

[原著論文]

## 投資SNSのネットワーク構造

末木 将史\*

### On Network structure of Investors' Social Network Service

Masachika SUEKI\*

#### Abstract

In this article, we will explore the network structure of SNS specialized in equity investment. We use network data from over 40,000 individual investors. This is part of a total number of over 450,000 investors in this SNS.

Unlike other popular social networks such as Facebook and LinkedIn, this network doesn't seem to be human relationship. However, it is important that investors interact to share investment ideas. This is a valuable feature of the investment community SNS.

First, we analyzed the entire network. As expected, the entire network consists of a small number of page owners with a very large number of links (hereinafter referred to as investors) and a large number of page owners with a small number of links. Except for the scale-free nature of this network, no network-wide features were found.

Next, we adopted the Modularity Q method and divided the entire network into several sub-networks. Subnetwork statistics are completely different when compared to network-wide statistics. Most subnetworks have several hubs or authorities. However, the reason for such a network formation is an open question and need to be explored in the next work.

**KEY WORDS :** Social network analysis, Individual investor, hub and authority

## 1. はじめに

「投資家」を主な利用対象者としたインターネット上のソーシャル・ネットワーク・サービス（以下、投資SNS）のネットワーク構造を分析する。

直接金融を通じたファイナンスの進展が望まれる今日において、個人の証券市場への参加は政策的課題でもある。その意味で、近年のインターネットの発達による金融取引や情報収集を容易にする各種システムの登場は望ましい方向と言える。さらに投資に関する情報交換も行える投資SNSも登場した。もともと個人投資家は孤立した存在であり、機関投資家のように社内や同業者間で情報交換することにより自らの判断の適切さを推測する手立てがなかったが、もし投資SNS上での情報交換が行われるならば、いわゆる集合知につながり、ひいては投資成果に寄与する可能性がある。

投資SNSにおける投資家とは、株式・債券を主に投資対象としたページ・オーナーを指す。ページ・オーナーとは、投資SNSに参加するにあたり、用意されたフォーマットに基づき自分でWebコンテンツを作成し、自分専用のページを開設している主体のことである。ただし、ページ・オーナーが個人であるか、組織であるのかを申告する義務がないため、個人か組織かの区別は不明である。投資SNS上で行われている情報交換の実態を分析する前に、ページ・オーナー間でのリンクの状況を整理する。投資SNS上では、一般的なSNSとは異なり、自分から他のページ・オーナーにリンクを申し込んで形成されるタイプのものと（以下、「Followings」と呼ぶ）と他のページ・オーナーから申し込まれた被リンクで形成されるタイプのもの（以下、「Followers」と呼ぶ）の2つのリンクがある。本稿では、Followersを取り上げ、投資SNSのネットワーク構造を明らかにするための出発点としたい。

## 2. 先行研究

以下では、SNSを含むインターネット上のメディアが投資コミュニティで利活用されている実態、個人投資家が用いているSNSやインターネット上の投資情報サービスに関する検討、および社会ネットワーク分析等に関する研究を概観する。

### 投資コミュニティにおけるソーシャル・メディア

証券市場において、情報が早く適切に市場参加者に伝わるという情報効率性は価格形成の必要条件である。情報伝達には適時開示やマス・メディアなど伝統的な

経路に加えて、企業が独自にソーシャル・メディアのアカウントを設定して一般および投資家に対して情報を提供することが増えてきた。米山（2014）は、企業がIR情報ツールとしてブログを含むネット上のソーシャル・メディアを用いている実態とその問題点について整理している。2013年時点の米国では、Twitter, YouTube, FacebookあるいはLinked inの順でコーポレート・コミュニケーションに用いられており、その中でもTwitterやSNSはアナリストおよび機関投資家の投資意思決定に利用されていることを報告している。ここで取り上げられているTwitterのツイートなどの情報発信母体は企業であることから、ネットワークとしては企業をハブとする投資家コミュニティである。ソーシャル・メディアは双方向性を有しているので、企業と投資家との間ではインタラクションが行われている可能性はおおいにありうるが、本研究では分析内容とはされていない。

一方、ソーシャル・メディア上での投資家が情報発信を行う場としてのソーシャル・メディアを対象として研究したものに、Giannini, et. al. (2015)がある。この研究は、投資家がTwitterに投稿した内容を自然言語解析して各銘柄のアナウンス前後における期待収益率とその後の収益率および取引高について分析している。投資家間での期待収益率に関する見解の相違度合（分散）が大きくなると、取引高が大きくなる一方、収益率は低くなった。なお、アナウンス前後における投資家の見解の相違度（分散）とマス・メディアのニュースの相関はなかったことから、投資家は独自に情報を分析している様子が窺われる。

この他にも、インターネットを利用して、投資家から予想投資収益率などを集計したウィスパー<sup>1</sup>と呼ばれるものを対象にした分析として、Bagnoli, et.al. (1999)やZaima and Harjoto (2006)がある。これらの研究は、一般の投資家から集めた予想情報を用いると投資収益率が得られることを報告している。つまり、この限りでは一般の投資家による予想情報であっても投資収益率を得るためには価値があることが窺える。

### 投資家の情報交換、ネットワーク

Chen, et.al. (2016)も、証券投資に特化したSNSであるSeeking Alphaへの投資家の投稿内容を自然言語解析して任意の銘柄の期待収益率に対するネガティ

<sup>1</sup> ウィスパーとは、サービス利用者が自主的に入力する予想投資収益率など投資関連予想のデータをまとめて、インターネット上で集計数値が公表されるものであり、今日では一般的になった星の数や5段階スコアなどのユーザー評価に類するものを指す。

ブまたはポジティブのセンチメント<sup>2</sup>を抽出している。なお、センチメントは初出の投稿だけではなく、他の投資家の投稿に基づく議論や対話も考慮されていることから、投資家間での情報交換が研究範囲に入っていることになる。そしてこれらのセンチメントを説明変数として超過収益率に対して回帰した。その結果、マイナスの超過収益率に対しては、ネガティブなセンチメントは有意な説明力を有していた。この有意差は、企業に関するマス・メディアの記事や財務情報でコントロールしても存在していた。このことから、投資家のSNS上での情報発信および相互の情報交換は、抽出されたセンチメントの限りでは情報生産していないとは言えないことになる。ただし、ポジティブなセンチメントと、それによるプラスの超過収益率については有意な結果が得られていないが、この理由の解明は今後の研究に残されている。

Ozsoylev et.al. (2011) は投資家のネットワークと投資収益率の分析をしているが、ここでのネットワークはSNS等ではなく、実取引ログのデータから投資家間にあると推測されるネットワークである。ログ上で取引を行うタイミングが近接している投資家間にはネットワークがあると仮定して、投資収益率を計測した。その結果、近接している投資家グループの投資収益率は、そうではない投資家のグループよりも優れていることが分かった。Ozsoylevらが主張するように、SNSにおけるネットワークが投資実行に対して影響を及ぼす情報のルートになっていたとしても、現実に売買行動にまで結びつかなければ、投資家ネットワークがコミュニティの構成員の投資収益率に有益であるとは言えない。とは言え、Ozsoylevらが実取引データから導き出したとする投資家ネットワークも、実際に確認できるリンクが存在するわけではないので、この分析アプローチの優越性が確認されるのはまだ先のことだと考えられる。

