかの仮定の下でネットワーク構造が時間とともにどのように変化したのかを推定する手法を提案した。推定したネットワーク構造の時間変化を分析した結果からは、コア構造中心部(友だち登録しあっているユーザーの多いユーザー群)の縮小がネットワーク全体の縮小よりも先行したことを明らかにした。

- Friendster のユーザーネットワーク衰退過程を説明するためのモデルとして、irSIR モデルという影響伝播モデルをネットワーク上の影響伝播モデルに拡張した「ネットワーク irSIR モデル」を提案した。
- ネットワーク irSIR モデルにいくつかの規則を追加することで、モデルの性質がどのように変化するかを示した。

本論文は以下のように構成されている。第 2 章では、本論文で使用する用語を定義する。第 3 章では、ネットワーク科学や影響伝播モデルを用いてオンラインコミュニティと SNS の成長・衰退の過程を分析した先行研究を概観する。第 4 章では、Friendster のユーザーネットワーク構造の時間変化を推定し、推定に基づいて Friendster が衰退した際にネットワークがどのように縮小したのかを導く。第 5 章では、Friendster のユーザーネットワークの時間変化を説明するためのモデルとしてネットワーク irSIR モデルを構築する。第 6 章では、一般の SNS などで見られる現象を単純化した規則を第 5 章のモデルに組み込み、組み込まれた規則と衰退の関係を分析する。最終章では本論文をまとめる。

第2章 用語の定義

2.1 SNS とオンラインコミュニティ

SNS は、「ソーシャル・ネットワーキング・サービス」「ソーシャル・ネットワーキング・サイト」「ソーシャル・ネットワーク・サイト」等の略称である。Boyd[8] は以下の三点を満たすウェブサービスを SNS と定義している。

1. そのシステムの内部で、公的あるいは半公的なプロフィールを作成できる。
2. 各ユーザーがどのユーザーと繋がりを持っているかが、明確に示されている。
3. 自分自身や他のユーザーが作成した繋がりのリストを閲覧し、その中を移動することができる。

SNS 上や他のウェブサービス上においては、オンラインコミュニティが形成されることが多い。オンラインコミュニティに関しては、多くの研究者が独自の定義を行っている。Lee[19] は、2003 年の時点において参照可能であった文献においてみられる定義を整理し、

計算機をベースとした情報技術に支えられたサイバースペースであり、メンバーによるコンテンツ作成のためのコミュニケーションや相互作用をその活動の中心とし、メンバー間の関係形成をもたらすもの

をバーチャルコミュニティと定義した。なお、Lee は「バーチャルコミュニティ」と「オンラインコミュニティ」が同義であるという前提のもとで既存文献の分析を行っている。SNS 上において形成されるユーザー集団は、Lee らのバーチャルコミュニティの定義を満たすものと考えられる。

SNS のある時刻におけるアクティブユーザーとは、その SNS 上において、その時刻付近に活発に活動しているユーザーとする。

2.2 Friendster

本節では、本論文で分析する Friendster という SNS の概要を記す。

2.2.1 Friendster の歴史

設立～急成長

Friendster は、Jonathan Abrams が 2002 年に開発を開始し、2003 年の 3 月に本格的に運営を開始した SNS である。Friendster という名称は、「friend」と「Napster¹」を合成したものである。当時、Abrams は Match.com

¹Napster は、1999 年に運営が開始されたファイル共有サービス。

という出会い系サイトに強く関心を寄せていた。そこで、各ユーザーが Match.com のようなページをそれぞれ持ち、それらの間にリンクを貼ってネットワークを形成する、という着想を得たことが、Friendster 開発のきっかけであった [10]。Abrams が出会い系サイトにソーシャルネットワークの要素を取り入れた背景には、『「友達の友達」は「赤の他人」よりも良い交際相手となりやすい』という仮定がある [7]。そのため、Friendster において各ユーザーは、自分からの距離が 4 未満であるユーザーのページへのみアクセスが許可されていた。

Friendster は運営開始後すぐに大きな反響を呼び、多くのユーザーを獲得した。登録ユーザー数は、運営開始 3 ヶ月後の 2003 年 6 月には 835000、その 4 ヶ月後には 200 万超に達した [10]。Time 誌は Friendster を「2003 年の最大の発明の一つ」と評し、Entertainment Weekly 誌は Abrams をその年の「最もフレンドリーな男 (Friendliest Man of the Year)」と名付けた [10]。このほかにも、多くの雑誌で取り上げられた [26]。こうした成功をうけて、2003 年末には Google が 3000 万ドルでの買収を申し出た。しかし、Abrams はこれを却下し、Kleiner Perkins Caufield & Byers から 1000 万ドルの資金援助を受けて運営を継続することを選択した [26]。

アメリカでの人気低下～アジア進出

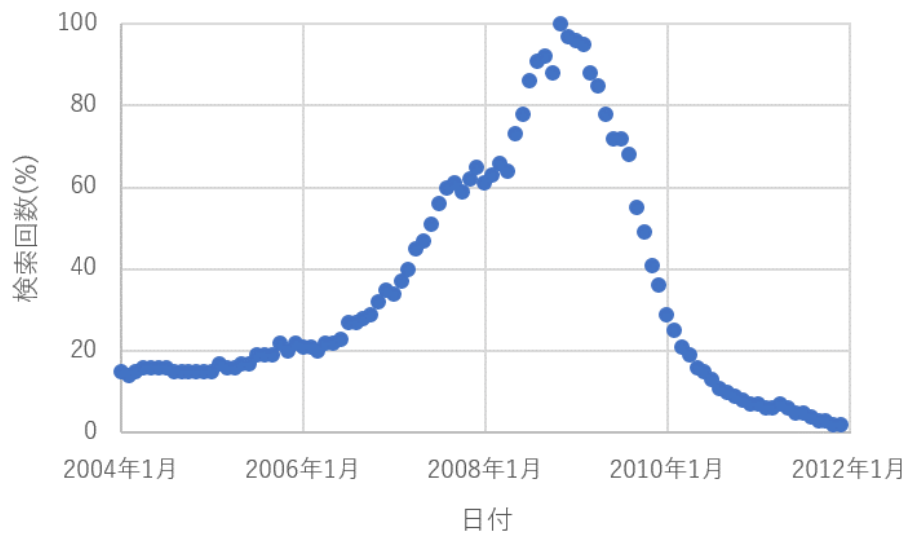
しかし、Friendster はすぐに大きな問題に直面した。前述したように、Friendster において各ユーザーは距離が 4 未満であるユーザーのページにのみアクセスが許されていた。時々刻々と構造を変化させるユーザーネットワークの中でこのような仕様を実現するためには、ユーザーが他のユーザーのユーザーページにアクセスしようとするたびに、2 ユーザー間の距離を計算する必要があった。ユーザー数の急増により、この計算に多大な時間を要するようになったため、ページの読み込み速度が低下し、ユーザーの不満が募っていった [10]。

こうした問題と、これに対処しようとした際に発生したさまざまな混乱、さらには MySpace など他の SNS の台頭により、Friendster は急速にアメリカにおける影響力を失っていった。2004 年末には、MySpace のアメリカ人アクティブユーザー数が 2200 万人に達したのに対して、Friendster は 110 万人に過ぎなかった [26, 10]。

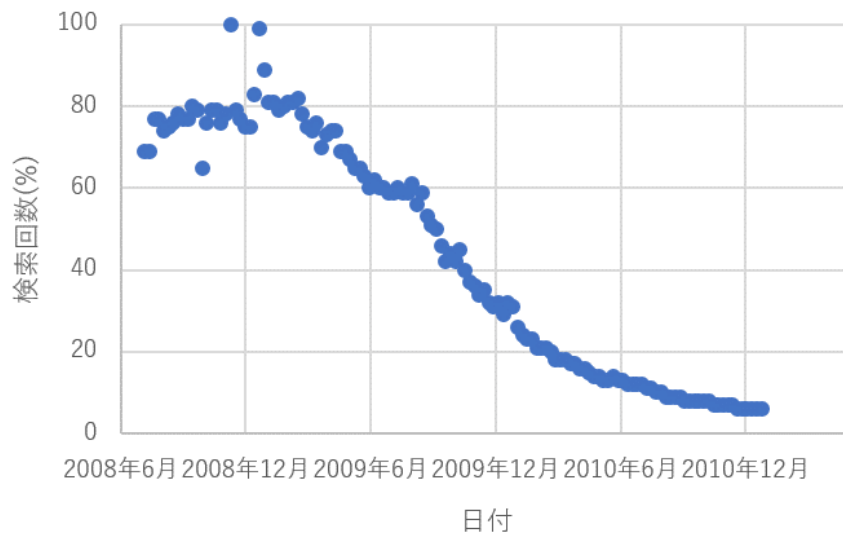
その一方で、Friendster は東南アジアにおいて人気を獲得しはじめていた。広告戦略や技術面での大幅な見直しが迫られるため、アジアからのアクセスは運営上必ずしも喜ばしいことではなく、Friendster はアジア人ユーザーを尊重するか切り捨てるかという選択に直面することとなった [10] が、その後の Friendster はしだいに東南アジアを中心とする成長戦略へと転換していった。特に、2008 年に最高経営責任者に就任した Richard Kimber は、アジア重視の姿勢を明確に打ち出している [32]。こうした戦略は功を奏し、2008 年 8 月にはマレーシアにおけるアクティブユーザー数が Facebook の 2.5 倍、MySpace の 3.5 倍に達し、全世界の累計ユーザー数は 8500 万人に達した [12]。

衰退～サービス終了

しかしながら、その後徐々に Friendster に衰退の兆候が見え始める。2009 年末に、Friendster はマレーシアに本社を置くインターネット企業である MOL Global に買収された。このときの買収額は 2640 万ドル



(a) 2004 年 1 月～2011 年 12 月



(b) 2008 年 6 月～2010 年 12 月

図 2.1: Google における「www.friendster.com」の検索回数の時間変化. (a) データが残っている 2004 年 1 月以降 (2011 年 12 月まで). (b) 衰退期. ただし, 対象期間を (a) では 1 か月ごと, (b) では 1 週間ごとの区間に分けて検索回数を集計し, 最大値を 100 %としたときの相対値をプロットしている.

であったが, 2009 年初頭の段階では 9800 万ドルから 2 億 7300 万ドルの価値を有するとみなされていたことから [3], 短期間で大幅に価値が低下したといえる [2]. また, Garcia[13] によれば, Google における「www.friendster.com」の検索回数は 2009 年の夏ごろから減少しはじめ, その後は減少の一途をたどった (図 2.1).

2011 年 6 月, Friendster はついに SNS 関連のサービスを終了し, オンラインゲームサイトとして再出発した. その後, 2015 年にはオンラインゲーム関連のサービスも終了し, その歴史に完全に幕を下ろした.

本論文では、2009年夏以降の衰退期を研究の対象とする。

2.2.2 Friendster の機能

Friendster の仕様や運営方針等の特徴について、二点補足する。

フレンドリストへの登録方法

「歴史」の節でも触れたように、Friendster において、各ユーザーは他のユーザーをフレンドリストに登録することができた。他のユーザーをフレンドリストに登録する手段としては、以下の三つが存在した [22].

1. サイト上に設置された招待ツールから、まだ Friendster に会員登録していない人に招待メールを送る。
2. サイト上に設置された検索窓から他のユーザーを検索し、友だちになることを申請する。
3. 他のユーザーのユーザーページに設置された "Add person as your friend" ボタンを押して、友だちになることを申請する。

いずれも、相手が招待・申請を受け入れた時点で、フレンドリストへの登録が完了した。

ユーザーの ID と登録時期

Friendster に会員登録したユーザーには、自動的にユーザーの ID が割り振られた。ID は連続的な整数値であり、会員登録が早かったユーザーほど小さな ID が割り振られていた。

2.3 ネットワーク科学の用語

本論文で用いる、ネットワーク科学の用語を定義する。

ネットワークは、頂点の集合 N と辺の集合 E から構成される (図 2.2)。すべての辺は、頂点集合 N に属する頂点のうちの 2 つ ($n_1, n_2 \in N$) を結んでいる。図 2.2 のネットワークに含まれる辺は向きを持たない無向辺であるが、向きを持つ有向辺を考えることもできる (図 2.3)。全ての辺が無向辺であるネットワークを無向ネットワーク、全ての辺が有向辺からなるネットワークを有向ネットワークと呼ぶ²。また、同じ 2 つの頂点 n_1, n_2 を結ぶ辺が複数存在する (多重辺) ことや、ある頂点から出た辺が同じ頂点に入るループは存在しないと仮定する。頂点 A と頂点 B を結ぶ無向辺が存在するとき、あるいは頂点 A から頂点 B に向かう有向辺が存在するとき、頂点 A は頂点 B と隣接していると呼ぶことにする。また、このとき頂点 B は頂点 A の友人である、と呼ぶことにする。

本論文では、SNS に属している各ユーザーを頂点として表し、ある 2 人のユーザー間にフォロー/被フォロー関係など何らかの関係が存在するとき、有向辺や無向辺を用いてこれを表現する。この頂点と辺からなるグラフをユーザーネットワークと呼ぶ。本論文では、「頂点」と「ユーザー」の 2 語を次のように使い

² 「有向辺と無向辺が混在するネットワーク」も有向ネットワークと呼ばれる場合があるが、本論文では有向辺と無向辺が混在するネットワークは扱わない。

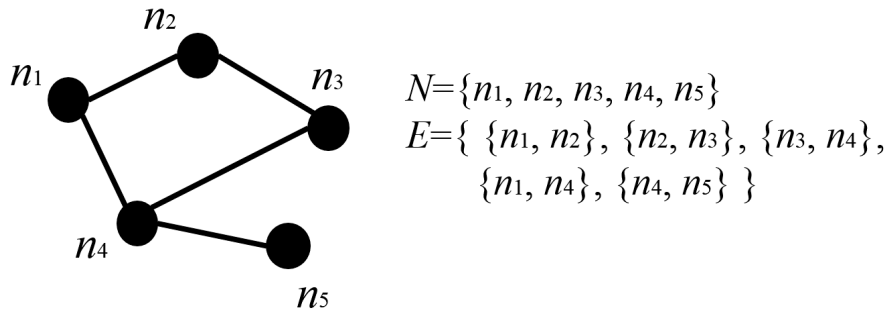


図 2.2: 無向ネットワークの例.

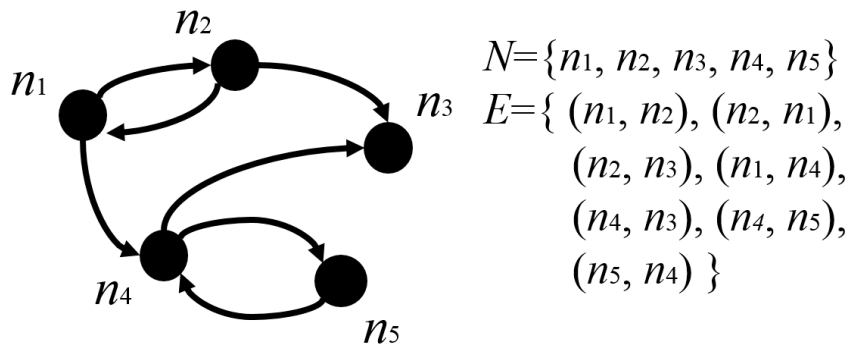


図 2.3: 有向ネットワークの例.

分ける。Friendster の実データを分析する際は「ユーザー」で統一する。コンピューターシミュレーションの際には、シミュレーションの方法や結果を説明する際には「頂点」を用い、得られた結果を SNS の言葉を用いて解釈する際には「ユーザー」を用いる。

無向ネットワークにおいて、ある頂点 n の次数 k_n とは、 n と他の頂点とを結ぶ辺の本数である。例えば、図 2.2 の頂点 n_4 の次数は 3 である。有向ネットワークにおいては、ある頂点 n から出ている辺の本数を n の出次数、 n に入ってくる辺の本数を n の入次数と呼ぶ。例えば、図 2.3 の頂点 n_1 の出次数は 2、入次数は 1 である。

与えられたネットワークのうちの頂点の部分集合 A の中から次数に比例する確率で 1 つの頂点を選ぶとは、頂点 $n (\in A)$ が選択される確率が $k_n / \sum_{i \in A} k_i$ となるように、 A の中から確率的に 1 つの頂点を選択することである。

ネットワーク中には多くの場合、頂点同士が密に辺で繋がり、強固なかたまりを形成している「コア」部と、頂点間のつながりが疎らな「周縁」部とが存在する。この「コア-周縁構造」を分析するための手法の一つが、 k -core decomposition 法 [29] である。 k -core decomposition 法は、無向ネットワークに対して定義される。 k -core decomposition 法ではまず、 k -core を定義する。

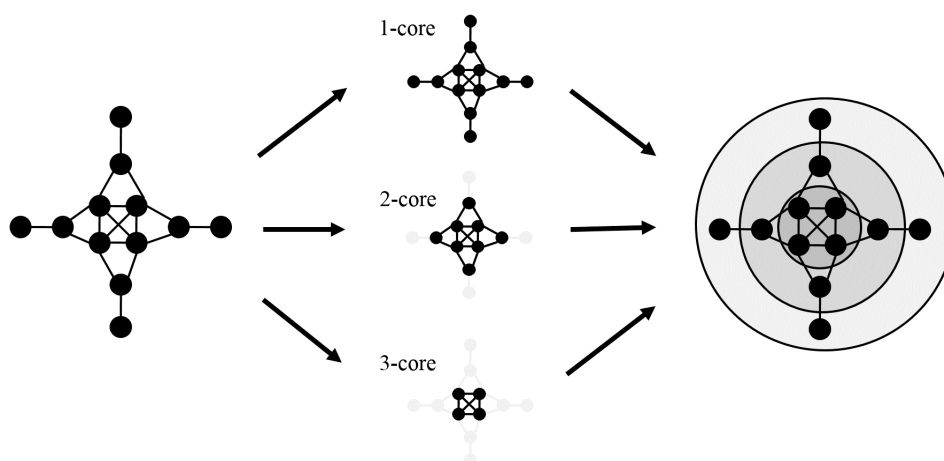


図 2.4: k -core decomposition 法により生成された階層構造（コア構造）。

定義 2.3.1 (k -core) 与えられたネットワークの部分ネットワークのうち、「そこに属するすべての頂点が、そこに属する k 個以上の頂点と隣接している」ような極大部分ネットワークを k -core とする。

定義より、与えられたネットワークの $(k+1)$ -core に属する頂点と辺はすべて k -core に含まれる。そのため、与えられたネットワークの 1-core, 2-core, 3-core, ... は階層構造（玉ねぎ構造）をなす（図 2.4）。

定義 2.3.2 (k -core decomposition 法) $k = 1, 2, 3, \dots$ に対応する k -core を順番に検出していき、これを用いてコアの階層構造をつくる手法を k -core decomposition 法という。

このような手法で生成されたコアの階層構造を、本論文ではコア構造と呼ぶ。

k -core を用いて、各頂点のコア値を定義する。

定義 2.3.3 (コア値) 頂点 i を含む k -core が存在し、かつ i を含む $(k+1)$ -core が存在しないとき、 i のコア値を k とする。

コア値の大きな頂点ほど、ネットワークのコア-周縁構造のうち、よりコアの中心部の近くに位置していると解釈する。

ネットワーク中の頂点全体のうち、次数が k であるものの割合 $p(k)$ を $k = 1, 2, 3, \dots$ について順に書き並べたものを次数分布という。次数分布がべき分布、すなわち $p(k) \sim k^{-\gamma}$ であるネットワークをスケールフリーネットワークと呼ぶ。次数分布がべき分布であるとき、両対数グラフに次数分布をプロットすると直線状となる。現実世界でみられる多くのスケールフリーネットワークでは $2 \leq \gamma \leq 3$ となっている [39]。

ネットワーク中の頂点全体のうち、次数が k 以上であるものの割合 $P(k)$ を $k = 1, 2, 3, \dots$ について順に書き並べたものを累積次数分布という。累積次数分布は次数分布を積分したものに相当するため、次数分布が $p(k) \sim k^{-\gamma}$ であるとき、 $P(k) \sim k^{-\gamma+1}$ となる。そのため、両対数グラフに累積次数分布をプロットすると直線状となる。積分することによりプロット列の細かな変動がならされて傾向を読み取りやすくなるため、得られた次数分布がべき分布であるか否かを判断する際には、次数分布の代わりに累積次数分布を用いることがある [39]。

スモールワールド性のあるネットワークとは、その中に埋め込まれた人々が「世間は狭い (It's a small world!)」と感じるようなネットワークである。厳密には、1) クラスタ係数が大きい、2) 平均距離が短い、の2点をともに満たすようなネットワークのことを指す [34].

クラスタ係数は以下のように定義される。まず、次数 k_i である頂点 i のクラスタ係数 C_i は、

$$C_i \equiv \frac{\text{頂点 } i \text{ を含む三角形の数}}{k_i(k_i-1)/2} \quad (2.1)$$

と定義される。ただし、「三角形」とは、3つの頂点が3本の辺でつながりあっている構造を指す。すなわち C_i は、頂点 i の次数 k_i が決定しているという条件下で、「頂点 i を含む三角形の、実際の数」を「頂点 i を含む三角形の数がとりうる値の最大値」で割った値である。以上の定義のもとで算出した各頂点のクラスタ係数を、ネットワーク中の全頂点について平均したものが、ネットワーク全体のクラスタ係数 C である。つまり、

$$C \equiv \frac{1}{N} \sum_{i=1}^N C_i \quad (2.2)$$

である。直感的には、クラスタ係数の大きなネットワークとは、「三角形が多いネットワーク」であり、「友達の友達が友達である傾向が強いネットワーク」である。

平均距離 L は、ネットワーク中のある頂点 i から別の頂点 j へ到達するまでに通る最小の辺の本数 $d(i, j)$ を、全ての2頂点について平均した値である。すなわち、

$$L = \frac{2}{N(N-1)} \sum_{\substack{i, j \\ i < j}} d(i, j) \quad (2.3)$$

である。

第3章 先行研究

本論文が対象とする Friendster は SNS の一種である。本論文ではネットワーク科学や影響伝播モデルによる知見を利用して分析を進める。そこで本章では、SNS・ネットワーク・伝播モデルに関する既存研究のうち、本研究との関わりが深いものを中心として概観する。

3.1 ネットワーク構造に着目した、SNS の成長・衰退過程の分析

ネットワーク科学の知見を用いて、SNS 全体や SNS 内で形成されたグループを分析する研究は多く行われている。

Viswanath[33] や Wilson[36] は、Facebook 上のあるグループが成長する過程におけるネットワーク構造の分析を行った。特に、「友達ネットワーク」と「実際にメッセージのやりとりを行っているユーザー間を辺で結ぶことによって構成したネットワーク (activity network)」の差異に着目して分析を行った。彼らが考えた activity network と本論文におけるアクティブネットワーク (4.2.1 節) との間には類似性がある。ただし、彼らは主に activity network 中の辺の強度・持続性や、activity network の定義のしかたの違いによるネットワークの性質の違いに注目して分析を行っている。そのため、Facebook 上において activity network ネットワーク全体の構造がどのように時間変化したのかは十分明らかにされていない。また、本論文で分析対象とする Friendster においては、ユーザー間のやりとりの履歴を示すようなデータセットが存在しないため、activity network の時間変化のようすを確かめることは難しい。

Kumar[18] は、Flicker と Yahoo!360 のネットワーク成長過程を分析し、ネットワークは「小さな連結成分が最大連結成分に吸収される」ことの繰り返しによって成長したと主張した。Qiu[23] は中国の SNS である WeChat 内において、開設直後にどのようなユーザーネットワーク構造を持つグループがメンバー数を大きく増加させていく傾向があるかを分析した。Qiu によれば、グループへの招待者と被招待者とを辺で結んだ木の高さがグループ開設直後から大きければ、ユーザー数はその後大きく増加する確率が高い。Kumar や Qiu による知見や分析手法は、SNS 上のネットワークが成長する過程や成長が鈍化する過程を分析する上では有効である可能性があるが、ネットワークからユーザーが離脱していく過程を理解する際に活用することは難しい。

Torkjazi[31] は MySpace という SNS が衰退する過程を分析した。MySpace のユーザーページには、そのユーザーが最後に MySpace にログインした日付が記載されている。ネットワーク構造と最終ログイン日を用いた分析により、友人関係にある 2 ユーザーは近い時刻に MySpace を離脱した傾向があることを、Torkjazi は示した。本論文で分析対象とする Friendster においては、一部のユーザーのユーザーページは参照可能であり、これを用いた分析も行っている (4.4.1 節)。しかし、Friendster の全ユーザーのユーザーページを参照することは不可能であり、またユーザーページには最終ログイン日が記載されていないため、彼らの手法

を用いて分析を行うことは難しい。

3.2 Friendster の衰退過程の分析

Friendster が衰退した過程について、ネットワーク構造に着目して分析した研究に、Garcia[13] と Yu[37] がある。

3.2.1 Garcia のモデル

Garcia[13] は、コスト・ベネフィット論的な枠組みを用いて、Friendster が衰退した過程を説明することを試みた。

Garcia は、各ユーザーは合理的に行動したと仮定した。すなわち、Friendster に留まることによって得られる利益が必要なコストよりも大きい場合には留まり、そうでない場合には離脱すると仮定した。ユーザー i が得る利益 $B_i(H)$ は、その時点でネットワークに留まっているユーザーの集合 H の関数として定義され、一方、コスト c は定数であると仮定した。Friendster において $B_i(H)$ がどのような形をとっていたかは明らかではないが、Garcia はひとまず

$$B_i(H) = bN_i \quad (3.1)$$

と仮定した。ただし、 N_i は「ユーザー i と隣接し、かつ H に属するユーザーの数」である。つまり、各ユーザーが得る利益は「その時点でネットワークに留まっている友人の数」に比例する、と仮定したことになる。

以上の仮定のもとでは、「離脱のカスケード (cascades of departures)」という現象が発生する。まず、 $bN_i < c$ であるユーザーが離脱する。その結果、その他のユーザーの bN_i の値が変化し、新たに $bN_i < c$ となるユーザーが出現する。そのため、それらのユーザーも離脱する。以降同様のことが繰り返されて、最終的にはある状態で安定する。このとき、留まっているユーザーからなるネットワークは、 $k = c/b$ の場合の k -core (第2章) となる。

Garcia がおいた仮定のもとでは、 b と c の値によって決定されるある閾値 $k_c = c/b$ が存在する。コア値が k_c 未満であるユーザーは、「離脱のカスケード」の結果として遠からずネットワークを離脱する。一方、コア値が k_c 以上であるユーザーは、ネットワークに留まり続ける。つまり、ユーザーの意思決定は、そのユーザーのコア値と閾値 k_c との関係によって決定づけられる。

Garcia は更に、閾値 k_c が一定のペースで増加し続けたという仮定を加えた。その原因は、Friendster 上におけるユーザーの活動の活発度が低下し、各々の繋がりが希薄化したことによって、利益の比例係数 b の値が低下したことであると述べた。このように k_c が上昇することで、Friendster に留まるユーザーは次第に減少していった、と Garcia は主張した。

Garcia は、以上のモデルによって Friendster の衰退過程をよく説明できると主張した。その根拠は、4.1 節で紹介するデータセットが示すユーザーネットワーク全体のコア構造を k -core decomposition 法で分析した上で、

- 衰退が開始した 2009 年 6 月には、データセット中のコア値が 3 以上であるユーザー全員が Friendster に留まっており (つまり、閾値 k_c が 3 であり)、

- その後一定のペースで k_c が増加していき（それに伴ってコア値が k_c 未満であるユーザーが離脱していき）,
- 2010年6月に $k_c = 67$ に達した.
- また、この期間において新規ユーザーの加入は一切発生しなかった.

