

インターネットのコミュニケーションモデルに関する一考察 (未成年保護のためのインターネットにおけるリスク軽減のモデル提唱)

田代光輝^{†1†5} 小松正^{†1} 江口清貴^{†2} 鳥海不二夫^{†3} 中西崇文^{†4}

本稿では、実際のソーシャルネットワークワーキングサービス(SNS)のデータを用いて、SNSの様相がどのようなになっているかを検証する。これまでSNSの様相の解釈にはスケールフリーモデルやCNNモデルなどが利用されてきた。今回、LINE株式会社のアバターサービスであるラインプレイ内のスクウェアのウィスパーサービス(LPW)のやりとりをSNSの実データとして用いて、SNS内のユーザのやりとりであるコミュニケーションの形を分析した。LPWはCNNモデルにあるような友達の友達を紹介したりつなげたりする機能がなく、スクウェア上で偶然出会うことでやり取りが発生するサービスである。インターネットコミュニティ上での出会いについての基礎的なアルゴリズムが確認できる。分析の結果、スケールフリーモデルではなく、ポアソン対数正規分布に近い形の分布になることが分かった。これはインターネットコミュニティ上で誰かと出会う確率がユーザの持つ「つながりやすさ」に比例し、さらにコミュニティ内の既存ユーザ同士も「つながりやすさ」に比例して出会うことを示している。今後「つながりやすさ」の研究を深めることで、犯罪の予兆やネットいじめの予兆などを検知し、インターネットコミュニケーションの安全性を高めることに寄与する。

A Study of the internet communication model (Research of risk reduction on the internet for minors' protection)

Mitsuteru TASHIRO^{†1†5} Tadashi KOMATSU^{†1} Kiyotaka EGUCHI^{†2}
Fujio TORIUMI^{†3} Takafumi NAKANISHI^{†4}

We analysis the data of the social networking service (SNS) .for clarify the aspects. We used the “whisper services” within the “line play” that is an avatar service LINE Ltd. We found the distribution in the data that sharpened close to a Poisson lognormal distribution. The potential of ease of connection is TSUNAGARIYASUSA. Nodes have the potential of ease of connection each other. Nodes will link each other proportionate the potential, and the links distribute close to a Poisson lognormal distribution. We are expecting that this research will lead to the safety of minors.

1. はじめに

本研究は、未成年がインターネットコミュニケーションサービスを利用して犯罪等に巻き込まれないよう、人・組織・システムの3つの要素からインターネットコミュニケーションの安全性を高めることを目的としている研究の一環として、誘い出しやいじめのシステムの感知を目的とした研究の予備調査である。

本研究はシステムの感知を目的とした研究の前段階として、ソーシャルネットワークワーキングサービス(SNS)上で行われるコミュニケーションの様相を推し量るために行った。コミュニケーションにはある一定の形が存在すると仮定している。その形が分かれば異常値を発見することで誘い出しや詐欺などの予兆を発見できるのではないかと考えられ

る。

ネットワークのモデルにはいくつかのモデルが提唱されている。例えば、スケールフリーネットワークモデルは、新しいノードが追加された時、リンク数の多いノードが優先的にリンクを獲得するというものであった。

我々は、LINE株式会社のアバターサービスであるラインプレイ内のスクウェアのウィスパーサービス(LPW)のやりとりをSNSの実データとして用いて、SNS内のユーザのやりとりであるコミュニケーションの形を分析したところ、ある着目しているノードの持つリンク数が、そのノードが今後リンクを持てる確率と関係していることを発見した。本稿では、ノードの持っているリンク数を、今後リンクを持てる確率と関係していることから、「つながりやすさ」という指標として提案する。また、ネットワークはこの「つながりやすさ」に比例してつながるために、ネットワークの形はべき乗ではなくポアソン対数正規分布に近い形の分布になることがわかった。これを応用することで今後、犯罪の予兆やネットいじめの予兆などを検知し、インターネットコミュニケーションの安全性を高めることに寄与したい。

†1 多摩大学情報社会学研究所
The New Institute for Social Knowledge and Collaboration, Kumon Center,
Tama University

†2 東京大学大学院
Graduate School of Education, The University of Tokyo

†3 東京大学
School of Engineering, The University of Tokyo

†4 国際大学グローバルコミュニケーション研究所
Center for Global Communications, International University of Japan

†5 青山学院大学大学院 社会情報学研究科
Graduate School of Social Informatics, Aoyama Gakuin University