投資電子掲示板に関するネットワーク分析を行った研究は少ないが、その中にDas and Sisk (2003) がある。この研究はYahoo!, Motley Fool, Raging Bull, Silicon Investorの投稿者ネットワークと銘柄の特性（価格変動や同分散など）を取り上げている。分析は、まず初めに投稿者が重複する銘柄間でネットワーク分析することにより銘柄間のコミュニティを抽出し、次いでコミュニティ内の価格変動の共分散や収益率を解明するものである。コミュニティの抽出方法は、

投稿者が重複する銘柄毎に固有ベクトル中心性を計算し、この値が高いものを選択することにより行った。分析の結果、多数の銘柄により構成される大規模コミュニティのリターンは、小規模のものよりも有意に大きく、共分散は有意に小さかった。また、高い固有ベクトル中心性スコアを有する株式は、そうでない株式よりも高い平均共分散を有する傾向があった。

#### SNSに対するネットワーク科学分析

投資家がTwitterあるいはSNSで発信する投稿の分析を通じて、安定的に超過収益率を説明する分析アプローチはまだ確立していないようである。ネットワークを構成する投資家、あるいは個々の投稿の中には、有意に超過収益率を説明する可能性があるものが含まれている一方、ノイズでしかないものが含まれている可能性も多分にある。

投資SNSではないが、松尾・安田 (2007) はmixiを対象にネットワーク科学の面から分析している。用いられた分析尺度は、次数分布、およびそのランクサイズ・プロット、クラスター係数等の標準的な分析を行い、mixiのネットワークにはスケール・フリー性やスモールワールド性が存在することを確認している。

石川・松尾 (2012) は、参加者数約37万人のSNSであるLococomを対象にしたネットワーク構造について分析している。用いた分析尺度は、次数中心性、近接中心性、媒介中心性および固有ベクトル中心性である。これら指標とネットワーク上での参加者のアクティビティである投稿との関連を分析している。また、最短経路長およびクラスター係数も、部分グラフごとに算出している。最大部分グラフの最短経路長は短い一方、クラスター係数は大きく、スモールワールドネットワークであることが確認された。

ネットワーク構造の把握から分析を参加者同士のつながりに置いた研究もある。丸井、他 (2010) はmixiを分析対象に、参加者の属性カテゴリー別に平均次数やクラスター係数等を算出している。さらにリンク両端のノードの属性およびその違い、つまり異質性ならびにアクセスなど行動に基づいてリンクの特徴をクラスター分析している。招待制による参加者で構築されるネットワークは、同質的な属性の参加者間でのリンクが形成される傾向があると推測されるが、分析の結果、リンクの特徴によりアクセスの頻度や、やり取りされる情報内容が異なることがうかがわれている。

<sup>2</sup> センチメントとは、当該銘柄に対する投資態度（売りたい / 買いたい / 保留）のことを指す。

Table. 1 オーナー・ページに含まれる主要なコンテンツ

カテゴリー	表題	内容説明	補足説明
投資行為に関する情報	予想	現在、投資中の銘柄	売買時点、現時点での収益率、投資実行時点および現時点での価格、当該投資行為に対する自分あるいは他のページ・オーナーからのコメント、等
	終了した予想	既に、売買が終了した銘柄	期間、収益率、価格、当該投資行為に対する自分あるいは他のページ・オーナーからのコメント、等
自由記述	プロフィール	投資歴、投資スタイル、情報源、利用証券会社、自由記述、等	---
	ブログ	自由記述	---
ネットワークに関する情報	リスペクト (Respect)	リスペクトしているページ・オーナー	---
	フォロー (Followings)	フォローしているページ・オーナー	---
	フォロワー (Followers)	フォローされているページ・オーナー	---
	コミュニティ	ページ・オーナー間での集まり	証券投資領域だけにとどまらず、多様なテーマがある

注1) 列名のうち、表題のみ運営会社が付けた名称。その他はすべて著者作成

### 3. 投資SNSの概要

本稿で対象とする投資SNSは、株式会社ミンカブ・ジ・インフォノイドが運営する「みんなの株式」である<sup>3</sup>。この投資SNSは、証券投資や外国為替証拠金取引 (FX) などに関する情報が交流の主な対象である。この投資SNSに参加する主体 (以下、この主体を「ページ・オーナー」と呼ぶ) は、アカウントを取得してWebページを開設する (以下、このページを「オーナー・ページ (OP)」と呼ぶ)。Webページには参加者の基本的情報 (性別や投資歴など) を記載するコンテンツのテンプレートや、行動情報 (投資対象や投資成果) についてのテンプレート、参加者間での交流を図るためのページ、ブログ等のコンテンツがある (Table. 1)。いずれのコンテンツもページ・オーナーの自己申告によるものであるが、一部は投資SNSのシステムにより自動処理されるものもある<sup>4</sup>。このうち本稿で対象とする情報内容は、他のページ・オーナーからリンクを申し込まれたことにより形成された入リンクを意味する Followers である。このリンクにより、任意のオーナー・ページと他のオーナー・ページとの間にネットワークが形成される<sup>5</sup>。

一方、当該ページ・オーナーが他のページ・オーナー

に対してリンクを申し込んだことにより形成された出リンクを意味する Followings もシステム上で用意されている。投資SNS上では、リンクを許諾したとしたら、相手にもリンクを張るのがネット上のマナーとして広がっているが、相互にリンクすることが必須ではないことから、Followers と Followings は出入方向が異なるだけの同じネットワークではない。両者には何かしらの形成理由の差異があると推測されることから、ネットワーク構造を理解するための対象として興味深いものの、本稿では Followers から取り上げることにする。

### 4. ネットワークの統計量

#### 4.1. サンプルの基本統計

投資SNSのデータ収集は、2016年4月から2018年4月30日まで行った。オーナー・ページには連番が付され、データ収集期間最終日における数値は453,573であった。この中から削除されて存在しないオーナー・ページ (45,292) と、投資対象がFXだけのオーナー・ページ (352,081) を除外した。以上、サンプル採用基準として、ページが現存しており、かつ株式の投資履歴が1件以上を満たすオーナー・ページ (第1次サンプル・セット) は41,599件であった<sup>6</sup>。

<sup>6</sup> 投資履歴には2種類ある。データ収集時点で、現物取引、先物取引のいずれにもかかわらず売買が完了して投資成果が確定したものに関する履歴情報と、当該時点で投資中のものに関する情報、である。

<sup>3</sup> 「みんなの株式」: <https://minkabu.jp/>

<sup>4</sup> 自動処理される例として、投資損益率 (パフォーマンス) や Followers (被リンク) 数、Followings (リンク) 数などがある。

<sup>5</sup> ここで述べた Followers や Followings 以外にも、実際にはリスペクト (Respect) 関係や“ブログ”に対するコメントなどでもネットワークは発生しうる。

Table. 2 ページ・オーナーの基本統計

variables	value
全ページ・オーナー数	41,599 件
買い銘柄数(算術平均)	5.4 銘柄
買い銘柄数(中央値)	2.0 銘柄
買い銘柄数(標準偏差)	41.1
最大買い銘柄数	2,996 銘柄
最少買い銘柄数	0.0 銘柄
全ページ・オーナー総計延べ買い銘柄数	225,677 銘柄
売り銘柄数(算術平均)	1.2 銘柄
売り銘柄数(中央値)	0.0 銘柄
売り銘柄数(標準偏差)	29.2
最大売り銘柄数	3,724 銘柄
最少売り銘柄数	0.0 銘柄
全ページ・オーナー総計延べ売り銘柄数	48,147 銘柄

