

スモールワールドとチャンス発見

Small World and Chance Discovery

松尾 豊
Yutaka Matsuo

産業技術総合研究所サイバーアシスト研究センター
Cyber Assist Research Center, National Institute of Advanced Industrial Science and Technology
y.matsuo@carc.aist.go.jp, <http://www.carc.aist.go.jp/~y.matsuo/>

Keywords: small world, chance discovery, social network, scale-free network, cluster

1. ま え が き

スモールワールドというグラフ構造が近年着目を集めている。スモールワールドとは簡単にいうと、ノードがクラスタ状に集まっているにも関わらず、ノード間のパスが短いという特徴を持つグラフである。自然界や人工物、社会におけるさまざまなネットワークがスモールワールドであり、実はチャンス発見とも密接な関連がある。スモールワールドは流行の伝播しやすいネットワークであり、適切な位置にアクションを起こせば効果的な影響を与えられるネットワークである。

本解説では、スモールワールドの由来から、スモールワールドとは何であるか、その性質や分類について紹介し、その後、チャンス発見との関連について述べる。スモールワールドとチャンス発見に共通するつながりの重要性を紐解いていこう。

2. Milgram の実験

スモールワールドは、もともとと社会心理学の分野で生まれた概念である。旅先やパーティーなどで、初めて会った人と思いがけず共通の知人を発見して「せまい世界ですね。(It's a small world!)」と驚いた経験は誰もがあろう。これがスモールワールド現象である。この現象はなぜこんなに頻繁に起こるのだろうか？

スモールワールドを定式化すると、「世界中の任意の2人 X と Z がどのくらいの確率で共通の知人 Y を持っているか」となる。より一般的に、X が a を知っていて、a が b を知っていて、...、y が Z を知っているという知人の鎖 (chain) を考えることもできる。では、任意の2人は平均何人を仲介すればつながっているのだろうか。

S. Pool と M. Kochen は、スモールワールドの理論的なモデルを作ることを試みた。まず、彼らは普通の人には何人くらい知人がいるのかを調べるため、様々な被検者が100日間に連絡をとった人数を数え、平均500人程度であるとした。米国のすべての人が米国の

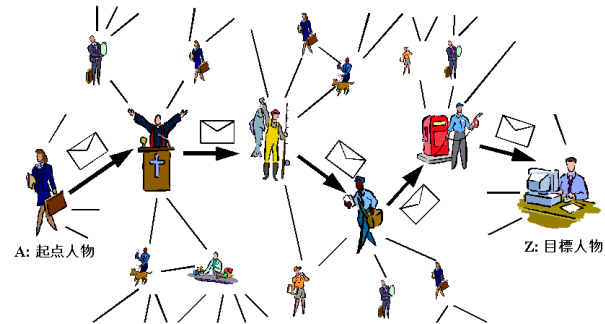


図1 起点人物から目標人物へのパス

中からランダムに500人の知人を持つとする単純なモデルでは、任意の2人が互いに知人である確率は20万分の1になる。そして、平均で、2人から3人を介せば、任意の2人はつながることになる。

しかし、知人の関係は本当にランダムに生成されるのだろうか。現実には、知人の知人は自分の知人と重なっている場合も多いだろう。貧乏な人の知人はどちらかというとな貧乏なことが多く、お金もちが話す相手は多くの場合、お金もちかもしれない。このような社会構造を考慮するためには、実際に確かめてみる必要がある。

そこで、1960年代、著名な社会心理学者である Stanley Milgram は、手紙を転送する実験を行った [Milgram 67]*1。ここで、A という人物から Z という人物へのパスを考えよう (図1)。A を起点人物 (start person)、Z を目標人物 (target person) と呼び、ランダムに選ぶ。起点人物は、目標人物を知っていればその人に手紙を転送する。目標人物と直接知り合いでない場合、自分よりも目標人物をよく知っていそうな知人 (ファーストネームで呼びあうような知人) に手紙を転送する。次々に手紙を転送していくことで、結果的に目標人物に到達することができる。

*1 Milgram は、自分の経験や思いつきから、とにかく実験による検証を行った人物であり、なかでもアイヒマン実験と呼ばれる「なぜ人は権威に服従するのか」という残酷な実験で有名である。

Milgram は起点人物をカンザス州ウィチター（1 回目の実験）とネブラスカ州オマハ（2 回目の実験）から 160 人選んだ。1 回目の実験の目標人物はケンブリッジに住む神学校の学生の妻であり、2 回目の実験では、目標人物は、ボストンに勤務し、マサチューセッツ州シャロンに住む株主買人である。