2. 先行研究等について

2.1 システム面からみたインターネットコミュニケーションの安全に関する研究

安全とは「“受容できないリスクが存在しないこと (Freedom from unacceptable risk)” [1]であり、ゼロリスクということはありません。そのためにリスクの見積もりを行い、受容可能なまでに抑える必要がある。リスクは人・組織・システムの3要素によって軽減できる。

本研究はシステムによるリスク軽減を図るための研究の予備研究である。システムによるリスク軽減につながるものとして中西らによる「A Feature Selection Method for Comparison of Each Concept in Big Data」 [2] 「A Discovery Method of Anteroposterior Correlation for Big Data Era,」 [3]や「Semantic Context-dependent Weighting for Vector Space Model,」[4], 岡田らによる「Method of Knowledge Creation and Knowledge Utilization by Generalized Inverse Operator」 [5], 鳥海らによる「SNSにおけるネットワーク成長モデルの提案」 [6]などの研究がある。

これら研究を応用し、実装可能なためのリスク感知のシステムを構築することでリスクを抑え、安全性を高めることを目的としている。

2.2 ネットワークとべき法則

バラバシ・アルバートによれば人のつながり、航空路のネットワーク、コンピュータネットワークなどはべき乗のスケールフリーネットワークになるとされている。スケールフリーとは、例えばコンピュータネットワークであればリンク数 k をもつ確率 $P(k)$ の確率分布が k の r 乗に比例するというものである。スケールフリーネットワークは成長と優先的選択によって誕生する [7]。成長とは与えられた期間毎に加えられる新しいノードである。優先的選択とは新しいノードがリンクを獲得する場合、すでにあるノードが持つリンクの多さに比例してリンクをするというものである。これを繰り返すことによりネットワーク内のノードがもつリンク数はべき乗になるとされている。べき指数 r は 2~3 になるとしている。

バスケスの CNN モデルでは「友達の友達は友達である」 [8] というアルゴリズムによってネットワークの成長を説明した。友達の友達 (NEAREST-NEIGHBORS) にはつながりやすいとして、べき乗になるつながりをモデリングした。

3. SNS コミュニケーションの分析

3.1 本研究で利用した SNS データについて

本研究において、LINE 株式会社のアバターサービスであるラインプレイ内のウィスパサービス (以下: LPW) のやり取りを分析した。LPW はラインプレイ上のスクウェアというコミュニティサービスにおけるチャットサービス

である。国籍を問わず集まることのできる NY タイムズスクウェアや、LINE カフェ、海賊島などのスクウェアがあり、ユーザは任意で参加できる。アバター同士では LINE 社が監視することを承諾したうえで、相手の承諾があれば LPW をすることが出来る。

LPW のやりかたは、スクウェア上の任意アバターと話したいと思ったら、その相手のアバター画像をタップする。タップすると選択肢 (図 1) が現れるので「ささやく」を選択する。そうすると規約への同意のダイアログが出るので同意すると (図 2) 相手に承諾するかどうかの確認ダイアログが開く。相手が承諾すれば LPW が開始される。会話は 1 対 1 のペアで行われるため、送信者 ID と受信者 ID が記録として残されている。さらに画面上のダイアログで友達申請もできるようになっている。そのため、LPW は深い会話をするというより友達になるきっかけとして利用される傾向がある。

またラインプレイは機能的に友達の数を表示する機能はあるものの、友達の友達を示唆するようなサービスはない。またスクウェアではスクウェアにたまたま一緒にいるアバター同士での会話になるために、CNN モデルが適用出来ない。スクウェアは偶然の出会いを誘発するサービスであり、LPW からは偶然の出会いからの会話を通じて友達関係になる過程を観測することが可能である。



図 1 スクウェアの様式: アバターをタップして出てくる選択画面 (一部加工)

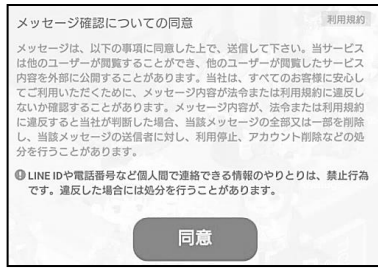


図 2 「ささやく」をタップした後に出てくる同意画面

3.2 分析対象について

本研究では LINE 社からの未成年保護を目的とした委託研究として提供された LPW のデータを分析した。提供されたデータの内、本研究においては 2015 年 2 月 1 日から 2 月 28 日までの LPW の送信者と受信者の ID、送信日時 of データを利用した。送信者と受信者の ID は匿名化処理されたもので、同一性は保たれつつも元がどの ID かは判らないように 18 ケタの数字に加工されて提供されたものである。本研究ではどの ID がどの ID に何日何時何分何秒にメッセージが送られたかという関係性のみを分析対象とした。