注1) 2018年4月30日現在の統計

投資SNSを概観する (Table. 2)。まず、2018年4月末時点でのネットワークの基本統計を述べた後、ネットワーク統計量を示す。サンプル・セットに含まれるオーナー・ページの平均買い銘柄数は5.4銘柄、平均売り銘柄数は1.2銘柄である。最大買い銘柄数は2,996銘柄であり、これらの数字は当該ページ・オーナーがこれまでに購入した延べ銘柄数を表しており、現時点で未売却のものも含む。最大売り銘柄数は3,724銘柄であり、これまで売却した延べ銘柄数を表したものである。

#### 4.2. ネットワークの基本統計量

次に、ネットワークの基本統計量を見てみる。前述したとおり、投資SNS上でのオーナー・ページ間において、リンクが張られることによりネットワークが発生する。このネットワークは、あるオーナー・ページからリンクが出て、別のオーナー・ページにリンクが入るといった方向性を持った有向グラフになる。以下で取り上げるのは、オーナー・ページ個別属性に類する統計量である次数中心性 (degree centrality) と、その分布度 (Rank Size Plot)、ページランク (Google PageRank)、および投資SNSネットワークに関するネットワーク統計量である密度 (density)、クラスター係数 (Cluster coefficient)、直径 (diameter)、平均パス長 (Avg. Path length) などである。

##### 4.2.1. リンクと次数中心性<sup>7</sup>

オーナー・ページ別にFollowers (入リンク) 数を見てみると、極めて偏った分布になっていることがわかる (Fig. 1)。最多のオーナー・ページはリンク数2位以下を大きく上回っている。2位および3位は

<sup>7</sup> 次数中心性のみを分析に用いる。現段階ではネットワーク構造を把握することが目的であり、ページ・オーナー間での情報の交流までは分析範囲とはしていない。そのため、他の中心性尺度である、情報中心性、媒介中心性などはここでは取り上げない。

1,000本台であり、これらリンク数が1,000本を超えるものは、3件しかない。また100本を上回るのは、53件 (第1次サンプル・セットの0.13%) にしか過ぎない。大多数のオーナー・ページは、リンク本数10本以下である。リンク本数が0本の孤立したオーナー・ページが最多層を形成しているが、これでは投資SNS上でネットワークを形成しているとはみなせない (Table. 3)。オーナー・ページ (第1次サンプル・セット) の41,599件によるFollowers平均は1.3本、中央値0.0本であるが、これら統計量ではリンク数の分布を適切に評価することは難しいと言える (Table. 4)。

##### 4.2.2. スケール・フリー性

前述の通り、リンク本数の分布はきわめて多いオーナー・ページが少数存在する一方、10本以下が大多数を占めるという歪んだものになっている。従って、基本的な統計量である平均値や中央値、標準偏差などで集団の分布特性を判断することはできないというスケール・フリーの可能性がある。そこで、ランクサイズ・プロットを行うことにより、基本的な統計量が分布特性を判断する上で意味があるかどうかを検討する。つまり、もしリンク数分布にスケール・フリー性があるならば、平均、分散、中央値といった統計量ではこの分布の特性を評価できない、ということである。

ランクサイズ・プロットは、縦軸にリンク本数を多い方からの順位の対数値を取ったものを、横軸にリンクの対数値を取ったものである (Fig. 2)<sup>8</sup>。図表中の右下端のプロットは、リンク数が最多のオーナー・ページであり、その左にある2つのプロットは同2位と3位のオーナー・ページであることが、ほぼ特定できる。順位4位以下から両対数が定率ではなくなってきたことから、スケール・フリー性が疑われる。これを検証するため、ランクサイズ・プロットの回帰を行った。ベキ乗モデルは $y$ をリンクの小さい順のランクに、 $x$ をリンク本数とすれば  $y = C x^{-\alpha}$  になるので、これを数値計算できるように両辺を対数化し、

$$\ln(\text{Rank\_Followers}) = C - \alpha \cdot \ln(\text{Link\_nobs\_Followers})$$

を計算する。ここで $\alpha$ が1に近ければスケール・フリーの分布とみなすことができる。なお、 $C$ は常数である。計算の結果、 $\log(\text{Link\_nobs\_Followers})$  の係数は有意に-0.93であった。このことから、 $\log(\text{Link\_nobs\_}$

<sup>8</sup> この場合、リンク数が最多のオーナー・ページは順位1位であることから、 $\log(\text{Rank\_Followers}) = \log(1) = 0$  となり、リンク数が最多の7,699本のオーナー・ページは  $\log(\text{Link\_nobs\_Followers}) = \log(7699) \approx 3.89$  になる。ゆえに、図表中の右下端のプロットは、リンク数が最多のオーナー・ページであると特定できる。

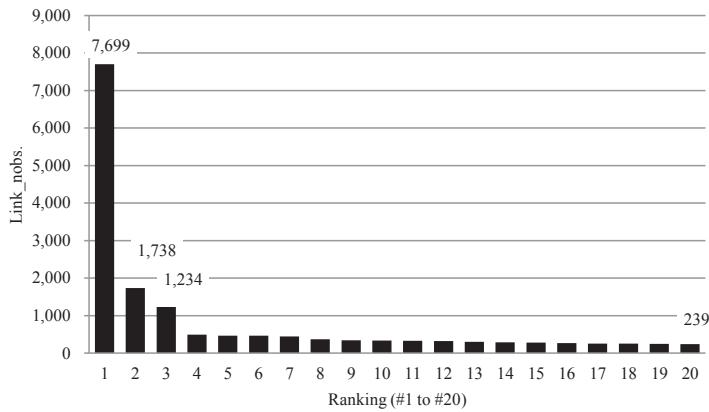


Fig. 1 リンク数順位上位オーナー・ページのリンク数

Table. 3 リンク階級別オーナー・ページ数

Link_nobs.	OP 数
1000～	3
501～1000	0
401～500	4
301～400	6
201～300	13
101～200	27
51～100	66
41～50	29
31～40	64
21～30	103
11～20	298
2～10	3,756
1	5,308
0	31,922