起点人物には、以下のものが入った封筒が送られる。

- 目標人物の名前とその人の情報。
- 目標人物に到達する方法！目標人物を個人的に知らないのであれば、彼と連絡をとろうとしないで下さい。そのかわり、目標人物を自分より良く知って、いそうな知人にこの封筒を送ってください。」
- 転送に参加した人のリスト。
- 15 枚の追跡カード。受け取った人は、追跡カードの一枚に書き込んで、Milgram まで送り返す。

さて、カンザスから始まった鎖が本当にマサチューセッツまで届くのだろうか？封筒を何人かの起点人物に送った 4 日後、神学の講師が通りで目標人物に「君あてだよ」と封筒を手渡した。封筒の中の名簿には、カンザスの小麦農家からカンザスの牧師に、そして、ケンブリッジの神学の講師を経由して、目標人物に渡ったことが書かれてあった。つまり、起点人物と目標人物を結ぶ仲介者の数は 2 であった。このケースは最も短い鎖のひとつであったが、鎖の長さはだいたい 2 から 10 で中間値は 5 であった。

「米国のどんな人もたった 6 人の向こう側にいる」という Milgram の実験の結果^{*2}は大きな驚きをもって迎えられ、その後、“six-degrees of separation”という言葉は米国で広く知られるところとなった。“Six-degrees of separation”（邦題：私に近い 6 人の他人）という映画も 1993 年に公開されている [Guare 90]。

また、「ベーコン数」や「Erdős 数」という遊びもよく知られている。ベーコン数は、映画俳優ケビン・ベーコンと共演した人は 1、共演した人と共演した人は 2 と数えるもので、ベーコン数が大きい俳優を見つけるのは非常に難しいと米国の話題になった。これを統計的に検証したデータもあり、50 万人以上の俳優についてのベーコン数の平均は 2.9 である^{*3}。Erdős 数は、数学者のコミュニティの中のゲームであり、多作な数学者 Erdős と共著関係にある人は 1、共著した人と共著した人は 2 と数えるもので、その平均値は 1.5 である^{*4}。いずれも、直観的な世界の大きさに比べて、その

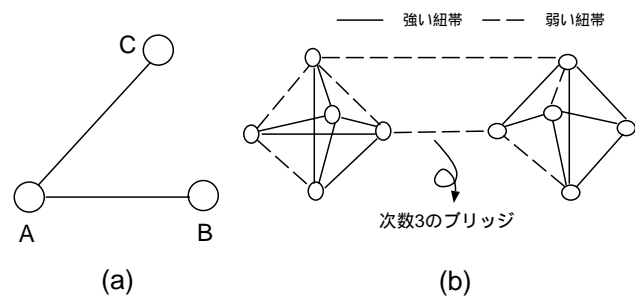


図 2 3 人の関係とブリッジ

つながりによる世界の小ささを示す数字である。

3. 強い紐帯、弱い紐帯

社会学者である Mark S. Granovetter は、2 人の人間の間を強い紐帯 (strong tie)、弱い紐帯 (weak tie) に分けて考えた [Granovetter 73]。

任意に選ばれた 2 人を考え、A と B としよう。さらに $S = \{C, D, E, \dots\}$ を A と B の少なくとも一方と紐帯をもつ人々とする。もし A と B の紐帯が強ければ、S 中の人 A と B 両方と紐帯を持つ確率が大きくなるだろう。この友人関係のオーバーラップは、紐帯がないときには小さく、紐帯が強いときには大きく、紐帯が弱いときにはその中間である。

A-B と A-C の紐帯が強ければ、A は B と C とともに多くの時間を過ごすはずであり、したがって B と C は知り合う確率が高い。したがって、A-B と A-C が強い紐帯ならば B-C は関係を持ちやすく、A-B と A-C が弱い紐帯ならば B-C が関係を持ちにくい。多くのソシオグラム (行為者間の関係を示したグラフ) で、「A-B と A-C は関係があるが B-C は関係がない」という 3 人間の関係 (図 2(a)) が起こる確率は、ランダムな期待値よりも小さい。

2 点間 A-B に、その紐帯以外のパスがないとき、A-B をブリッジという。A につながるどんな人から B につながるどんな人への情報も A-B のブリッジを経由しなければならない。大きなネットワークでは、ある紐帯が 2 点間のただひとつのパスとなることは滅多に起こらないので、その紐帯を除いた 2 点間の最短パスの長さ n を使って、次数 n の局所的ブリッジとよぶ (図 2(b))。次数が大きければ大きいほど、多くの人にとって唯一の効率的なパスであるという点で、重要となる。このようなブリッジは、上記の性質から強い紐帯ではなく、弱い紐帯である。