3.3 ID と発言数に関する分析

LPW のデータを 2015 年 2 月 1 日 0 時 0 分 0 秒～28 日 23 時 59 分 59 秒まで分析した。発言数や ID 数や送受信ペアの数は

ID 総数 : 42,713
 発言数 : 45,788,789
 送信 ID 数 : 41,839
 受信 ID 数 : 41,899
 送受信ペア : 99,620

となっている。

まずは ID ごとの発言数(以下: ID 発言数)を分析する。図 3 は ID 発言数の割合をグラフにしたものである。X 軸が発言数、Y 軸がその発言数をした ID の数である。ID ごとの発言数の最頻値は 1 回で 2,386ID である。次が 2 回で 2,047ID、その次が 3 回で 1,606ID となっている。発言の回数が増えるごとに ID の数は減っていく。最小値は 0 回で 872ID がまったく発言していない。これは話しかけられても無視をしているケースになる。いわゆる「ナンパ」されても無視している割合考えると割合は少なめである。これは LPW が話しかける際に相手の承諾が必要であるために、無視したい場合は承諾しなければよい。そのため発言が 0 である割合が少なめになっていると考えられる。最大値は 9,168 回である。この ID は 1 か月の間に 104ID とチャットをしている。2 番目は 6763 回で 108ID と、3 番目が 6,750 回で 244ID とチャットをしている。

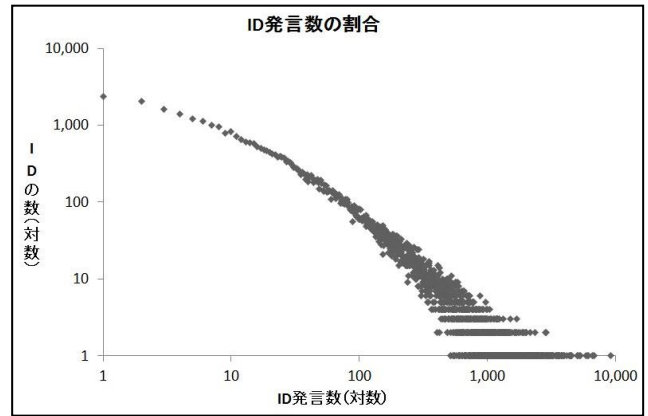


図 3 ID 発言数の割合 (対数)

3.4 ペア発言数の分析

次に 1 ペアあたりの発言数の総数 (以下: ペア発言数) を分析する。図 4 はペア発言数の割合をグラフにしたものである。X 軸がペア発言数、Y 軸がその発言数をしたペアの数である。ペア発言数の最頻値は 9 回で 2,775 ペアである。次に多いのが 7 回で 2,708 ペア、その次が 8 回で 2,625 ペアとなっている。9 回をピークに 5 回～10 回までがペア数の 2,500 を超えている。最小値は 1 回で 2,142 ペア。0 回の場合は記録が残らないため、最小値は 1 回ということになる。ID 発言数は 1 回が最頻値だったが、ペアで集計すると 9 回となる。これは PLW の開始が相手の承諾が必要であるために、ある程度チャットを楽しみたいという人同士の会話となると考えられる。最大値は 5,673 回で、3,012 回の返信に対して 2,661 回の返信となっている。返信の ID は 1 か月あたり 3,424 回の発信をしていて、1 か月の 88.0% を返信相手とのチャットに費やしている。次が 3,539 回、返信が 1,960 回で返信が 1,589 回である。その次が 3,146 回で 2,003 回の返信に対して 1,143 回の返信がある。

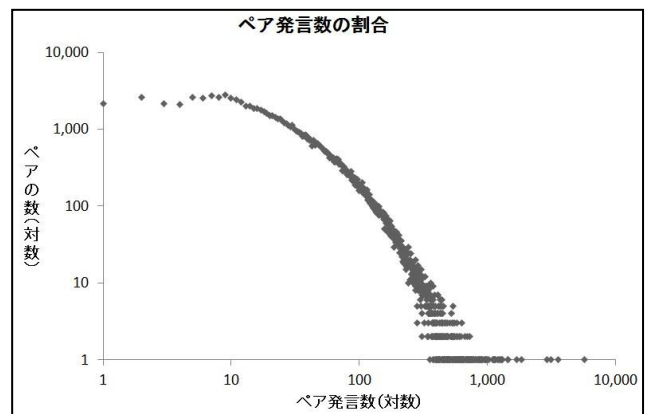


図 4 ペア発言数の割合

3.5 ペア数

最後に ID ごとにペアとなった ID 数 (以下: ペア数) がどれだけいるかを分析した。図 5 はペア数をグラフにした

ものである。X軸がペア数、Y軸がIDの数である。ペア数の最頻値は1ペアで19,502IDである。その次が2ペアの6,840ID、その次が3ペアの3,614IDである。最小値も1ペアで、1ペアから数字が増えるごとに徐々にIDの数も減っていく。最大値は585IDである。