と仮定すると、ネットワークに留まり続けているユーザー数の時間変化の様子と、この期間における実際のアクティブユーザー数の時間変化とがよく一致することである。ただし、Friendsterのアクティブユーザー数の推移は公表されていないため、Googleにおける「www.friendster.com」の検索回数をアクティブユーザー数とみなした。

3.2.2 Yuのモデル

Yu[37]は、FriendsterにおいてGarciaがおいた「 $k(=c/b)$ が一定のペースで増加した」という仮定の妥当性を示す根拠が十分でないとして、このような仮定を置くことなくネットワークの縮小を説明できるモデル(KQ-cascade model)を提案した。

Yuは、Friendster上の友人のうちアクティブに活動しているユーザーの「数」と「割合」が、各ユーザーの意思決定に影響を与えたと考えた。具体的には、次の2つの条件のいずれかを満たすとき、ユーザーは確率 f でFriendsterを離脱（非アクティブ化）すると仮定した。

1. アクティブであるユーザーとの繋がり数がある閾値 k_s 以下であるとき（Friendsterに留まっても、十分大きな利益を得られないため）
2. 隣接しているユーザーのうち割合 q 以上が離脱したとき（Friendsterに留まることで、突然のサービス終了による不利益を被る危険や、他のよりよいサービスを利用する機会を逸する危険が高まっているため）

Yuは、パラメータが $k_s = 20$, $q = 0.2$, $f = 0.15$ の場合に、衰退期におけるアクティブユーザー数の時間変化がよく説明できると主張した。

3.3 影響伝播モデルを用いたオンラインコミュニティの分析

オンラインコミュニティのライフサイクルを理解するうえで有益なもう一つの道具が、影響伝播モデルである。影響伝播モデルとは、「各ユーザーは各時刻に有限個の状態のうちの一つをとり」、「各ユーザーは周囲のユーザーの影響を受けることによって状態を遷移させる」と仮定するようなモデルを指す。伝播現象に関しては、既に多くの学問分野で膨大な研究成果が蓄積されている。例えば、SISモデルやSIRモデルなどの感染症伝播モデル[15]、新製品や新たなアイデアが普及する過程を模した普及モデル[27]などが挙げられる。そのため、ユーザーが他のユーザーの影響を受けてコミュニティへ加入するという現象を「感染」や「普及」の一形態とみなして、既存の影響伝播モデルに関する知見を適切に活用することができれば、オンラインコミュニティに関する理解を深めることができる。

影響伝播モデルを用いて SNS のユーザーネットワークの時間変化を分析した研究としては、以下のものが挙げられる。Cannarella[9] は MySpace という SNS の成長と衰退の過程を、SIR モデルという感染症伝播モデルをもとにした irSIR モデルを用いて説明した。Zhu[38] は、感染症伝播モデルにいくつかの規則を加えることで、中国の SNS である Renren Network と Sina Weibo のアクティブユーザー数の時間変化を説明できると主張した。Bauckhage[6] は、数多くの SNS の成長と衰退の過程が、いくつかの単純な統計モデルによって説明可能であると主張した。Tanaka[30] は、日本における Mixi から Facebook へのユーザー流出の過程を、SIR モデルを用いて論じた。Ribeiro[24] が提唱した影響伝播モデルは、活発な状態が長続きする SNS と、すぐに衰退する SNS との相違点を明らかにした。Ribeiro[25] のモデルは、Ribeiro[24] のモデルにさらにいくつかの仮定を加え、モデルを精緻化することにより、様々な SNS のアクティブユーザー数の時間変化を同時に説明し、また 2008 年ごろから Facebook が一人勝ちし始めたメカニズムを説明した。Nazir[21], Liu[20] は、Facebook 上の贈り物授受アプリによって形成された、贈り物授受ネットワークの時間変化を説明するためのモデルを提案した [21, 20].

以上で挙げた先行研究のモデルのうち、本論文では特に

- irSIR モデル (Cannarella[9]) (第 5 章, 第 6 章)
- Ribeiro のモデル (Ribeiro[25]) (第 6 章)

を参考とする。そこで、以下にこれらのモデルの概要をまとめる。なお、これらのモデルはいずれもネットワーク構造を考慮せず、全ての人がある他の人と等しく接触し、影響を及ぼし合う状況を仮定しているため、本論文ではネットワーク用に拡張したモデルを定義する。

3.3.1 SIR モデル

感染症の伝播現象を説明するためのモデルとして広く知られているものの一つが、SIR モデル [17] である。SIR モデルでは、以下のような仮定をおく。

- 状態
「まだ病気にかかっていない」か「現在病気にかかっている」か「既に治癒し、免疫を持っている」という 3 つの状態を考える。全ての人、この 3 つの状態のいずれかにあると仮定する。3 つの状態をそれぞれ S (Susceptible, 感受性保持者), I (Infected, 感染者), R (Recovered, 免疫保持者) と名付ける。
- 状態遷移
以下の規則による状態遷移が発生すると仮定する。
 - 状態 S の人と状態 I の人が接触すると、状態 S の人は確率 β で状態 I に遷移する。
 - 状態 I から状態 R への遷移は、他の人との接触機会の多寡によらず、つねに一定の確率 γ で発生する。

状態遷移は一方向に進行し、状態 S から状態 I, 状態 I から状態 R への遷移は発生するが、これ以外の遷移は発生しない。

状態 S から状態 I への遷移には状態 I との接触が必要だが、状態 I から状態 R への遷移には状態 R との接触を必要としない。これは、このモデルが感染症の伝播現象を模したものであるためであり、非感染者が伝染病に感染するためには周囲の感染者との接触が必要であるが、感染者が治癒するために周囲の「既に治癒した」人と接触する必要はないためである。

SIR モデルの性質のうち、本論文で使用するものを記述する。

状態 S から状態 I への遷移規則から、状態 S である人が時刻 A から時刻 B までのあいだに状態 I である人と x 回接触するならば、時刻 B までに状態 I に遷移する確率 P は、

$$P = 1 - (1 - \beta)^x \quad (3.2)$$

$$\approx \beta x \quad (\beta \ll 1, x \ll \infty) \quad (3.3)$$

となる。つまり、 β の値が十分小さいとき、 x が小さい範囲においては、状態 I に遷移する確率は状態 I である人との接触回数に比例する。

全ての人が他の全ての人と等しい頻度で接触するという仮定のもとでは、SIR モデルを表現する微分方程式を立てることができる。 $S(t), I(t), R(t)$ をそれぞれ、時刻 t における状態 S, I, R の頂点数とする。時刻 t に状態 S から状態 I に遷移する頂点数は $S(t)$ と $I(t)$ の積に比例し、時刻 t に状態 I から状態 R に遷移する頂点数は $I(t)$ に比例すると仮定する。このとき、

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \quad (3.4)$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \quad (3.5)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad (3.6)$$

という微分方程式を用いて、 $S(t), I(t), R(t)$ の時間変化を表現できる [9]。ただし、 β, γ はパラメータである。

3.3.2 irSIR モデル

Cannarella[9] は SIR モデル (3.3.1 節) をもとに、MySpace のアクティブユーザー数の時間変化を説明することを試みた。

Cannarella は、SIR モデルにおける 3 つの状態を以下のように解釈しなおした。

S : まだその SNS に会員登録していない状態

I : その SNS でアクティブに活動している状態

R : すでにその SNS を離れた状態

次に、SNS の性質に合うように SIR モデルを修正した。SIR モデルでは、状態 I から状態 R への遷移確率に対して、他者との接触の有無は影響を及ぼさないと仮定していた。一方、SNS においては、すでにその SNS を離脱した友人から触発されることが、アクティブユーザーが非アクティブ化する原因となりうると考えた。そこで、状態 I から状態 R への遷移も状態 S から状態 I への遷移と同様に、状態 R の人との接触によって引き起こされると仮定した。このように修正したモデルは次のような微分方程式で表現できる。

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \quad (3.7)$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t)R(t) \quad (3.8)$$

$$\frac{dR(t)}{dt} = \gamma I(t)R(t) \quad (3.9)$$

このモデルを irSIR モデル (infectious recovery SIR model) と呼ぶ。Cannarella は、パラメータ値を適切に調整したとき、この連立微分方程式の解として求まる状態 I のユーザー数の時間変化と、実際の MySpace のアクティブユーザー数の時間変化とがよく一致することを示した。

3.3.3 Ribeiro のモデル

Ribeiro[24] のモデルにおいて、人々は 3 つの状態 U, A, I のうちのいずれかをとる。

U： 未だその SNS のメンバーではない状態 (非ユーザー)

A： その SNS のメンバーとしてアクティブに活動している状態 (アクティブユーザー)

I： その SNS のメンバーではあるが非アクティブな状態 (非アクティブユーザー)

状態 U, A, I の意味するものはそれぞれ、irSIR モデルにおける状態 S, I, R に近いといえる。ただし、irSIR モデルにおける状態 R が「既に SNS を離脱した」状態であったのに対し、Ribeiro のモデルにおける状態 R は「SNS のメンバーではあるが非アクティブ」な状態を表すという差異は存在する。

状態遷移規則として、Ribeiro は次の 4 つを設定した。

規則 1 非アクティブユーザーとアクティブユーザーが接触すると、ある確率で非アクティブユーザーがアクティブユーザーとなる ($A+I \rightarrow 2A$)

規則 2 アクティブユーザーはある確率で非アクティブユーザーとなる ($A \rightarrow I$)

規則 3 アクティブユーザーと非ユーザーが接触すると、ある確率で非ユーザーが会員登録してアクティブユーザーになる ($U+A \rightarrow 2A$)

規則 4 宣伝等の効果により、ある確率で非ユーザーが会員登録してアクティブユーザーになる ($U \rightarrow A$)

irSIR モデルでは、状態遷移が $S \rightarrow I \rightarrow R$ と一方向に進行するが、Ribeiro のモデルでは状態 I から状態 A への遷移を仮定した点 (規則 1) は大きな違いである。Ribeiro のモデルにおいては、状態 I から状態 A への遷移が十分多く発生するようにパラメータ値を設定したとき、状態 A であるユーザーの数は正のある値に収束し、アクティブユーザーは完全には消滅しないことが示されている。

第4章 Friendsterのコア構造の分析

本章では、FriendsterのユーザーネットワークがGarciaやYuらのモデルに従って縮小したのかを検証するため、ユーザーネットワーク構造の時間変化を分析し、考察する。

GarciaとYuのモデル(3.2節)はいずれも、これまで十分に解明されてこなかったFriendsterの衰退のメカニズムに関して新たな示唆を与えたという点で意義がある。しかし、GarciaとYuのモデルによりアクティブユーザー数の推移がよく説明できたとしても、実際のユーザーネットワーク構造の時間変化に関しては確かめられていない。

そこで本章では、ユーザーネットワークの時間変化、特にコア構造の時間変化について分析する。

4.1 データセット

Friendsterのユーザーネットワークの構造を示すデータセットを、Internet Archive [16]が公開している。Friendsterにおいて、各ユーザーは他のユーザーを自らのフレンドリストに追加することができた。つまり、全ユーザーのフレンドリストを得ることができれば、Friendster上に形成されたユーザーネットワークの構造を知ることができる。そこで、Internet ArchiveはFriendsterがSNS関連のサービスを終了する直前の2011年6月に、全ユーザーのフレンドリストを探索し、その結果をデータセットとして公開した。このデータセットにおいて、各ユーザーはIDによって表わされている。2.2.2節で述べたように、このIDはユーザーが会員登録を行った順番に割り当てられた、連続的な整数値である。つまり、Friendsterの設立直後に会員登録したユーザーには小さなID(例:1000)が、閉鎖直前に会員登録したユーザーには大きなID(例:100000000)が割り振られている。

このデータセットが表わすネットワークは、辺に向きのあるネットワークである。ユーザーAがユーザーBをフレンドリストに登録しているとき、AからBへ向かう辺が存在するとみなしている。Internet Archiveによれば、Friendsterにおいて仮にAのフレンドリスト内にBが存在していても、Bのフレンドリスト内にAが存在しているとは限らない[16]ため、このように有向ネットワークとみなすことが妥当であり、以下では有向ネットワークとみなして分析を進める。ただし、ほとんどの場合において辺 $A \rightarrow B$ が存在すれば辺 $B \rightarrow A$ も存在しているので、無向ネットワークとみなした場合の分析も行った(4.4.4節)。

データセットは124,836,751のユーザー(頂点)と、2,586,147,869の友だち登録(有向辺)から構成される。このうちの5.8%にあたる7,248,621のユーザーは既にユーザーページを削除しているため、データセットにはそれらのユーザーの辺情報が含まれていない。また、全体の11.2%にあたる14,001,031のユーザーはユーザーページを非公開に設定していたため、それらのユーザーから出た辺の情報はデータセットに含まれていない。なお、シリアルIDの小さいユーザー(=早期にFriendsterに会員登録したユーザー)ほど、ユーザーページを非公開にしている比率が大きい(図4.1)。ユーザーページを非公開にしていたユーザー

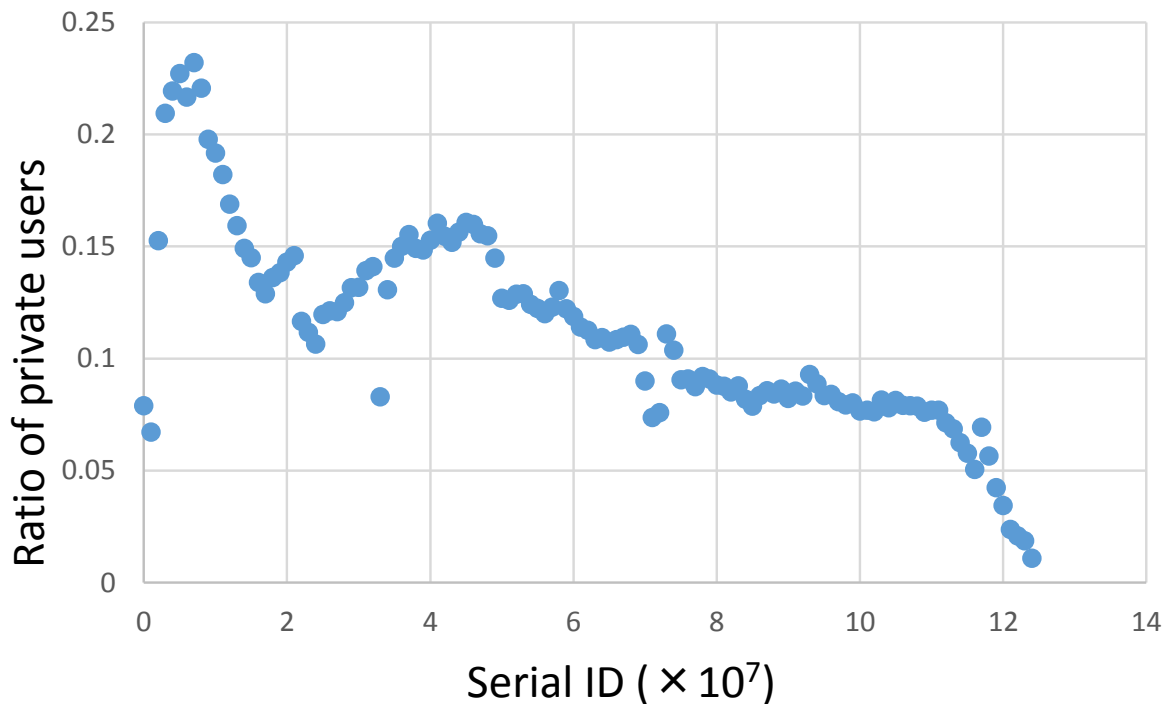


図 4.1: ID と、ユーザーページを非公開にしているユーザーの割合との関係。

はユーザーネットワークにおいて重要な役割を果たしていた可能性もあり、これらのユーザーの動向をより正確に把握し分析することは今後の課題である。

このデータセットが収集された 2011 年 6 月の時点では、大部分のユーザーはすでに Friendster 上での活動を行っていなかった。しかし、上述したようにユーザーページが削除されているアカウントは 5.8% に留まっており、大部分のユーザーはアカウントを放置したまま Friendster を離脱したものと考えられる。そのため、このデータセットが示すネットワークは、「Friendster の初期から末期までに形成されたアカウントや辺の情報が全て含まれたネットワーク」とみなすことにする。

SNS 上のユーザーネットワークに関するデータセットは、ほかにも多数存在する。しかしその中でも、InternetArchive によって収集されたこの Friendster のデータセットは貴重である。なぜならば、これほど大規模な SNS において、

- ほぼ全てのユーザーを含むユーザーネットワークの構造が示されており、かつ、
- (このあとの 4.2 節で説明するように) ネットワーク構造の時間変化の様子を推定できる。

ようなデータセットは、著者の知る限りほかに皆無だからである。

Garcia[13] が指摘したように、このような有益なデータセットが公開されているのは、Friendster がすでにサービスを終了しており、データセットを公開しても利害問題が発生しないためであると考えられる。皮肉にも、Friendster が閉鎖されたおかげで、われわれは Friendster が閉鎖されるまでの過程に関する理解を深め、SNS の運営に関する教訓を得ることができるのである。

4.2 分析方法

本章で用いる分析方法は、次の二段階に分けられる。

1. Friendster の運営が開始されてからサービス終了までの各時刻におけるアクティブネットワークを推定する。
2. アクティブネットワークのコア構造を分析する。

アクティブネットワークとは、与えられたネットワーク全体のうち、その時刻のアクティブユーザーのみからなる部分ネットワークである。

4.1 節のデータセットは Friendster のサービス終了直前の、ある一つの時刻におけるネットワーク構造が記録されたものである。このデータセットには、ネットワーク中の各ユーザー・辺が生成された時刻などを示す情報は含まれていない。しかし、各ユーザーの ID から、ユーザーが会員登録を行った順序を知ることが可能である。これを利用して、ネットワーク構造が時間とともに変化した様子を推定する。以下の分析ではユーザーの ID を用いて、時刻を表している。例えば、「時刻 10000」は、「ID10000 を割り当てられたユーザーが会員登録した時刻」の意味で用いている。ID を用いた時刻と実際の時刻とのズレについては 4.4.1 節で考察する。

分析の詳細な手順を以下に示す。

4.2.1 各時刻におけるアクティブネットワークの推定

まず、時刻 T におけるアクティブユーザーを推定し、それを用いて、ユーザーネットワーク全体 (N, E) のうち、時刻 T におけるアクティブネットワークを推定する。

ここでは、以下の 2 つの条件をともに満たすとき、ユーザー i は時刻 T にアクティブであったと仮定する。

1. ユーザー i は時刻 T にすでに会員登録済みである。
2. ユーザー i は時刻 T 付近に、Friendster 上でなんらかの活動を行っている。

Friendster において、ユーザーはさまざまな種類の活動（つながりの形成、メッセージのやりとり、コンテンツのアップロード等）を行っていた。しかし、分析対象としているデータセットから、全ての種類の活動の様子を窺い知ることはできない。そこで、「ユーザー i が友人 A をフレンドリストに登録している」ときに、「友人 A が会員登録した時期に、ユーザー i も活発に活動していた可能性が高い」と推定することにする。ユーザー i が友人 A をフレンドリストに登録した時刻は、 A が会員登録を行った直後とは限らないが、他の情報がないため、ここではこのように仮定する。

仮定 4.2.1 (アクティブユーザー) 次の 2 つの条件をともに満たすとき、ユーザー i は時刻 T におけるアクティブユーザーとする。

1. i の ID は $T + \delta$ 未満とする。
2. i は ID が T 以上 $T + \delta$ 未満である 1 人以上のユーザーへの辺を持つ。

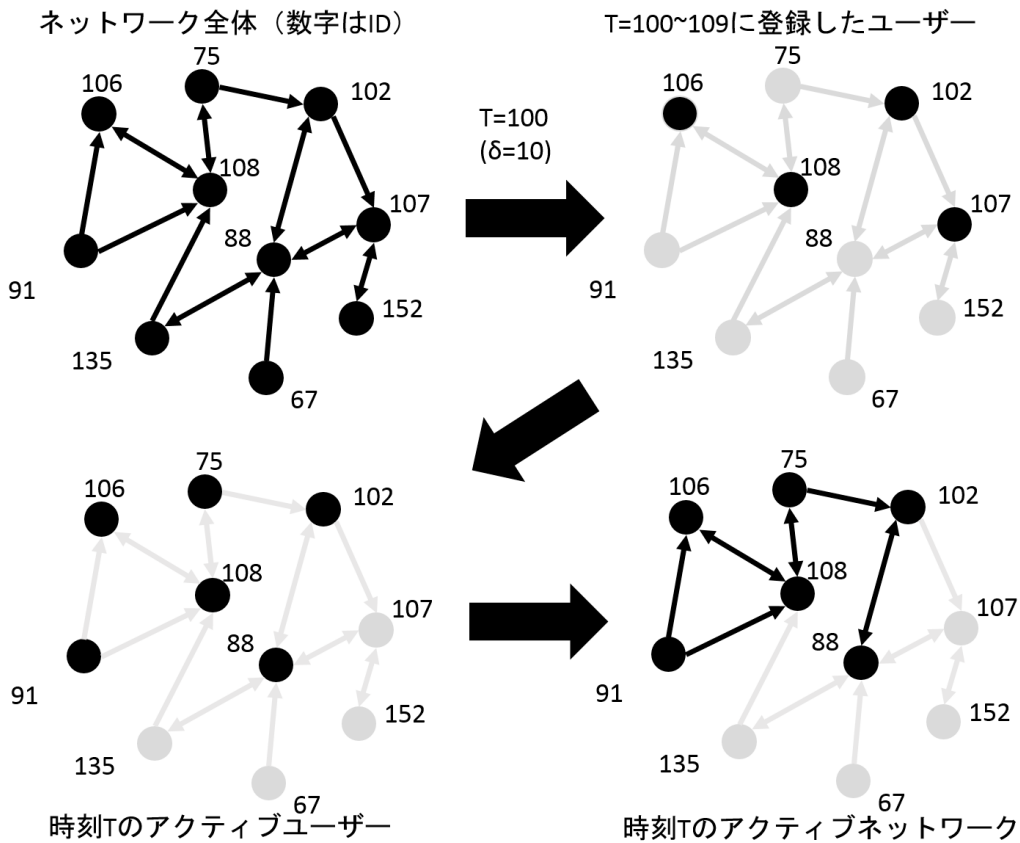


図 4.2: アクティブネットワークの推定方法 (時刻 $T = 100$, $\delta = 10$) .

ただし, δ はパラメータである.

本論文では断りがない限り $\delta = 10^6$ として分析を行う. 10^6 は, 単位時間あたりの会員登録者数が最も多かった時期でおよそ 10 日に相当する (4.4.1 節). なお, δ の値をある程度変更しても同様の結果が得られることを, 4.4.2 節で確かめる.

次に, 各時刻におけるアクティブな辺と, アクティブネットワークの条件を仮定する.

仮定 4.2.2 (アクティブな辺) データセット中の辺 e が結ぶユーザーが, とともに時刻 T にアクティブであるとき, 辺 e は時刻 T におけるアクティブな辺と仮定する.

仮定 4.2.3 (アクティブネットワーク) 時刻 T におけるアクティブユーザーとアクティブな辺から構成される部分ネットワークを, 時刻 T のアクティブネットワークと仮定する.

図 4.2 は, 以上で仮定したアクティブネットワークの推定方法を図示したものである.

4.2.2 有向 k -core decomposition 法

アクティブネットワークのコア構造の分析には, k -core decomposition 法 [29] を用いる. ただし, ここでは辺に向きのあるネットワークを扱っているため, 無向ネットワークに対して定義された k -core decomposition

法をそのまま用いることはできない。そこで、有向ネットワーク用の k -core と k -core decomposition 法を定義する。

定義 4.2.1 (有向 k -core) 与えられたネットワークの部分ネットワークのうち、「そこに属するすべての頂点が、そこに属する k 個以上の頂点への辺を持っている」極大部分ネットワークを有向 k -core と定義する。

定義 4.2.2 (有向 k -core decomposition 法) $k = 1, 2, 3, \dots$ に対応する有向 k -core を順番に検出していき、これを用いてコアの階層構造をつくる手法を有向 k -core decomposition 法と定義する。

4.3 分析結果

本節では 4.2 節の手法を用いて、Friendster のコア構造の時間変化を分析する。4.3.1 節では、コア構造の時間変化を可視化し、おおまかな傾向を把握する。4.3.2 節ではいくつかの指標を導入し、定量的に傾向を確かめる。

4.3.1 コア構造の時間変化の可視化

与えられたネットワークの k -core を有向 k -core decomposition 法によって検出すると、「 k -core には含まれるが $(k+1)$ -core には含まれないユーザー」の数がわかるので、これを左から順に並べ、一本の帯として表すことによって、ネットワーク中のユーザーのコア値の分布を視覚的に表す (図 4.3. 図では k を 5 刻みで表わしている)。帯の全長から、ネットワークに含まれる全ユーザー数を読み取ることができる。

以上の描画方法によって、Friendster のアクティブネットワークのコア構造が時間とともに変化していく様子を示したものが、図 4.4 である。ここで、各々の帯はそれぞれの時刻におけるコア構造を示している。最も上の帯は時刻 0.5×10^7 におけるコア構造、次の帯は時刻 1×10^7 におけるコア構造、その次の帯は時刻 1.5×10^7 におけるコア構造、... を表している。なお、コア値の高いユーザー数の時間変化を読み取りやすくするため、グラフは右揃えで描画している。

図 4.4 によれば、コア構造の中心部 (コア値の高いユーザー群) に位置するユーザーの数は時刻 8.5×10^7 付近で減少し始める。例えば、時刻 8.5×10^7 の帯の右端に存在する水色の断片と灰色の断片との境界 (100-core) や、その左側の境界 (95-core) は、次の時刻 9×10^7 には右に移動している。その一方で、アクティブユーザー数の合計値は時刻 9×10^7 付近でピークを迎えている。つまり、コア中心部の縮小はアクティブユーザー数の減少よりも前に始まっている。

この付近を詳細に検討するために、時刻 8.5×10^7 から 10×10^7 までの範囲におけるコア構造の時間変化を、 0.1×10^7 刻みで図 4.5 に示す。図中の右側に位置する境界線ほど、早い時刻から右に移動し始めていることが読み取れる。つまり、より内側のコアほど早く、拡大から縮小に転じていることが読み取れる。ネットワーク周縁部に属するユーザー数の時間変化を確かめるため、図 4.5 と同じ分析結果を左揃えで描画したものを図 4.6 に示す。最も左側の断片や、左から二番目の断片の長さがほぼ一定であることから、ネットワークの周縁部 (コア値の低いユーザー群) に位置するユーザーの数は、この期間内においてはほぼ一定であることがわかる。

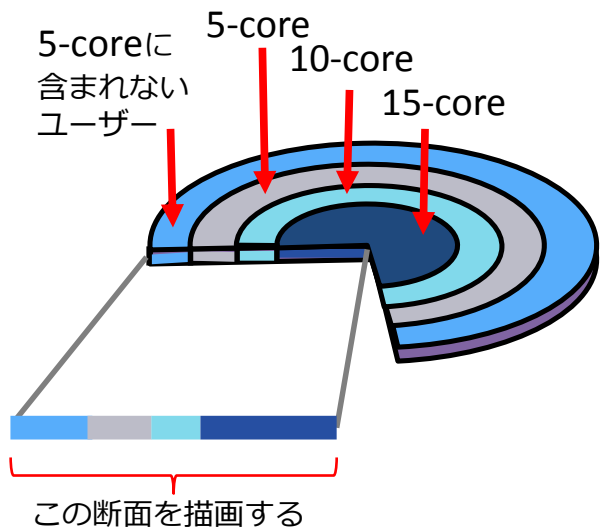


図 4.3: コア構造の描画方法. 各コアに属するユーザー数を, 円の半径によって表わす.