普及 (Diffusion) の研究で有名な Rogers は、何が普及するとき革新者 (innovator) がまずそれを採用し、初期少数採用者、前期多数採用者、後期多数採用者、採用遅滞者が次々に追従するというモデル [Rogers 62] を提唱した。このような普及は、弱い紐帯を通じて起

^{*2} Milgram の実験で明らかになったのは、「カンザス州やネブラスカ州の 160 人からマサチューセッツ州の目標人物までのパスの中間値が 5 であった」ことだけであり、「米国の任意の 2 人が 5 人を介してつながっている」というのは拡大解釈しすぎだという議論もある。その後、本格的な実証実験はされず、それを Web 上で確かめようというプロジェクト (<http://smallworld.sociology.columbia.edu/>) も行われている。

^{*3} <http://oracleofbacon.org/>

^{*4} <http://www.acs.oakland.edu/grossman/erdoshp.html>

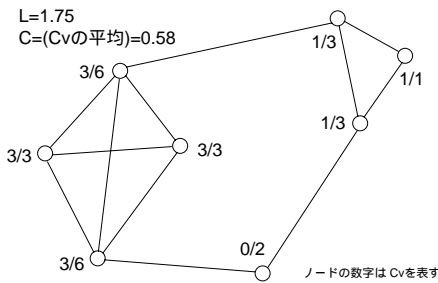


図3 L と C の計算の例.

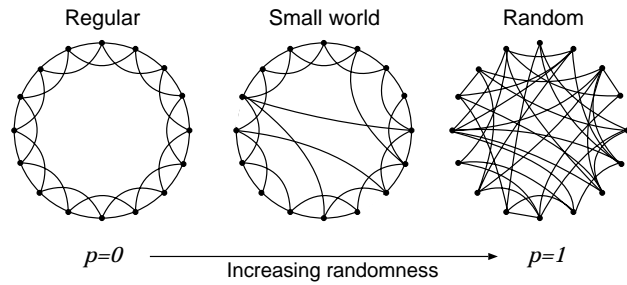


図4 regular ring lattice のランダムなつながりかえ.

こる．ある人が噂を自分の近い友人に話すとしよう．その友人も同じように噂を広めると、強い紐帯は友人を共有するので、多くの人が噂を何度も聞くことになる．したがって、強い紐帯を経由する噂は、弱い紐帯を経由する噂より広がりが限られてしまう．このような意味で、弱い紐帯には強い力がある．逆にいうと、たくさんの弱い紐帯を持っている人は、何かを普及させるのに最も適しているというのが Granovetter の主張であった．

4. Watts の定式化

スモールワールドは長らく社会心理学的な研究対象であったが、1998年、Cornell 大学で応用力学の博士課程の学生であった Duncan Watts はスモールワールドをグラフにおける特徴量として鮮やかに定式化し、Nature 誌に発表した [Watts 98]．以来、スモールワールドは計算機科学の分野で急速に注目を集めるようになった．

Watts が定式化したグラフにおける 2 つの特徴量とは、以下である．

- L (characteristic path length) : グラフ中のすべてのノードの組についての最短パスの長さの平均．
- C (clustering coefficient) : グラフ中のノード v が k_v 個のノードと隣接しているとき、 k_v 個のノード間に存在する $k_v C_2 = k_v(k_v - 1)/2$ 本のエッジに対して、実際に存在するエッジの割合を C_v とする．すべてのノード v について C_v の平均をとったものが C である．人間関係のネットワークでいうと、共通の知人を持つ 2 人が直接の知人である確率を表す．

計算の例を図 3 に示す．ノード数、エッジ数が一定である場合、 C が大きいグラフは、近傍同士でのエッジが多いので L は大きくなる傾向がある．また、 C が小さいランダムなグラフでは L は小さい．したがって、 L と C はある程度対応しているのだが、その中間領域では、 L が小さいにも関わらず C の大きなグラフが存在する．それがスモールワールドである．

図 4 は、 β -Graph と呼ばれるスモールワールドのモ

デルである [Watts 99a]．まず、 n 個（ここでは 16 個）のノードがそれぞれ近傍の k ノード（ここでは 4 ノード）にエッジが張られている規則的なグラフ (regular ring lattice) を考える．次に、各エッジに対して、確率 p でエッジのつながりかえを行う．エッジのつながりかえとは、ランダムに選んだノードへとエッジを張りかえる操作である．これを全てのエッジに対して 1 回ずつ適用する． $p=0$ のときは、全くつながりかえを行わないことに相当し、 $p=1$ のときは、すべてのエッジをつなぎかえたランダムグラフとなる．この中間の値では、近傍を結ぶ規則的なエッジとランダムな長いエッジが混在するグラフが得られる．このとき、 C が大きく L が小さいスモールワールドとなる．ランダムなエッジがショートカットの役割を果たし、急激に L を減少させるのである．このような、近いノードを結ぶローカルなエッジと、少数の長いショートカットというのがスモールワールドの典型的な例である．