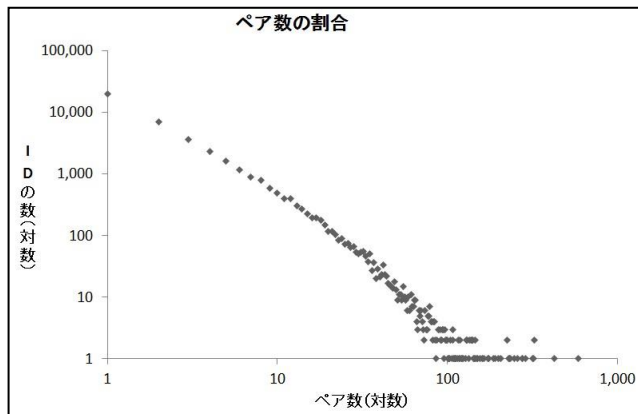


図5 IDごとのペア数の割合

4. コミュニケーションの形の考察

4.1 コミュニケーションの形の考察

前章において分析した結果からコミュニケーションの形について考察する。

ID発言数、ペア発言数やペア数は先行研究等で「べき乗」になるとされていた分布である。これらをまとめて図にしたのが図6である。特にペア数に関しては対数で直線状になっており、いままでべき乗であるとされていたグラフに形が近い。しかしいずれも完璧なべき乗ではなく緩やかにカーブを描いている。最も直線的に見えるペア数についても補助線として直線を引いてみると緩やかなカーブを描いていることが見てとれる。

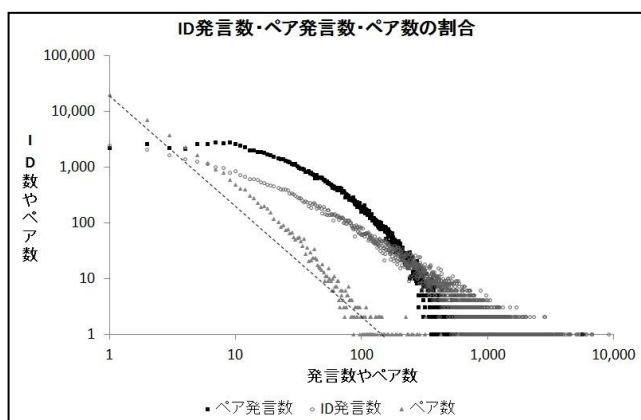


図6 IDとペアごとの発言数とペア数の対数グラフ

このような緩やかなカーブを描く分布は対数正規分布に見られる形である。特に値が離散的で平均値が1を下回るよ

うなときに見られる形である。ID発言数、ペア発言数、ペア数をそれぞれ対数にして正規検定を行ったが、いずれもP値は1%以下で正規分布であるとは言えなかった(表1)。特にペア数に関しては1の数が多く、全体の平均や分散などに影響をあたえと考えられる。しかし分位数をplotしてみると(図7~9)直線に近い形になっていることが確認できる。さらに理論値と実数を比較すると(図10)数値の低いところでの乖離が大きい。これらはポアソン対数正規分布の特徴である。ポアソン対数正規分布とは、分布の平均が極めて小さく離散的に発生するポアソン分布が対数的に形成されたものである。これより、ID発言数、ペア発言数、ペア数のそれぞれの対数の分布は、ポアソン対数正規分布に近い分布だったのではないかと推測できる。

	ID発言数・対数	ペア発言数・対数	ペア数・対数
平均値	1.08819723649	1.37133774085	.36015719370
平均値の下限	1.08594282563	1.36804466264	.35597775721
平均値の上限	1.09045164735	1.37463081906	.36433663020
5%トリム平	1.09304091066	1.38248021361	.31697826409
中央値	1.11394335200	1.39794000900	.30102999600
分散	.259	.281	.190
標準偏差	.509002393685	.530297895892	.436135971230
最小値	0.000000000	0.000000000	0.000000000
最大値	3.478854968	3.753812784	2.767155866
範囲	3.478854968	3.753812784	2.767155866
4分位範囲	.748188027	.706795342	.602059991
歪度	-1.07	-2.70	1.189
	.006	.008	.012
	-.372	-.101	.869
尖度	.011	.016	.024
	.035	.026	.262
Kolmogorov-Smirnovの正規性の検定(探索的)	統計	.035	.026
	自由度	195828	99619
有意確率	0.000	.000	0.000