Table. 4 ネットワーク基本統計量

Stats.	value
平均入リンク数	1.3 本
入リンク数(中央値)	0.0 本
最大入リンク数	7,699 本
最少入リンク数	0.0 本

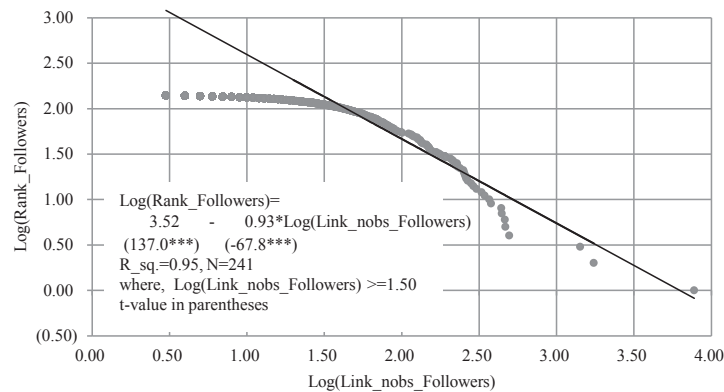


Fig. 2 Followers のリンク本数に関するランクサイズ・プロット

注1) 図表中の直線は-1の傾きを示す。

Followers) が1.5～2.5のレンジにおけるリンク数分布は、スケール・フリー性を有している可能性がある<sup>9</sup>。

#### 4.3. ネットワーク統計量

##### 4.3.1. ネットワーク密度, 平均クラスター係数, 平均パス長

ここからは、リンクが結ばれたネットワークを対象とするため、リンク本数が0本の孤立したオーナー・ページを除外した第2次サンプル・セットを用いてネ

<sup>9</sup> 参考のための傾き -1 の直線を図中に示してある。

ットワーク統計量を計測する。Followersのリンクは有向グラフであるため、オーナー・ページが相互にリンクしている場合は、片方がキャンセルされた数のオーナー・ページ数になる。オーナー・ページの数には15,944件であり、リンク本数は60,313本である。

ネットワーク密度<sup>10</sup>は0.000、平均クラスター係数<sup>11</sup>は0.183、平均パス長<sup>12</sup>は2.999であった。これらの尺度を組み合わせて考えると、以下のように述べられるだろう。まず、ネットワーク密度が低いことから、オーナー・ページ間でリンクが多く張られているとは言えず、それに応じて平均クラスター係数も低いということになり、リンクを相互に結ぶというネットワーク上での活動が活発でない様子がうかがえる。平均パス長もサンプルサイズの規模の割には小さいので、1つの巨大なネットワークなのではなく、いくつかの部分ネットワークに分離しているのかもしれない。あるいは、スモールワールド性を備えているのかもしれない<sup>13</sup>。

<sup>10</sup> ネットワーク密度とは、ネットワーク構成者（オーナー・ページ）間でのつながりを全体的に評価するための尺度である。ネットワーク内に現実のリンクの数に対して、もし全ての構成者（オーナー・ページ）がその他すべての構成者とリンクを持つネットワーク（これを完全グラフと呼ぶ）と想定した場合のリンク数に対する比率である。 $D$ をネットワーク密度、 $G$ を現実のリンク数、 $K$ を完全グラフであったと仮定した場合のリンク数とすると、 $D=G/K$ である。最大値は1であり、1に近いほどネットワーク内のリンクが密接である。

<sup>11</sup> 平均クラスター係数とは、ネットワークがどの程度、集中、凝集しているのかを判断する尺度である。ネットワーク上で、あるオーナー・ページ（A）にリンクで隣接している別のオーナー・ページ（B）と、そのBにリンクしているオーナー・ページ（C）という3者がいたとして、もしAとCの間にもリンクがあるならば、三角形ができ、これをクラスターと呼ぶ。クラスターは、構成者A,B,Cの3者全てがリンクされているため、密接な関係をもつネットワークと言える。構成者数が増えても全ての構成者が他の全ての構成者とリンクをもつネットワーク（完全グラフ）は存在しうるが、原理的にクラスターは最少規模の完全グラフである。そこで、このクラスターがどの程度ネットワーク内に観察しうるかにより、集中、凝集の判断ができることになる。個々のオーナー・ページについてのクラスターを計算する式は、 $C_i = v_i / \{k_i(k_i - 1) / 2\}$ である。ここで、 $C_i$ はオーナー・ページ*i*のクラスター係数、 $v_i$ はオーナー・ページ*i*を含む三角形の数、 $k_i$ はオーナー・ページ*i*のリンク本数である。この $C_i$ を全オーナー・ページで平均したものが平均クラスター係数であり、高いほどネットワーク内の集中、凝集性が高いとみなせる。

<sup>12</sup> 平均パス長とは、ネットワークの広がり、長さなどを把握するための尺度である。全てのオーナー・ページ毎に、ネットワーク上でつながっているその他すべてのオーナー・ページまでの距離（何個のリンクを経由しているか）を合計し、平均したもの。ノード（オーナー・ページ）数が少ないネットワークならば平均パス長が小さくなるのは当然だが、巨大なネットワークでも凝集性が高かったり、スモールワールド性が備わっていたりする場合は小さくなることもある。

<sup>13</sup> スモールワールド性とは、相対的な概念だが、以下の3つの要件に該当するものをいう。まず、全オーナー・ページ数に対してリンク数が少ないこと、次に、ある任意のオーナー・ページと別のものとの2点間の距離が小さいこと、そして、クラスターになっている度合いが大きいこと、である。

なお、証券投資に関するSNSであるという本質を考えれば、仲間づくり・友人作りのSNS（Facebookなど）や仕

#### 4.3.2. ページランク

ネットワーク上における個別オーナー・ページの重要性や、オーナー・ページ間の関係性を計測するには、次数中心性だけでは不十分である。リンクはシステム上、無許可で一方的に張ることも可能だからである。従って、次数が高いからと言って、当該のオーナー・ページがネットワーク上で重要であることを意味しない。もし重要なオーナー・ページであるならば、同様なオーナー・ページとリンクが張られているかを考慮するのも一案である。そこで、ページランク（Google PageRank）を用いて計測してみる<sup>14</sup>。

個別オーナー・ページに関してページランクを算出してみたところ、値は全般的に0.1程度であり、高くはない。分布もオーナー・ページ別にみても、極めて偏っていることが分かる（Fig. 3）。以上から、ページランクを用いてネットワーク上における個別オーナー・ページの重要性の評価を試みても、次数中心性での試み同様に上手く行かないことが分かった。

しかし、ページランクとリンク本数を対比してみると、それぞれの相対的な違いが発見できる可能性がある。散布図を描いた（Fig. 4）。両者の相関は、 $r=0.69$ であり相応に高いことから、リンク数が多いオーナー・ページのページランクが高くなっている可能性があるかと推測できる。そこで、リンク数に比べてページランクが高くなっているオーナー・ページを見つ

事関係のSNS（LinkedInなど）、あるいは現実社会でのオフライン関係とはリンクを張る動機が異なることが影響しているのかもしれない。この点は詳しい分析が必要である。

<sup>14</sup> ページランクとは、入リンクと出リンクの比率に基づくウェイト付けしたものであり、この数値が高いほどネットワーク上での重要性あるいは存在が大きいのである。原初的な算出式は、 $r(OP_i)$ をオーナー・ページ*i*のページランク、 $r(OP_j)$ を同*j*のもの、 $|OP_i|$ を同*i*からの出リンク、 $B(OP_j)$ を $OP_i$ に向けた出リンクとすれば、 $r(OP_i) = \sum_{P_j \in B(OP_i)} \{r(OP_j) / |OP_j|\}$ で表される。左辺および右辺の分子を見ればわかる通り、 $OP_j$ を計算するためには $OP_j$ が用いられていることから、値が収束するためには繰り返しが必要となる。しかし、ネットワークの中に出リンクを持たない（これをランクシンクと呼ぶ） $OP_j$ が存在している場合、分母がないことを原因としてページランクが論理整合性なく高まるという問題が生じる。そこでランクシンクのオーナー・ページから別のオーナー・ページへジャンプすることにより本問題を回避する改善を施した。オーナー・ページ間のリンク関係を行列で表し、それに改善を施した行列を $S$ 、繰り返しのためのオペレーターの全要素が1の行ベクトルを $e$ 、調整パラメーターを $\alpha$ 、 $G$ を中間算出値とすると、 $G = \alpha S + (1 - \alpha) \cdot \frac{1}{nee^T}$ と表せる（ $n$ は計算量のオーダー）。この $G$ を用いて $r(OP)$ に類するGoogleベクトル $\pi$ を算出する。 $\pi$ の算出式は $\pi^{(k+1)T} = \pi^T G$ であり、ここで $k$ は繰り返し回数を表す。より詳しい説明はLangville & Meyer（2006）を参照のこと。