Watts によるスモールワールドの正確な定義は以下のようにになっている [Watts 99b] ．

Definition 1 スモールワールドは、ノード数 n が大きく、疎に結合した分散したグラフ ($n \gg k_{max} \gg 1$) であり、 L がランダムグラフと近く ($L \sim L_{rand}$)、 C がランダムグラフに比べて極めて大きい ($C \gg C_{rand}$) ものである (ただし、 L_{rand} 、 C_{rand} は同じノード数、エッジ数のランダムグラフにおける L と C の値であり、 k_{max} はグラフ中において、ノードから出ているエッジ k の最大値である．)

5. さまざまなスモールワールド

Watts は、映画俳優の共演グラフ、送電網、神経回路網を例にとりて C と L を調べた．映画俳優の共演グラフは、インターネット上の映画のデータベース^{*5}から作ったものであり、ノードは映画俳優を、エッジは同じ映画に一度でも共演した関係を表す．送電網は、米国西部の送電網で、ノードは発電所、変圧器、変電所を表し、エッジは高電圧送電線を表す．また、神経回

*5 <http://www.us.imdb.com/>

表 1 スモールワールド構造を持つ様々なグラフの L, C [Watts 98]

	n	k	L	L_{rand}	C	C_{rand}
Film actor	226,000	61	3.65	2.99	0.79	0.00027
Power grid	4,941	2.94	18.7	12.4	0.080	0.005
<i>C. elegans</i>	282	14	2.65	2.55	0.28	0.05

路網は、*C. elegans* という線虫の神経ネットワークであり、エッジは 2 つのニューロンをつなぐシナプスもしくはギャップ結合を表す。

表 1 は、これら 3 つのグラフに対しての特徴量である。いずれのグラフにおいても、 L は L_{rand} と同程度（もしくは少し大きい）であるが、 C は C_{rand} よりかなり大きく、スモールワールドの特徴を備えていることが分かる。

その後、さまざまなネットワークがスモールワールドの性質を持っていることが報告されている。WWW(ページをノード、ハイパーリンクをエッジとする) は、直径がおよそ 19 から 21 クリック程度のスモールワールドである [Adamic 99, Albert 99, Kaiser 99]。また、インターネット [Faloutsos 99]、食物連鎖のグラフ*6、科学論文の共著関係 [Newman 01]、生物の代謝ネットワーク [Jeong 00]、航空路線 [Amaral 00]、自然言語文書における語の共起関係 [Ferrer 01, 松尾 01]、JAVA の class diagram [Valverde 02] などさまざまなネットワークがスモールワールドであると報告されている。また、Walsh は、探索問題の中でも特に難しい問題は変数間の依存関係がスモールワールドの特徴をもっていることを示した [Walsh 99]。

このように、自然界、人工物、社会におけるさまざまネットワークにスモールワールドの性質があることが分かり、多くの研究者の興味を刺激している。

6. スモールワールドの 3 タイプ

Albert-László Barabási は、Web ページのリンク関係において、各ページがどのくらい他のページからリンクされているかを調べた。ほとんどのページは数個以下のリンクしか集めないが、Yahoo! などのごく少数のページは 100 万個もの大量のリンクを集める。そして、他のページからのリンク数が k 本であるページの数に関して、面白い現象を発見した。ページ数はべき分布 (power law)

$$P(k) \sim k^{-\gamma}$$

(ただし γ は定数でこの場合 2.1) に従うのである。つまり、 k が小さいページは多く、 k が大きいページは少ないのだが、 k が非常に多いページも存在し得るのである。そして、このようにノードの次数の分布がベ

*6 ただし、[J.A 02] には、 C がランダムグラフと同程度に小さいためスモールワールドではないとしている。

キ法則に従うネットワークを scale-free (尺度のない) ネットワークと名付けた。極端に多くのエッジを持つノードも存在し、系に特徴的なスケールを決めることができないためである [Barabási 02]。