表1 ID発言数・ペア発言数・ペア数の対数のKolmogorov-Smirnovの正規性の検定の結果

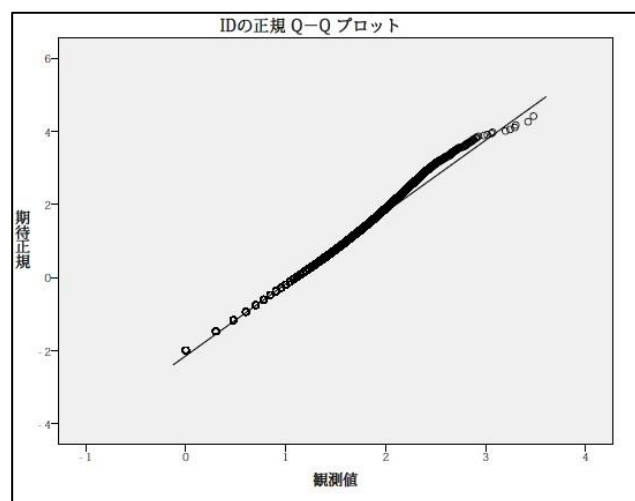


図7 ID発言数の対数の分位数plot

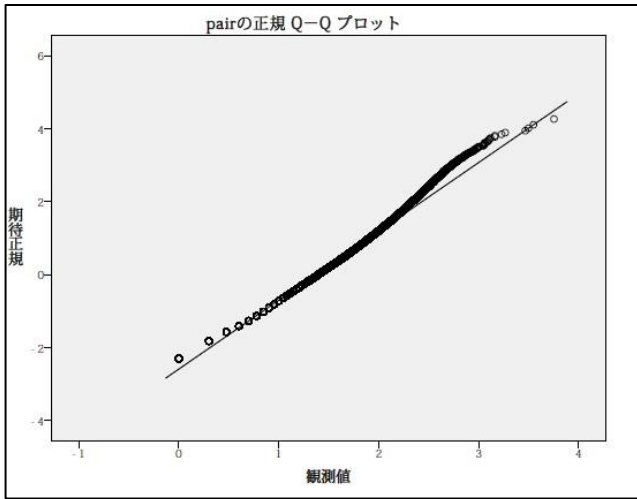


図 8 ペア発言数の対数の分位数 plot

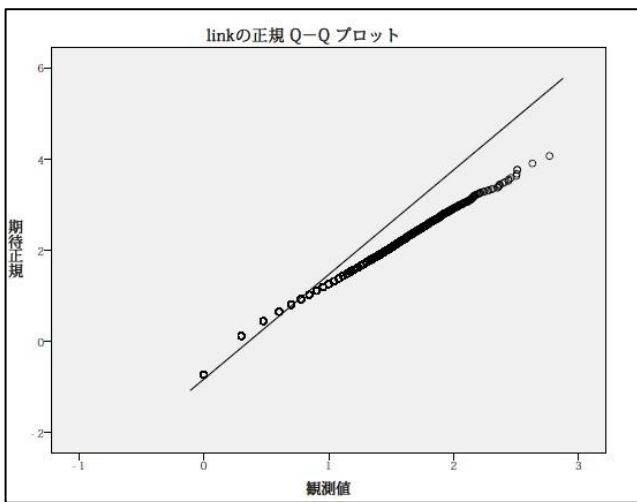


図 9 ペア数の対数の分位数 plot

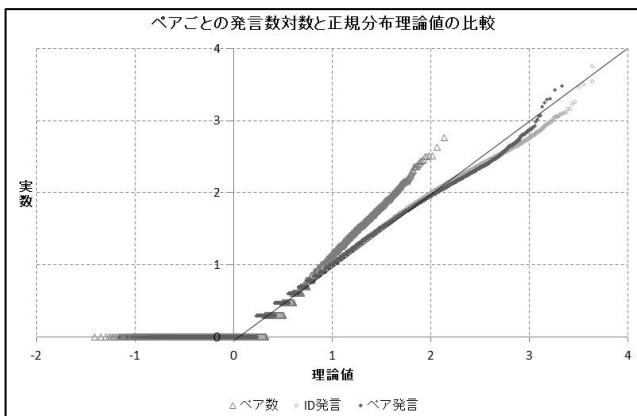


図 10 理論値と実数の比較

4.2 ポアソン対数正規分布について

なぜ LPW の分析結果の数値はべき乗のような対数にして直線ではなく、ポアソン対数正規分布のような対数にしてゆるやかなカーブに近い形になるのであろうか。

例えば 1 人とペアになれる確率を a とすると k 人と友達

になれる確率はポアソン分布になるというのがこれまでの考え方である。しかし a はすべての人に均一ではない。1 人しかペアになれない人が 2 人に増やせる確率と、10 人とペアになっている人が 11 人とペアになる確率と、100 人とペアになれた人が 101 人とペアになる確率は違うはずである。仮にプレイ時間あたりでペアが増える確率が一定だとすると、ペアの数はその人のプレイ時間に比例する。例えば 1 日 5 分しかプレイしない人が 1 人と 1 日 50 分プレイする人が 10 人のペアを獲得しているのを比較すれば、その後前者が 1 人から 2 人へ、後者が 10 人から 20 人に増える機会は同等である。ペアになる確率がある一定の平均があると考え、機会の多さがその後のペアとなる割合の多さにつながる。そうするとそれまでペアになった数が今後ペアを獲得できる割合に反映されると考えられる。つまり変化の率がペアの数に比例しているため、ポアソン分布が対数正規的に分布したときと同じ分布となることが予測される。