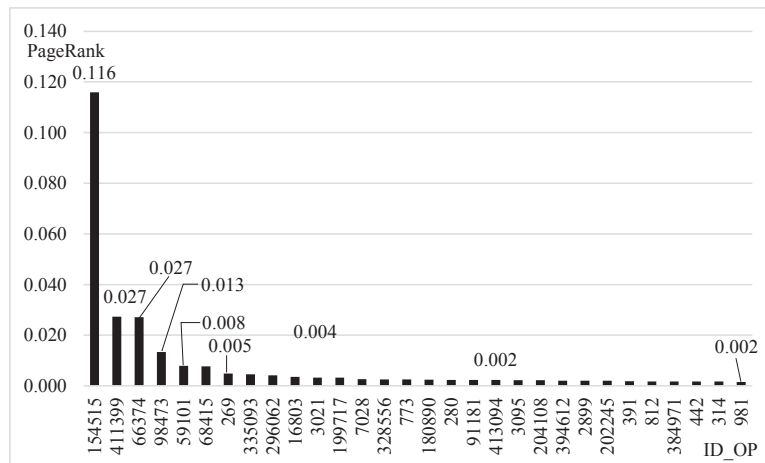


Fig. 3 オーナー・ページ別ページランクの分布

- 注1) オーナー・ページの番号(ID\_OP)は区別のために便宜的につけられたものであり、分析上の意味はない。
- 注2) PageRank<0.002 のオーナー・ページは図表から省略してある。

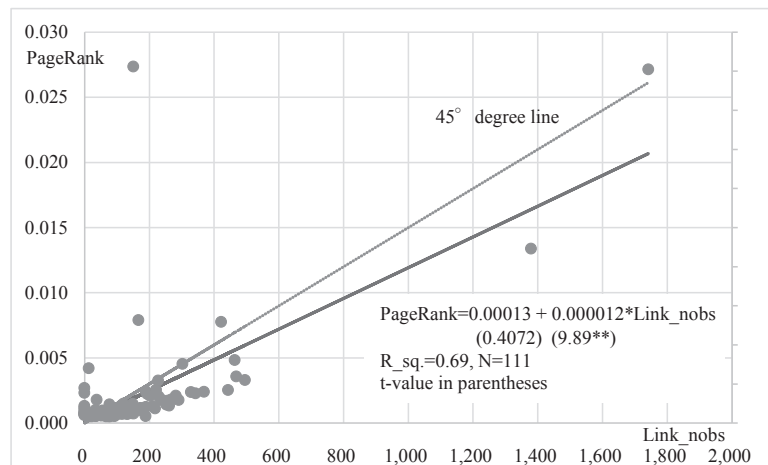


Fig. 4 ページランクとリンク数の相関

- 注1) 点線は45度線(ページランクとリンク数が完全相関していると仮定した場合の線)を、実線は通常最小二乗法(OLS)による回帰線である。今回の分析では、リンク本数を分散で基準化していないため、推計結果と検定統計量は分散不均一性による影響を受けている可能性が高い。本来ならば、加重最小二乗法(WLS)などによる推計が望ましい。
- 注2) 分散不均一性による影響を減じるため、最大ページランクとリンク数のサンプルを除いてある。分析に用いたサンプル・セットは、ページランク 0.027 以下かつ 0.001 以上の N=111 である。

けるため、回帰分析による理論値よりもページランクが大きいもの(図表中の回帰線よりも左上にプロットされるもの)を探したところ、36件のオーナー・ページ(全111サンプルの36.4%)が見出された<sup>15</sup>。さらに、完全相関を意味する45度線よりもページランクが大きい(図表中の45度線よりも左上にプロットされるもの)を探したところ、件数は不明ながらも幾つかの

<sup>15</sup> ただし、リンク本数の変数を正規化していないため、推計結果と検定統計量は分散不均一性による影響を受けている可能性が高い。このため、見出された36件のサンプルのページランクが理論値を上回っているのか検討の余地がある。

オーナー・ページが見つかった<sup>16</sup>。これらのオーナー・ページを特定して近隣のネットワーク構造を分析してみる価値はあるだろう。

## 5. ネットワーク構造

### 5.1. ネットワーク・グラフ

サンプル採用基準を満たすオーナー・ページ全41,599

<sup>16</sup> ページランクのスコアの方がリンク数よりも高いオーナー・ページを探し出す方法として、回帰直線を用いるよりも45度線を用いた方が論理的に正しいと思われるが、サンプルを検出するのは困難であり、今後の課題である。



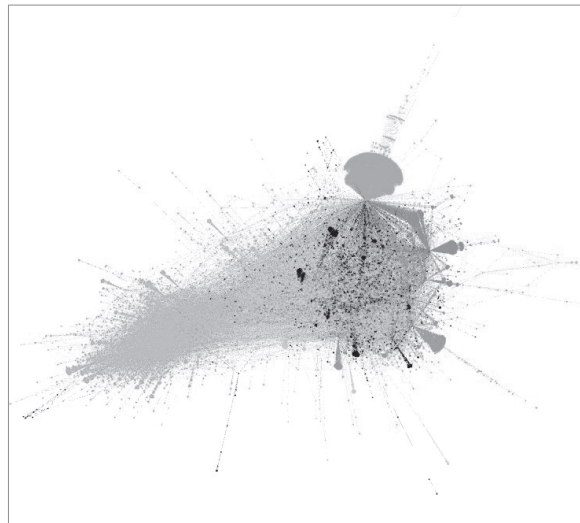


Fig. 5 Followers のネットワーク・グラフ

注1) 濃度は関連性が強いと認められたサブ・ネットワーク(コミュニティ)の区別を表す(後述)。

件の中で、リンクを有するオーナー・ページは15,946件であり、これらによるリンク数は60,318本である。これをForce Atlas2アルゴリズムにより2次元のグラフを描いた (Fig. 5)<sup>17</sup>。一見したとおり、全リンクでグラフを描いても、混沌としていてネットワークの特徴が判断できない。

### 5.2. コミュニティ (サブ・ネットワーク)

全リンクのグラフを描いても、混沌としていてネットワークの特徴が判断できなかった。そこで、判別できるように部分ごとサブ・ネットワーク (以下からはネットワーク分析の通例に従い、“コミュニティ”と呼ぶ) に分割してみる。モジュラリティーQに基づきコミュニティを分割した<sup>18</sup>。その結果、131のコミュニティに分けることができたが、これらコミュニティも、オーナー・ページ毎のリンク数の格差同様に、含まれるオーナー・ページの数に格差がある (Fig. 6)。コミュニティ4は6,000近いオーナー・ページにより構成されている一方、構成数1,000以下のものはコミュニティ8を含め125個となっている (なお、コミュニティの番号は区別のために便宜的

