

LJ-008

中心性の相関によるIRCユーザネットワークの分析 Analysis of IRC User Network by Correlation of Centrality

松澤 智史*

井出 和典†

山崎 誠‡

武田 正之*

Tomofumi Matsuzawa

Kazunori Ide

Makoto Yamazaki

Masayuki Takeda

1. まえがき

近年、P2P ネットワークを代表とする、ノードの配置やエッジの増減が頻繁に変動する動的なネットワークに注目が集まっている。動的なネットワーク上で新しいサービスの実現や有効なシステム構築をするためには、どのような構造を取るのかを明らかにすることが重要である。特に通信にサーバを介さない P2P ネットワークでは、ノードは情報の検索、受信、送信、中継を担うため、ノードの数が増えると通信量が加速度的に増える。したがってノード間の情報伝播の効率化が求められる。

一方で、マーケティングや強固な通信ネットワークの構築などに活用が期待されている複雑ネットワークの研究が盛んに行われている。複雑ネットワークとはグラフ理論の一分野で、現実世界に存在する複雑で巨大なネットワークの性質について研究する学問である。複雑ネットワークには次の性質があることがわかっている。

スモールワールド ノードがクラスタ状であるにもかかわらずノード間のパス長が短い特徴を示すグラフ構造の性質。 [1][2]

スケールフリー 一部のノードが極端に多くのリンクを持ち、一方で大多数のノードがわずかなリンクしか持っていないグラフ構造の性質。 [3][4]

これまでの研究でスモールワールド、またはスケールフリーの性質を持つネットワークが多数示されている。これらの性質は感染症の予防やマーケティング、強固な通信ネットワークの構築といった分野において活用が期待されている。

友部ら [5] は、後述する中心性を用いて、JSAI での発表文献の共著関係から業績の高い研究者を抽出しているが、これらの研究の多くは静的なネットワークを対象としているおり、ノードやエッジがアクティブに増減(特に減少)を繰り返す動的ネットワークを対象とした分析はほとんど行われてこなかった。P2P のネットワークや、インターネット上のチャットや掲示板などのコミュニケーションのネットワークは、ある一定の期間だけリンクを張っている状態が多く、このようなネットワークはリンクがはらわれている短時間でネットワーク全体の構造を取得することが難しく、これまでほとんど分析対象にはされてこなかった。

そこで我々は、IRC(Internet Relay Chat) という古来から存在するチャットシステムに着目した。IRC はリア

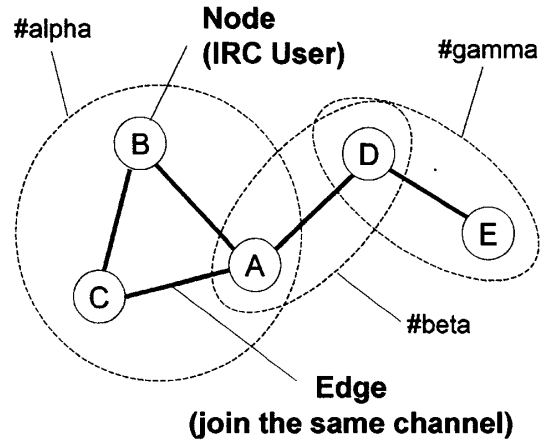


図 1: IRC のユーザネットワークの例

ルタイムチャットシステムであり、そこに入出入りするユーザで構成されるネットワークを時系列に抽出することで、現実世界で形成されるリアルタイムな複雑ネットワークとすることができると考える。

2. IRC ユーザネットワークの構築とその特徴

IRC(Internet Relay Chat) は古来から存在するリアルタイムチャットシステムであり、各ユーザをノードとし、各ユーザが同じチャンネルに参加しているかどうかの有無をリンクとして表現したネットワークは、リアルタイムに変化する。本稿では、このネットワークを IRC のユーザネットワークと呼ぶ。

IRC のユーザネットワークの例を図 1 に示す。図 1 のようにユーザ A やユーザ B をノードで表現し、ノード A, B, C もしくはノード A, D などの同一チャンネルに参加している関係の有無をエッジで表現する。図 1 でいえば、ユーザ A, B, C が参加しているチャンネルが #alpha, ユーザ A, D が参加しているチャンネルが #beta, ユーザ D, E が参加しているチャンネルが #gamma ということになる。このように、ユーザ間の参加チャンネル関係をネットワークとして表現することで、ユーザ間の参加チャンネル関係の構造を分析することができる。