実数と理論値の分布に関して決定係数を割り出してみた(表 2)(図 11)。ポアソン対数はいずれの場合も 0.983~0.998 と高い数値を出している。べき乗グラフは観測値のペア数では 0.912 と高い決定係数となったが、ID 発言では 0.769、最頻値が 1 以外になるペア発言では 0.398 となった。

決定係数	ID 発言数	ペア発言数	ペア数
べき乗	0.796	0.398	0.912
ポアソン対数	0.998	0.983	0.992

表 2 実数と理論値を比較した決定係数

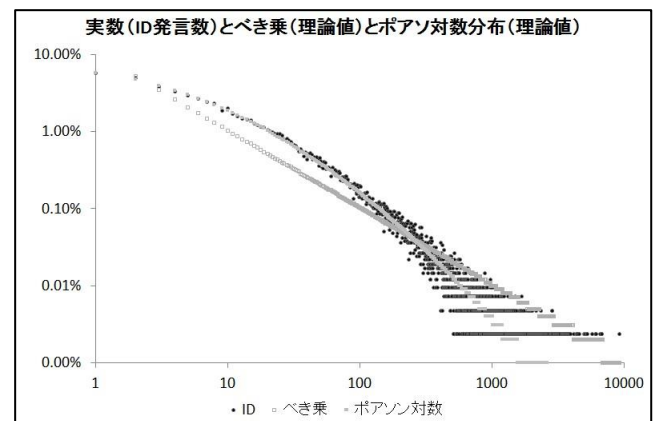


図 11 観測値とべき乗とポアソン対数の比較

5. コミュニケーションの形に関する考察

5.1 ネットワークの成長モデルの提案

スケールフリーモデルと本研究におけるモデルの違いはなんだろうか。それはスケールフリーモデルが「新しいノード一定時間で一定数追加され、リンク数の多いノード

ドが優先的にリンクされる」というものであるのに対して本研究の提案するモデルは「リンクができる確率は極めて小さい(ポアソン分布的)」としたうえで「ノードの持つリンク数が、そのノードが今後リンクのもてる確率を反映している」ということである。小さい確率ながらもノードによってリンクができる確率がちがう。現在ノードが持っているリンク数は過去の「つながりやすさ」を反映したものであり、今後の「つながりやすさ」を予測する値として利用できる。

スケールフリーモデルでは常に新しいノードが追加されることを前提としているが、本研究の提案するモデルは新しいノードの追加を前提としていない。前者は一定時間に新しいノードが一定数追加されるモデルとしてべき乗のネットワークモデルを提唱している。本研究ではポアソン分布をベースとして考えている。そもそもリンクができる確率が非常にまれであり、起こるときは1か0の離散数として発生する。つながる割合はノードの持つ「つながりやすさ」に比例する。

5.2 度数相関の確認とつながりやすさの考察

リンクが出来る確率が極めて小さく、かつ「つながりやすさ」に比例してリンクを持てるとすると、「つながりやすさ」が低いノード同士はつながらず、「つながりやすさ」の高いノードが低いノードとはつながりやすくなるはずである。