につけられたものであり、分析上の意味はない。また、構成数3以下のモジュラリティーは図表から省略してある)。

### 5.3. コミュニティ毎のネットワーク統計

ある程度の規模を持つ主要なコミュニティを取り上げてネットワーク構造を分析する。対象とするのは、オーナー・ページ数100以上により構成されるコミュニティで、多い順にコミュニティ4,1,10,9,3,2,6,8,0,5の10個である。基本的なネットワーク統計量 (Table. 5) からは、全般的に密度 (Density) も低く、このことから予想されるように平均クラスター係数 (Avg. Cluster\_coef.) も低くなっている。その一方で、コミュニティ内におけるさらに小さなコミュニティ数 (nobs\_SubCommunity) が存在しており、それらの平均オーナー・ページ数は、コミュニティ (Fig. 6参照) の上位8番目 (コミュニティ8) から12番目 (コミュニティ8) に該当する規模である。このように、さらに小さなサブ・コミュニティが構成されていることが分かる。

ネットワーク・グラフを見てみると (Fig. 7)、幾つかのカテゴリーに分けられることが分かる。サブ・コミュニティが混然としているもの (Community\_1,10)、混然としているが、次数中心性が高いオーナー・ページが見受けられるもの (Community0,2,3,5,6,8,9)、極めて高い次数中心性のオーナー・ページが存在するもの (Community\_4) などである。

コミュニティに関するネットワーク統計量間での相関行列を見る (Table. 6)。正の相関が高い (+0.7

<sup>17</sup> Force Atlas 2 アルゴリズムを用いた理由は、計算速度が他のアルゴリズム (Fruchterman-Rheingold, YifanHu など) と比較して早いことによる。詳しくは Jacomy et.al (2014) 参照のこと。

<sup>18</sup> モジュラリティーQとは、サブ・グラフがランダムグラフから異なることを示す指標であり、この値が高いほど良いとする分割手法である。今、あるサブ・ネットワークをC、コミュニティ*i* からコミュニティ*j*へ接続されているオーナー・ページ数を  $e_{ij}$ 、ネットワーク全体に含まれるオーナー・ページの総数を  $k$  とすると、 $Q = \sum_{i \in C} \{e_{i,j}/2k - (\sum_{i \in C} e_{i,j}/2k)^2\}$  で表される。

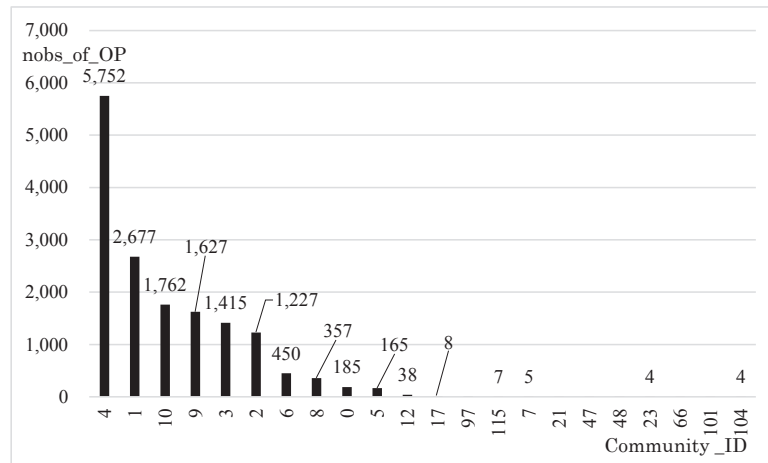


Fig. 6 コミュニティー別の構成オーナー・ページ数

- 注1) コミュニティーの番号は区別のために便宜的につけられたものであり、分析上の意味はない。  
 注2) 構成オーナー・ページ(OP)数3以下のコミュニティ(109個)は図表から省略してある。

Table 5 主要コミュニティのネットワーク統計量

Community_ID	OP_nobs 注1)	Link_nobs	Avg. Degree	Density	nobs_Sub Community 注2)	Avg. Cluster coeff.	Diameter	Avg. Path_ length
0	232 (185)	341	1.470	0.006	21 (11.0)	0.066	4	1.932
1	3,082 (2,677)	16,630	5.396	0.002	12 (256.8)	0.262	7	3.024
2	2,019 (1,277)	3,961	1.962	0.001	28 (72.1)	0.122	10	2.441
3	1,679 (1,415)	2,692	1.603	0.001	23 (73.0)	0.106	8	2.858
4	8,327 (5,752)	11,155	1.340	0.000	128 (65.1)	0.018	5	2.018
5	287 (165)	494	1.721	0.006	11 (26.1)	0.078	9	3.324
6	725 (450)	1,514	2.088	0.003	11 (65.9)	0.088	9	3.366
8	685 (357)	1,434	2.093	0.003	14 (48.9)	0.104	11	3.530
9	2,596 (1,627)	6,282	2.420	0.001	13 (199.7)	0.152	8	2.917
10	2,872 (1,762)	15,413	5.367	0.002	8 (359.0)	0.141	9	3.575

- 注1) オーナー・ページ数(OP\_nobs)は、出リンクと入リンクそれぞれが異なるコミュニティに属するオーナー・ページも含めてあるため、Fig. 6 で示したオーナー・ページ数と違いがある。Fig. 6 と同じ OP\_nobs は下段のカッコの中に示してある。  
 注2) サブ・コミュニティ(nobs\_SubCommunity)数の下段のカッコ内は、サブ・コミュニティ1つあたりの平均オーナー・ページ数(OP\_nobs)である。

以上) の関係は、ある意味当然であるので、細かくは取り上げない<sup>19)</sup>。一方、相関係数としては低いものの、サブ・コミュニティ (nobs\_Sub\_Community) と、平均クラスター係数 (Avg. Cluster\_coeff.)、直

<sup>19)</sup> 例えば、リンク数 (Link\_nobs) が多ければ、平均次数 (Avg. Degree) が高くなるのは当然であるし、直径 (Diameter) が大きければ平均パス長 (Avg. Path\_length) も長くなるのも当然だからである。また、密度 (Density) とオーナー・ページ数 (OP\_nobs)、およびリンク数 (Link\_nobs) には -0.5 以下の負の相関があるが、これも密度の計算式を踏まえれば当然の傾向である。

径 (Diameter)、平均パス長 (Avg.Path\_length) の間には負の相関がある。これらから推測されることは、コミュニティ内には、さらに小さなサブ・コミュニティが存在し、そこでは多くのオーナー・ページがそのようなサブ・コミュニティに分かれてまとまっていて、他のサブ・コミュニティとの間のネットワークが疎になっている様子である。直径 (Diameter) も小さいので平均パス長 (Avg.Path\_length) も短いものになっているが、相関係数としては高いものにな