Amaral らは、この scale-free ネットワークを拡張することで、スモールワールドに 3 つのタイプがあることを示した [Amaral 00]*7。

- scale-free ネットワーク。極端に多くのエッジを持つノードも存在する。
- ノードの次数が、先端にカットオフのあるべき分布にしたがう broad-scale ネットワーク。 $P(k) \sim k^{-\gamma}$ だが、ある k 以上では頭打ちになっており、極端にエッジの多いノードは存在しない。
- ノードの次数が急激に減退する single-scale ネットワーク。例えば、 k の分布が平均値を中心とするポアソン分布にしたがうもの。大きな k に対しては $P(k) \sim e^{-k}$ であり、エッジの多いノードは存在しない。

Amaral らのデータの範囲においては、WWW、インターネット、論文の引用ネットワーク、食物連鎖は scale-free であり、共演関係のネットワークは broad-scale のネットワークである。また、送電網や航空路線のネットワーク、神経回路網やポリマー鎖の構造空間、人間関係のネットワークは single-scale ネットワークであった。

さて、このような違いはなぜ生まれるのだろうか。Barabási によると、scale-free ネットワークは、

- 新しくノードが追加される成長するネットワークであること
 - 新しいノードは、degree の大きい (エッジを多く張られた) ノードにエッジを張る選好があること
- という 2 つの条件が満たされるときに現れる [Barabási 99]。例えば、Web ページを作ったとき、すでにある有名なサイトにリンクを張ることが多いだろう。論文の引用をするにも、すでに多く引用されている有名な論文を引用することが多いだろう。このように、次数が多いノードがさらに優先的に次数が多くなる、つまり「金持ちはもっと金持ちに」なるネットワークは scale-free ネットワークを構成する。

一方、broad-scale なネットワークは、ノードが年を取るという要因によって現れる。つまり、有名な俳優もいつか引退しなければならない。たくさんのノードとつながったノードであっても、最終的にエッジを集めるのをやめ、そのままの形でネットワークに残ることに

*7 scale-free なネットワークは、 L が小さく C が大きいという特徴があるが、これをスモールワールドと呼ぶかどうかに関しては今のところ意見が統一されていないようである。Watts は [Watts 99b] の中では、スモールワールドの定義として、分散したグラフと記述しており、scale-free のネットワークは含めないことを意図しているが、スモールワールドとして scale-free なネットワークを含めている論文も多い。

なる．そして、scale-free の分布に制限を与え、broad-scale ネットワークを構成する．

また、single-scale なネットワークはノードにエッジを張るコストやノードあたりのエッジの数の制限が要因となって現れる．極めて少数のハブ空港があった方が効率的であるが、それほど多くの滑走路は一ヶ所に集められない．このような場合に single-scale ネットワークを構成するのである．

7. 流行の伝わりやすいスモールワールド、攻撃に対するロバスト性と脆弱性

Watts は、スモールワールドでは流行や伝染病が伝わりやすいことも示した．彼が用いたモデルでは、時間 $t=0$ でひとりの人が病気になる．病気になった人は、近傍の人に確率 r で病気を伝染し、ネットワークから除かれる．(免疫ができる、または死ぬ．)最終的に、病気の流行が終わるか、病気が広がって人がいなくなるまでこれをくり返す．

このモデルを用いたシミュレーションの結果、規則的なグラフに対するランダムなつながりかえの割合 p (図 4 参照) を高くすると、感染力の弱い病気でも伝わりやすくなることが分かった．また、全ての人口に感染が広まるまでの時間は L と対応していることが分かった．つまり、 L が小さいスモールワールドでは、伝染は短時間で広まるのである．

他にも、スモールワールドの性質として、ダメージに対して強いことが挙げられる．ランダムにノードを除去しても、グラフにおけるノード間の距離はそれほど変わらず、影響は少ない．しかし興味深いことに、scale-free ネットワークにおいて、エッジが集中しているノードを選択的に攻撃した場合には、ダメージが大きいのである [Albert 00, Jeong 00]．例えば、食物連鎖のグラフでは、ある種(あるノード)が絶滅するとそれを捕食している種(グラフ中で隣接したノード)が絶滅することもある．絶滅させる種をランダムに選んだ場合、グラフ全体に対する影響はほとんどない [Solé 00]．しかし、たくさんのエッジを持つ種が絶滅すると、その影響は非常に大きく、たくさんの種が絶滅してしまう．つまり、ランダムに攻撃に対するロバスト性と、狙いをつけた攻撃に対する脆弱性の両面がスモールワールドの特徴である．

8. なぜスモールワールドが現れるのか

では、なぜスモールワールドが自然界や人工物に遍在するだろうか．Marchiori らは、ネットワークにおける情報伝達という点からスモールワールドを捉え、局所的な情報伝達効率と大域的な情報伝達効率の両面で有利であるから、スモールワールドが出現する [Marchiori