まずリンク数1のノード、つまり本研究においてはペア数1のIDがどんな傾向にあるかを分析した。ペア数1のIDは19,501IDある。そのうちペア数1とペアになっているものが2,965IDで最頻値となっている。ペア数2とペアになっているものが1,429ID、ペア数3とペアになっているものが1,089IDと相手のペア数に応じてID数も減っていく。これを全体のペア数の割合と比較したのが図12のグラフである。△がペア数1とペアになっているIDのペア数の割合。●が全体のIDのペア数の割合である。△と●には大きな差がある。これはペア数が多いということ自体がペアを獲得している割合が多いためである。この差を修正するために全体のIDのペア数の割合をそのペア数で割ったのが□である。逆にペア数1とペアになったIDのペア数の割合にそのペア数をかけたのが×である。

同様にペア数100以上のIDのペアのIDがどのようなペア数を持っているかを分析した。ペア100以上のIDは793IDあり、12,290IDとペアになっている。ペアの内訳はペア1のものが1,482IDで最頻値となっている。ペア数2が1,029ID、ペア数3のIDが764IDで、ペア数に応じてID数も減っていく。これを全体のペア数の割合と比較したのが図13のグラフである。ペア数の多いもの同士があまりつながっていないことがうかがえる。

ペア数によってペアが持っているペア数にどのような

傾向があるだろうか。ペア数をリンク数として度数相関をとると負の相関になっている(図14)。人間関係はペア数の多いもの同士やペア数の少ないもの同士、つまり似た同士が繋がることで正の相関になるとされているがLPWでは負の相関になった。本来リンクを持ちにくい「つながりやすさ」の小さいノードが、リンクを持ちやすい「つながりやすさ」の大きいノードによってリンクを持たされていると考えられる。LPWが片方からチャットの申請を送り、相手が応諾することで会話が発生することから、より積極的なノードが消極的なノードを誘いやすいからだと考えられる。