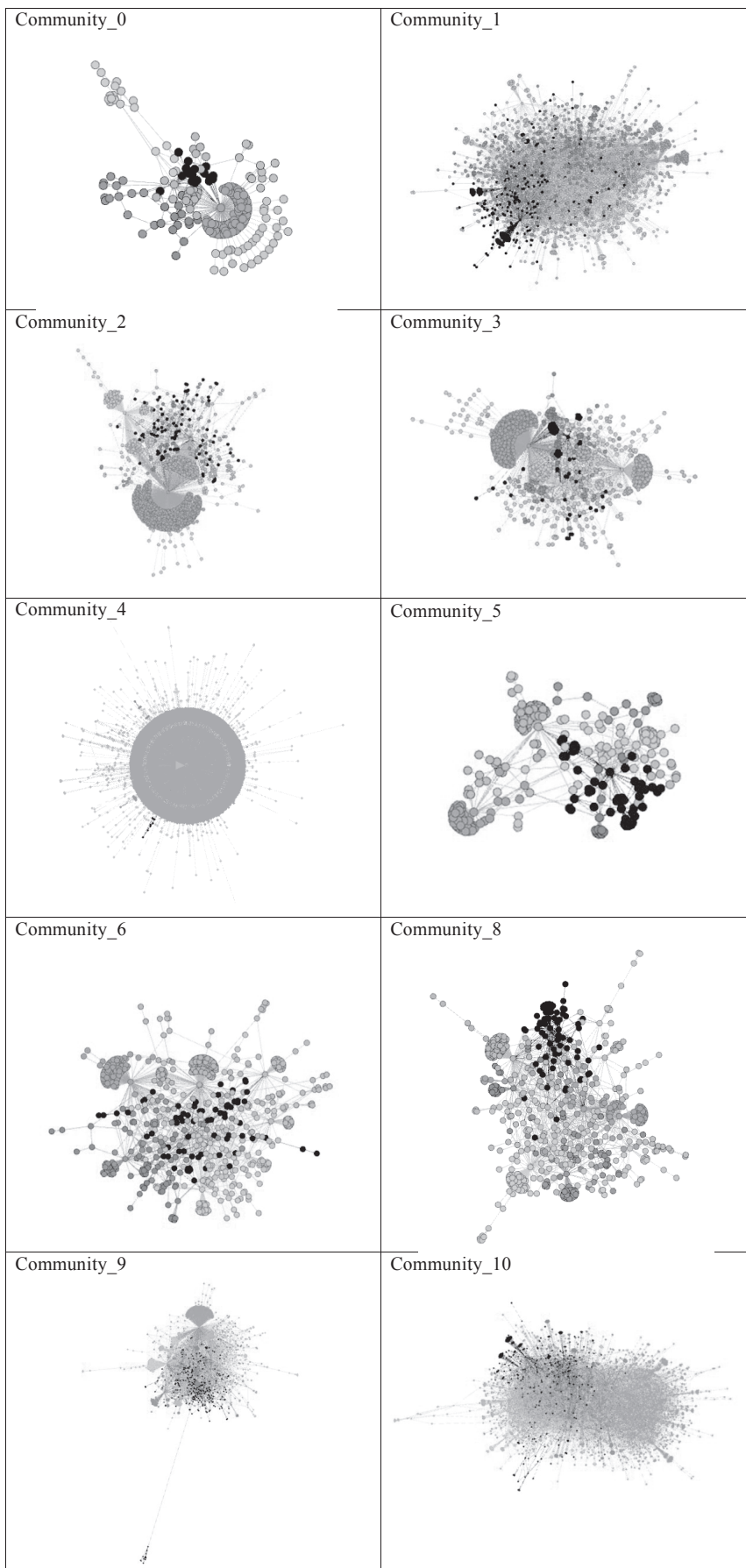


Fig. 7 コミュニティ毎のネットワーク・グラフ

Table 6 ネットワーク統計量の相関係数行列

	Link_nobs	Avg. Degree	Density	nobs_Sub_Community	Avg. Cluster_coeff.	Diameter	Avg. Path_length
OP_nobs	0.636	0.080	-0.692	0.871	-0.147	-0.410	-0.413
Link_nobs		0.816	-0.510	0.220	0.549	-0.181	0.072
Avg. Degree			-0.156	-0.363	0.796	0.137	0.466
Density				-0.449	-0.179	-0.099	0.106
nobs_Sub_Community					-0.547	-0.502	-0.625
Avg Cluster coeff.						0.225	0.378
Diameter							0.771

注1) 相関係数が±0.7以上のものと、-0.5以下のものには網掛けを施してある。

っていない。

幾つかのサブ・コミュニティにおいても、度数中心性が高いオーナー・ページが存在していることが分かったが、そのようなオーナー・ページがネットワーク上でどの様な重要性を持っているのかは分からない。そこで、各サブ・コミュニティにおけるハブとオーソリティーの関係を見てみる (Fig. 8)<sup>20</sup>。ネットワーク・グラフ (Fig. 7) と併せて考慮してみると、ネットワーク構造は判然としないながらもハブとオーソリティーの関係が整合しているもの (コミュニティ 1,8,10)、オーソリティーのスコアに偏ってオーナー・ページが分布するもの (コミュニティ 0,3)、逆にハブに偏って分布するもの (コミュニティ 2,5,6,9)、ハブもしくはオーソリティーのいずれかのスコアだけが高いものと、いずれも低いスコアによる実質的に3つのタイプしかないもの (コミュニティ 4) に分類できるだろう。

以上、コミュニティに分割してみたところ、全オーナー・ページでネットワークを観察したよりは見出せるものがあつた。全般的に各コミュニティでも密度 (Density) は低いものの、サブ・コミュニティが形成されており、平均パス長 (Avg.Path\_length) も短い。これらの特徴からは、スモールワールド性を備えているように思えるが、投資SNSの特質を考慮に入れば、もしスモールワールド性があつたとしても見かけ上のことに過ぎないかも知れない (脚注13参照)。各コミュニティのネットワーク・グラフを厳密にカテゴリー分けすることはできないが、大きく3種類に分けられ、さらにハブとオーソリティーの統計量を用いれば4種類に分類できる余地があることが分

<sup>20</sup> ページランクでも個々にオーナー・ページの重要性を計測することはできるが、ここでは入リンクと出リンクのバランスについて把握することを目的とするため、ハブとオーソリティーを用いた。

かった。ただし、現段階では分類が仮に論理整合的にできたとしても、なぜそのようなネットワーク構造に形成されるのかという原因については未解である。

## 6. 今後の課題

本稿では、投資SNSにおけるネットワーク構造の分析を行った。全オーナー・ページ (第2次サンプル・セット) では、ネットワーク統計量を算出してみても判然としなかったが、コミュニティ (サブ・ネットワーク) に分割してみたところ、幾つかのカテゴリーに分類できそうなことが分かった。しかし、なぜ各カテゴリーに分類できそうなのかの原因については不明なままである。加えて、各コミュニティ内における特徴的なオーナー・ページについての分析にも踏み込んでいない。これらが本稿の延長線上にある分析課題である。

ところで、今回分析対象とした投資SNSなどに対するページ・オーナーの参加動機や目的は様々だろうが、SNSのテーマが投資である以上、投資収益率の向上に資する情報の収集、発表、交流が根底にあると推測される。このため、本稿で行ったような社会ネットワーク分析も、ページ・オーナーの投資収益率に貢献しているのか否かを明らかにできてこそ意味を持つものだろう。この点が、本稿の延長線上にある分析課題が済んだ後に、考慮して分析する必要がある研究課題である。