00] という指摘をしている．

また、Mathias らは、グラフの連結性の最大化とエッジを維持するコストの最小化というトレードオフによってもたらされると指摘している [Mathias 01]．ここでは、ノード間の距離として、物理的な距離(例えば航空路でいえば、2 地点間の物理的な距離)とグラフ上の距離(飛行機を何回乗り継がなければいけないか)の 2 つを考える．信号や情報、物質などの伝達効率からは、グラフ上の平均距離 L は短い方が望ましい．一方、エッジを張るコストはノード間の物理的な距離の平均 W に比例すると考えると、 W は短い方が望ましい． L と W はトレードオフの関係にあるが、[Mathias 01] では関数

$$E = \lambda L + (1 - \lambda)W$$

(ただし、 λ は 0 以上 1 以下の重みを表すパラメータ) を想定し、これを最小化するようなエッジの張り方を求めると、 λ が 0 のときには regular lattice が、 λ が 1 のときはランダムグラフが得られ、その中間ではスモールワールドが出現することを明らかにした．多くのグラフでは、連結性 (L が小さいこと) とコストの最小性 (W が小さいこと) という相反する両方の要求があるため、スモールワールドが現れるのである．

また、Valverde らは、ソフトウェアデザインの結果としての JAVA の class diagram がスモールワールドであることを示している．あまり大きなクラスを作てはいけないという整合性の調整の結果としてスモールワールドが現れる [Valverde 02]．また、自分の知り合いを人に紹介するという局所的なインタラクションによってスモールワールドが現れると報告する論文もある [Davidsen 02]．

9. スモールワールドとチャンス発見

さて、本節では、スモールワールドとチャンス発見について述べる．まず、チャンス発見において重要なツールのひとつとして用いられている KeyGraph [大澤 99] について簡単に紹介する．

9.1 KeyGraph

KeyGraph [大澤 99] は、もともと自然言語文書を対象にした重要語抽出法である．一般的なバスケットデータにも適用することができ、販売データ、地震発生の時系列データ、Web ページのリンクの共起関係のデータなど、さまざまなデータに用いられている．

そのアルゴリズムは、グラフ構造に基づき、大きく 2 段階から構成される．

- (1) 頻出アイテムをノード(黒ノード)として取り出す．

- (2) 共起関係*⁸が強いノード間にエッジを張る．
- (3) 不要なエッジを取り除き、クラスタ（連結部分グラフ）を抽出する．

次の段階では、こうして得たクラスタに着目する．

- (4) 2つ以上のクラスタと共起関係の強いアイテムを赤ノードとして抽出する．クラスタとの共起とは、クラスタに含まれるいずれかのノードとの共起である．

赤ノードは、複数のクラスタの間に位置する頻出でないノードであり、自然言語の文書の場合には複数の概念をつなぐ文書の主旨であり、POS データの場合には複数の購買要因の間に位置する商品であり、Web ページの場合には複数のコミュニティの中間に位置するページとなる．

(1)(2)の処理は、バスケットデータからのグラフの構成としては一般的な手法である*⁹．しかし、KeyGraph の場合には、明示的にクラスタの抽出 (3) と、クラスタ間に位置するノード (4) の抽出まで行っているところが特徴的である．

9.2 スモールワールドとチャンス発見：共通点と相違点

チャンス発見では、特に、クラスタやコミュニティをつなぐ部分、境界となる部分に着目し、将来の動きを理解したり、有効なアクションを探る研究が行われてきた．3 節で述べた Granovetter の研究では、弱い紐帯のもつ情報伝達における重要な役割について述べられており、4 節のスモールワールド構造においても、ショートカットがスモールワールドの成立に大きな役割を果たす．このように、スモールワールドとチャンス発見のいずれにおいても、クラスタをつなぐノードやショートカットはそのネットワークごとに特有の重要な意味を持つ．両者の差異のひとつは、チャンスディスカバリーはその効果を静的に測るだけでなく、それを積極的に人間が利用していくところにある．

5 節では、スモールワールドがさまざまな現象の中に遍在することを述べた．一方、KeyGraph もさまざまなデータに対して有効な汎用性を示している．[Ohsawa 02]．これは、多くのネットワークが、自然な性質として複数のクラスタとその間の相互関係という形で機能していることが多いためであろう．8 節で述べたように、ノード間のローカルな相互作用、そしてグラフ全体として何らかの機能を果たすような場合に、スモールワールドが出現する．そういったデータやネットワークに対しては、KeyGraph による解析が有効であると考えられる．