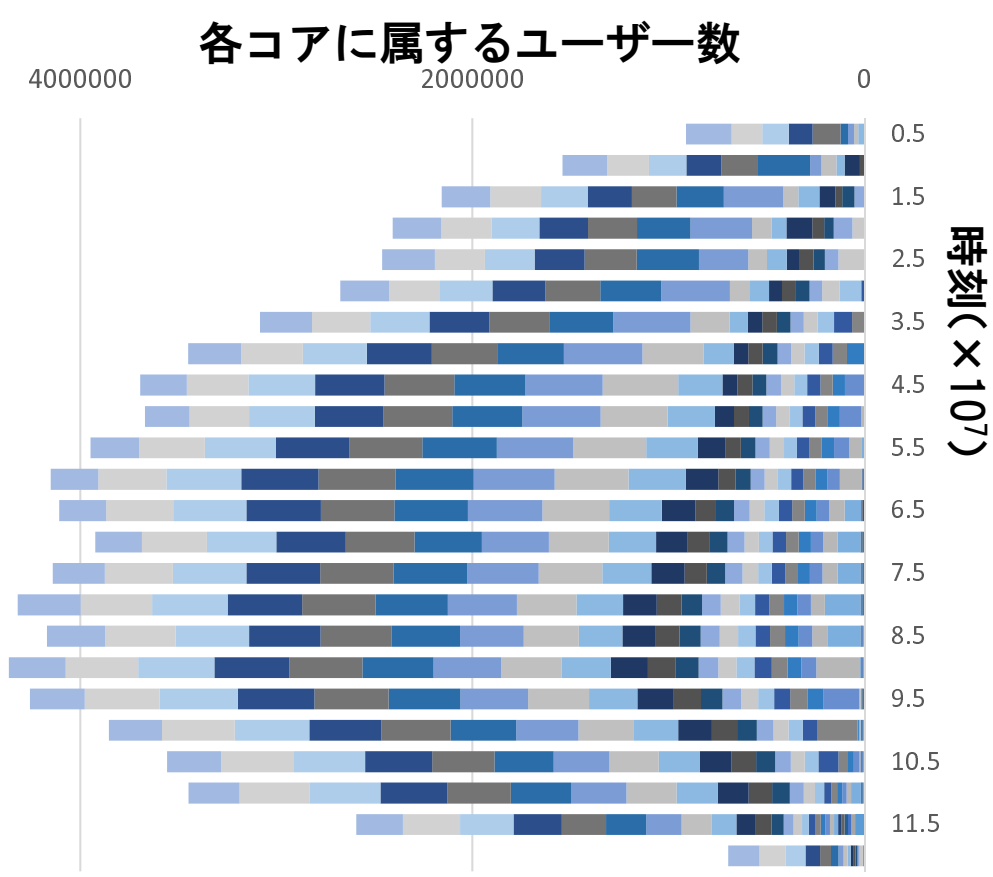


図 4.4: Friendster のユーザーネットワークにおける有向 k -core 構造の時間変化.

4.3.2 指標に基づいた評価

コア構造の時間変化のしかたに関して, いくつかの指標を用いて, 定量的に傾向を確かめる.

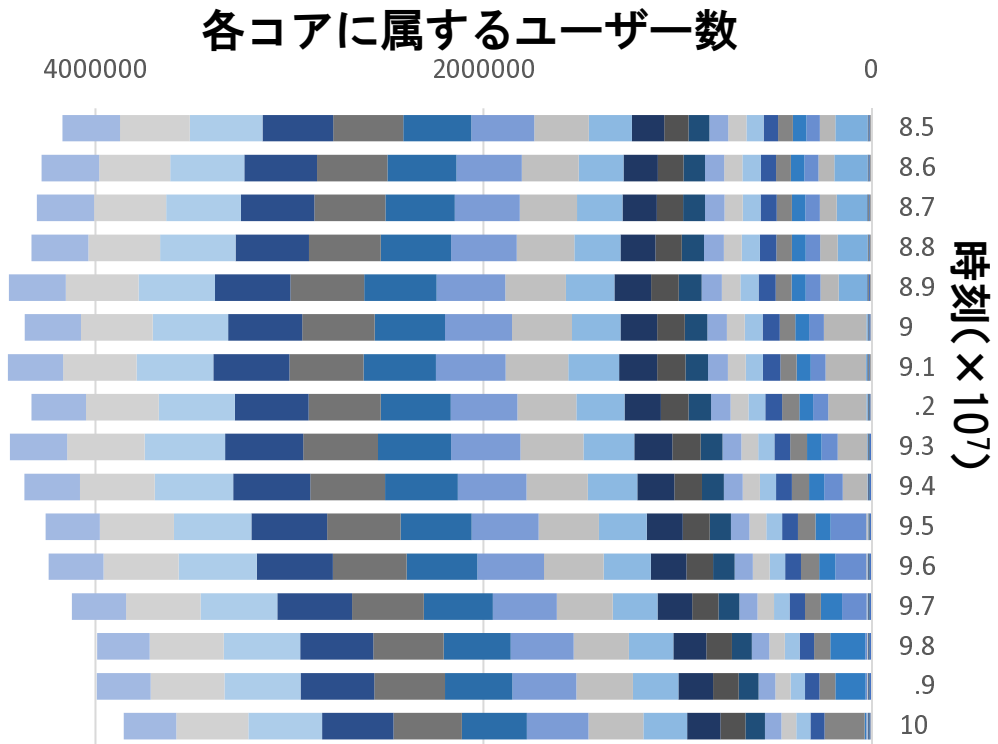


図 4.5: 時刻 8.5×10^7 から 10×10^7 までの、アクティブネットワークのコア構造の時間変化.

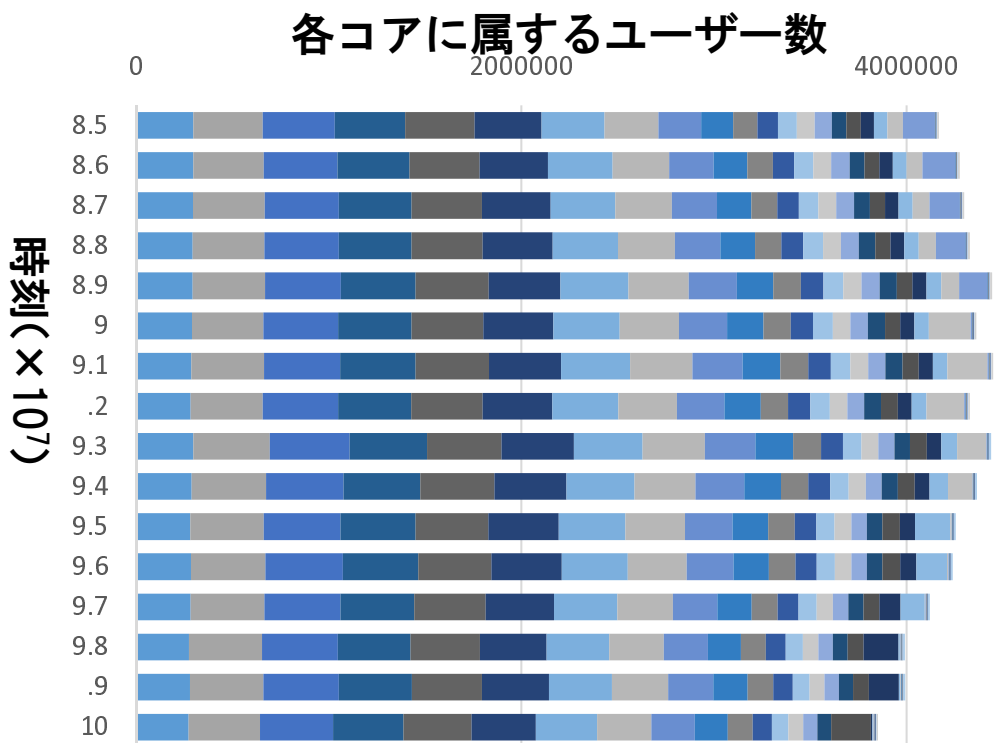


図 4.6: 時刻 8.5×10^7 から 10×10^7 までの、アクティブネットワークのコア構造の時間変化 (左揃えで描画).

| | 最大値をとる時刻 ($\times 10^7$) |
|------------|----------------------------|
| アクティブユーザー数 | 9.1 |
| 20-core | 9.0 |
| 40-core | 8.9 |
| 60-core | 8.7 |
| 80-core | 8.7 |

表 4.1: アクティブユーザー数, 各コアに属するユーザー数の時間変化を近似した 2 次式が最大値をとる時刻.

k -core に属するユーザー数を NUM_k とすると, NUM_{60} は時刻 8.9×10^7 に最大値 (876990) をとり, NUM_{90} は時刻 8.6×10^7 に最大値 (346965) をとり, アクティブユーザー数は時刻 9.1×10^7 に最大値をとる. このように, 中心部の縮小はアクティブユーザー数の減少に先立って始まる.

アクティブユーザーのうち k -core に属するものの割合を RAT_k とすると, RAT_{60} は時刻 8.6×10^7 に最大値 (20.11%) をとり, RAT_{90} は時刻 8.1×10^7 に最大値 (8.27%) をとる. これらはいずれも, アクティブユーザー数が最大値をとる時刻 (9.1×10^7) よりも早い.

各 core に属するユーザーの数はピーク付近で何度か増加と減少を繰り返す場合があるので, 各 core に属するユーザー数の時間変化を二次曲線で近似して, ピークをとる時刻を比較してみる. 図 4.5 と同じ時刻の範囲における, アクティブユーザー数, 20-core, 40-core, 60-core, 80-core のサイズの時間変化を, 2 次曲線で近似した結果を図 4.7 に示す. 各近似曲線が最大値をとる時刻を表 4.1 に示す. この表からは, k の値が大きいほど, 早い時刻に最大値をとることが分かる.

以上から, Friendster のユーザーネットワークにおいて, アクティブユーザー数の減少よりも先にコア中心部の縮小が始まったことと考えられる.

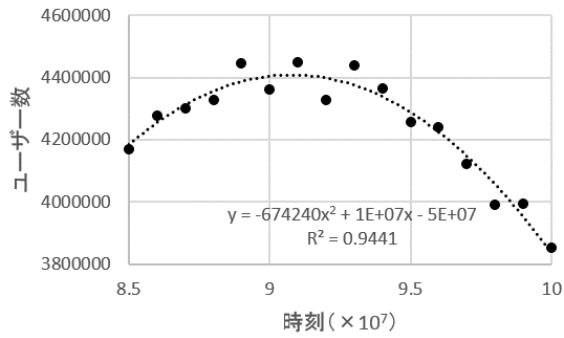
4.4 分析の仮定の妥当性の検討

4.2 節と 4.3 節の分析ではさまざまな仮定を置いた. 本節では, 4.3 節で述べた結果に対するそれらの仮定の影響について検討する.

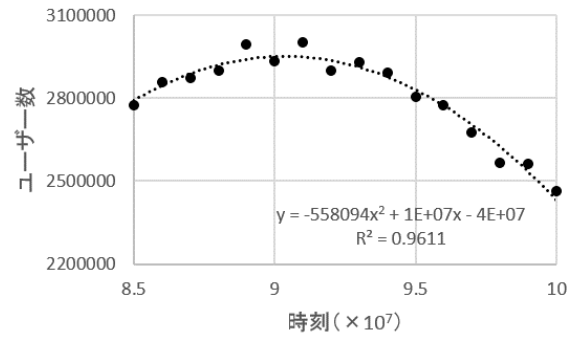
なお, 以下の項目は断りが無い限り独立である. 例えば, 4.4.3 節で使用するアルゴリズムは 4.2 節で定義したアルゴリズムに対して変更を加えたものであり, 直前の 4.4.2 節で変更したアルゴリズムをさらに変更するものではない.

4.4.1 ID を時刻とみなしたことの影響

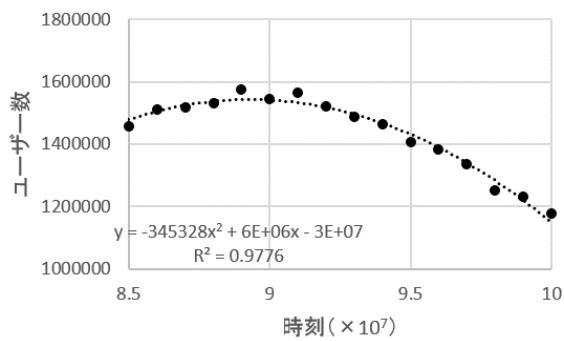
4.3 節の分析では, ユーザーの ID を時刻と仮定した. しかし, ID を時刻として用いると, 新規ユーザーの加入ペースが速いほど, 時間の経過も速くなる. その結果, 「ユーザーの ID を用いて表した時刻」が T から $T + \delta$ まで経過するまでに費やされた「実際の時間の長さ」は短くなり, 各コアのサイズが過小に見積もられてしまう.



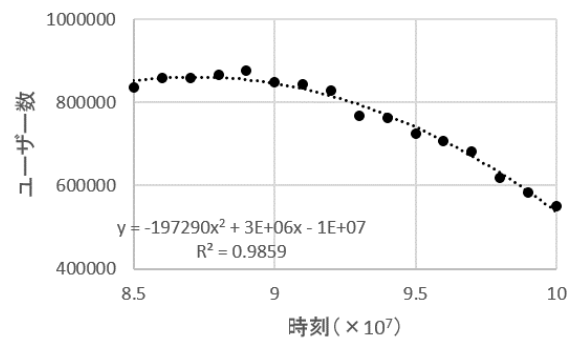
(a) アクティブユーザー数 (最大値をとる時刻: 9.1×10^7)



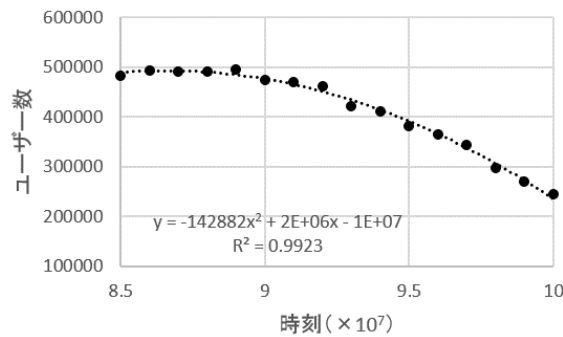
(b) 20-core (最大値をとる時刻: 9.0×10^7)



(c) 40-core (最大値をとる時刻: 8.9×10^7)



(d) 60-core (最大値をとる時刻: 8.7×10^7)



(e) 80-core (最大値をとる時刻: 8.7×10^7)

図 4.7: アクティブユーザー数や各コアに属するユーザー数の時間変化を 2 次式で近似した結果。

この影響の大きさを確かめるため、Internet Archive¹ を用いて Friendster の累計ユーザー数が時間とともにどのように変化したかを調べた。Internet Archive には、Friendster の一部のユーザーのユーザーページが保存されている²。ほとんどのユーザーのユーザーページには、そのユーザーが会員登録した年と月が記載されている。これらの情報をもとに、2004 年 4 月以降を 3 か月ごとの期間に区切り、各期間の最初に会員登録したユーザーの ID を求めた結果をプロットしたものが、図 4.8 である。表 4.2 には、確認できた範囲における、各月の最初に登録したユーザーの ID と最後に登録したユーザーの ID を示す。4.3 節では、アク

¹<https://archive.org/>

²<http://profiles.friendster.com/> (ユーザーの ID)

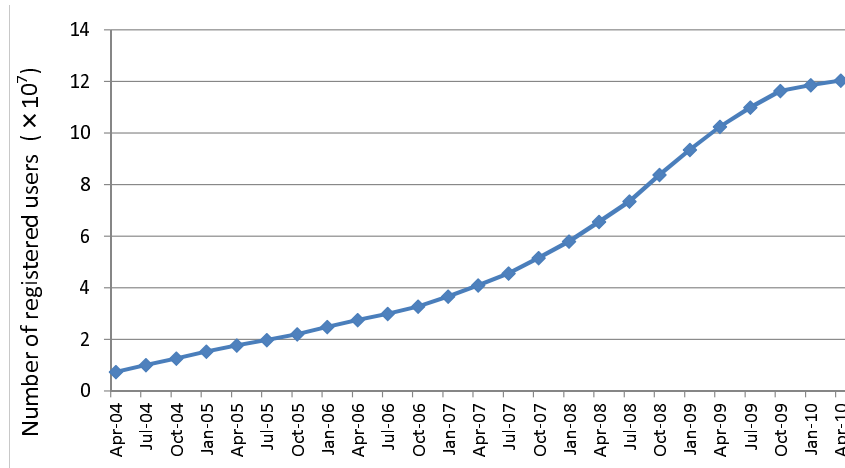


図 4.8: Friendster の累計ユーザー数の時間変化.

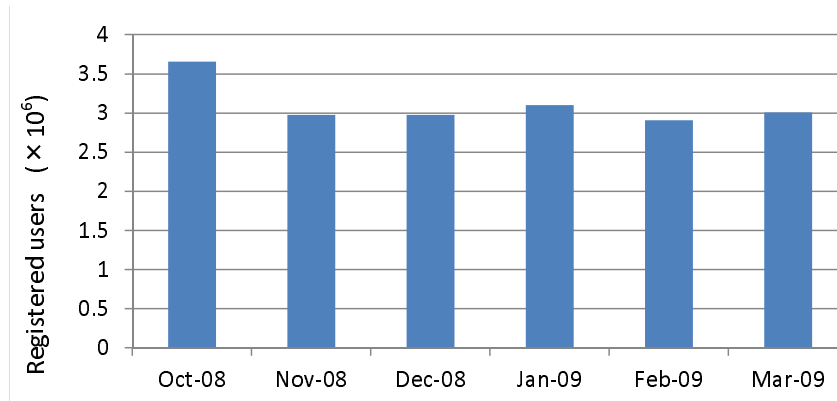


図 4.9: Friendster の一ヶ月あたりの新規登録ユーザー数 (2008 年 10 月から 2009 年 3 月まで).

ティブユーザー数の減少が時刻 8.5×10^7 から 10×10^7 ごろにかけて進行したことをみた. 累計ユーザー数の時間変化 (図 4.8) によれば, ID 8.5×10^7 から 10×10^7 のユーザーは, 2008 年 10 月頃から 2009 年 3 月頃にかけて会員登録している. そこで, この期間における, 1ヶ月ごとの新規登録ユーザー数を図 4.9 に示す. 1ヶ月ごとの新規登録ユーザー数は, 2008 年 11 月から 2009 年 3 月にかけてはほぼ一定である. したがって, 少なくとも時刻 8.8×10^7 から 10×10^7 の間では, 時間の進み方の違いによる結果の歪みはほとんど存在しないと考えられる.

なお, 本章の分析では $\delta = 1 \times 10^6$ と置いてきたが, 図 4.9 に示したデータを使うと, 1×10^6 人のユーザーが新規登録するまでに要した実時間は, 単位時間あたりの新規登録者数が最も多かった 2008 年後半から 2009 年前半頃には 10 日程度と求められる.

4.4.2 δ の値による影響

ここまでの分析では, δ の値を 1×10^6 に固定してきた. しかし, この値によって得られる結果が変わる可能性がある. そこで, δ の値を変えても同様の結果が得られるかを確かめる.

$\delta = 0.1 \times 10^6$ の場合 (図 4.10a) NUM_{60} (RAT_{60}) と NUM_{90} (RAT_{90}) はそれぞれ時刻 8.6×10^7 (8.6×10^7)

| | 最初のユーザーの ID | 最後のユーザーの ID | | 最初のユーザーの ID | 最後のユーザーの ID |
|--------|-------------|-------------|--------|-------------|-------------|
| Apr04- | 7450157 | 10131108 | Jul07- | 45652752 | 51655751 |
| Jul04- | 10137107 | 12712151 | Oct07- | 51663042 | 58131618 |
| Oct04- | 12719250 | 15307093 | Jan08- | 58140008 | 65695813 |
| Jan05- | 15313092 | 17756161 | Apr08- | 65703018 | 73665188 |
| Apr05- | 17762112 | 19898123 | Jul08- | 73673338 | 83951111 |
| Jul05- | 19906227 | 22037024 | Oct08- | 83959107 | 93568474 |
| Oct05- | 22043634 | 24864181 | Jan09- | 93575517 | 102592689 |
| Jan06- | 24870173 | 27554150 | Apr09- | 102600119 | 110054091 |
| Apr06- | 27562039 | 30007376 | Jul09- | 110062734 | 116495399 |
| Jul06- | 30015409 | 32859011 | Oct09- | 116504249 | 118718106 |
| Oct06- | 32867288 | 36700088 | Jan10- | 118726088 | 120454633 |
| Jan07- | 36706376 | 40972778 | Apr10- | 120459842 | 121674473 |
| Apr07- | 40981631 | 45646374 | | | |

表 4.2: InternetArchive で確認できた範囲における, 各期間に登録した最初のユーザーの ID と, 最後のユーザーの ID.

と 8.9×10^7 (8.9×10^7) で最大値をとり, これらはいずれもアクティブユーザー数が最大値をとる時刻 (9.1×10^7) よりも前である. ただし, 図 4.10a の期間におけるアクティブユーザー数, 20-core, 40-core, 60-core, 80-core のサイズの時間変化を二次曲線で近似したとき, 最大値をとる時刻はそれぞれ 8.76×10^7 , 8.19×10^7 , 3.81×10^7 , 9.58×10^7 , 7.75×10^7 , となり, 一定の傾向はみられない. この δ は約 1 日に相当するので, アクティブユーザーといえど友人を登録する間隔としては短すぎ, 抽出するには適していないと考えられる.

$\delta = 0.5 \times 10^6$ の場合 (図 4.10b) $NUM_{60} (RAT_{60})$ と $NUM_{90} (RAT_{90})$ はそれぞれ時刻 8.1×10^7 (8.2×10^7) と 8.1×10^7 (8.1×10^7) に最大値をとり, これらはいずれもアクティブユーザー数が最大値をとる時刻 (9.1×10^7) よりも前である. 図 4.10b の期間におけるアクティブユーザー数, 20-core, 40-core, 60-core, 80-core のサイズの時間変化を二次曲線で近似したとき, 最大値をとる時刻はそれぞれ 9.04×10^7 , 9.01×10^7 , 8.86×10^7 , 8.66×10^7 , 8.69×10^7 となり, 内側のコアほど早い時刻に縮小に転じる.

$\delta = 2 \times 10^6$ の場合 (図 4.10c) $NUM_{60} (RAT_{60})$ と $NUM_{90} (RAT_{90})$ はそれぞれ時刻 8.8×10^7 (8.5×10^7) と 8.8×10^7 (8.5×10^7) で最大値をとり, これらはいずれもアクティブユーザー数が最大値をとる時刻 (9.3×10^7) よりも前である. 図 4.10c の期間におけるアクティブユーザー数, 20-core, 40-core, 60-core, 80-core のサイズの時間変化を二次曲線で近似したとき, 最大値をとる時刻はそれぞれ 9.06×10^7 , 9.03×10^7 , 8.96×10^7 , 8.77×10^7 , 8.76×10^7 となり, 内側のコアほど早い時刻に縮小に転じる.

$\delta = 5 \times 10^6$ の場合 (図 4.10d) $NUM_{60} (RAT_{60})$ と $NUM_{90} (RAT_{90})$ はそれぞれ時刻 8.8×10^7 (8.6×10^7) と 8.7×10^7 (8.6×10^7) で最大値をとり, これらはいずれもアクティブユーザー数が最大値をとる時刻 (9.0×10^7) よりも前である. 図 4.10d の期間におけるアクティブユーザー数, 20-core, 40-core, 60-

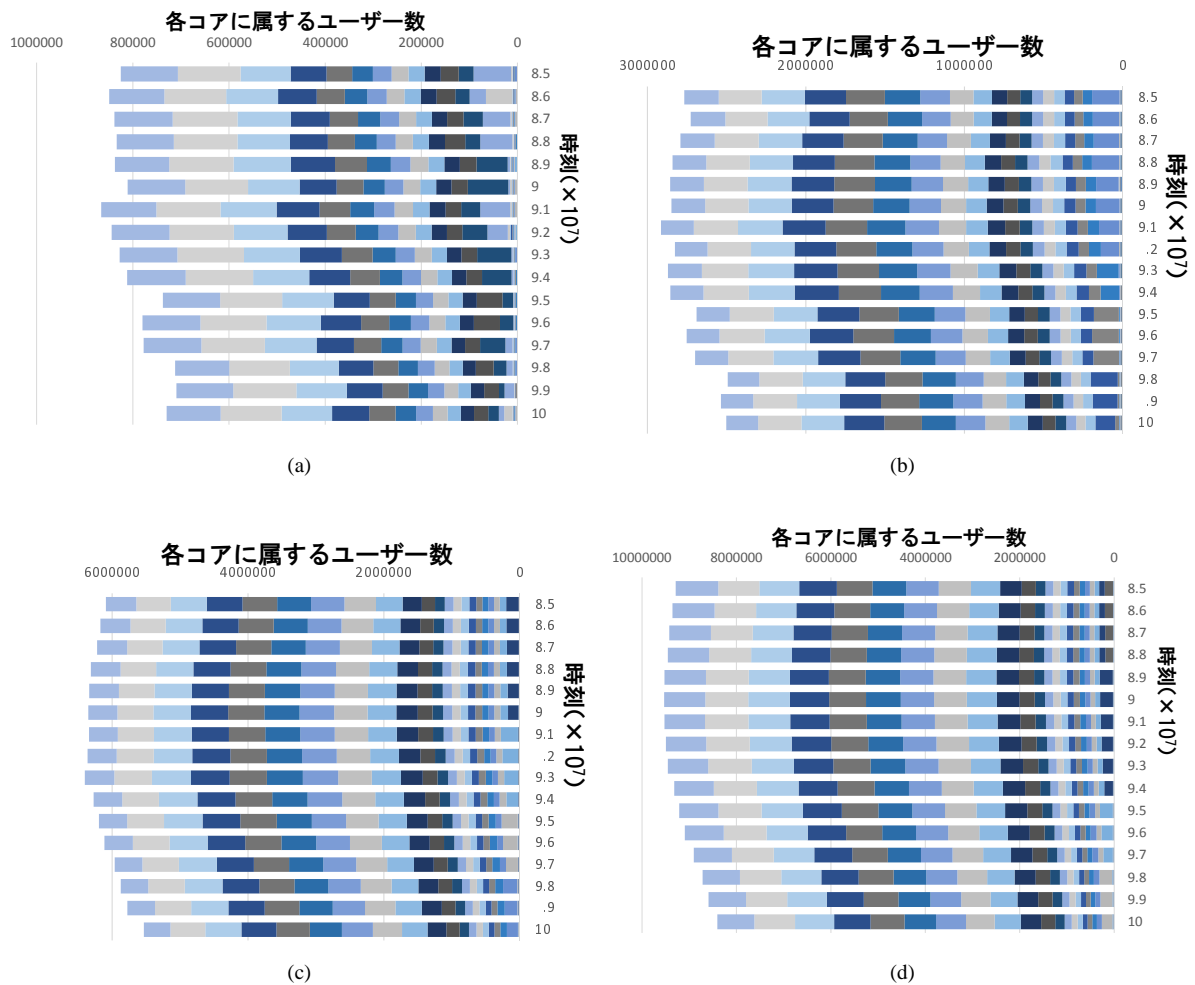


図 4.10: 時刻 8.5×10^7 から 10×10^7 までの、アクティブネットワークのコア構造の時間変化. (a) $\delta = 0.1 \times 10^6$, (b) $\delta = 0.5 \times 10^6$, (c) $\delta = 2 \times 10^6$, (d) $\delta = 5 \times 10^6$.

core, 80-core のサイズの時間変化を二次曲線で近似したとき, 最大値をとる時刻はそれぞれ 8.98×10^7 , 8.94×10^7 , 8.88×10^7 , 8.73×10^7 , 8.72×10^7 となり, 内側のコアほど早い時刻に縮小に転じる.

以上から, $0.5 \times 10^6 \leq \delta \leq 5 \times 10^6$ の範囲において, δ の選択は結果に影響しないことが確かめられた. 0.5×10^6 は約 5 日, 5×10^6 は約 50 日に対応するので, ユーザーの活動としてはかなり広い時間の範囲で成立しているといえる.