このネットワークが現実世界で形成される複雑ネットワークの特徴であるスモールワールドやスケールフリーの特性を持つことは、我々の過去の分析により判明している。 [6][7]

本稿では、IRC のユーザで形成されるネットワークの中心性を調べて、その中心性の相関を分析することで、IRC のユーザネットワークの特徴を分析することを目的

*東京理科大学 理工学部 情報科学科

†(株) リコー

‡東京理科大学大学院 理工学研究科 情報科学専攻

とした。

3. データの取得方法

本稿では代表的な IRC ネットワークのうち、Friend-Chat[§]を対象とした。データを取得する IRC クライアントを実装し、ユーザ情報とチャンネル情報を取得した。データの取得には、IRC プロトコル [8] の LIST[¶], NAMES^{||}, WHO^{**}, WHOIS^{††}, WHOWAS^{‡‡} コマンドを使用した。

データの取得は 10 分ごとに行い、今回調査する対象のチャンネルとユーザは情報を公開している public なものに限った。チャンネルに関してユーザ数が 1 のものはグループとして機能を果たさないため処理の対象からは除外している。

なお、データの取得期間は 2006 年 10 月 3 日から 2006 年 10 月 28 日までである。

4. ネットワークの中心性とその指標

ネットワークの中心性 (以下、中心性と呼ぶ) とはネットワーク内でノードがどれだけ中心的な存在かを計る尺度である。これまでも多くの中心性の指標が提案されてきたが、代表的な 3 つの中心性の指標を用いて評価する。

以下、各指標の定義について説明する。IRC のユーザネットワークを重みなし無向グラフ G とし、ノード数を N 、エッジ数を K 、ノード i とノード j をつなぐエッジを a_{ij} とする。

ノードの持つ紐数 (Degree)

ノードの持つ紐数 (Degree) とはノードのエッジ数、つまり直接つながっているノードの数によって評価する指標である。エッジ数が多ければ多いほど中心性が高いと考える。ノード i と隣接するノードのエッジ数 k_i が Degree の値となる。正規化はノード i を除く $N - 1$ で除す。数式で表すと次のようになる。

$$C_i^D = \frac{k_i}{N-1} = \frac{\sum_{j \in G} a_{ij}}{N-1}$$

ノードの持つ媒介性 (Betweenness)

あるノードが他の全てのノードのペアの最短パスに含まれる総数で評価する指標である。媒介性とも呼ばれ、そのノードがいなければ情報が伝わらない、または遅れるといった情報伝達の重要性を示す。ノード j とノード k 間の最短経路数を n_{jk} 、そのうちノード i を含む経路数を $n_{jk}(i)$ とすると Betweenness

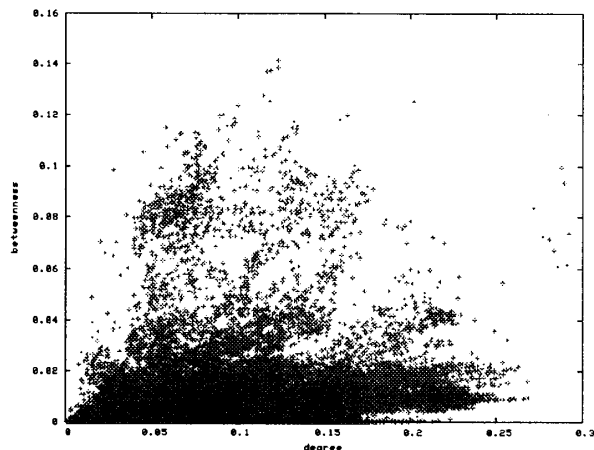


図 2: Degree と Betweenness の相関

の値は次の式で求めることができる。

$$C_i^B = \frac{2 \sum_{j < k \in G} n_{jk}(i) / n_{jk}}{(N-1)(N-2)}$$

ノードの間の距離 (Closeness)