もちろん、スモールワールドでないネットワークが

あるように、KeyGraph による分析が適切でないようなデータやネットワーク構造が存在するだろう．6 節では、スモールワールドの 3 つのタイプについて述べた．Watts が当初スモールワールドとして想定していたのは、いくつかのクラスタがつながれた single-scale ネットワークである．そして、KeyGraph が主に想定しているのもこの single-scale ネットワークである．しかし、解析するデータによっては、scale-free なネットワーク、broad-scale なネットワークの性質をもつこともある．このような場合は、通常、KeyGraph では、頻度の多すぎるアイテムを取り除く操作を前処理として行う．これは、次数の多いノードを取り除くことによって多くのノードがつながれてしまうのを避け、クラスタの抽出処理がうまく働くようにするためである．このような前処理をインタラクティブに行うことで、single-scale でないタイプのネットワークに対しても、適切な処理が可能となっている．

7 節の前半では、流行や情報の流れという点におけるスモールワールドの特徴を述べた．チャンス発見においても、コミュニティの内外における情報の流れは重要な研究テーマのひとつである．流行や情報が流れやすいコミュニティをどう作るか、逆にグラフの構造からコミュニティを特徴づけることができるか [Ohsawa 02] といったテーマが研究されている．

7 節の後半では、攻撃に対するグラフの耐性についての文献を紹介した．スモールワールドでは、ランダムな攻撃に対してはロバスタだが、狙いをつけた攻撃には脆い．チャンス発見でも、同様の考えに基づいたデータ解析と意思決定のプロセスを重視している．つまり、クラスタをがっちりと構成しているノードに対しては、たとえ人間がアクションを行っても他のノードとの関係を変えるのは難しく、大きな効果を生み出すのは難しい．しかし、クラスタの中間にあるノードやエッジに対して人間がアクションを起せば、両方のクラスタに対して大きな影響を与えられる可能性がある．このようなノードを KeyGraph は抽出しようと試みており、それをスモールワールドにおける重要なノードおよびエッジという観点からより直接的に抽出しようとする試み [松尾 01] も行われている．

10. おわりに

“It’s a small world!” という現象、そして Milgram の実験から始まったスモールワールドは、自然界、人工物、社会における多くのネットワークに共通する性質として、大きく注目を集めている．誌面の都合で取り上げなかったが、最近では統計物理とスモールワールドの関連も数多く研究されている．

それとは全く別に、データマイニングという地点からスタートしたチャンス発見であるが、その考え方や

*8 共起頻度や Jaccard 係数などによって測る．

*9 例えば [Guha 00, Becker 96, Strehl 00] など

思想は、スモールワールドと共通した部分が多い。両者の関連を考えることは、チャンス発見の理解を深める上でも大変重要であるし、チャンス発見の研究に参考になる部分もある。

ただし、スモールワールドは純粹に世界を抽象化した後のグラフの解析や諸性質の話であるので、比較的、誰がみても明らかなネットワークが対象となる。一方、チャンス発見は、さまざまな実世界のデータをどのような観点で取得し、どのように加工してグラフを作るか、さらに解析した結果を人間の意思決定にどのように生かすかといった、データから意思決定に関わるプロセス全体の研究である。この点はひとつの大きな立場の違いであろう。

◇ 参 考 文 献 ◇

[Adamic 99] Adamic, L. A.: The Small World Web, in *Proc. ECWL'99*, pp. 443–452 (1999).

[Albert 99] Albert, R., Jeong, H., and Barabási, A.: Diameter of the World-Wide Web, *Nature*, Vol. 401, No. 6749 (1999).

[Albert 00] Albert, R., Jeong, H., and Barabási, A.: Error and attack tolerance of complex networks, *Nature*, Vol. 406, pp. 378–382 (2000).

[Amaral 00] Amaral, L. A. N., Scala, A., Barthélémy, M., and Stanley, H. E.: Classes of small-world networks, *Proceedings of the National Academy of Sciences*, Vol. 97, No. 21 (2000).

[Barabási 99] Barabási, A.-L., Albert, R., and Jeong, H.: Mean-field theory for scale-free networks, *Physica A*, Vol. 272, pp. 173–187 (1999).

[Barabási 02] Barabási, A.-L.: 新ネットワーク思考, NHK 出版 (2002).

[Becker 96] Becker, S.: Mutual information maximization: Models of cortical selforganization, *Network: Computation in Neural Systems*, Vol. 7, pp. 7–31 (1996).

[Davidsen 02] Davidsen, J., Ebel, H., and Bornholdts, S.: Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks, *Physical Review Letters*, Vol. 88, No. 12 (2002).