今後の分析では、ネットワーク構造のより詳しい特徴を抽出するとともに、それがページ・オーナーの投資成果にどう結びついているのかまでを分析する。そこで、各ページ・オーナーの投資銘柄やその投資収益率、運用スタイルなどの情報も含めての分析を行いたい。また、今回は当該ページ・オーナーに対する他の

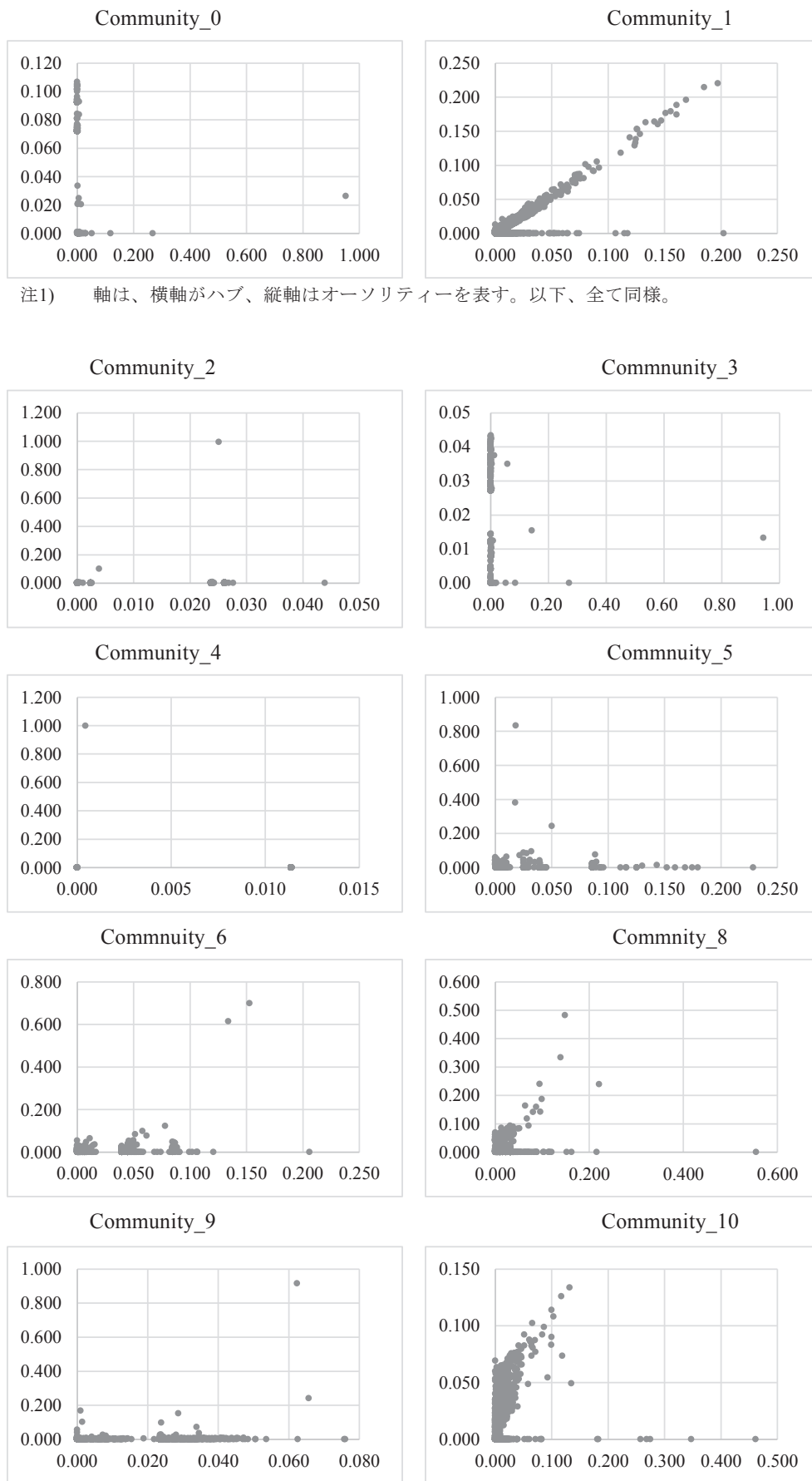


Fig. 8 コミュニティ毎のハブとオーソリティーの相関

ページ・オーナーからのフォロワー (Followers) のみをネットワーク分析の対象とした。他にも、当該ページ・オーナーの他のページ・オーナーに対するフォロワー (Followings) や、何かしらの理由に基づき注視している他のページ・オーナー (Respect), ブログや投資行動に対する他のページ・オーナーからのコメントなどにより形成されるネットワークもある。これらを併せて分析することにより、投資SNSにおけるネットワークと投資成果についての関係について解明したい。

### 参考文献

- 飯田恭弘, 岸本康成, 藤原靖宏, 塩川浩昭, 鬼塚真 (2014) 「大規模グラフ構造データからのコミュニティ抽出と重要度計算—高速化への取組みと応用—」, 人工知能, 第29巻
- 石川智康, 松尾和洋 (2012) 「SNSコミュニティの社会ネットワーク分析とその応用」, 情報処理学会第74回全国大会
- 松尾豊, 安田雪 (2007) 「SNSにおける関係形成原理—mixiのデータ分析—」, 人工知能学会雑誌
- 丸井淳己, 加藤幹生, 松尾豊, 安田雪 (2010) 「mixiのネットワーク分析」, 情報処理学会第72回全国大会
- 米山徹幸 (2014) 「IRとソーシャルメディアの進展」, 証券経済研究, 第85号
- Bagnoli, Beneish and Watts (1999) , “Whisper Forecasts of Quarterly Earnings per Share” , *working paper*
- Blondel, V. J. Guillaume, R. Lambiotte and E. Lefebvre (2008) “Fast unfolding of communities in large networks” , *Journal of Statistical Mechanics: Theory and Experiment* (10) ,
- Brandes, U. and T. Erlebach (1998) , “*Network Analysis*” , Springer
- Chen, Hl, P. De, Y. Hu, and B. Hwang (2016) , “Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media” , *working paper*
- Das and J. Sisk (2003) ,” Financial Communities” , *working paper*
- Giannini, R., P. Irvine and T. Shu (2019) ,” The Convergence and Divergence of Investors’ Opinions around Earnings News: Evidence from a Social Network” , *Journal of Financial Markets*, Vol.42
- Jacomy, J., T. Venturini, S. Heymann and M. Bastian (2014)” ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software” , *PLOS ONE*, Vol.9, Issue 6
- Konke, D. and S. Yang (2008) “*Social Network Analysis 2<sup>nd</sup> ed.*” , Sage publications
- Langville, A. and C. Meyer (2006)” *Google’s PageRank and beyond*” , Princeton Univ. Press, (邦訳 岩野和生, 黒川利明, 黒川洋, 訳『Google PageRankの数理』, 共立出版, 2009年
- Missaoui, R. and I. Sarr (2014) , “*Social Network Analysis – Community Detection and Evolution*” , Springer
- Ozsoylev, H., J. Walden, M. Deniz Yavuz, and R. Bıldik (2011) “Investor Networks in the Stock Market” , *working paper*
- Scott, J. (2000) , “Social Network Analysis” , Sage
- Zaima, J. K. and M. A. Harjoto (2006) , “Conflict in Whispers and Analyst Forecasts: Which One should be Your Guide?” *working paper*

Received date 2020年1月10日

Accepted date 2020年1月27日