あるノードから他の全てのノードまでの最短距離の平均で評価する指標である。この指標が高ければ、他のノードへの到達が容易になる。ノード i とノード j 間の最短経路の距離を d_{ij} とすると、 d_{ij} の平均の逆数で求めることができる。また、ノード i とノード j が非連結の場合に距離が ∞ となってしまう測定できないが、本稿では $d_{ij} = N - 1$ としている。式で表すと次のようになる。

$$C_i^C = (L_i)^{-1} = \frac{N-1}{\sum_{j \in G} d_{ij}}$$

5. 中心性の各指標の相関関係

IRC のユーザネットワークを、3 つの中心性の指標で分析し、その相関を調べた。

Degree と Betweenness の相関

Degree と Betweenness の相関図は図 2 のようになった。Betweenness の値が低いノードは Degree に依存することなく分布しているが、Betweenness の高いノードは Degree の中位に多く分布している。

Degree と Closeness の相関

Degree と Closeness の相関図は図 3 のようになった。Degree の値は Closeness の値に依存することなく分布しているが、Degree の高いノードは例外なく Closeness が低くない。

Betweenness と Closeness の相関

Betweenness と Closeness の相関図は図 4 のようになった。Betweenness, Closeness とともに大きい

[§]<http://www.friend.td.nu/>

[¶]現在の IRC ネットワークに存在するチャンネル名の一覧とそのチャンネルに JOIN しているユーザ数を得ることができる。

^{||}、チャンネル名の一覧とそのチャンネルに JOIN しているユーザのニックネームを得ることができる。

^{**}IRC サーバに接続しているユーザのニックネームの一覧を得ることができる。

^{††}そのユーザの IP アドレス、アイドル時間などを得ることができる。

^{‡‡}現在存在していないニックネームに対して一番最後に存在していたそのニックネームのユーザの IP アドレス、接続していた IRC サーバ名を得ることができる。

値をとるノードも見られる。一方で、Closenessは大きいとBetweennessが小さい値をとるノードも多数存在する。

6. 考察

それぞれの相関図より以下のようなことが考察できる。
ハブノードの数が比較的少ない

多数の友人知人を持つノード (Degreeの高いノード) は、自然とネットワークにおける媒介性が高くなる傾向にあると予想していたが、図2ではBetweennessの値はDegreeに依存しない結果となった。Betweennessが高くDegreeも高いノードは、ネットワークのハブ的存在と言え、このノードが多数失われるとネットワークが分裂する可能性もあるアキレス腱のようなノードであるが、IRCユーザネットワークには、それほど多数存在せず、Degreeのそれほど高くないノードにもBetweennessの高いノードが多数存在する。この結果は、アキレス腱となるノードに依存しているネットワークの中央に行くほど密になるネットワークではなく、適度にクラスタが分散しているネットワークの構造を示していると言える。また、アキレス腱となるノードがそれほど存在しないため、特定のノードの消滅によるネットワークの分断や崩壊などは起こりにくいと考える。

リンクの多いノードが必ずしも最短経路上に存在しない

Degreeの値が低い、リンク数の少ないノードは、当然Betweennessも低くなるが、Degreeの値が高いノードの中にもBetweennessの値が0に近いノードが多数存在する。これは、情報伝搬の最短距離を通ろうとした場合に、リンク数の多いノードを頼るのが必ずしも最短経路にはならないことを示している。

リンクの少ないノードでも各ノードへの距離は比較的短いケースが多い

図3より、Degreeが高くなれば、相対的にClosenessも高くなったが、Degreeが0に近いようなノードでもClosenessの高いノードが多数存在した。これは、友人知人をたくさん持っていないノードでも、ネットワーク内の各ノードへの距離が比較的近い位置にいるケースが少なくないことを示している。

最短経路上に位置することのないノードでも各ノードへの距離は比較的短いケースが多い

図4より、Betweennessの低いノードでもClosenessが非常に高いノードが数多く存在することがわかる。これは自身が他ノード間の最短経路上に存在するケースが少ないようなノードでも、各ノードへの距離は短いケースが多々あることを示している。

一般的にスケールフリー性を持つネットワークは、ハブノードへの攻撃耐性が低いとされている。[4] IRCのユーザネットワークは、スケールフリー性を持つが、中心的存在となるハブノードは存在せず、中規模のハブノードが多く存在している。また、ハブノードとはならないようなリンクの少ないノードがネットワーク上の任意の2ノード間の最短経路上にいることも多い。そして最短経路上に存在しにくい位置に存在するノードにも、各ノードへの距離が比較的短いケースも多い。