4.4.3 時刻 T から $T + \delta$ に登録したユーザーは時刻 T にアクティブであるとみなすか否かによる影響

4.3 節の分析においては, 時刻 T にアクティブであるユーザーとは, 「時刻 T から $T + \delta$ の間に新規登録したユーザー」への「辺を 1 本以上持つ」ユーザーであると定義した. この定義のもとでは, 時刻 T から $T + \delta$ の間に「新規登録した」ユーザーであっても, 時刻 T に非アクティブであったと判定されるものが存在する. しかし, それらのユーザーは少なくとも会員登録という活動を時刻 T 付近に行っていることから,

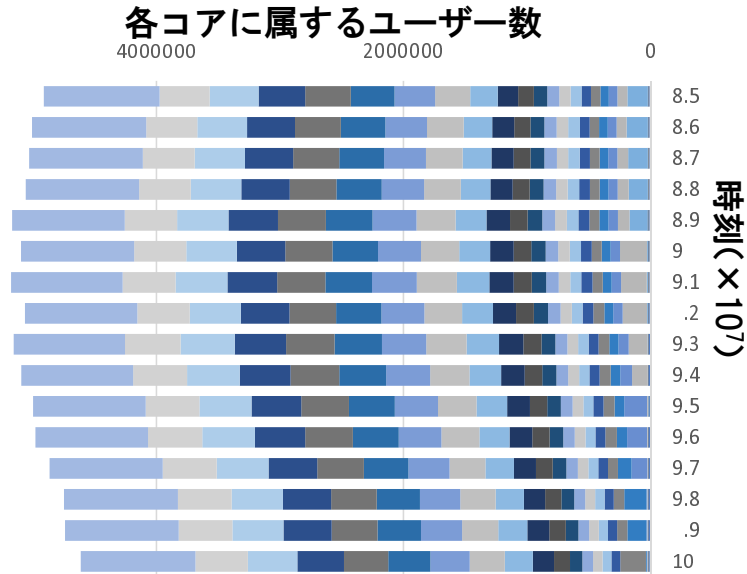


図 4.11: 時刻 8.5×10^7 から 10×10^7 までの、アクティブネットワークのコア構造の時間変化 (T から $T + \delta$ の間にユーザー登録したユーザーはすべて時刻 T にアクティブであるとみなす場合)。

アクティブであるとみなすべきであるとも考えられる。そこで、これらのユーザーをアクティブであるとみなしたときの、コア構造の時間変化を調べる。

図 4.11 から、コア中心部の縮小がアクティブユーザー数の減少に先立って始まっていることが読み取れる。 NUM_{60} (RAT_{60}) と NUM_{90} (RAT_{90}) はそれぞれ時刻 8.9×10^7 (8.6×10^7) と 8.1×10^7 (8.1×10^7) に最大値をとる。これらはいずれも、アクティブユーザー数が最大値をとる時刻 (9.1×10^7) よりも早い。図 4.11 の期間におけるアクティブユーザー数、20-core, 40-core, 60-core, 80-core のサイズの時間変化を二次曲線で近似したとき、最大値をとる時刻はそれぞれ 9.08×10^7 , 9.04×10^7 , 8.93×10^7 , 8.71×10^7 , 8.65×10^7 となり、内側のコアほど早い時刻に縮小に転じる。以上から、「時刻 T から $T + \delta$ に登録したユーザーは時刻 T にアクティブである」とみなしても結果に影響しない。

4.4.4 データセットを有向ネットワークとみなすか無向ネットワークとみなすかの影響

ここまでは、データセットを有向ネットワークとみなして分析してきた。しかし、例えば自らの行動を他のユーザーに見られることによって満足感を得る場合などを考えると、他のユーザーから自ら「へ向かう」辺の影響も無視は出来ない。そこで、各時刻におけるアクティブネットワークを無向ネットワークとみなした場合の影響について分析する。

4.2 節で抽出した各時刻の有向アクティブネットワークから、「ユーザー i からユーザー j 」または「ユーザー j からユーザー i 」の有向辺が存在する場合に i と j を結ぶ無向辺が存在すると考えて無向アクティブネットワークを構成し、生成されたネットワークのコア構造を従来の（無向） k -core decomposition 法で分析した。その結果、 NUM_{60} (RAT_{60}) と NUM_{90} (RAT_{90}) はそれぞれ 8.9×10^7 (8.6×10^7) と 8.1×10^7 (8.1×10^7) に最大値をとり、これらはいずれもアクティブユーザー数が最大値をとる時刻 (9.1×10^7) よりも前である（図 4.12）。

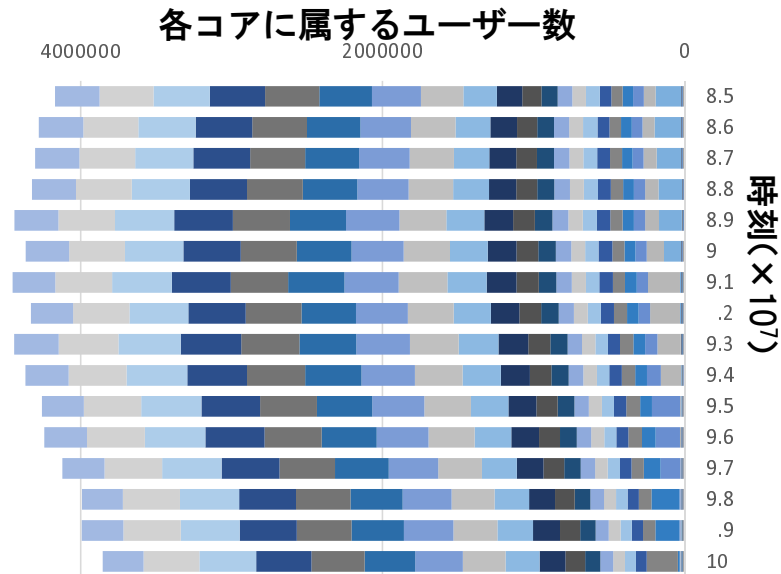


図 4.12: 時刻 8.5×10^7 から 10×10^7 までの、無向ユーザーネットワークのコア構造の時間変化.

4.4.5 k -core を用いるか k -辺連結成分を用いるかの影響

ここまでの分析では、 k -core を用いてコア構造の時間変化を調べてきたが、 k -core よりも k -辺連結成分の方がより細かい構造を抽出できるため、コア構造の分析に適しているとする主張もある [1]. k -辺連結成分の定義は次の通りである.

定義 4.4.1 (k -辺連結成分) その内部のどの $k-1$ 本の辺を除去しても非連結にできないような、与えられたネットワークの極大部分ネットワークを k -辺連結成分と呼ぶ.

例えば、図 4.13 のようなネットワークのコア構造を分析する際、 k -辺連結成分を用いれば左右の 2 つの集団をそれぞれ k -辺連結成分として検出することが可能であるが、 k -core を用いるとネットワーク全体を 1 つの大きな k -core として検出することしかできない.

ここでは、4.4.4 節で生成した無向ネットワークの k -辺連結成分の時間変化を分析した. 分析には、Akiba らによって公開されているオープンソースソフトウェア [1] を利用した. 各時刻における有向 80-core, 有向 100-core と, 80-辺連結成分, 100-辺連結成分のサイズを図 4.14 に示す. 計算時間の制約上, これより小さな k に対する k -辺連結成分のサイズを得ることはできなかったが, 少なくとも k がある程度大きな範囲では, k -辺連結成分に属するユーザー数と有向 k -core に属するユーザー数がほぼ等しい. したがって, k -辺連結成分からコア構造を構成した場合でも, コア構造中心部の縮小はアクティブユーザー数の減少よりも先に始まる.

4.5 本章のまとめ

本章では, いくつかの仮定の下で Friendster の歴史の各時点におけるアクティブネットワークを推定し, それらのコア構造を分析し, Friendster 衰退の過程でユーザーネットワークのコア構造がどのように変化した

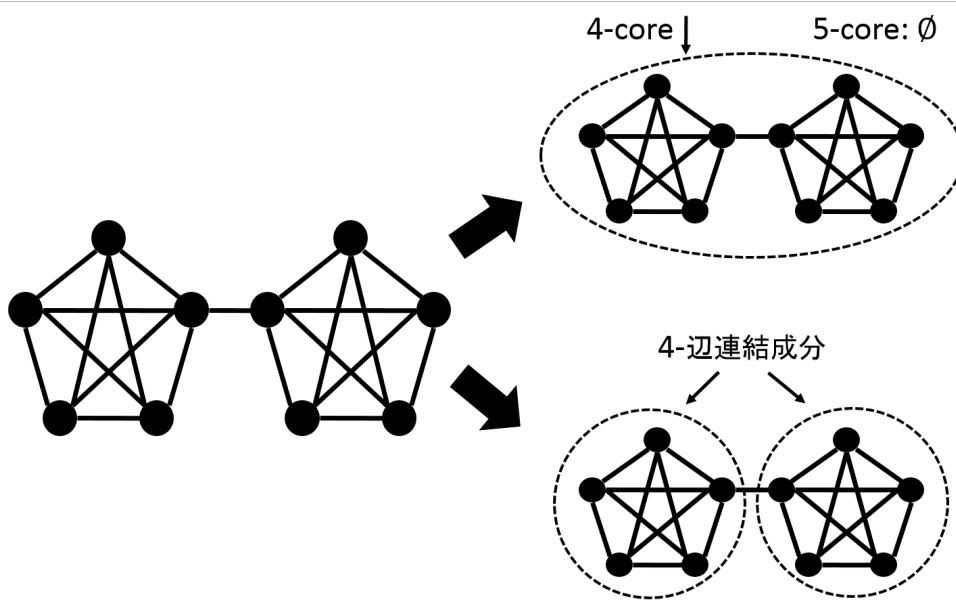


図 4.13: k -core よりも k -辺連結成分の方が直感に合致するコア検出を行えるネットワークの例.

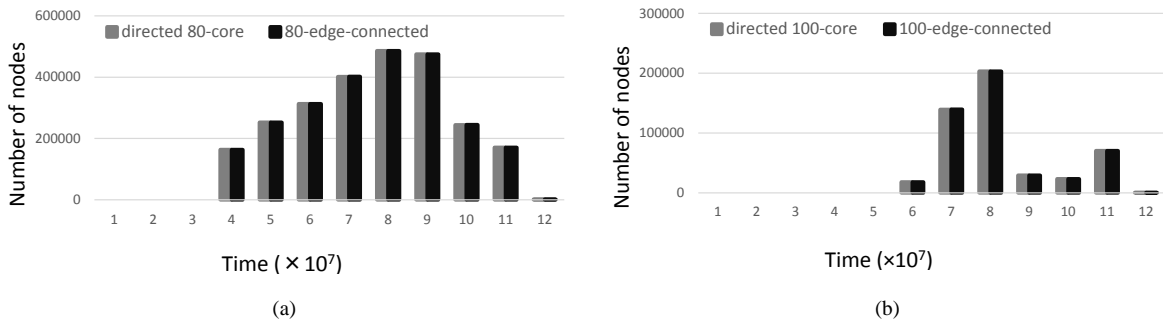


図 4.14: (a) 有向 80-core, 80-辺連結成分, (b) 有向 100-core, 100-辺連結成分のサイズの時間変化.

たかを調べた。その結果、いくつかの仮定の下でアクティブユーザー数の減少よりも先に、コア値の高いユーザー数の減少が開始したことが示された。さらに、得られた結果が仮定を少々変更しても影響を受けないものであることを示した。

第5章 Friendsterのユーザーネットワークがコア構造の内側から縮小したメカニズム

本章では、Friendsterのネットワークの縮小メカニズムを説明するためのモデルを構築する。

5.1 本章で解明したい問題

第4章では、いくつかの仮定の下で、Friendsterの衰退過程において「アクティブユーザー数の減少よりも先に、コア中心部の縮小が始まった」ことが示された。本章では、このような現象を説明できるモデルを構築する。

5.2 ネットワーク irSIR モデル

Cannarellaによれば、MySpaceというSNSの成長と衰退の過程を説明する上では、irSIRモデル(3.3.2節)が有効であった。しかし、irSIRモデルはネットワーク構造を考慮していない。本論文ではネットワーク構造に関する考察を行っているため、irSIRモデルをネットワーク構造を考慮したネットワーク irSIRモデルに修正する。

ネットワーク irSIRモデルにおいては、(各時刻のアクティブネットワークではなく)与えられたネットワークの上で、以下のように影響伝播が発生すると仮定する。

- 状態 S のユーザーと状態 I のユーザーとの間での影響伝播は、「状態 S のユーザーから状態 I のユーザーへ向かう(有向)辺」上でのみ発生する(逆向きの辺では発生しない)と仮定する。これは、状態 S であるユーザー i に影響を及ぼすのは主に、「 i が友人であるとみなしている」人々であると考えられるからである。単位時間あたりに影響伝播が発生する確率は、条件を満たす全ての辺において等しい値 β であると仮定する。
- 状態 I から状態 R への遷移については、「状態 I のユーザーから状態 R のユーザーへ向かう辺」上で発生すると仮定する。影響伝播の確率は γ とする。

β , γ は全てのユーザーが等しい値をとると仮定する。

以上の仮定を具体的に式で書くと以下ようになる。あるユーザー i が単位時間あたりに状態 S から状態 I へ遷移する確率 $P(S \rightarrow I)$ は、 i に隣接する状態 I のユーザー数を I_{adj} とすると

$$P(S \rightarrow I) = 1 - (1 - \beta)^{I_{adj}} \quad (5.1)$$

$$\approx \beta I_{adj} \quad (\beta \ll 1, I_{adj} \ll \infty) \quad (5.2)$$

と近似できる。つまり、 β の値が十分小さいとき、 I_{adj} が小さい範囲においては、遷移確率は I_{adj} の値に比例する。

状態 I から状態 R への遷移確率も同様に、隣接する状態 R の頂点数を R_{adj} とすると、 γ の値が十分小さく、 R_{adj} が小さい範囲において、

$$P(I \rightarrow R) \approx \gamma R_{adj} \quad (\gamma \ll 1, R_{adj} \ll \infty) \quad (5.3)$$

と近似できる。

つまり、ネットワーク irSIR モデルには、

- 「単位時間あたりに状態 S の頂点が状態 I に遷移する確率」は「隣接する状態 I の頂点の数」にしたがって上昇する
- 「単位時間あたりに状態 I の頂点が状態 R に遷移する確率」は「隣接する状態 R の頂点の数」にしたがって上昇する

という性質がある。そこで、実際の SNS のユーザーネットワークにおいて、状態遷移確率がこの性質を満たすとき、そのユーザーネットワークの状態 S, I, R のユーザー数はネットワーク irSIR モデルに従って時間変化するとみなす。

5.3 Friendster のユーザーネットワークの時間変化はネットワーク irSIR モデルの遷移確率と同じ式を満たすかの検証

本節では、Friendster の状態間の遷移確率は、ネットワーク irSIR モデルの遷移確率と同じ式（式 5.2, 式 5.3）を満たすかを検証する。

まず、Friendster ネットワークにおける状態 S, I, R を定義する。

S： まだ Friendster に会員登録していない。

I： Friendster に会員登録済みであり、かつ離脱していない。

R： Friendster を離脱済みである。

ネットワーク irSIR モデルでは、状態 S から状態 I, 状態 I から状態 R への状態遷移のみが発生する。第 4 章でコア構造を分析した際には、あるユーザーが時刻 T にアクティブであるのは「時刻 T 付近に Friendster 上でアクティブに活動していた」場合であると仮定した。この仮定のもとでは、ユーザーは「アクティブ状態」と「非アクティブ状態」の間を何度でも遷移できることになる。ここでは簡単なモデルとするために、「アクティブ状態」から「非アクティブ状態」方向への遷移のみ起こるとして、上記のように状態を定義した。なお、「非アクティブ状態」から「アクティブ状態」への遷移を組み込んだモデルについても、6.3 節から 6.5 節で検討している。

次に、状態遷移が発生する時刻を定義する。ここでは時刻としてユーザーの ID を使う。まず、ユーザー i が状態 S から状態 I に遷移する時刻は、 i が会員登録した時刻、つまり i 自身の ID と定義する。

ユーザー i が状態 I から状態 R に遷移する時刻 $L(i)$ は、「 i からの辺が存在するユーザー (= i の友人) のうち、最後に会員登録した者が会員登録した時刻」と定義する。このように定義する理由は、 i からそのユーザーへの辺は必ず時刻 $L(i)$ 以降に形成されたものであり、 i は時刻 $L(i)$ にはまだネットワークを離脱してはいなかったことが保証されるためである。式で書けば、

$$L(i) = \max_{j \in A(i)} (\text{ユーザー } j \text{ の ID}) \quad (5.4)$$

である。ただし、 $A(i)$ はユーザー i のネットワーク上の友人の集合である。

以上のように Friendster の各ユーザーの状態と状態遷移を定義したとき、状態 S から状態 I、状態 I から状態 R の遷移確率がネットワーク irSIR モデルと同様の遷移確率を持つかを確認する。

5.3.1 状態 S → I の遷移確率

Friendster における状態 S から I への遷移確率が、ネットワーク irSIR モデルの遷移確率の式 5.2 に従うかを確認する。Friendster においてユーザー A から他のユーザー B への辺が実際に形成されたのは A の会員登録後であるので、A が会員登録した時点では A から状態 I のユーザーへの辺は一つもない。したがって、ネットワーク irSIR モデルに従うとすると、状態 S から状態 I への遷移は起こらないことになってしまう。そこでここでは、「A が状態 R に遷移するまでに形成した友人のうち、その時刻において状態 I であったユーザーの数」に比例する数の友人が (Friendster 外において) A に Friendster への参加をうながしたと仮定し、アクティブでないユーザーも含むネットワーク全体を使って分析することにする。具体的には、ある 2 つの時刻 A , $B (> A)$ を設定し、時刻 A に状態 S であった全ユーザーを対象として、「時刻 A における、状態 I の友人との繋がり数」と「そのユーザー自身が時刻 $B (> A)$ までに状態 I に遷移したか否か」の関係を調べる。この関係をプロットしたものが図 5.1 である。図 5.1 の (a) から (d) は異なる A と B に対するプロットである。友人数が 200 以下の範囲では、時刻 A における状態 I の友人の数が多ければ、時刻 B までに状態 I に遷移する確率が高くなる傾向があることが読み取れる。これはネットワーク irSIR モデルの遷移確率の式 5.2 と近い。

「時刻 A における状態 I の友人数」と「時刻 B までに状態 I に遷移した確率」との関係を示す。隣接する状態 I の友人数が 200 以下である範囲に限定して線形回帰した決定係数を表 5.1 に示す。決定係数は Friendster の運営が開始された直後 ($A = 1 \times 10^7$, $B = 2 \times 10^7$) とアクティブユーザー数の減少が始まった時刻 (9×10^7) 以降を除く幅広い期間において、高い値 (> 0.5) を示している。したがって、隣接する状態 I の友人数が 200 以下かつアクティブユーザー数の減少が始まる前に限れば、状態 S から状態 I への遷移確率はネットワーク irSIR モデルの遷移確率の式 5.2 に従っているといえる。

状態 I の友人数が 200 を超えるユーザーの数は、状態 I であった友人数の頻度分布 (図 5.2) から分かるように少なく、いずれの条件においても全体の 1% 未満である ((a) 0.07%, (b) 0.2%, (c) 0.4%, (d) 0.7%)。そのようなユーザーを考慮に入れたモデルは 6.1 節で検討する。

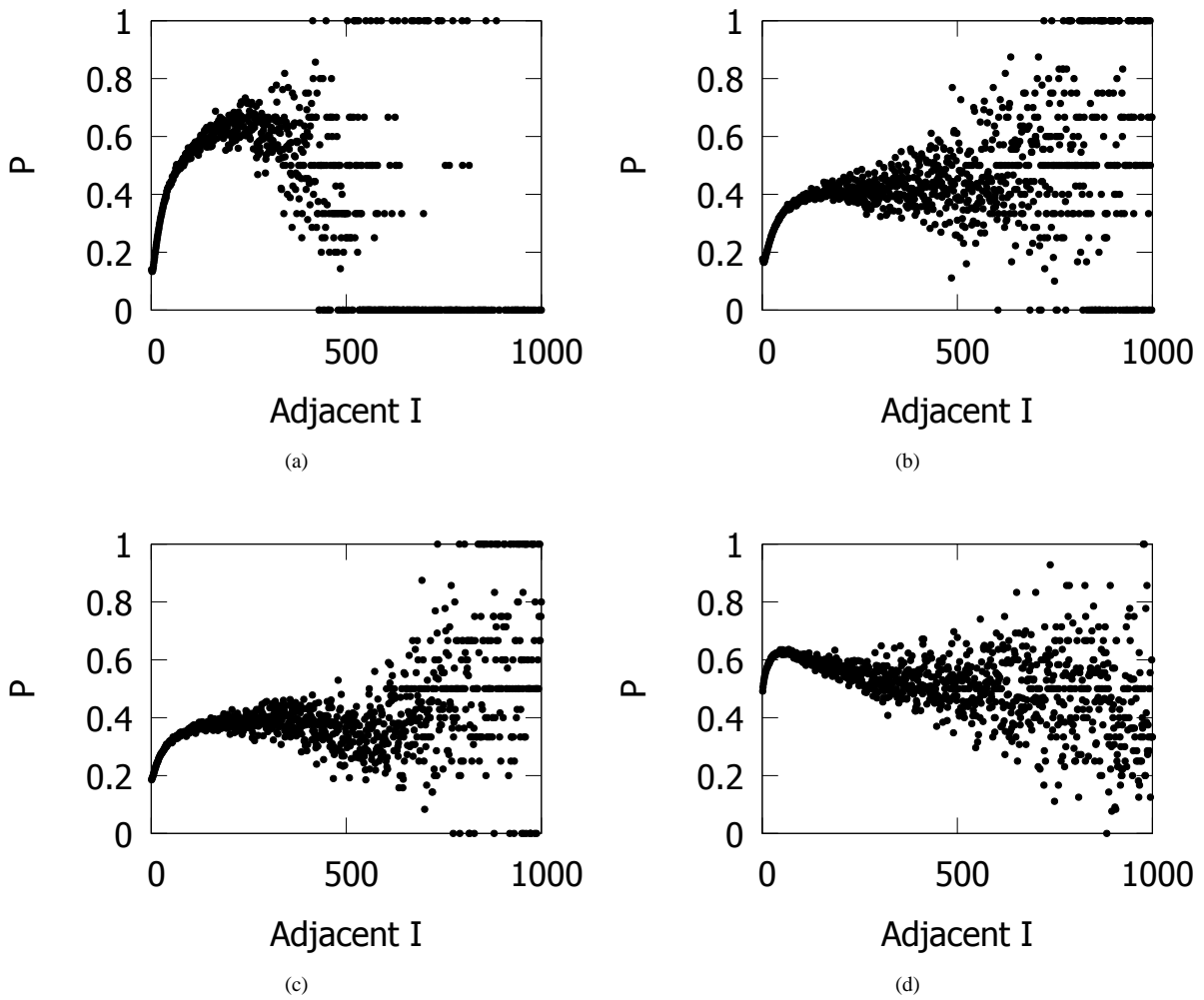


図 5.1: 「時刻 A において、状態 I であった友人数」と「時刻 $B(> A)$ までに状態 I に遷移した確率」との関係.
 (a) $A : 2 \times 10^7, B : 3 \times 10^7$, (b) $A : 5 \times 10^7, B : 6 \times 10^7$, (c) $A : 7 \times 10^7, B : 8 \times 10^7$, (d) $A : 10 \times 10^7, B : 11 \times 10^7$.

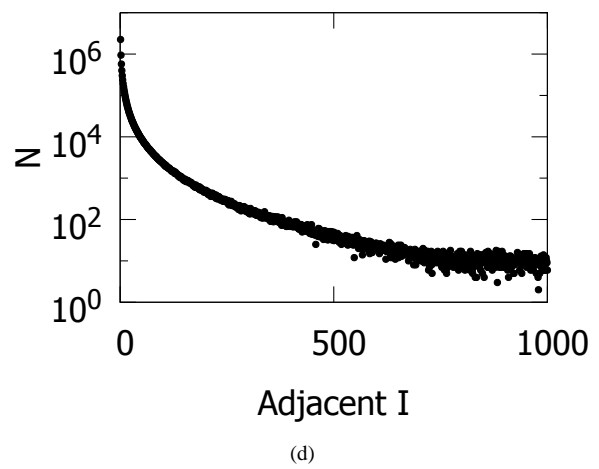
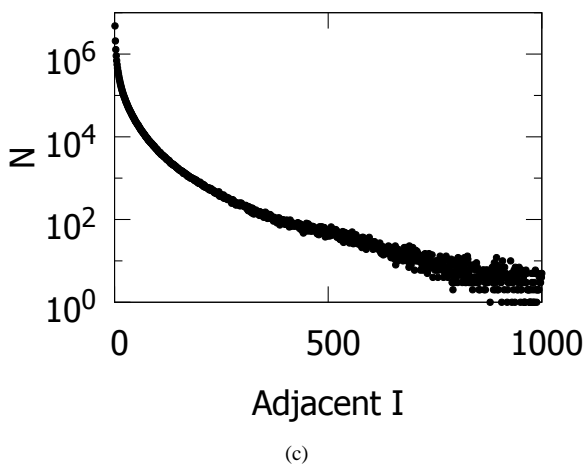
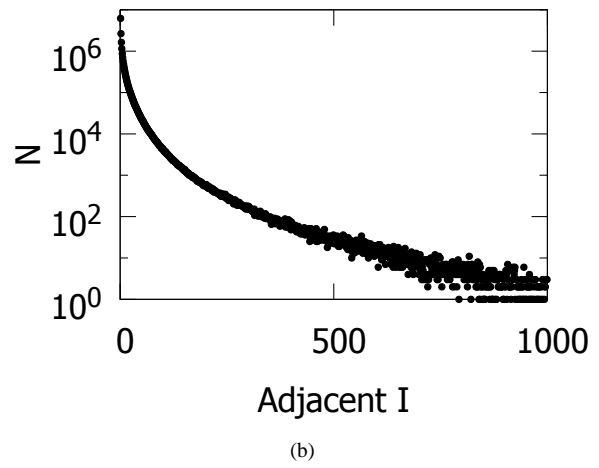
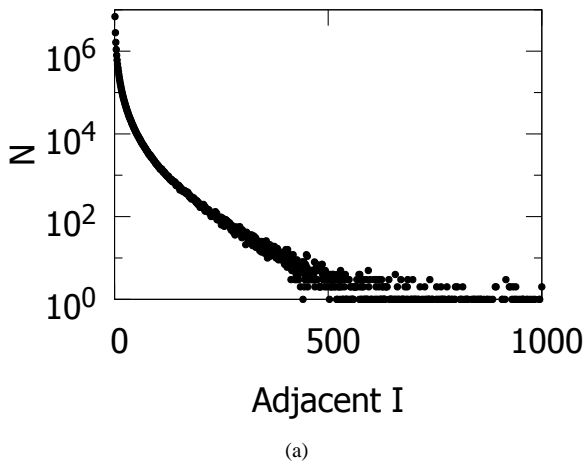


図 5.2: 時刻 A において、状態 I であった友人数の頻度分布. ただし、時刻 A に状態 S であったユーザーのみを集計対象とした. (a) $A : 2 \times 10^7$, (b) $A : 5 \times 10^7$, (c) $A : 7 \times 10^7$, (d) $A : 10 \times 10^7$.

| $A, B (\times 10^7)$ | 決定係数 |
|----------------------|------|
| $A = 1, B = 2$ | 0.23 |
| $A = 2, B = 3$ | 0.82 |
| $A = 3, B = 4$ | 0.80 |
| $A = 4, B = 5$ | 0.77 |
| $A = 5, B = 6$ | 0.77 |
| $A = 6, B = 7$ | 0.82 |
| $A = 7, B = 8$ | 0.77 |
| $A = 8, B = 9$ | 0.59 |
| $A = 9, B = 10$ | 0.03 |
| $A = 10, B = 11$ | 0.09 |