極めて小さい確率でできあがるリンクが「つながりやすさ」の低いノードでも、高いノードによってリンクが作成されていることがうかがえる。

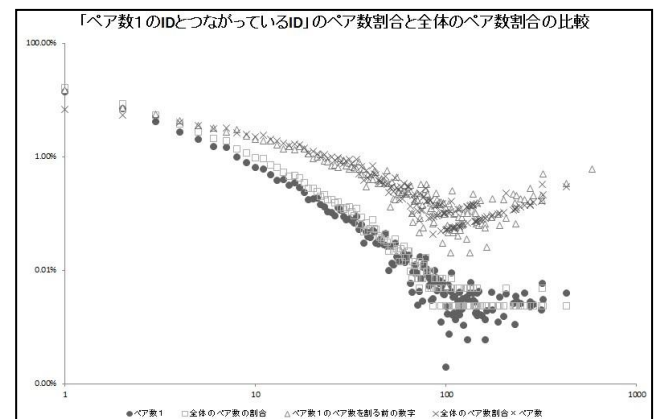


図12 「ペア数1とIDとつながっているID」のペア数割合と全体のペア数割合の比較

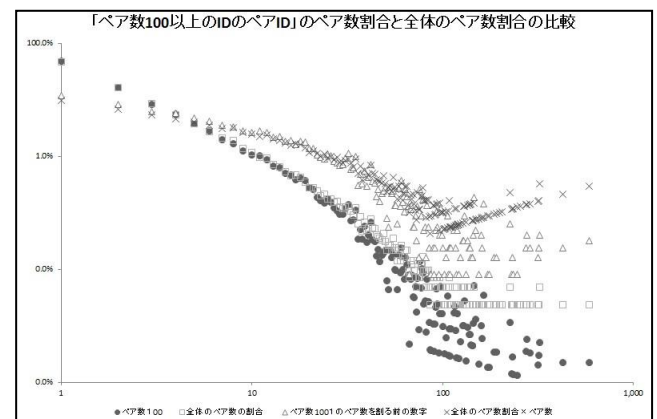


図13 「ペア数100とIDとつながっているID」のペア数割合と全体のペア数割合の比較

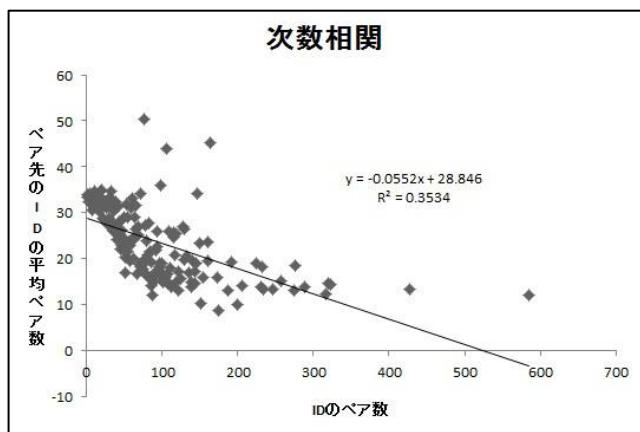


図 14 ペア数でみた次数相関

5.3 「つながりやすさ」の意味合いと今後課題

LPW のように偶然の出会いにおいて、リンク数や発言数がポアソン対数正規分布に近い分布になるとを前提とすると、つまりリンクがノードの持つ「つながりやすさ」に比例して増えるとすると、これを基本的な概念として異常値をみつけることで犯罪被害の予防につながられるかもしれない。

例えば「つながりやすさ」を無視してランダムにつながるノードを発見することで犯罪者予備軍を見つけられるかもしれない。もしくは「つながりやすさ」が大きい未成年のノードを発見することで、実際につながりが発生する前にリスクの高いユーザを発見し、未成年の非行や誘い出しなどの犯罪被害の予防につながられるかもしれない。

今回は 2015 年 2 月のデータを用いたが、LINE 社からは 2 月および 3 月のデータが提供されている。今後 2 月のデータと 3 月のデータを比較することで「つながりやすさ」の理論をさらに深める。異なる 2 つの期間を比較することでネットワークの成長だけではなく、ネットワークの衰退などのモデルも考察できる可能性がある。

今後はこの研究をさらに深め、未成年を犯罪に巻き込もうとする動きを系統的に閼知し、犯罪を未然に防ぐためのシステム実装を目指す。

6. まとめ

データ分析の結果から「つながりやすさ」を考慮することでネットワークの形はべき乗ではなくポアソン対数正規分布に近い形であることを示した。本研究ではポアソン分布を基にして、リンク（本研究ではペア）を多く持つノード（本研究では ID）は「つながりやすさ」の値が高く、新たなリンクを持つ確率が高いことを前提に、ポアソン対数正規分布に近い分布から、実際のデータから見られる「つ」の字型の緩やかなカーブを説明できた。

今回導き出した「つながりやすさ」を応用し、インターネットコミュニティ上で未成年を犯罪に巻き込もうとする

ノードの発見や、犯罪者と「つながりやすさ」によってつながってしまう確率の高い未成年のノードの発見につなげたい。

謝辞 本研究は LINE 株式会社からの委託研究として、インターネットコミュニケーションにおける犯罪からの未成年保護の研究の一環として行った。データ提供をいただいた LINE 株式会社の皆様に感謝申し上げます。またデータ分析において国際大学 Glocom の豊福様にサポートいただきました。感謝申し上げます。

参考文献

- 1) SO/IEC ガイド 51 (JIS Z 8051) : 安全側面一規格への導入指針
- 2) Takafumi Nakanishi, "A Feature Selection Method for Comparison of Each Concept in Big Data," In Proceedings of 4th IEEE/ACIS International Conference on Computer and Information Science 2015, IEEE, 2015.
- 3) Takafumi Nakanishi, "A Discovery Method of Anteroposterior Correlation for Big Data Era," Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Studies in Computational Intelligence Volume 569, pp 161-177, 2015.
- 4) Takafumi Nakanishi, "Semantic Context-dependent Weighting for Vector Space Model," In proceedings of 2014 IEEE International Conference on Semantic Computing (ICSC2014), pp.262-266, 2014[Acceptance rate 35%].
- 5) Ryotaro Okada, Takafumi Nakanishi, Takashi Kitagawa, "A Method of Knowledge Creation and Knowledge Utilization by Generalized Inverse Operator," In Proceedings of 2014 IIAI 3rd International Conference on Advance Applied Informatics, IEEE, pp.253-258, 2014.
- 6) 鳥海不二夫, 石田健, 石井健一郎, "SNS におけるネットワーク成長モデルの提案" 電子情報通信学会論文誌 D Vol.J93-D No.7 pp.1135-1143 (07/2010)
- 7) アルバート・ラズロ・バラバシ (著), 青木 薫 (翻訳) 新ネットワーク思考—世界のしくみを読み解く NHK 出版 (2002/12/26)
- 8) Vazques, A : Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlation, Phys.Rev Vol67 p.056104(2003)