[Faloutsos 99] Faloutsos, M., Faloutsos, P., and Faloutsos, C.: On Power-law Relationships of the Internet Topology, in *SIGCOMM*, pp. 251–262 (1999).

[Ferrer 01] Ferrer, R. and Sole, R. V.: The small world of human language, in *Proceedings of The Royal Society of London. Series B, Biological Sciences*, Vol. 268, pp. 2261–2265 (2001).

[Granovetter 73] Granovetter, M.: Strength of Weak Ties, *American Journal of Sociology*, Vol. 78, pp. 1360–1380 (1973).

[Guare 90] Guare, J.: *Six Degrees of Separation: A Play*, Vintage Books, New York (1990).

[Guha 00] Guha, S., Rastogi, R., and Shim, K.: ROCK: A Robust Clustering Algorithm for Categorical Attributes, *Information Systems*, Vol. 25, No. 5, pp. 345–366 (2000).

[J.A 02] J.A, D., R.J., W., and Martinez, N.: Small networks but not small worlds: unique aspects of food web structure, Technical Report Working Paper 02-03-10, Santa Fe Institute (2002).

[Jeong 00] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A. L.: The large-scale organization of metabolic networks, *Nature*, Vol. 407, pp. 651–654 (2000).

[Kaiser 99] Kaiser, J.: It's a Small Web after all, *Science*, Vol. 285, p. 1815 (1999).

[Marchiori 00] Marchiori, M. and Latora, V.: Harmony in the small-world, *Physica A*, Vol. 285, pp. 539–546 (2000).

[Mathias 01] Mathias, N. and Gopal, V.: Small worlds: How and why, *Physical Review E*, Vol. 63, No. 2 (2001).

[松尾 01] 松尾, 大澤, 石塚: Small World 構造に基づく文書からのキーワード抽出, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1825–1833 (2001).

[Milgram 67] Milgram, S.: The small-world problem, *Psychology Today*, Vol. 2, pp. 60–67 (1967).

[Newman 01] Newman, M. E. J.: The structure of scientific collaboration networks, *Proceedings of the National Academy of Sciences*, Vol. 98, No. 2, pp. 404–409 (2001).

[大澤 99] 大澤, ネルス E., 石塚: KeyGraph: 語の共起グラフの分割・統合によるキーワード抽出, 電子情報通信学会誌, Vol. J82-D-I, No. 2, pp. 391–400 (1999).

[Ohsawa 02] Ohsawa, Y.: Chance Discoveries for Making Decisions in Complex Real World, *New Generation Computing*, Vol. 20, No. 2 (2002).

[Rogers 62] Rogers, E.: *Diffusion of Innovations*, Free Press, New York (1962).

[Solé 00] Solé, R. V. and Montoya, J. M.: Complexity and Fragility in Ecological Networks, Technical Report Working paper 00-11-060, Santa Fe Institute (2000).

[Strehl 00] Strehl, A. and Ghosh, J.: A Scalable Approach to Balanced, High-dimensional Clustering of Market-baskets, in *Proceedings of the Seventh International Conference on High Performance Computing (HiPC 2000), 17-20 December 2000, Bangalore, India*, Vol. 1970 of *Lecture Notes in Computer Science*, pp. 525–536, Springer (2000).

[Valverde 02] Valverde, S., Cancho, R. F., and Solé, R. V.: Scale-Free Networks from Optimal Design, Technical Report Working paper 02-04-019, Santa Fe Institute (2002).

[Walsh 99] Walsh, T.: Search in a Small World, in *Proc. IJCAI-99*, pp. 1172–1177 (1999).

[Watts 98] Watts, D. and Strogatz, S.: Collective dynamics of small-world networks, *Nature*, Vol. 393, pp. 440–442 (1998).

[Watts 99a] Watts, D.: *Small worlds: the dynamics of networks between order and randomness*, Princeton (1999).

[Watts 99b] Watts, D. J.: Networks, Dynamics and The Small World Phenomenon, *American Journal of Sociology*, Vol. 105, No. 2, pp. 493–527 (1999).

〔担当委員： 〕

年月日 受理

— 著 者 紹 介 —

松尾 豊 (正会員)

1997年東京大学工学部電子情報工学科卒業。2002年同大学院博士過程修了。工学博士。同年より、産業技術総合研究所サイバースタディーズセンター勤務。仮説推論、重要語抽出、マイニング、ユーザモデリング等に興味がある。ユーザにとって価値の高い情報とは何だろうかというのが研究テーマである。情報処理学会、AAAIの各会員。