表 5.1: 図 5.1 を線形回帰したときの決定係数（状態 I である友人数が 200 以下である範囲のみを対象とする場合）。

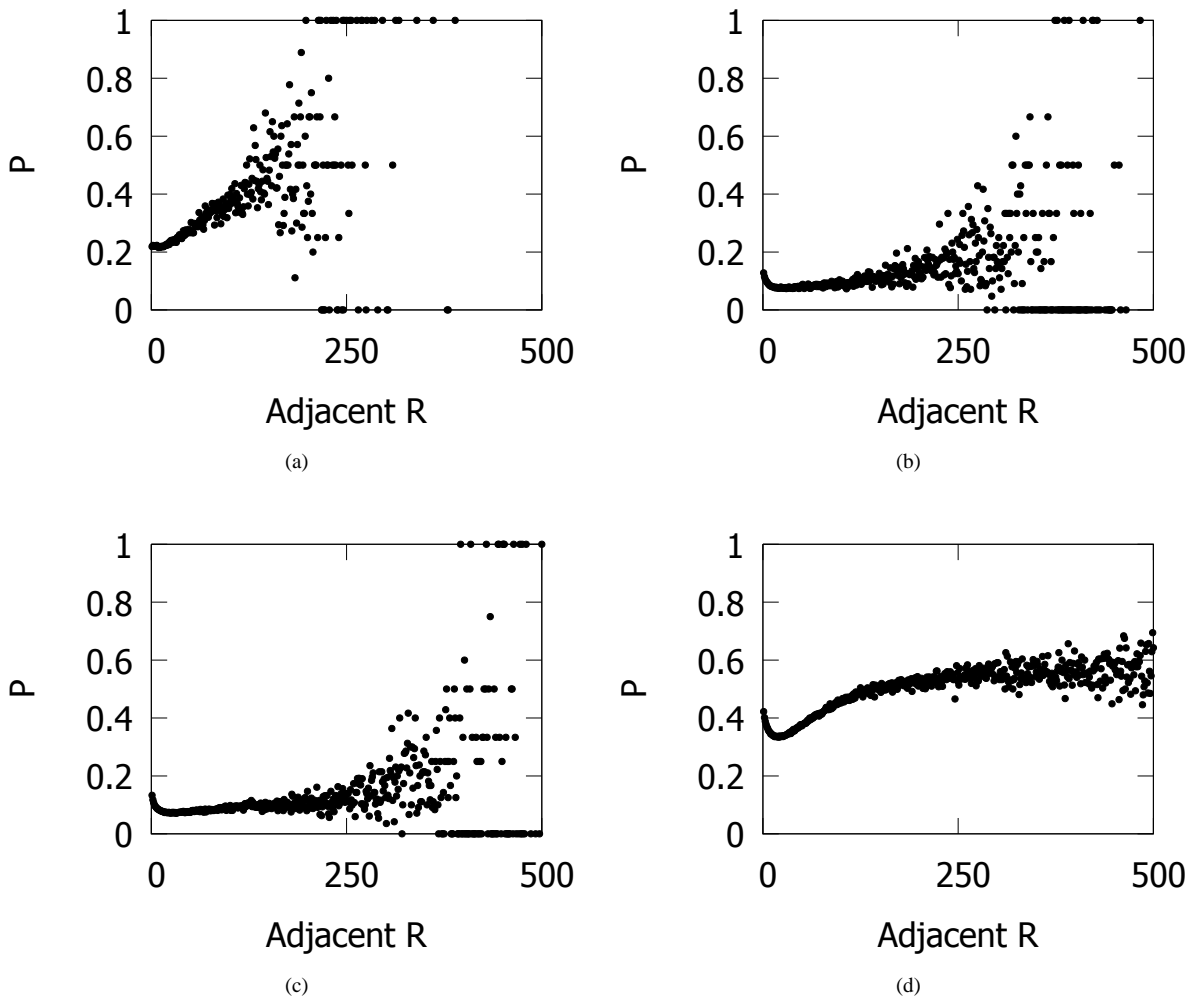


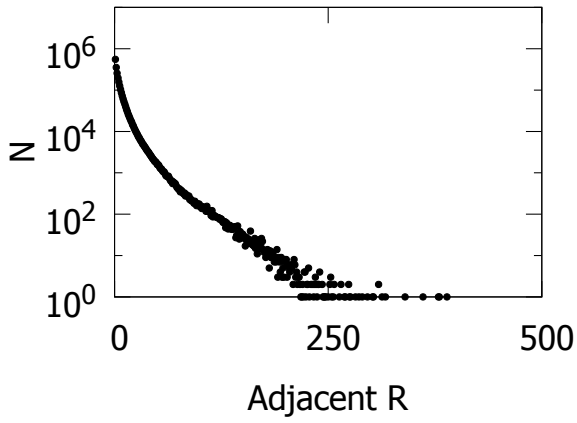
図 5.3: 「時刻 A において、状態 R であった友人数」と「時刻 $B(>A)$ までに状態 R に遷移した確率」との関係。 (a) $A: 2 \times 10^7, B: 3 \times 10^7$, (b) $A: 5 \times 10^7, B: 6 \times 10^7$, (c) $A: 7 \times 10^7, B: 8 \times 10^7$, (d) $A: 10 \times 10^7, B: 11 \times 10^7$.

5.3.2 状態 $I \rightarrow R$ の遷移確率

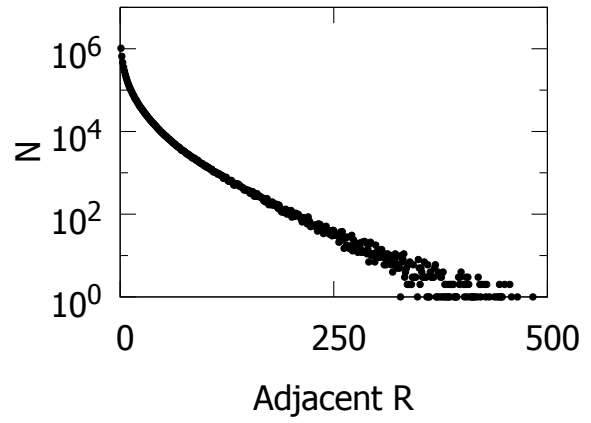
次に、Friendster における状態 I から R への遷移確率が、ネットワーク irSIR モデルの遷移確率の式 5.3 に従うかを確認する。図 5.3 は、「時刻 A における状態 R の友人数」と「時刻 B までに状態 R に遷移した確率」との関係をいくつかの A と B の組合せに対してプロットしたものである。各時刻における「状態 R である友人数」の分布は図 5.4 に示した。表 5.2 には、状態 R の友人数が 200 以下のユーザーに対して、線形回帰の決定係数を示した。Friendster の開設直後 ($A = 1, B = 2$) の決定係数は 0.002 と低いものの、それ以降はかなり高い値 (≥ 0.4) を示す。

ただし、「時刻 A に状態 R である友人をほとんど持たない (y 軸付近) ユーザー」の遷移確率がかなり高い。本節では簡単のために状態 R である友人をほとんど持たないユーザーは無視する。この点に関しては、6.2 節で考察する。

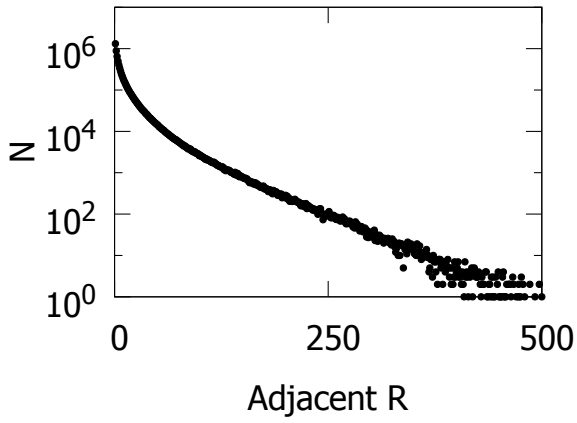
状態 S から状態 I 、状態 I から状態 R の遷移については、遷移先の状態の友人数が 200 以下で、特定の時期 (時刻 2×10^7 までと時刻 9×10^7 以降) を除き、ネットワーク irSIR モデルと同様の遷移確率の式を満



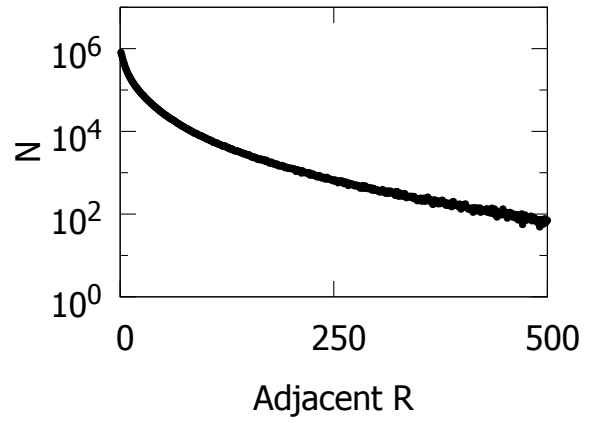
(a)



(b)



(c)



(d)

図 5.4: 時刻 A において、状態 R であった友人数の頻度分布. ただし、時刻 A に状態 I であったユーザーのみを集計対象とした. (a) $A : 2 \times 10^7$, (b) $A : 5 \times 10^7$, (c) $A : 7 \times 10^7$, (d) $A : 10 \times 10^7$.

たす.

| $A, B (\times 10^7)$ | 決定係数 |
|----------------------|-------|
| $A = 1, B = 2$ | 0.002 |
| $A = 2, B = 3$ | 0.54 |
| $A = 3, B = 4$ | 0.68 |
| $A = 4, B = 5$ | 0.72 |
| $A = 5, B = 6$ | 0.47 |
| $A = 6, B = 7$ | 0.43 |
| $A = 7, B = 8$ | 0.40 |
| $A = 8, B = 9$ | 0.70 |
| $A = 9, B = 10$ | 0.89 |
| $A = 10, B = 11$ | 0.92 |

表 5.2: 図 5.3 を線形回帰したときの決定係数（状態 R である友人数が 200 以下である範囲のみを対象とする場合）。

5.4 ネットワーク irSIR モデルにしたがって時間変化するネットワークではコアの中心部が全体より先に縮小するかの検証

本節では、「ネットワーク irSIR モデルに従って時間変化するネットワークはコア構造の中心部から縮小する」かをシミュレーションを用いて検証する。具体的には、人工的に生成したネットワーク上でネットワーク irSIR モデルをシミュレーションし、アクティブネットワークのコア構造の時間変化を確かめる。

5.4.1 ネットワーク生成アルゴリズム

ネットワーク irSIR モデルでシミュレーションを行うために、現実の SNS 上のユーザーネットワークに近い性質を持つネットワークを生成する。現実のユーザーネットワークは次数分布がべき分布のような裾の広い分布であり、また「友達の友達が友達である」というクラスター性や、ネットワーク中からランダムに選択した 2 頂点間の平均距離が短いなどの性質を持つことが知られている [39]。次数分布がべき分布で、2 頂点間の平均距離が短いネットワークを生成するモデルには、「成長¹」と「優先的選択²」の規則に基づく Barabási-Albert モデル [4] がある。クラスター性を持たせるために、このモデルに「A と B, B と C に繋がりが存在することで A と C との繋がりの形成が誘発される」とする「三者閉包」の規則 [14] を加えた以下のモデルを提案する。

- **Input:** 初期状態の頂点数 n_0 , 最終状態の頂点数 N_{total} , 規則 1 による辺生成本数 m_1 , 規則 2 による辺生成本数 m_2
- **Output:** 生成されたネットワーク
- 初期状態: n_0 個の頂点からなる完全グラフ。
- 頂点数が N_{total} になるまで以下の規則 1 と規則 2 を繰り返し、各ステップで頂点 1 個と辺 $m_1 + m_2$ 本を追加する。

規則 1 新たな頂点を 1 個追加し、既にネットワーク中に存在する頂点の中から次数に比例する確率で選ばれた m_1 個の頂点と辺で結ぶ。

規則 2 以下を m_2 回繰り返し、ネットワーク内の辺を m_2 本増やす。

1. ネットワーク中の辺を 1 本ランダムに選ぶ。
2. その辺が結んでいる 2 つの頂点のうちの 1 つをランダムに選ぶ (選ばれた頂点を A, もう 1 つの頂点を B と呼ぶ)。
3. 頂点 B と隣接している頂点のうち、頂点 A 以外から 1 つランダムに選ぶ (選ばれた頂点を C と呼ぶ。B と隣接している頂点が A のみの場合は、1. からやり直す)。
4. 頂点 A と頂点 C を辺で結ぶ (既に AC が辺で結ばれている場合は 1. からやり直す)。

¹時間経過とともに新たな頂点が追加され、ネットワークのサイズが増加していく。

²すでに多くの辺を所持している頂点ほど、新たな辺を獲得しやすい。

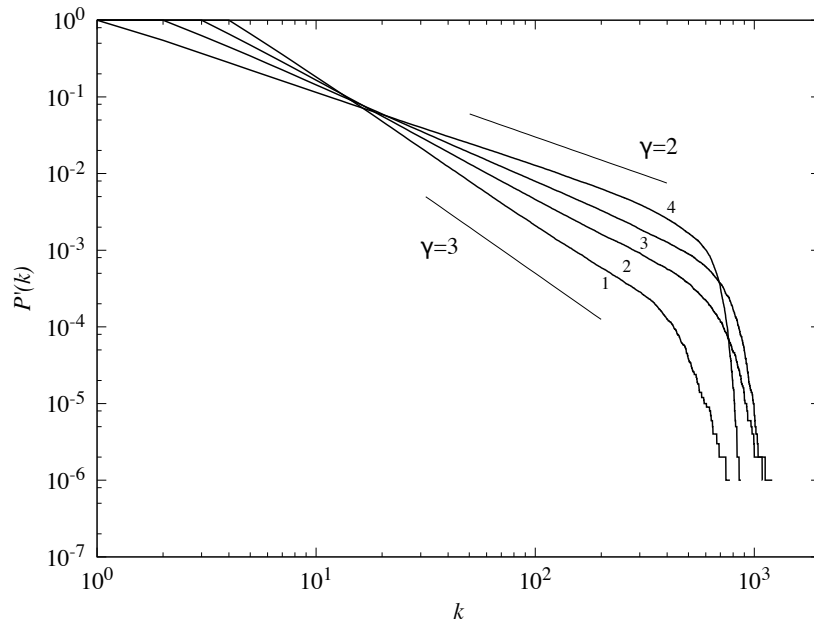


図 5.5: 辺活性化モデルで生成したネットワークの累積度数分布. 1: $m_1 = 4, m_2 = 0$, 2: $m_1 = 3, m_2 = 1$, 3: $m_1 = 2, m_2 = 2$, 4: $m_1 = 1, m_2 = 3$. $N = 25000$ として, 50 回の試行の平均をとった.

規則 2 において, ネットワーク中のある辺の活性化 (規則 2 の 1.) をきっかけとして新たな辺が生成されることから, このネットワーク生成モデルを辺活性化モデルと呼ぶことにする.

辺活性化モデルで生成されるネットワークの主な性質は以下の通りである.

- 頂点数 N_{total} , 平均次数 $2(m_1 + m_2)$.
- 度数分布はべき分布であり (図 5.5), べき指数は $(3m_1 + 4m_2)/(m_1 + 2m_2)$ である (度数分布の解析的な導出は付録 A 節参照).
- クラスタ係数が高く, 平均距離が短いスモールワールドネットワークである (図 5.6).

生成されたネットワーク中の頂点は, 幅広いコア値を持つ. これは Barabási-Albert モデルによって生成されたネットワークとは対照的である. Barabási-Albert モデルとは, 規則 1 のみを用いて頂点・辺を追加し, ネットワークを生成するモデルである. Barabási-Albert においては, 新たにネットワークに加わる頂点を持つ辺を m_1 本とするとき, (初期状態において存在していた n_0 個の頂点を除く) 全ての頂点のコア値は m_1 となる.

5.4.2 ネットワーク irSIR モデルのアルゴリズム

生成されたネットワーク上において, ネットワーク irSIR モデルのシミュレーションを行い, アクティブネットワークのコア構造の時間変化を確かめる. 5.2 節で見たように, ネットワーク irSIR モデルにおいて, 隣接する状態 I の頂点数が x であるときの状態遷移確率は $P = 1 - (1 - \beta)^x$ である. しかし, ここでは単純化のため, $\beta \ll 1, x \ll \infty$ であるときの近似値である $P = \beta x$ を用いる. ただし, βx が 1 を超える場合は, 状態遷移確率は 1 とする. 状態 I から状態 R への遷移確率も同様に扱う.

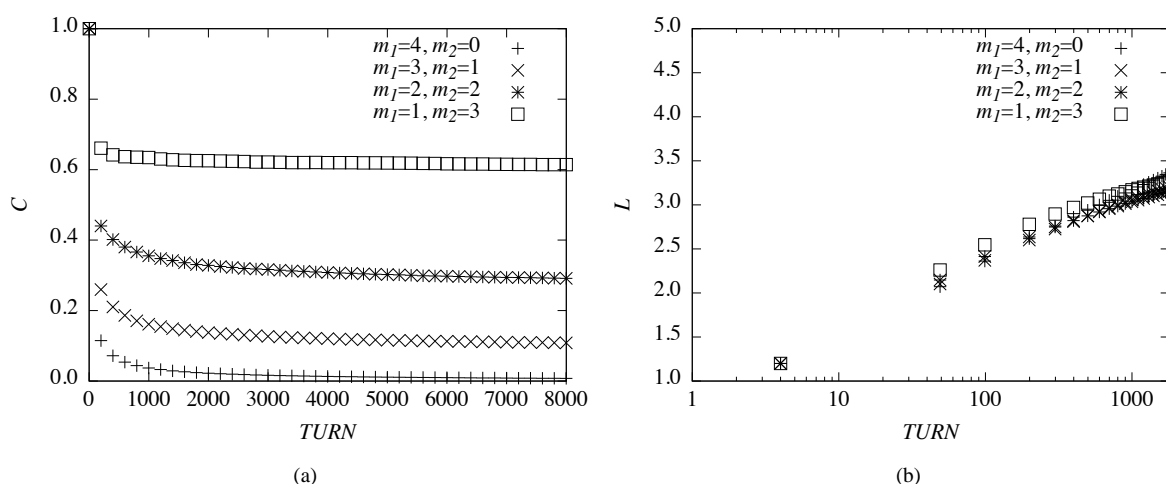


図 5.6: 辺活性化モデルによるネットワークの生成するときの, (a) ステップ毎のクラスター係数の変化, (b) ステップ毎の平均距離の変化. 50 回の試行の平均.

「状態 I から状態 R への遷移確率は隣接する状態 R の頂点数に比例する」という仮定をそのまま用いると, 初期状態において状態 R の頂点が存在しない場合, 全ての頂点の状態 I から状態 R への遷移確率は 0 なので, 状態 I から状態 R への遷移は一切発生しなくなる. Cannarella はこの問題に対処するため, 初期状態において少数の頂点が状態 R をとると仮定したが, ある SNS が誕生したときからその SNS を離脱したユーザーが存在するとは不自然である. そこで, 状態 I の頂点は状態 R の頂点と一切隣接していない場合であっても, ある確率で状態 R に遷移するものと仮定する.

以上より, シミュレーションのアルゴリズムを以下のように定義する.

- **input** : 辺活性化モデルによって生成されたネットワーク, 初期状態において状態 I である頂点の割合 I_0 , 状態 S から I への遷移確率を決めるパラメータ β , 状態 I から R への遷移確率を決めるパラメータ γ, ν
- **output** : 各ステップにおける, ネットワーク中の各頂点の状態
- **初期状態** : 各頂点は状態 I か状態 S のいずれかの状態をとる. 状態 I をとる確率を I_0 とする.
- **各ステップで**, 各頂点は以下の規則で確率的に状態遷移する.

状態 S の頂点: 確率 βI_{adj} で状態 I に遷移する.

状態 I の頂点: 確率 $\gamma R_{adj} + \nu$ で状態 R に遷移する.

状態 R の頂点: 状態遷移しない.

ただし, I_{adj} は隣接している状態 I の頂点数, R_{adj} は隣接している状態 R の頂点数である.

ここでは, ネットワーク全体のうち, 「状態 I である頂点」と, 「状態 I である 2 つの頂点を結ぶ辺」から構成される部分ネットワークを, その時刻におけるアクティブネットワークと定義する. このとき, アクティブネットワークのコア構造が, 時間とともにどのように変化していくかを, シミュレーションによって調べる.

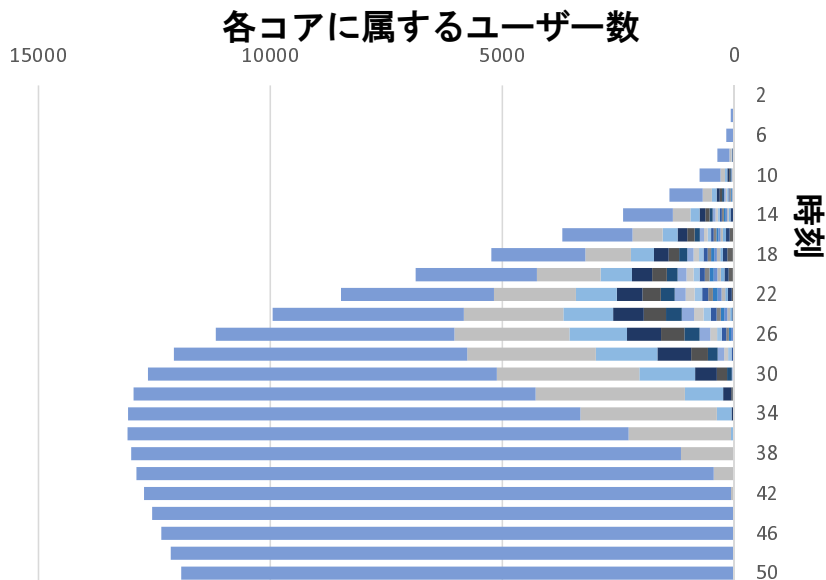


図 5.7: ネットワーク irSIR モデルに従って成長・衰退するネットワークのコア構造の時間変化. $I_0 = 0.001$, $\beta = \nu = \gamma = 0.002$. 50 回の試行の平均値.

5.4.3 パラメータがコア中心部の縮小に与える影響

5.4.1 節のアルゴリズムを用いてシミュレーションを行った. コア構造の時間変化のしかたの一例を図 5.7 に示す. ネットワークとしては, 5.4.2 節の辺活性化モデルで, パラメータを $N_{total} = 10^5$, $n_0 = 10$, $m_1 = m_2 = 5$ として生成したものを用いた. コア構造の描画の方法は, Friendster ネットワークのコア構造の描画 (4.3.1 節) と同様である. この図からは, コア中心部の縮小が状態 I の頂点数の減少よりも先に始まることが読み取れる. これは, Friendster のユーザーネットワークの分析から得られた結論と一致している.

パラメータによる影響を調べるために, 状態 I の頂点数の数が減少し始める時刻 t_{active} が, k -core に含まれる頂点数が減少し始める時刻 t_k より後であるかを調べる. そのために, いくつかの β , γ , ν の値の設定のもとで $t_{active} - t_{60}$ と $t_{active} - t_{80}$ を求めた (図 5.8). 図から, $\nu = 0.001, 0.005, 0.025$ のいずれの場合でも, $0.002 \leq \beta \leq 0.05$, $0.002 \leq \gamma \leq 0.05$ の範囲において $t_{active} - t_{60} > 0$ かつ $t_{active} - t_{80} > 0$ である. ただし, コア値が 60, 80 に到達しない条件を除く.

ネットワーク irSIR モデルにおいて, アクティブネットワークがコア構造の中心部から縮小する理由は次のように説明できる. (アクティブネットワークではなく与えられたネットワーク全体のうちの) 中心部に位置する (= コア値の高い) 頂点ほど, 相対的に早く影響の伝播を受けやすい. コア値の高い頂点は次数が大きく, かつ辺密度の高い部分ネットワークに埋め込まれているため, 早い段階で状態 S から状態 I への遷移確率が大きくなりやすい. その結果, それらの頂点はシミュレーション開始後すぐに状態 I に遷移する. 一方, 状態 I から状態 R への遷移についても同様のことが言えるため, すぐに状態 I に移り, コア中心部が縮小する.

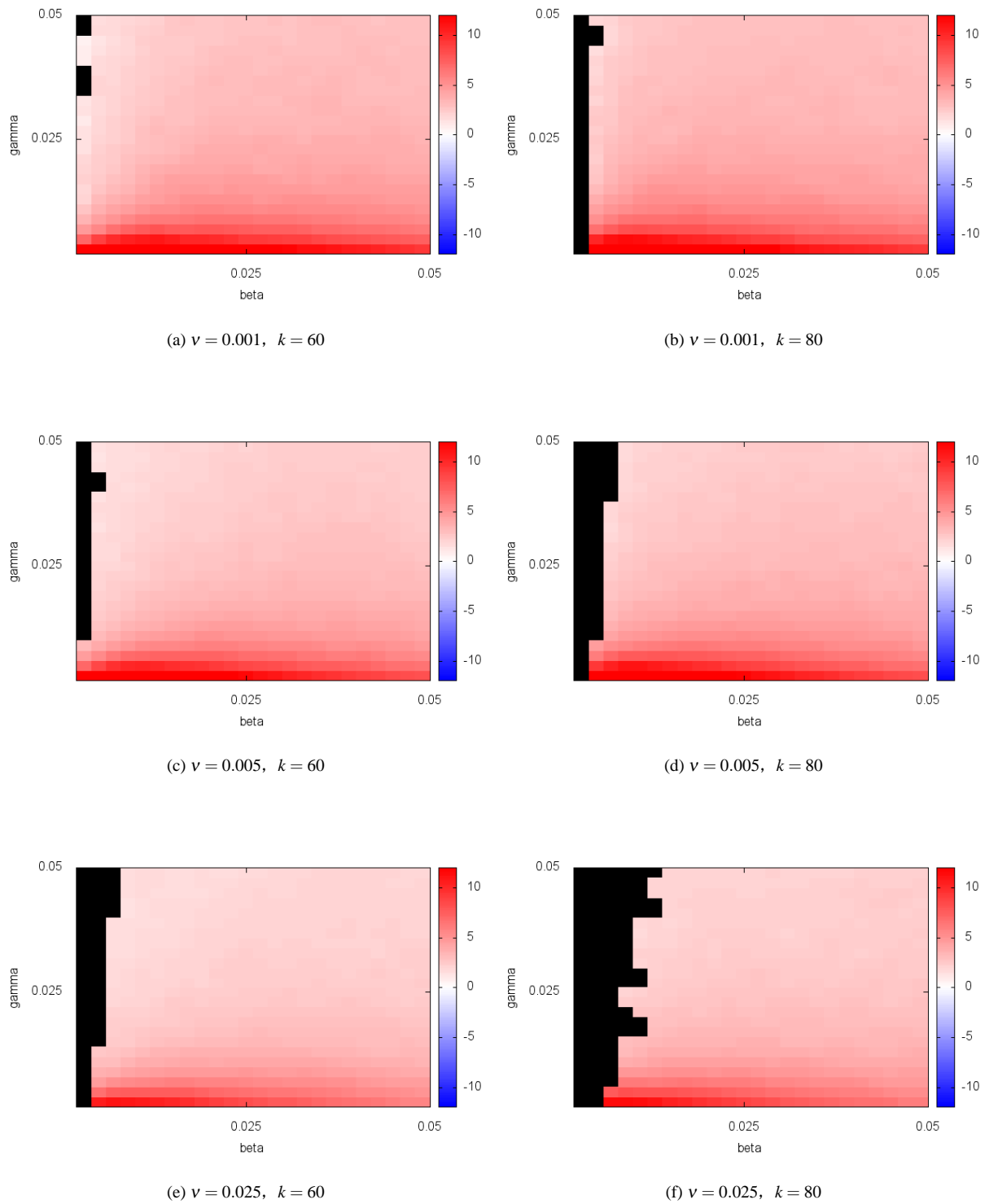


図 5.8: パラメータ β, γ, ν に対する $t_{active} - t_k$ の値. 10 回の試行のうち, アクティブネットワークにおけるコア値が k に達した頂点が存在する試行の平均値. 10 回の試行の半分以上でコア値が k に到達する頂点が存在しなかった場合は黒塗りとした.