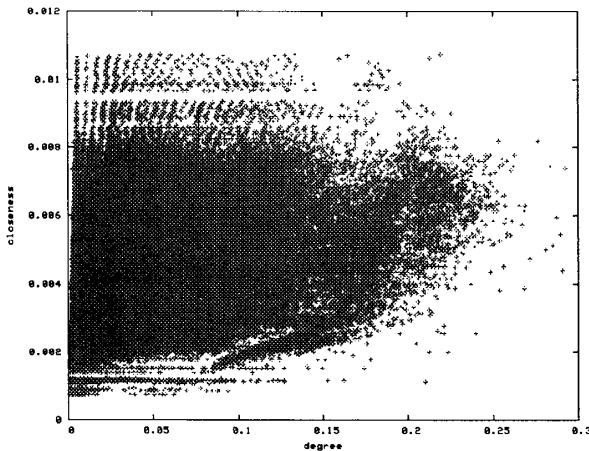


図3: DegreeとClosenessの相関

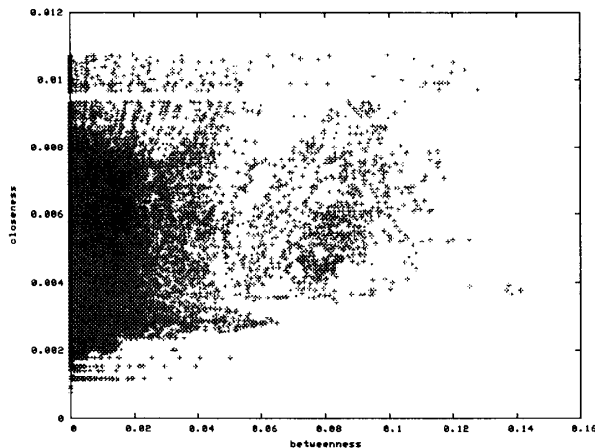


図4: BetweennessとClosenessの相関

- [8] J.Oikarinen, D.Read:Internet Relay Chat Protocol, RFC1459 (1993)

結論として、IRCのユーザネットワークはスケールフリーの脆弱性と言われているハブノードへの攻撃への耐性が比較的高いネットワークだと考えられる。

リアルタイムに形状を変化させる動的なネットワークは、ノードやリンクの増減が多いネットワークである。ハブノードと呼ばれる中心的なノードが半永久的に存在すると仮定される静的ネットワークとは違い、ハブノードですら一定時間後には消滅する可能性がある。このようなノードの消滅による被害が少ないようなネットワークが構成されていることは、非常に興味深い結果と言える。

7. まとめ

本稿では、動的ネットワークの分析を行うため、IRCのユーザネットワークを構築し、ネットワークの中心性の分析によって構造の特徴を分析した。

IRCのユーザネットワークはスモールワールド性とスケールフリー性といった現実世界の複雑ネットワークと同じ構造を持つネットワークであるが、代表的な中心性の指標 Degree, Betweenness, Closeness の各指標の相関図から、ネットワークのアキレス腱となるような中心的なハブノードはほとんど存在せず、中規模のハブノードが分散しており、スケールフリーの脆弱性と言われている特定ノードへの攻撃には、比較的耐性の強いネットワークであることが判明した。

参考文献

- [1] Watts, D.and Strogatz, S.: Collective dynamics of small-world networks, *Nature*, Vol.393, pp.440-442 (1998)
- [2] Watts, D.J.: Small Worlds: The Dynamics of Networks Between Order and Randomness, *Princeton University Press* (1999)
- [3] Barabási, A.L., R. Albert: "Emergence of scaling in random networks," *Science*, vol.286, pp.509-512 (1999)
- [4] Albert, R. et al.: "Error and attack tolerance of complex networks," *Nature* 406, pp. 378-382 (2000)
- [5] 友部博教, 松尾豊, 武田秀明, 安田雪, 橋田浩一, 石塚満: Semantic Web のための人の社会ネットワーク抽出と利用, 情報処理学会論文誌 Vol.46 No.6, pp.1470-1479, June 2005
- [6] 松澤智史, 中山雅哉: IRC におけるグループの生起時間と規模の研究, 情報処理学会 研究報告「高品質インターネット」No.007, May, 2003
- [7] 井出和典, 松澤智史, 武田正之: IRC におけるユーザ関係の分析とその応用例, 情報処理学会 電子情報通信学会 FIT2005, Sep, 2005