5.5 パラメータが状態 I の頂点数の時間変化に与える影響

ネットワーク irSIR モデルにおいて用いたパラメータ β , γ の値が、状態 I の頂点数の時間変化にどのような影響を与えるのかを調べる。図 5.9 に、パラメータ β , γ の値と、状態 I の頂点数の時間変化との関係を示す。図 5.9a からは、 β の値が大きいほど

- 状態 I の頂点数の最大値は大きくなるが、
- 頂点数が減少に転じる時刻も早くなる。

ことが読み取れる。一方、図 5.9b からは、 γ の値が小さいほど

- 状態 I の頂点数の最大値は大きくなり、
- 頂点数が減少に転じる時刻も遅くなる。

ことが読み取れる。

β の値の大きさは、まだその SNS に加入していない人々から見た、その SNS の魅力の大きさを表していると解釈することが可能である。なぜならば、ネットワーク irSIR モデルにおいては β の値が大きいほど、アクティブユーザーとの接触頻度が小さくても（＝アクティブユーザーから強く勧誘されなくとも）、「会員登録をして活動を開始する」という意思決定がなされる確率が高くなるためである。この解釈のもとでは、 β の値が大きく「まだその SNS に加入していない人から見た、その SNS の魅力の大きさ」が大きいほど、SNS の最大アクティブユーザー数は大きくなり、かつアクティブユーザー数の多い状態を保つ期間も長くなると考えるのが自然であろう。シミュレーションの結果によれば、最大アクティブユーザー数は、 β の値が大きいほど大きくなる。しかし、アクティブユーザー数の多い期間は、 β の値を大きくしても長くはならない。

γ の値の大きさは、既にその SNS で活動している人々から見た、その SNS の魅力の小ささを表していると解釈することが可能である。なぜならば、ネットワーク irSIR モデルにおいては γ の値が大きいほど、既に離脱したユーザーとの接触頻度が小さくても（＝離脱済みユーザーから強く説得されなくとも）、「その SNS を離脱する」という意思決定の確率が高くなるためである。このような解釈のもとでは、 γ の値が小さく「SNS で活動している人から見た、その SNS の魅力の大きさ」が大きいほど、SNS の最大アクティブユーザー数は大きくなり、アクティブユーザー数の多い状態を保てる期間も長くなると考えるのが自然であろう。シミュレーション結果は、この直感に合致している。

5.6 本章のまとめ

本章では、「特定の時期（時刻 2×10^7 までと時刻 9×10^7 以降）を除いて Friendster の時間変化がネットワーク irSIR モデルに従う」ことを示し、「 $v = 0.001, 0.005, 0.025$ のとき、 $0.002 \leq \beta \leq 0.05$, $0.002 \leq \gamma \leq 0.05$ の範囲においてネットワーク irSIR モデルではアクティブユーザー数の減少がコア中心部の縮小よりも先に始まる」ことを示した。これは、Friendster のアクティブネットワークがコア構造の中心部から縮小したのと合致する。このことは、Friendster のネットワーク構造の時間変化について分析するモデルとして、ネットワーク irSIR モデルが使える可能性があることを示している。

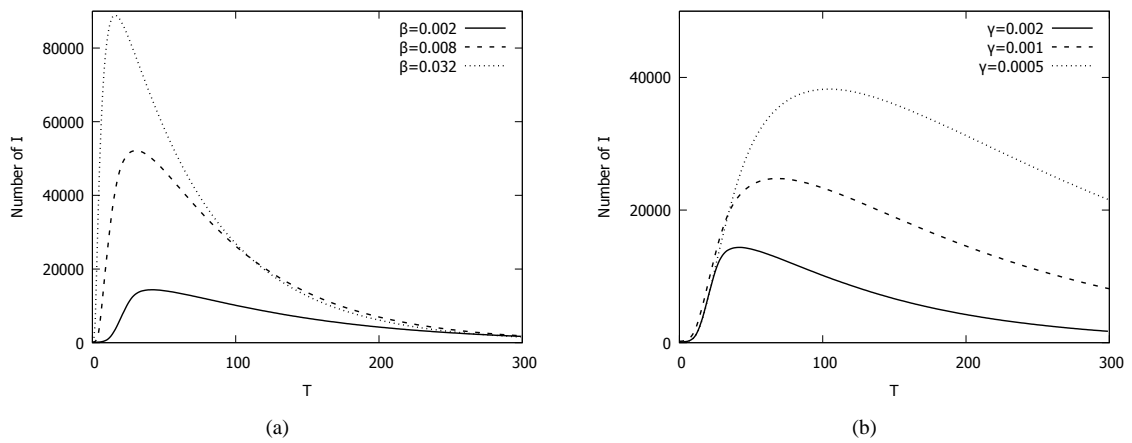


図 5.9: ネットワーク irSIR モデルで, (a) β , (b) γ の値を変えたときの, 状態 I の頂点数の時間変化. (a) $I_0 = 0.001, \nu = \gamma = 0.002$. (b) $I_0 = 0.001, \nu = \beta = 0.002$. 50 回の試行の平均値.

第6章 ネットワーク irSIR モデルの拡張

第5章では、特定の時期（時刻 2×10^7 までと時刻 9×10^7 以降）を除いて Friendster の時間変化がネットワーク irSIR モデルに従うことを示し、ネットワーク irSIR モデルではパラメータが $\nu = 0.001, 0.005, 0.025, 0.002 \leq \beta \leq 0.05, 0.002 \leq \gamma \leq 0.05$ の範囲で、コア中心部の縮小がアクティブユーザー数の減少よりも先に始まることを示した。本章ではネットワーク irSIR モデルを拡張し、遷移規則の違いでネットワークの衰退がどのように変化するかについて分析する。

各節の概要は以下の通りである。6.1 節では、5.3.1 節で得られた知見をもとに状態 S から状態 I への遷移確率に変更を加えた場合でも、5.4 節で得られた結論に影響が及ばないことを確かめる。6.2 節では、ネットワーク irSIR モデルに「次数の小さな頂点は、短期間のうちに状態 I から状態 R へ遷移する確率が高い」という規則を加えることでどのような変化が生じるかを調べる。6.3 節では、一度離脱したユーザーが友人の影響によって復帰するという現象を模すため、ネットワーク irSIR モデルに状態 R から状態 I への遷移規則を加えたネットワーク irSIRI モデルの性質を調べる。6.4 節では、SNS の管理者等が離脱したユーザーを直接説得して復帰させるという現象を模すため、ネットワーク irSIRI モデルにさらに「状態 R の頂点の中から直接選択して状態 I に遷移させる」という規則を加えた直接遷移ネットワーク irSIRI モデルの性質を調べる。6.5 節では、一人のユーザーの復帰をきっかけとして他のユーザーが連鎖的に復帰するという現象を模すため、ネットワーク irSIR モデルに「状態 R の頂点を状態 I に遷移させると、状態 R から状態 I への遷移の連鎖反応が発生する」という規則を加えた復帰の連鎖反応モデルの性質を調べる。

本章で検討するモデルの間の相関を図 6.1 に示す。6.1 節と 6.2 節ではネットワーク irSIR モデルの状態遷移確率に変更を加えた場合の結果を検討し、6.3 節から 6.5 節ではネットワーク irSIR モデルに状態 R から状態 I への遷移規則を加えることによる影響を検証する。

6.1 状態 I への遷移確率が隣接する状態 I の頂点数に比例しない場合

Friendster の状態 S から I への遷移確率のプロット（図 5.1）では、隣接する I の数が多くなるに連れて傾きがしだいに緩やかになったり、負に転じたりしている。そこで、S から I への状態遷移確率にこのような性質をもたせたモデルでも、5.4 節と同様に「アクティブネットワークはコア中心部から縮小する」という結果が得られることを確認する。

図 6.2 は、ネットワーク irSIR モデルにおいて状態 S から状態 I への遷移確率が状態 I である隣接頂点数 I_{adj} の平方根に比例すると仮定した場合の、コア構造の時間変化である。具体的には、遷移確率を $\beta \times \sqrt{I_{adj}}$ と定義し、 $\beta = 0.01$ とした。また、図 6.3 は状態 I である隣接頂点数がある値を超えると遷移確率が減少に転じるとする場合のコア構造の時間変化である。具体的には、遷移確率は隣接頂点数がある閾値 n_θ に達するまでは傾き β で線形的に増加し ($\beta \times I_{adj}$)、閾値を超えたあとは傾き $-\beta$ で線形的に減少すると仮定し

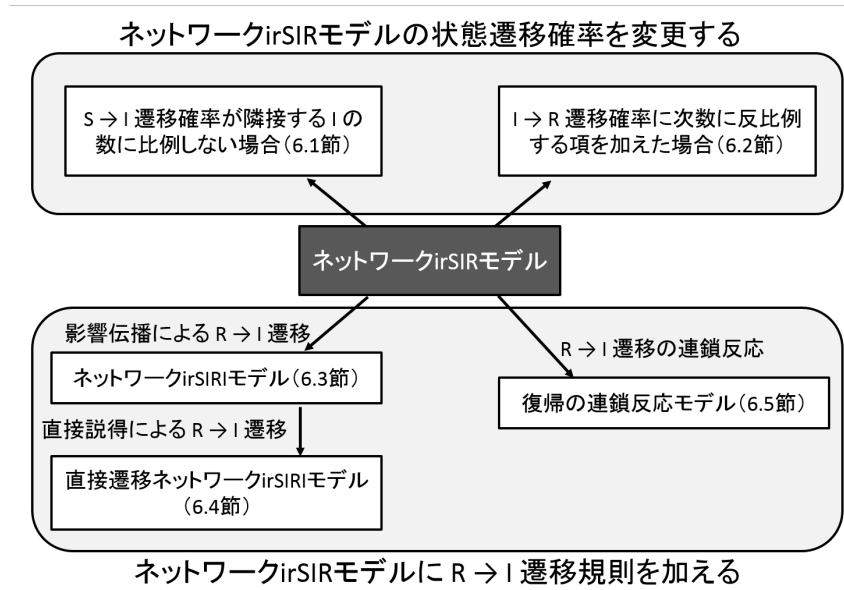


図 6.1: 本章で検討するモデルの間の相関.

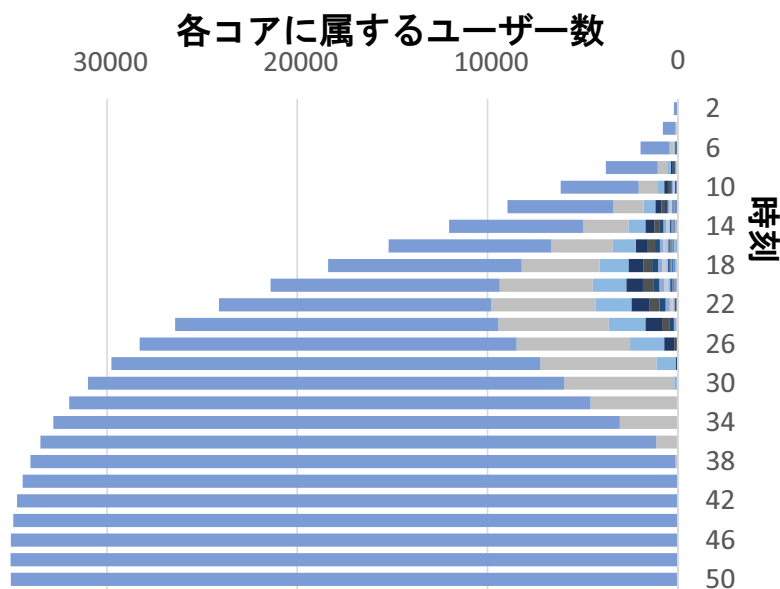


図 6.2: 状態 S から I への遷移確率が、隣接する状態 I の頂点数の平方根に比例すると仮定した場合の、コア構造の時間変化. 50 回の試行の平均値.

た. β は 0.01 に設定し, 図 6.3a では $n_\theta = 20$, 図 6.3b では $n_\theta = 80$ とした. ネットワークとしては, 5.4.2 節の辺活性化モデルで, パラメータを $N_{total} = 10^5$, $n_0 = 10$, $m_1 = m_2 = 5$ として生成したものを用いた. いずれの場合においても, コア中心部の縮小は状態 I である頂点数の減少に先立って発生していることが図から読み取れる.

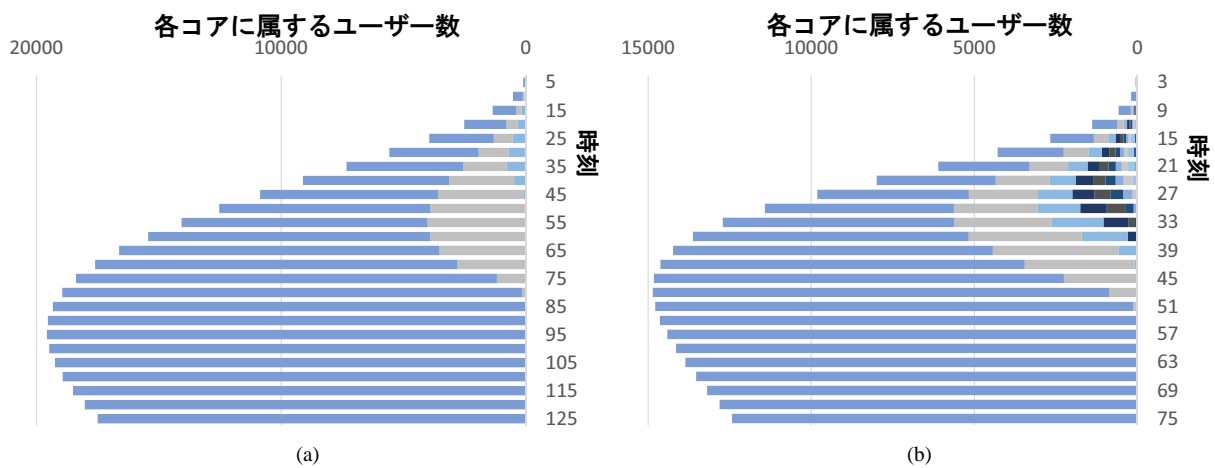


図 6.3: 状態 S から状態 I への遷移確率は, ある閾値 $I_{adj} = n_\theta$ までは線形的に増加し, その後は線形的に減少すると仮定した場合のコア構造の時間変化. (a) $n_\theta = 20$, (b) $n_\theta = 80$. 50 回の試行の平均値.

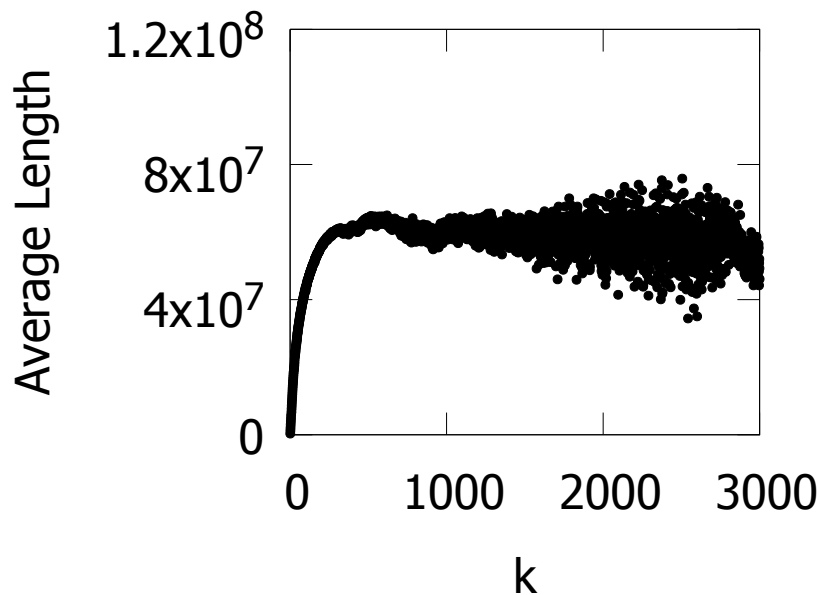


図 6.4: ネットワーク全体の中における出次数と, 状態 I であった期間の長さとの関係.

6.2 状態 R への遷移確率に次数に反比例する項を加えた場合

Friendster の全ユーザーについて, 「出次数」と「状態 I である期間の長さ」との関係を図 6.4 に示した. 状態 I である期間の長さは, 出次数が 200 以上では $6 \times 10^7 \pm 2 \times 10^7$ の範囲に入っている. 一方, 出次数が 200 未満の範囲では, 出次数が少ないほど状態 I である期間が短い.

そこで, このように状態遷移確率を変更する. 修正モデルでは, 「次数の小さなユーザーは状態 R に遷移しやすい」という傾向を具体的に μ/k という項で表現し, ネットワーク irSIR モデルの状態 I から状態 R への遷移確率を $\nu + \gamma R_{adj} + \mu/k$ とする. ただし, k はその頂点の次数, μ はパラメータである.

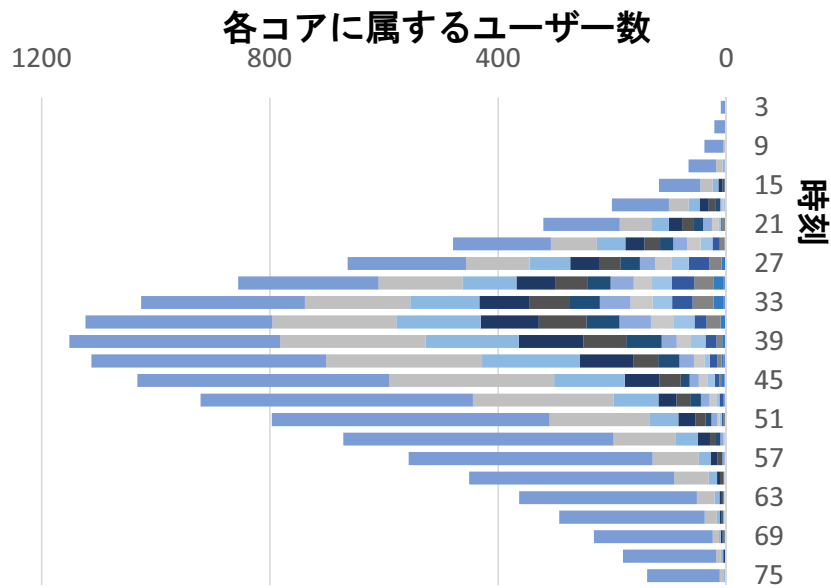


図 6.5: 状態 R への遷移確率に度数に反比例する項を加えた場合の, アクティブネットワークのコア構造の時間変化. $I_0 = 0.001$, $\beta = \nu = \gamma = 0.002$, $\mu = 5$. 50 回の試行の平均値.

修正モデルを用いてシミュレーションした結果の一例を図 6.5 に示す. 修正前のモデルのシミュレーションの結果 (図 5.7) と Friendster のコア構造の時間変化 (図 4.4) を比べると, Friendster においてはアクティブネットワークの縮小がある程度進行した後でもコア値の高いユーザーがある程度存在し続けるのに対し, 修正前のモデルにおいては状態 I の頂点数の減少傾向が顕著になる前に, コア値が 5 以上である頂点はほぼ完全に消滅している. これに対し, 修正モデル (図 6.5) では, 状態 I の頂点数が最大値をとる時刻 39 付近においても明確なコア構造が残っている.

状態 I の頂点数が最大値をとる時刻における, 状態 I の頂点全体のうちコア値が 5 以上であるものの割合を $Prop_5$ と定義する. 図 6.6 では, 修正前のモデルと修正モデルについて, $Prop_5$ の値を $0.002 \leq \beta \leq 0.05$, $0.002 \leq \gamma \leq 0.05$ の範囲で比較している. パラメータ領域の大部分において, $Prop_5$ は修正モデルのほうが高い値を示す. つまり, 修正モデルでは, 修正前のモデルにみられた, 「状態 I の頂点数が減少に転じる前にコア中心部が急激に縮小する」という傾向が緩和されている.

6.3 ネットワーク irSIRI モデル

実際のネットワークでは, すでにその SNS を離れたユーザーが復帰することがある. これを組み込んだモデルとしては Ribeiro のモデル (3.3.3 節) があるが, ここではネットワーク構造を考慮して, ネットワーク irSIR モデルに復帰の規則を組み込んだネットワーク irSIRI モデルを定義する. ネットワーク irSIRI モデルは, ネットワーク irSIR モデルに

- 状態 R である頂点は, 状態 I の頂点と多く隣接しているほど高い確率で状態 I へ遷移する.

という規則を加えたものである.

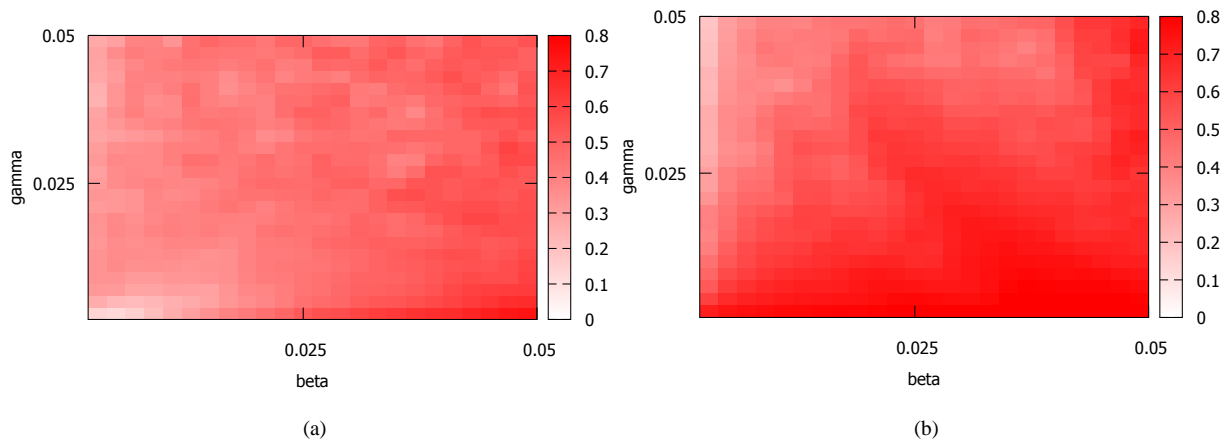


図 6.6: $0.002 \leq \beta \leq 0.05$, $0.002 \leq \gamma \leq 0.05$ の範囲における, $Prop_5$ の値の比較. (a) ネットワーク irSIR モデル, (b) 状態 R への遷移確率に次数に反比例する項を加えた場合 ($\mu = 5$). $\nu = 0.002$, 20 回の試行の平均値.

6.3.1 ネットワーク irSIRI モデルのアルゴリズム

ネットワーク irSIRI モデルのアルゴリズムを以下のように定義する. ここでは便宜上, 頂点の状態を S, I_1 , R, I_2 の 4 つに分ける. 状態 I を状態 I_1 と状態 I_2 の 2 つに分けたのは, 状態 I への初めての遷移 (I_1) と 2 度目以降の遷移 (I_2) とを分けて分析することを可能とするためであり (図 6.7), 2 つの状態を区別する必要がない場合は状態 I_1 と状態 I_2 を合わせて状態 I と呼ぶ.

- **Input**: 生成されたネットワーク, 初期状態において状態 I_1 である頂点の割合 I_0 , 状態 S から I_1 への遷移確率を決めるパラメータ β_1 , 状態 I から R への遷移確率を決めるパラメータ γ , ν , 状態 R から I_2 への遷移確率を決めるパラメータ β_2
- **Output**: 各ステップにおける, ネットワーク中の各頂点の状態
- **初期状態**: 各頂点は状態 I_1 か状態 S のいずれかの状態をとる. 状態 I_1 をとる確率を I_0 とする.
- **各ステップで**, 各頂点は以下の規則で確率的に状態遷移する.

状態 S の頂点: 確率 $\beta_1 I_{adj}$ で状態 I_1 に遷移する.

状態 I_1, I_2 の頂点: 確率 $\gamma R_{adj} + \nu$ で状態 R に遷移する.

状態 R の頂点: 確率 $\beta_2 I_{adj}$ で状態 I_2 に遷移する.

ただし, I_{adj} は隣接する状態 I の頂点数, R_{adj} は隣接する状態 R の頂点数である.

6.3.2 ネットワーク irSIRI モデルのシミュレーションの結果

ネットワーク irSIRI モデルにおいて, パラメータ β_2 と γ の値の大小関係が状態 I の頂点数の時間変化に与える影響を調べるため, 次の条件で実験を行った. ネットワークとしては, 5.4.2 節の辺活性化モデルで,

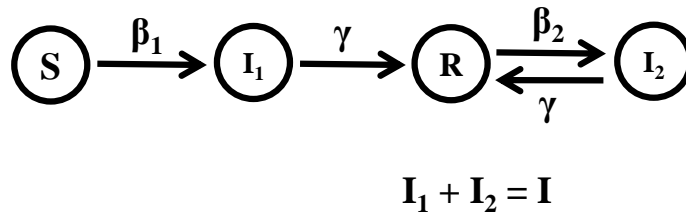


図 6.7: ネットワーク SIRS モデルの遷移規則.

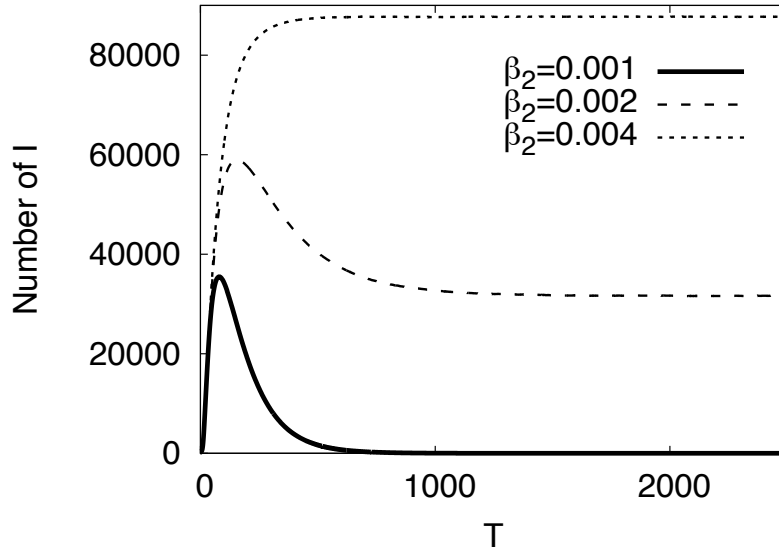


図 6.8: ネットワーク irSIRS モデルにおける, 状態 I の頂点数の時間変化. $I_0 = 0.001, \nu = \beta_1 = \gamma = 0.002$. 50 回の試行の平均値.

パラメータを $N_{total} = 10^5$, $n_0 = 10$, $m_1 = m_2 = 5$ として生成したものを用いた.

実験 1 $I_0 = 0.001, \nu = \beta_1 = \gamma = 0.002$ とし, $\beta_2 = 0.001, 0.002, 0.004$ のときの, 状態 I の頂点数の時間変化を比較する. ただし, 状態遷移のステップ数は 2500 とし, 50 回の試行の平均をとる.

実験 1 の結果を図 6.8 と図 6.9 に示す. 図 6.8 は状態 I である頂点数の時間変化を, 図 6.9 では状態 I を細分化して, 状態 I_1 の頂点数と状態 I_2 の頂点数の時間変化をそれぞれ示している.

図 6.9 によれば, いずれの条件においても状態 I_1 である頂点の数はパラメータ値によらず, ある時刻に最大値をとり, その後は減少の一途をたどる. 一方, 状態 I_2 の頂点数は, $\beta_2 \gg \gamma$ のときは時刻とともに単調に増加し, 正のある値に収束する. $\beta_2 \ll \gamma$ のときは, 状態 I_2 の頂点は最終的に消滅する. したがって, 状態 I である頂点の数は初めのうちは単調に増加していくが, その後はパラメータの値によって正のある値に収束する場合と, 0 に収束する場合とに分かれる.

次に, パラメータ β_1, ν が状態 I の頂点数の時間変化に与える影響を調べるため, 以下の実験を行った.

実験 2 $I_0 = 0.001, \nu = \beta_2 = \gamma = 0.002$ とし, $\beta_1 = 0.001, 0.002, 0.004$ のときの, 状態 I の頂点数の時間変化を比較する. ただし, 状態遷移のステップ数は 2500 とし, 50 回の試行の平均をとる.

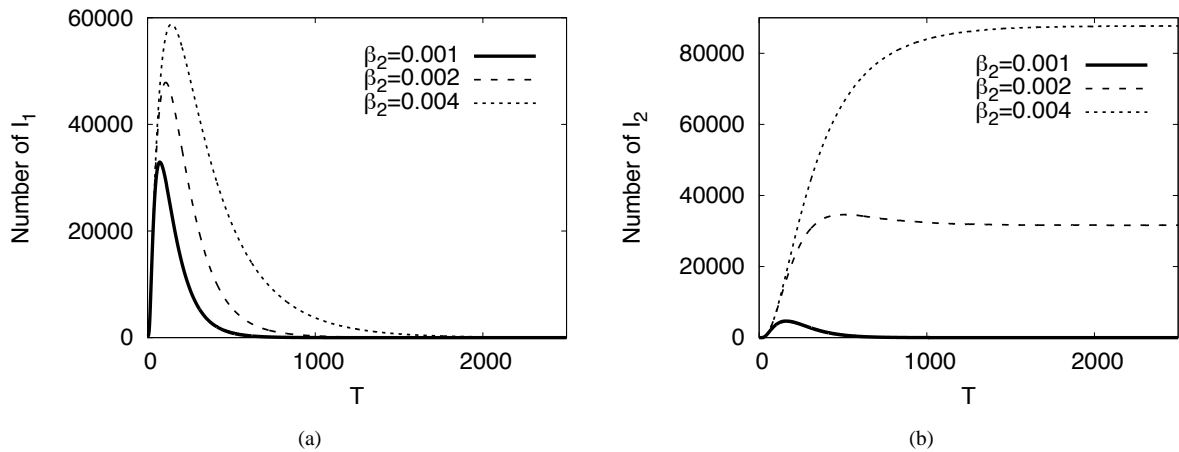


図 6.9: ネットワーク irSIRI モデルにおける, 状態 I_1 , 状態 I_2 の頂点数の時間変化. (a) I_1 , (b) I_2 . $I_0 = 0.001, \nu = \beta_1 = \gamma = 0.002$. 50 回の試行の平均値.

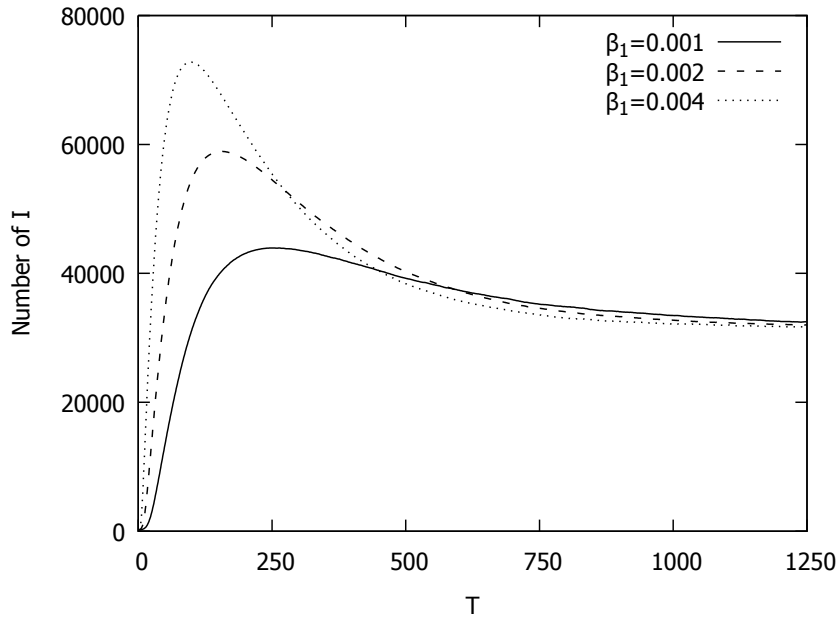


図 6.10: いくつかの β_1 の値のもとでの, 状態 I の頂点数の時間変化. $I_0 = 0.001, \nu = \beta_2 = \gamma = 0.002$. 50 回の試行の平均値.

実験 3 $I_0 = 0.001, \beta_1 = \beta_2 = \gamma = 0.002$ とし, $\nu = 0.0004, 0.002, 0.01$ のときの, 状態 I の頂点数の時間変化を比較する. ただし, 状態遷移のステップ数は 2500 とし, 50 回の試行の平均をとる.

実験 2 の結果を図 6.10 に示す. このモデルでは, 最終的には状態 S である頂点が消滅し, 状態 S から状態 I への遷移が一切発生しなくなるため, いずれの条件のもとでも状態 I の頂点数は等しい値に収束する.

実験 3 の結果を示した図 6.11 によれば, 平衡状態における状態 I の頂点数は ν の値の大小によっても影響を受けることがわかる.

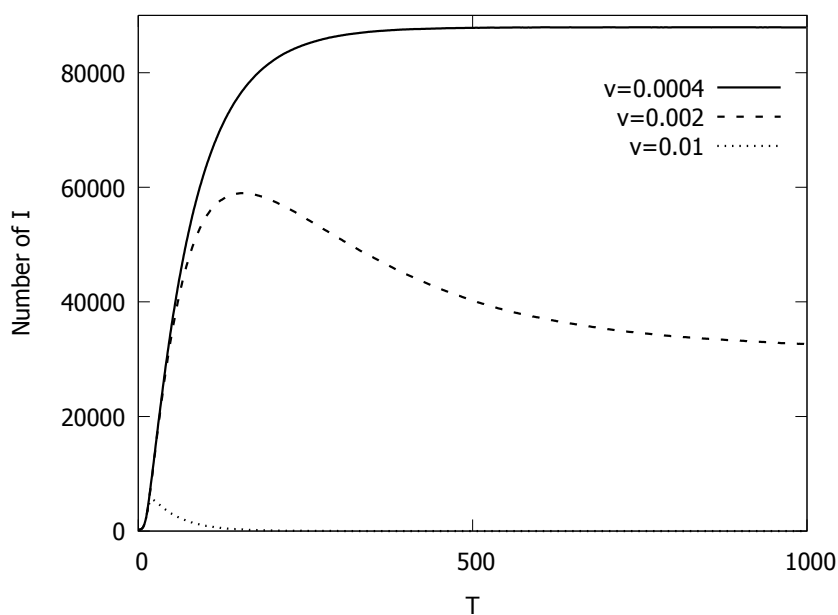


図 6.11: いくつかの v の値のもとでの，状態 I の頂点数の時間変化. $I_0 = 0.001$, $\beta_1 = \beta_2 = \gamma = 0.002$. 50 回の試行の平均値.

6.3.3 実験結果の解釈

6.3.2 節の実験結果は，ネットワーク irSIRI モデルにしたがう SNS において，SNS の言葉では次のように解釈できる.

- 非アクティブユーザーのアクティブ化の確率 β_2 が十分高い場合，アクティブユーザーは消滅しない.
- 非ユーザーの会員登録の確率 β_1 が増加しても，平衡状態におけるアクティブユーザー数は増加しない.
- パラメータ β_2 , γ のみならず，状態 R に遷移する確率を決めるパラメータ v の値も，平衡状態におけるアクティブユーザー数に影響を与える.

6.4 直接遷移ネットワーク irSIRI モデル

本節では，「SNS を離脱した人々を，SNS の運営者等が直接説得して引き戻す」という現象を再現するため，直接遷移ネットワーク irSIRI モデルを定義し，状態 I の頂点数の時間変化にどのような変化が生じるかを調べる.

6.4.1 直接遷移ネットワーク irSIRI モデルのアルゴリズム

直接遷移ネットワーク irSIRI モデルのアルゴリズムを以下のように定義する.

アルゴリズム 1 (遷移させる頂点をランダムに選択する場合)

- **Input** : 生成されたネットワーク, 初期状態における状態 I の頂点の割合 I_0 , 状態 S から I への遷移確率を決めるパラメータ β_1 , 状態 I から R への遷移確率を決めるパラメータ γ , ν , 状態 R から I への遷移確率を決めるパラメータ β_2 , 各ステップで状態 R から状態 I に遷移させる頂点数 n_p
- **Output** : 各ステップにおける, ネットワーク中の各頂点の状態
- **初期状態** : 各頂点は状態 I か状態 S のいずれかの状態をとる. 状態 I をとる確率を I_0 とする.
- 各ステップで,
 - 各頂点は以下の規則で確率的に状態遷移する.
 - 状態 S の頂点: 確率 $\beta_1 I_{adj}$ で状態 I に遷移する.
 - 状態 I の頂点: 確率 $\gamma R_{adj} + \nu$ で状態 R に遷移する.
 - 状態 R の頂点: 確率 $\beta_2 I_{adj}$ で状態 I に遷移する.
 - 状態 R である頂点の中からランダムに n_p 個を選択し, 状態 I に遷移させる.

アルゴリズム 2 (遷移させる頂点を度数に比例する確率で選択する場合)

- **Input** : 生成されたネットワーク, 初期状態における状態 I の頂点の割合 I_0 , 状態 S から I への遷移確率を決めるパラメータ β_1 , 状態 I から R への遷移確率を決めるパラメータ γ , ν , 状態 R から I への遷移確率を決めるパラメータ β_2 , 各ステップで状態 R から状態 I に遷移させる頂点数 n_0
- **Output** : 各ステップにおける, ネットワーク中の各頂点の状態
- **初期状態** : 各頂点は状態 I か状態 S のいずれかの状態をとる. 状態 I をとる確率を I_0 とする.
- 各ステップで,
 - 各頂点は以下の規則で確率的に状態遷移する.
 - 状態 S の頂点: 確率 $\beta_1 I_{adj}$ で状態 I に遷移する.
 - 状態 I の頂点: 確率 $\gamma R_{adj} + \nu$ で状態 R に遷移する.
 - 状態 R の頂点: 確率 $\beta_2 I_{adj}$ で状態 I に遷移する.
 - 状態 R である頂点の中から度数に比例する確率で n_p 個を選択し, 状態 I に遷移させる.

ただし, I_{adj} は隣接する状態 I の頂点数, R_{adj} は隣接する状態 R の頂点数である

6.4.2 直接遷移ネットワーク irSIRI モデルのシミュレーションの結果

直接遷移ネットワーク irSIRI モデルにおいて, 1) n_p の値と, 2) 遷移させる頂点の選択方法とが結果に与える影響を確かめるため, 以下の実験を行った. ネットワークとしては, 5.4.2 節の辺活性化モデルで, パラメータを $N_{total} = 10^5$, $n_0 = 10$, $m_1 = m_2 = 5$ として生成したものを用いた.

実験1 アルゴリズム1において、 $n_p = 0, 5, 10, 15, 20$ に対してシミュレーションを行う。

実験2 $n_p = 10, 30, 50$ として、アルゴリズム1とアルゴリズム2を用いてシミュレーションを行う。

パラメータ β , γ の値としては、Friendster において実際に観測された値を使用することも考えられる。しかし、Friendster において β , γ の値は一定ではなく、時刻によって様々な値をとる (図 5.1, 図 5.3)。そこで、ここでは単純化のため、影響伝播による状態 S から I, I から R, R から I への遷移のしやすさが同程度となるよう、いずれの実験においても、 $I_0 = \beta_1 = \beta_2 = \gamma = \nu = 0.002$ とした。

実験1の結果を図 6.12 に示す。図 6.12 からは、 n_p の値を大きくするほど、状態 I の頂点数は大きな値に収束することが読み取れる。

実験2の結果を図 6.13 に示す。この図からは、状態遷移させる頂点を等確率で選ぶか次数に比例する確率で選ぶかは、平衡状態における状態 I の頂点数にほとんど影響を及ぼさないことが読み取れる。

6.4.3 実験結果の解釈

6.4.2 節の実験結果は、直接遷移ネットワーク irSIRI モデルに従う SNS において、SNS の言葉では次のように解釈できる。

1. 単位時間あたりに SNS に復帰させるユーザーの数が多いほど、平衡状態におけるアクティブユーザー数は多くなる。
 2. SNS に復帰させるユーザーをランダムに選択するか、影響力の大きなユーザーに優先的にはたらしかけるかは、平衡状態におけるアクティブユーザー数に影響しない。
2. は、「影響力の大きなユーザーを呼び込めば、その影響を受けて多くの人が SNS に加わるため、アクティブユーザー数を増加させる上で効果的だろう」という直感には反した結果である。

6.5 復帰の連鎖反応モデル

現実の SNS では、次のような現象がしばしば発生する。

あるところに、一つの SNS があった。その SNS は、かつては盛り上がっていたのだが、最近ではアクティブユーザー数の減少が顕著である。そんなある日、盛り上がっていた頃の中心人物であったユーザーが、数年ぶりに姿を現した。その噂を聞きつけたかつてのユーザー達が一人また一人と集まってきて、SNS は一時的にかつての盛り上がりを取り戻した。

そこで、そのような現象を生じるモデルを考える。そのために、ネットワーク irSIRI モデルで加えた状態 R から状態 I への遷移規則の代わりに、1) ランダムまたは次数に比例する確率での状態 R \rightarrow I 遷移と、2) 「復帰の連鎖反応」という規則を加えた、復帰の連鎖反応モデルというモデルを定義する。

「復帰の連鎖反応」とは、次のような規則である。

- 連鎖反応を開始する頂点の選択 ネットワーク中の状態 R である頂点の中から 1 つ 1) ランダムに、または 2) 次数に比例する確率で選択し、状態 I に遷移させる

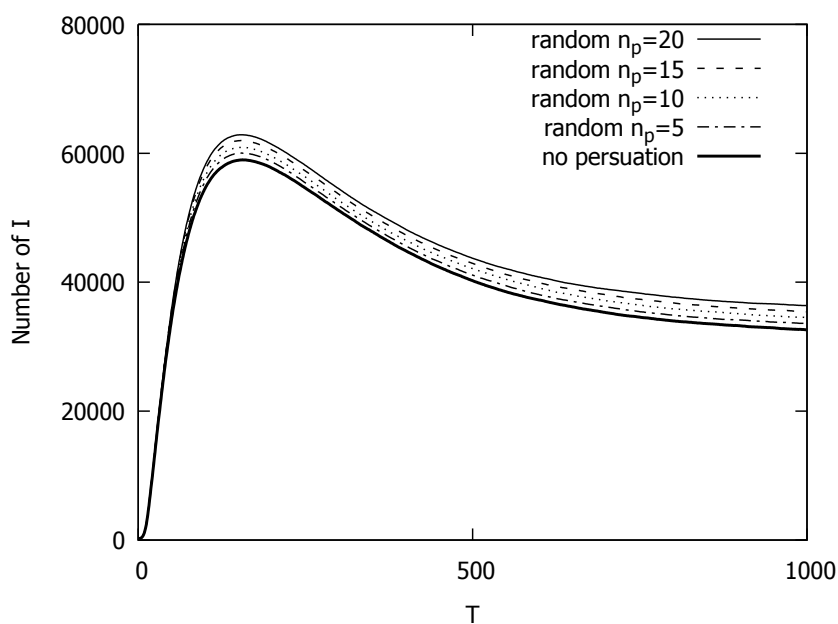


図 6.12: 直接遷移ネットワーク irSIRI モデル (ランダム) の, n_p が状態 I の頂点数の時間変化に及ぼす影響. $I_0 = 0.002$, $\nu = \beta_1 = \beta_2 = \gamma = 0.002$. 50 回の試行の平均値.

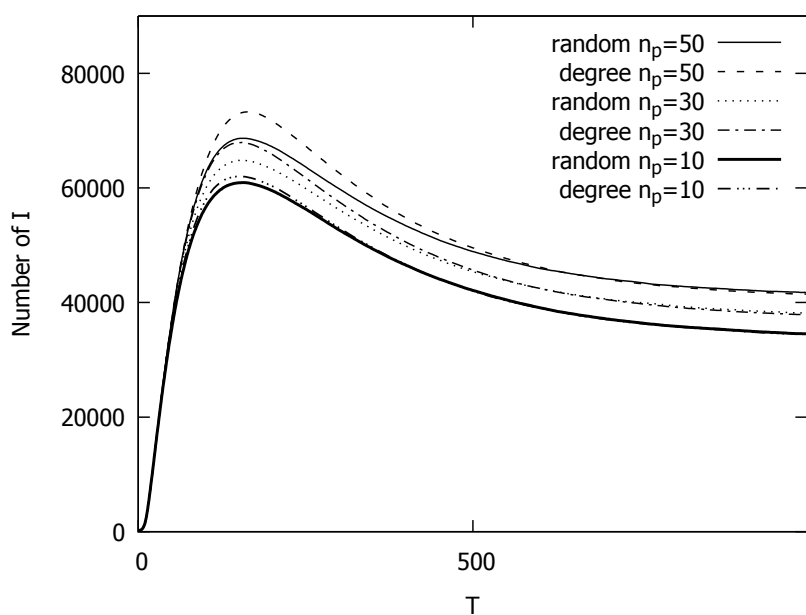


図 6.13: 直接遷移ネットワーク irSIRI モデルで, 状態遷移させる頂点をランダムに選択するか度数に比例する確率で選択するかの違いが, 状態 I の頂点数の時間変化に及ぼす影響. $I_0 = 0.002$, $\nu = \beta_1 = \beta_2 = \gamma = 0.002$. 50 回の試行の平均値.

連鎖反応 状態 R である頂点は, 隣接する頂点のうちの 1 つが状態 R から状態 I に遷移するごとに,

- 確率 β_3 で状態 I に遷移し,
- 確率 $1 - \beta_3$ で状態 R に留まり続ける.

新たな状態遷移が発生しなくなるまで、これが連鎖的に繰り返される。

ネットワーク irSIRI モデルとの大きな違いは、隣接する頂点の状態ではなく状態遷移によって、状態遷移が引き起こされる点である。

6.5.1 復帰の連鎖反応モデルのアルゴリズム

復帰の連鎖反応モデルのアルゴリズムを以下のように定義する。

アルゴリズム

- **Input** : 生成されたネットワーク, 初期状態における状態 I の頂点の割合 I_0 , 状態 S から I への遷移確率を決めるパラメータ β_1 , 状態 I から R への遷移確率を決めるパラメータ γ , ν , 「復帰の連鎖反応」の起こりやすさを決めるパラメータ β_3 , 各ステップで状態 R から状態 I に遷移させる頂点数 n_0
- **Output** : 各ステップにおける, ネットワーク中の各頂点の状態
- **初期状態** : 各頂点は状態 I か状態 S のいずれかの状態をとる. 状態 I をとる確率を I_0 とする.
- **各ステップ** で, 各頂点は以下の規則によって確率的に状態遷移する.

状態 S の頂点: 確率 $\beta_1 I_{adj}$ で状態 I に遷移する.

状態 I の頂点: 確率 $\gamma R_{adj} + \nu$ で状態 R に遷移する.

状態 R の頂点: 「復帰の連鎖反応」サブルーチン 1 または 2

「復帰の連鎖反応」サブルーチン 1 (連鎖反応の開始点をランダムに選択する場合)

1. 状態 R である頂点の中から 1 つランダムに選択し, 状態 I に遷移させる. また, 選択された頂点はキューに入れられる.
2. キューが空でない間, 以下を繰り返す.
 - キューから頂点を 1 つ取り出す. 取り出された頂点を i とする.
 - i と隣接しており, かつ状態 R である全ての頂点について, 以下の規則により確率的に状態遷移を行う.
 - 確率 β_3 : 状態 I へ遷移し, キューに入れられる.
 - 確率 $1 - \beta_3$: 状態遷移は発生しない.

「復帰の連鎖反応」 サブルーチン 2 (連鎖反応の開始点を次数に比例する確率で選択する場合)

1. 状態 R である頂点の中から 1 つ次数に比例する確率で選択し、状態 I に遷移させる。また、選択された頂点はキューに入れられる。
2. キューが空でない間、以下を繰り返す。
 - キューから頂点を 1 つ取り出す。取り出された頂点を i とする。
 - i と隣接しており、かつ状態 R である全ての頂点について、以下の規則により確率的に状態遷移を行う。
 - 確率 β_3 : 状態 I へ遷移し、キューに入れられる。
 - 確率 $1 - \beta_3$: 状態遷移は発生しない。

6.5.2 復帰の連鎖反応モデルのシミュレーションの結果

1) 連鎖反応を開始する頂点の選択のしかたと、2) パラメータ β_3 の値とが状態 I の頂点数の時間変化のしかたに与える影響を確かめるため、以下の実験を行った。ネットワークとしては、5.4.2 節の辺活性化モデルで、パラメータを $N_{total} = 10^5$, $n_0 = 10$, $m_1 = m_2 = 5$ として生成したものをを用いた。

実験 1 $\beta_3 = 0.01$ として、サブルーチン 1 を用いたシミュレーションとサブルーチン 2 を用いたシミュレーションを 1 回ずつ行い、状態 I の頂点数の時間変化を求めた。

実験 2 実験 1 の各条件におけるシミュレーションを 50 回ずつ行い、状態 I の頂点数の平均値の時間変化を求めた。

実験 3 $\beta_3 = 0.001, 0.1$ として、サブルーチン 1 を用いたシミュレーションとサブルーチン 2 を用いたシミュレーションを 50 回ずつ行い、状態 I の頂点数の平均値の時間変化を求めた。

実験 4 $\beta_3 = 10^{-0.5}, 10^{-1}, 10^{-1.25}, 10^{-1.5}, 10^{-1.75}, 10^{-2}, 10^{-2.25}, 10^{-2.5}, 10^{-3}$ として、サブルーチン 1 を用いたシミュレーションとサブルーチン 2 を用いたシミュレーションを 20 回ずつ行い、1000 ターン経過後の状態 I の頂点数を求めた。

なお、影響伝播による状態 S から I, I から R への遷移のしやすさを同程度とするため、いずれの実験においても、 $I_0 = \beta_1 = \gamma = \nu = 0.002$ とした。

図 6.14a は、実験 1 の結果を示している。連鎖反応の出発点となる頂点をランダムに選択するとき、状態 I の頂点数は低い値に抑えられている (例えば、図 6.14a の時刻 380 付近から時刻 450 付近までは、大規模な連鎖反応が発生せず状態 I の頂点数が単調に減少している)。一方、次数の大きな頂点を優先的に選択するとき、状態 I の頂点数は長期間にわたって高い値に保たれている。これは、次数の大きな頂点を優先的に選択するほうが、大規模な連鎖反応が頻繁に発生するためであると説明できる。

図 6.14b は実験 2 の結果を示している。50 回の試行の平均をとった場合であっても、次数の大きな頂点を優先的に選択した場合のほうが、状態 I の頂点数は高い値を保つことがわかる。この原因は、実験 1 で観

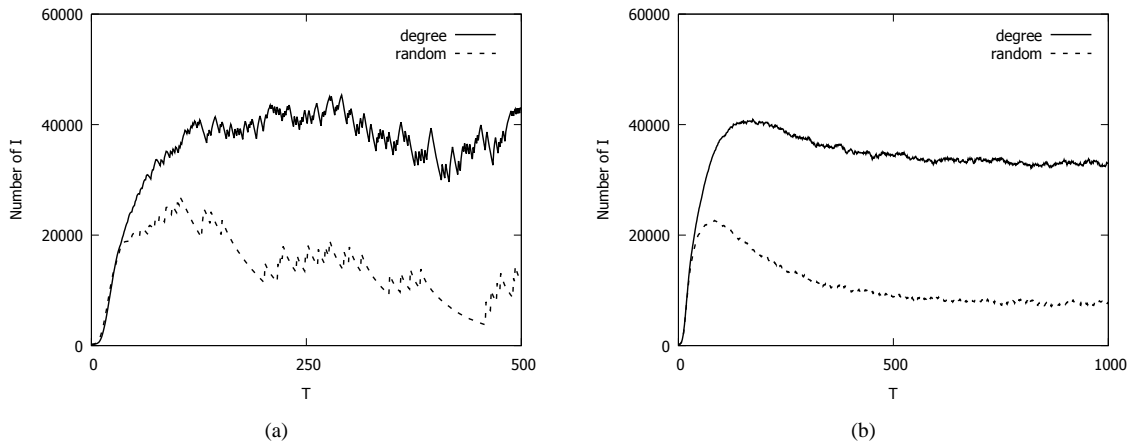


図 6.14: 復帰の連鎖反応モデルのシミュレーション結果. $I_0 = 0.002$, $\nu = \beta_1 = \gamma = 0.002$, $\beta_3 = 0.01$. (a) 1 回の試行の結果 (実験 1), (b) 50 回の試行の平均値 (実験 2).

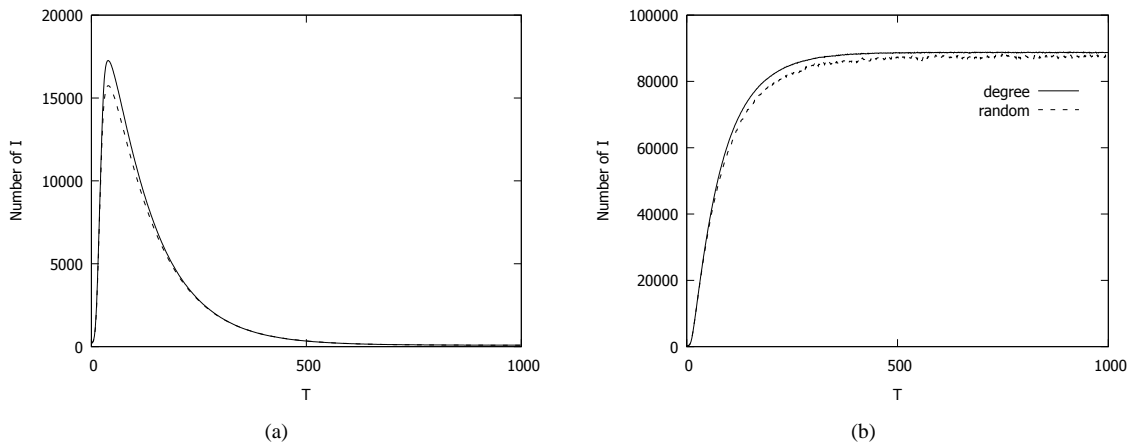


図 6.15: 復帰の連鎖反応モデルにおける, 実験 3 の結果. (a) $\beta_3 = 0.001$, (b) $\beta_3 = 0.1$. $I_0 = 0.002$, $\nu = \beta_1 = \gamma = 0.002$.

察したように, 次数の大きな頂点を優先的に選択した場合のほうが, 大規模な連鎖反応が発生しやすいためであると説明できる.

図 6.15 は実験 3 の結果を示している. 実験 3 で値を変更したパラメータ β_3 は, 連鎖反応の広がりやすさを司るパラメータである. β_3 の値が小さいとき (図 6.15a) は, 連鎖反応の開始点の選び方によらず, 状態 I の頂点はきわめて小さな値に収束することが分かる. 一方, β_3 の値が大きいつき (図 6.15b) は, 連鎖反応の開始点の選び方によらず, ネットワーク中の大部分の頂点が状態 I となった状態で安定することが分かる.

β_3 の値と平衡状態における状態 I の頂点数との関係をより詳細に調べた, 実験 4 の結果を図 6.16 に示す. $10^{-2} \leq \beta_3 \leq 10^{-1.5}$ の範囲では, 次数の大きな頂点を優先的に選ぶ場合のほうが, 平衡状態における状態 I の頂点数が 20000 個 (20%) 以上多い. $\beta_3 < 10^{-2}$, $10^{-1.5} < \beta_3$ の範囲では, 二つの条件による結果の差は小さい.

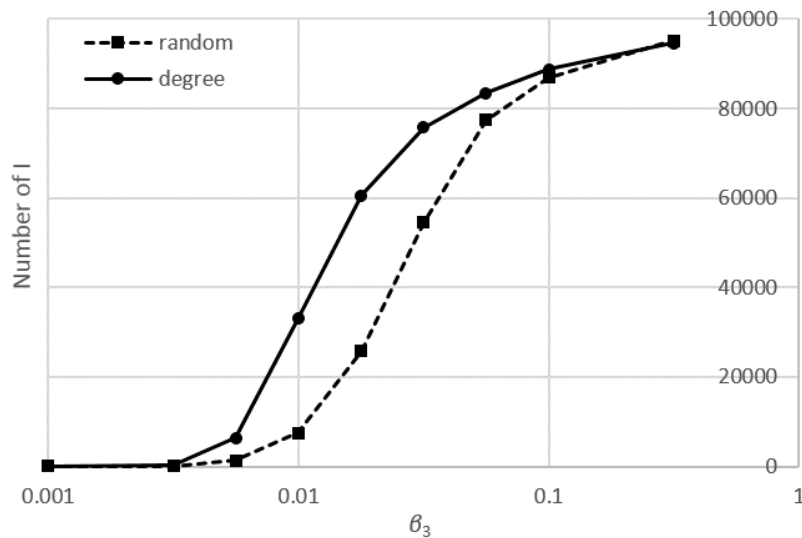


図 6.16: 復帰の連鎖反応モデルにおける, いくつかの β_3 の値と連鎖反応の開始点の選択方法のもとでの, 平衡状態における状態 I の頂点数.

6.5.3 実験結果の解釈

6.5.2 節の実験結果は, 復帰の連鎖反応モデルに従う SNS において, SNS の言葉では次のように解釈できる.

- SNS への復帰が連鎖反応を起こすと仮定するとき, 次数の大きな (影響力の大きな) ユーザーを優先的に復帰させることは, ランダムに選択したユーザーを復帰させるよりも効果的である.
- ただし, 次数の大きなユーザーを優先的に復帰させることが効果的であるのは, 連鎖反応の広がりやすさを司るパラメータ β_3 の値が $10^{-2} \leq \beta_3 \leq 10^{-1.5}$ の範囲にある場合に限られる.

6.6 本章のまとめ

本章では, ネットワーク irSIR モデルにさまざまな規則を加えたモデルの性質を分析した. また, これらのモデルのシミュレーション結果を解釈することを通して, SNS のさまざまな性質が成長と衰退の過程に与える影響を捉えることを目指した.

状態 S から状態 I への遷移確率に変更を加えたシミュレーションからは, 加えた変更が前章の結論に影響を及ぼさないことが確かめられた.

状態 I から状態 R への遷移確率に次数に反比例する項を加えたシミュレーションからは, ネットワーク irSIR モデルに「次数の小さな頂点は, 短期間のうちに状態 I から状態 R へ遷移する確率が高い」という規則を加えることで, コア中心部の急激な縮小が抑えられることが確かめられた.

ネットワーク irSIRI モデルのシミュレーションからは, 状態 R から状態 I への遷移確率を決定するパラメータ γ の値を十分に大きくすれば, 状態 I の頂点数は正のある値に収束することが確かめられた.

直接遷移ネットワーク irSIRI モデルのシミュレーションからは, 次のことが明らかになった.

- 単位時間あたりに状態 R から状態 I に遷移させる頂点数を多くするほど、平衡状態における状態 I の頂点数は多くなる。
- 状態 R から状態 I に遷移させる頂点をランダムに選んでも、次数に比例する確率で選んでも、平衡状態における状態 I の頂点数はほとんど変化しない。

復帰の連鎖反応モデルのシミュレーション結果からは、次のことが明らかになった。

- 状態 R から状態 I への遷移が連鎖反応を起こすと仮定するとき、状態 I へ遷移させる頂点を次数に比例する確率で選択するほうが、ランダムに選択する場合よりも、平衡状態における状態 I の頂点数は大きくなる。
- 上のような現象がみられるのは、連鎖反応の起こりやすさを司るパラメータの値 β_3 の大きさが $10^{-2} \leq \beta_3 \leq 10^{-1.5}$ であるときに限られる。

以上のモデルのシミュレーション結果を、SNS の用語で解釈すると以下のようなになる。

- 離脱したユーザーの復帰がある閾値以上の確率で発生するとき、アクティブユーザー数は正のある値に収束する。
- 復帰の連鎖反応がなければ、SNS の管理者などが直接復帰を働きかけるとき、働きかけるユーザーをランダムに選択する場合と、次数の高いユーザーを優先的に選択する場合との間で、平衡状態におけるアクティブユーザー数には差がない。ただし、あるユーザーが SNS に復帰することによって復帰の連鎖反応が引き起こされると仮定すると、次数の高いユーザーに優先的に働きかけることは効果的である。

ただし、これはその SNS が本章のモデルに沿っている場合であることには留意する必要がある。

第7章 おわりに

7.1 結論

本論文では、Friendster という SNS が衰退する過程におけるネットワーク構造の時間変化を分析し、ネットワーク irSIR モデルを用いてメカニズムの分析を行った。

初めに、Friendster 上のユーザーネットワークが縮小する過程を分析した。各時刻におけるアクティブネットワークの構造を推定し、有向 k -core decomposition 法を用いてコア構造を分析したところ、コア構造中心部の縮小がアクティブユーザー数の減少よりも先に開始したことが明らかになった。

そこで次に、Friendster のユーザーネットワークがコア構造の内側から縮小したメカニズムを分析するために、ネットワーク irSIR モデルを構築した。ネットワーク irSIR モデルに従って時間変化するネットワークはコア構造の内側から縮小することを、シミュレーションによって示した。

次に、ネットワーク irSIR モデルにさまざまな追加規則を加えたモデルのシミュレーションを行った。この結果から、これらのモデルに従う SNS において、SNS が衰退する過程や SNS の衰退が防止される条件についての知見を得た。

本論文の貢献をまとめると以下ようになる。

- ユーザーの会員登録の順序のみを追加情報として持つ、ある一時点におけるネットワーク構造を示すデータセットから、その時点以前のネットワーク構造の時間変化を推定する手法を考案した (4.2.1 節)。これにより、先行研究では行われていなかった、Friendster の衰退過程におけるネットワーク構造の時間変化の分析を行った。
- Friendster において、いくつかの仮定の下でコア構造の中心部の縮小がアクティブユーザー数の減少に先行したことを示した (4.3 節)。
- コア構造が時間とともに変化していく様子を描写する手法を考案した (4.3.1 節)。この手法は、Friendster に限らずネットワークのコア構造分析の際に適用可能である。
- 各ユーザーに割り振られた ID と、Internet Archive に保存されたユーザーページの情報とを組み合わせることによって、Friendster のユーザー数の時間変化を明らかにした (4.4.1 節)。ユーザー数に関する情報が運営者から公式に発表されることはほとんど無かったため、本論文の手法によって初めて情報が得られたと言える。
- irSIR モデルを、ネットワーク構造を考慮したネットワーク irSIR モデルへと拡張した (5.2 節)。これによって、SNS 上のユーザーネットワークの時間変化と irSIR モデルとを関連させて論じることが可能となった。

- Friendster の状態遷移確率を単純化すると、ネットワーク irSIR モデルで表せることを示した (5.3 節)。
- ネットワーク irSIR モデルによるネットワークのコア構造の変化が、Friendster のネットワークのコア構造を説明できることを示した (5.4 節)。更にその結果が、ある範囲で仮定を変更しても変わらないことを示した。
- ネットワーク irSIR モデルに修正を加えたモデルをシミュレーションすることにより、「一度離脱したユーザーが SNS に復帰する」という現象などが SNS にもたらす影響を調べた (第 6 章)。

7.2 今後の展望

本論文では多くの知見を得ることができたが、以下のような課題も残されている。今後はこれらの課題に取り組むことによって、本論文で得られた成果を更に発展させることが可能であると考えられる。

数か割合か 本論文で用いたネットワーク irSIR モデルにおいては、隣接するある状態のユーザーの「数」によって、状態遷移確率が決定すると仮定している。一方、これとは異なる仮定をおく影響伝播モデルも存在する。例えば、隣接するある状態のユーザーの「割合」が、より強く意思決定に影響を与えている可能性がある。「隣接する状態 I のユーザー数」と「次の時刻までに状態 I に遷移する確率」との関係プロットしたときと同様に、「隣接する状態 I のユーザーの割合」と「次の時刻までに状態 I に遷移する確率」との関係プロットしたものが図 7.1 である。いずれの条件においても、概ね明確な右上がり傾向がみられることがわかる。つまり、Friendster のネットワーク構造の時間変化は、割合型の影響伝播モデルによっても説明される可能性があり、どちらがより適切なものか検討する必要がある。

ほとんどの頂点が状態 R に遷移する ネットワーク irSIR モデルを用いた説明の問題点は、(流行が発生した場合¹には)最終的にはネットワーク中の大部分の頂点が状態 R となることである。つまり、ネットワーク中のユーザーの大部分は、一度はその SNS に会員登録し、アクティブに活動していた時期が存在するということになる。しかし、実際にはいかに規模の大きな SNS であっても、社会のすべての人々が会員登録済みであるという状況は考えにくい。この点を考慮に入れたモデルを構築し、比較する必要がある。

非アクティブユーザーの詳細な分析 ユーザーが SNS を利用する目的は、「その SNS に興味を持って会員登録した」時刻から「その SNS に飽きて離脱する」時刻までの間に、段階的に変化していくと言われている [11]。例えば、SNS に加入した直後には積極的に新たな友人を作ろうとするが、その後はしだいに新たな繋がりの形成に消極的になり、既存の人間関係の中だけで活動を行うようになっていく傾向がある。

すなわち、SNS のユーザーは会員登録からの時間が経過するとともに「活発に活動しているが、新たな辺の形成には消極的な」状態へと変化していきやすいと考えられる。これらのユーザーの中には、既にその SNS において多くの繋がりをもち、大きな影響力を示すものも存在することが予想される。

¹ ネットワーク構造を考慮しない Cannarella らの元のモデル (式 3.8-3.9) では、初期状態における dI/dt が負であれば状態 I の頂点はすぐに消滅して影響伝播が停止する。そうでないときには、集団の大部分に流行が伝播していく [9]。

しかし、「辺を形成したか否か」を重視する本論文のアクティブユーザーの定義では、そのようなユーザーは「非アクティブ」であると判定されてしまう。このようなユーザーを検出する手法と、それを組み込んだモデルを構築する必要がある。

ユーザーページ非公開ユーザーの影響 本論文で分析対象としたデータセットには、ユーザーページを非公開としているユーザーから出る辺の情報が含まれていなかった。そのため、それらのユーザーは辺を一切持たないユーザーとして扱った。しかし、ユーザーページを非公開にしているユーザーこそが、Friendster のユーザーネットワークにおいて重要な役割を果たしていたという可能性もある。Wilkinson[35]によれば、SNSにおいてユーザーページを非公開としているユーザーは、そうでないユーザーよりも活発に活動している傾向がある。その理由は、多くのSNSにおいて初期状態ではユーザーページは公開状態となっているため、ユーザーページを非公開としているユーザーは「少なくともユーザーページを非公開に設定しようと考え、行動を起こす程度にはアクティブである」ことが保証されるためである。Friendsterにおいても初期状態ではユーザーページは公開状態であったことから、非公開ユーザーは公開ユーザーよりもアクティブであった可能性がある。したがって、ユーザーネットワーク構造の時間変化をより正確に理解するためには、それらのユーザーの動向を把握することが必要である。

ネットワーク成長過程の分析 本論文では、ユーザーネットワークの成長過程については立ち入らなかった。ネットワークの成長過程においてどのような傾向がみられるのか、また、それはネットワーク irSIR モデルや他の既存のモデルで説明できるのか、といった点は今後の検討課題である。

より詳細な影響伝播モデル 本論文で定義したモデルでは頂点がとりうる状態は S, I, R の3つのみであると仮定したが、これ以外の状態を加えることで SNS におけるユーザーの挙動をより詳細にシミュレーションできる可能性がある。一例としては、Lurker の存在をモデルに反映させることが考えられる。Lurker とは英語圏におけるネットスラングであり、「掲示板等における議論の成り行きを見守ってはいるが、自分から積極的に意見を発信することは無い人々」のことを指す。Schneider[28]によれば、オンラインコミュニティの参加者の大部分は Lurker である。したがって、状態 I を細分化し、Lurker と活発に発言しているユーザーとの違いをモデルに反映させることが考えられる。

ネットワーク irSIR モデルを用いた他の SNS の分析 ネットワーク irSIR モデルは、Friendster 以外の SNS の成長や衰退の過程を説明する上でも有用なのかを検証する必要がある。他の SNS において有用であることが確認された場合には、本論文の第 6 章で行った考察を手がかりとして、その SNS の衰退の防止に貢献することが目指される。

これらの課題に取り組み、解決することによって、Friendster や他の SNS の成長と衰退の過程への理解をさらに深め、SNS の予期せぬ衰退を防ぎ、活発で価値の高い状態を長く維持するための手掛かりを得ていきたい。

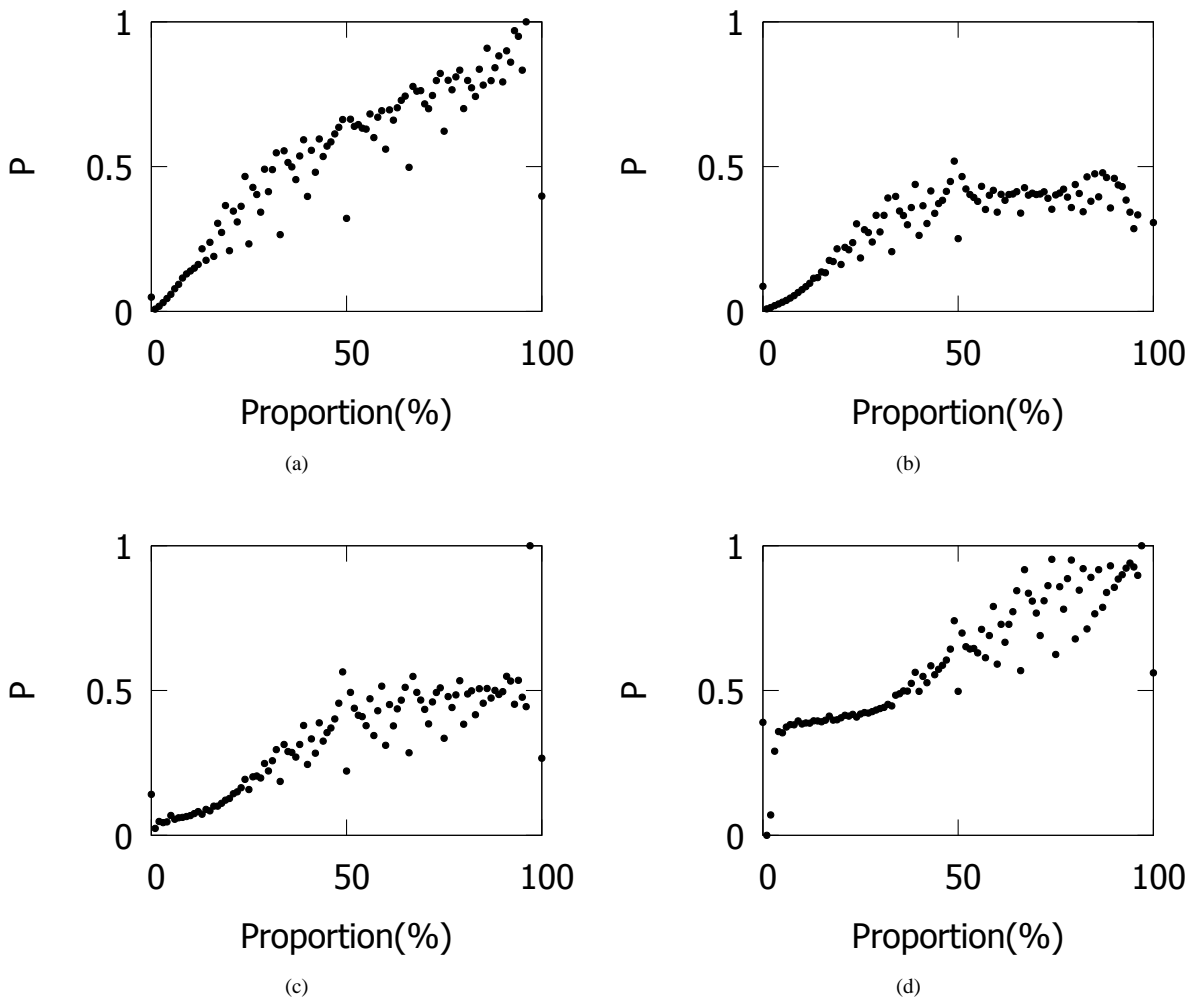


図 7.1: 「時刻 A において, 状態 I であった友人の割合」と「時刻 $B(> A)$ までに状態 I に遷移した確率」との関係. (a) $A : 2 \times 10^7, B : 3 \times 10^7$, (b) $A : 5 \times 10^7, B : 6 \times 10^7$, (c) $A : 7 \times 10^7, B : 8 \times 10^7$, (d) $A : 10 \times 10^7, B : 11 \times 10^7$.

付録A 辺活性化モデルで生成したネットワーク の次数分布の解析的導出

辺活性化モデルで生成したネットワークの次数分布を、解析的に導出する。ここでは導出のため、平均場近似 [5] を用いる。すなわち、実際のネットワーク生成過程においては辺の生成が確率的に行われるが、これを単純化し、全ての過程が「期待値どおりに」進行したと仮定して、次数分布を導出する。

時刻 t_i にネットワークに加わった頂点の、時刻 t における次数 k_i の時間変化は、次のように近似できる。

$$\frac{\partial k_i}{\partial t} = m_1 \frac{k_i}{\sum_j k_j} + m_2 \frac{k_i}{\sum_j k_j} + m_2 \frac{\sum_{j \in \Gamma(i)} k_j \frac{1}{k_j}}{\sum_j k_j} \quad (\text{A.1})$$

ただし、 $\Gamma(i)$ は頂点 i に隣接している頂点の集合であり、右辺の各項の分母ではその時点でネットワーク中に存在する全頂点についての総和をとっている。

ここで、右辺の第一項は、規則 1 による次数の増加量の期待値である。第二項はアルゴリズム中の「頂点 A」として選ばれることによる次数の増加量の期待値である。第三項は「頂点 C」として選ばれることによる次数の増加量の期待値であり、この値は「頂点 i と隣接している頂点のうちの 1 つが『頂点 B』として選ばれる確率」と「頂点 B と隣接している頂点の中から頂点を 1 つ選ぶ際に、頂点 i が選ばれる確率」とを掛け合わせたものとなっている。

$\sum_j k_j \approx 2(m_1 + m_2)t$ と近似して式変形すると次のようになる。

$$\frac{\partial k_i}{\partial t} = \frac{m_1 + 2m_2}{2(m_1 + m_2)} \frac{k_i}{t} \quad (\text{A.2})$$

頂点 i はネットワークに加わると同時に m_1 本の辺を持つことから、初期条件を $k_i(t_i) = m_1$ としてこれを解くと、

$$k_i(t) = m_1 \left(\frac{t}{t_i} \right)^A \quad (\text{A.3})$$

となる。ただし $A = (m_1 + 2m_2)/(2(m_1 + m_2))$ である。

式 A.3 を変形することにより、時刻 t に「次数が k 未満である」頂点の割合 $P(k_i < k)$ は、時刻 $(m_1/k)^{1/A}t$ 以降にネットワークに加わった頂点の割合に等しいことがわかる。つまり、

$$P(k_i < k) = P\left(t_i > \left(\frac{m_1}{k}\right)^{\frac{1}{A}} t\right) = \frac{1}{t} \left(t - \left(\frac{m_1}{k}\right)^{\frac{1}{A}} t\right) \quad (\text{A.4})$$

これを k で微分することにより、時刻 t に「次数が k である」頂点の割合、つまり次数分布 $p(k)$ が得られる。

$$p(k) = \frac{\partial P(k_i < k)}{\partial k} = \frac{m_1^{\frac{1}{A}}}{tA} k^{-(\frac{1}{A}+1)} \quad (\text{A.5})$$

となる。よって、辺活性化型モデルで生成したネットワークの次数分布は、 $\gamma = \frac{1}{A} + 1 = \frac{3m_1 + 4m_2}{m_1 + 2m_2}$ のべき分布となる。

Barabási-Albert モデルと等価なモデルとなる $m_2 = 0$ のとき、 $\gamma = 3$ となり、Barabási-Albert モデルにおける結果と一致する。また、 $m_1 \ll m_2$ のときは $\gamma \approx 2$ となる。 m_1, m_2 の値によって、 γ は 2 から 3 の間を動く。

謝辞

本研究を進め、論文を書き上げる過程においては、多くの方々にお世話になりました。

山口和紀教授には急遽主査をお願いすることとなりましたが、ご多忙の中きめ細かなご指導を賜りました。森畑明昌准教授、田中哲朗准教授、山口泰教授、植田一博教授には、審査の過程で多くの的確な指摘と前向きなアドバイスをいただきました。学部の卒業研究以来長きにわたってご指導いただいた中村政隆准教授には、研究生活の基礎を教えてくださいました。山口和紀研究室、中村政隆研究室の皆様には多くの有益なアドバイスをいただきました。家族からの物心両面のサポートがなければ論文を書き上げることは出来ませんでした。

以上の皆様に厚く御礼申し上げます。

参考文献

- [1] Takuya Akiba, Yoichi Iwata, and Yuichi Yoshida. Linear-time enumeration of maximal k -edge-connected subgraphs in large networks by random contraction. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, pages 909–918. ACM, 2013.
- [2] Michael Arrington. Friendster valued at just \$26.4 million in sale, 2009. <https://techcrunch.com/2009/12/15/friendster-valued-at-just-26-4-million-in-sale/>.
- [3] Michael Arrington. Modeling the true value of social networks: 2009 edition, 2009. <https://techcrunch.com/2009/12/15/friendster-valued-at-just-26-4-million-in-sale/>.
- [4] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [5] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1):173–187, 1999.
- [6] Christian Bauckhage and Kristian Kersting. Strong regularities in growth and decline of popularity of social media services. *arXiv preprint arXiv:1406.6529*, 2014.
- [7] Danah M Boyd. Friendster and publicly articulated social networking. In *Extended abstracts of the 2004 conference on Human factors and computing systems*, pages 1279–1282, 2004.
- [8] DM Boyd and NB Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13:210–230, 2008.
- [9] John Cannarella and Joshua A Spechler. Epidemiological modeling of online social network dynamics. *arXiv preprint arXiv:1401.4208*, 2014.
- [10] Max Charfkin. How to kill a great idea!, 2007. <https://www.inc.com/magazine/20070601/features-how-to-kill-a-great-idea.html>.
- [11] Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd international conference on World Wide Web*, pages 307–318. ACM, 2013.
- [12] Friendster. Friendster is the 1 social network for adults and youth in Malaysia, 2008. <https://web.archive.org/web/20081219023712/http://www.friendster.com/info/presscenter.php?A=pr48> (Internet Archive).

- [13] David Garcia, Pavlin Mavrodiev, and Frank Schweitzer. Social resilience in online communities: The autopsy of Friendster. In *Proceedings of the First ACM Conference on Online Social Networks*, pages 39–50. ACM, 2013.
- [14] Mark S Granovetter. The strength of weak ties. *American Journal of Sociology*, pages 1360–1380, 1973.
- [15] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.
- [16] InternetArchive. Friendster social network dataset: Friends, 2011. <https://archive.org/details/friendster-dataset-201107>.
- [17] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 115, pages 700–721. The Royal Society, 1927.
- [18] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. Structure and evolution of online social networks. In *Link Mining: Models, Algorithms, and Applications*, pages 337–357. Springer, 2010.
- [19] Fion SL Lee, Douglas Vogel, and Moez Limayem. Virtual community informatics: A review and research agenda. *JITTA: Journal of Information Technology Theory and Application*, 5(1):47, 2003.
- [20] Han Liu, Atif Nazir, Jinoo Joung, and Chen-Nee Chuah. Modeling/predicting the evolution trend of OSN-based applications. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 771–780. ACM, 2013.
- [21] Atif Nazir, Alex Waagen, Vikram S Vijayaraghavan, Chen-Nee Chuah, Raissa M D’Souza, and Balachander Krishnamurthy. Beyond friendship: modeling user activity graphs on social network-based gifting applications. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 467–480. ACM, 2012.
- [22] David Pogue and JD Biersdorfer. *The Internet: the missing manual*. O’Reilly Media, Inc., 2006.
- [23] Jiezhong Qiu, Yixuan Li, Jie Tang, Zheng Lu, Hao Ye, Bo Chen, Qiang Yang, and John E Hopcroft. The life-cycle and cascade of WeChat social messaging groups. In *Proceedings of the 25th International Conference on World Wide Web*, pages 311–320. International World Wide Web Conferences Steering Committee, 2016.
- [24] Bruno Ribeiro. Modeling and predicting the growth and death of membership-based websites. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 653–664. ACM, 2014.
- [25] Bruno Ribeiro and Christos Faloutsos. Modeling website popularity competition in the attention-activity marketplace. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 389–398. ACM, 2015.
- [26] Gary Rivlin. Wallflower at the web party, 2011. <http://www.nytimes.com/2006/10/15/business/yourmoney/15friend.html>.

- [27] E.M. Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, 2003.
- [28] Andreas Schneider, Georg Von Krogh, and Peter Jäger. “What ’ s coming next? ” epistemic curiosity and lurking behavior in online communities. *Computers in Human Behavior*, 29(1):293–303, 2013.
- [29] Stephen B Seidman. Network structure and minimum degree. *Social networks*, 5(3):269–287, 1983.
- [30] Moena Tanaka, Yusuke Sakumoto, Masaki Aida, and Konosuke Kawashima. Study on the growth and decline of SNSs by using the infectious recovery SIR model. In *Proceedings of the 10th Asia-Pacific Symposium on Information and Telecommunication Technologies*, pages 1–3. IEEE, 2015.
- [31] Mojtaba Torkjazi, Reza Rejaie, and Walter Willinger. Hot today, gone tomorrow: On the migration of myspace users. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, pages 43–48. ACM, 2009.
- [32] Jessica E Vascellaro. New friendster CEO has asia focus, 2008. <https://www.wsj.com/articles/SB121790017932212287>.
- [33] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, pages 37–42. ACM, 2009.
- [34] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘ small-world ’ networks. *Nature*, 393(6684):440–442, 1998.
- [35] David Wilkinson and Mike Thelwall. Social network site changes over time: The case of MySpace. *Journal of the American Society for Information Science and Technology*, 61(11):2311–2323, 2010.
- [36] Christo Wilson, Bryce Boe, Alessandra Sala, Krishna PN Puttaswamy, and Ben Y Zhao. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European Conference on Computer Systems*, pages 205–218. Acm, 2009.
- [37] Yi Yu, Gaoxi Xiao, Jie Zhou, Yubo Wang, Zhen Wang, Jürgen Kurths, and Hans Joachim Schellnhuber. System crash as dynamics of complex networks. *Proceedings of the National Academy of Sciences*, page 201612094, 2016.
- [38] Xiang Zhu, Yuanping Nie, and Aiping Li. Demographic prediction of online social network based on epidemic model. In *Asia-Pacific Web Conference*, pages 93–103. Springer, 2014.
- [39] 増田 直紀 and 今野 紀雄. 複雑ネットワーク : 基礎から応用まで. 近代科学社, 2010.