

The Structure and function of complex networks¹ Mark E. J. Newman

概要

この論文は、Newman による複雑ネットワークに関するサーベイ論文である。基本的に、ネットワークのトポロジーやノードの分布などネットワーク全体が示す特徴に着目しており、ネットワーク分析が主に対象とするそれぞれのノードの特性や位置づけには感心を払っていない。これは、Newman 自身がこの論文の中で、小さなネットワークではノードの位置づけや特徴が重要であるが、その規模が大きくなるとそれは些末な事だと言っていることから分かる。本論文の内容は、主にランダムグラフ、Small-world、成長するグラフのモデルの3つから構成されている。

1 Introduction

ネットワークとは、ノードとエッジから構成されている。そして、各ノードとエッジは属性を持つことがある。このようなネットワークをもつシステムは多く存在する。例えば、インターネットや、WWW、Social Networks of acquaintance or other connections between individuals、組織、network of business relations between companies、ニューラルネットワーク、血管網、郵便配達経路、論文の参照関係、など。

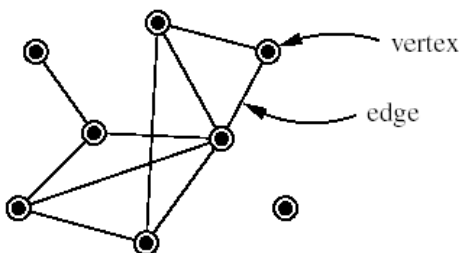


図 1: A small example network with eight vertices and ten edges

ネットワークに関する研究は、いわゆるグラフと呼ばれる数学の一分野である。始まりは、1735年に Euler によって解かれた Königsberg bridge problem である。また、ネットワークの研究は社会科学の分野でも広く研究されている。典型的な社会ネットワークの研究は中心性²と関係性³を明らかにする事にある。

近年の研究は、小さなグラフの分析と個々のノードやエッジのもつ特性から、大規模なネットワークの統計的特性を考慮するように変遷した。これらは、コンピュータの利用が可能となったことと大規模なコミュニケーションネットワークを収集できるようになった事にある。また、ネットワークの規模の小さいときに得られた答えは、その規模が大きくなったときには役に立たない事もある。例えば、社会ネットワークでは、「どのノードが取り除いたときにネットワークの接続性に被害を及ぼすのか」という問題に取り組んだが、1万ノードもあるネットワークでは、その問題に意味は余りない。

このレビュー論文では、主に三つのことに取り組む。まず第一に、ネットワークの振る舞いや構造を特徴づけるパスの長さや次数分布のような統計的特性を明らかにすることと、これらの特性のおおよその計測方法を提案する。二つ目に、ネットワークのモデルを構築すること。三つ目に、ネットワーク化されたシステムの振る舞いを予測することである。

¹The structure and function of complex networks, SIAM Review, pp.167-256

²どのノードが一番つながっているかもしくは影響を与える存在である

³どのような形で他のノードと接続しているのか

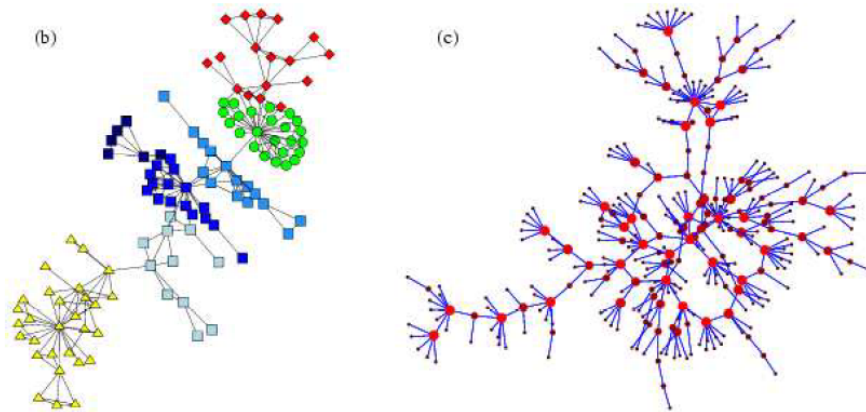


図 2: (b) The network of collaborations between scientist at a private research institution, (c) A network of sexual contacts between individuals

1.1 Other resources

レビューの理解に役立つ参考文献としてあげられているものは、Albert and Barabási のレビュー⁴と Dorogovtsev and Mendes のレビュー⁵が物理学の分野では広く受け入れられている。

- Albert-László Barabási. *Linked*: Barabási 自身のネットワークの研究に特に焦点を当てた自叙伝
- Duncan Watts *Six Degrees*: 社会学的な視点を与えられており、ネットワークの研究の始まりの過程を知ることができる。
- Mark Buchanan *Nexus*: 科学ライタの視点でまとめられた読み物
- 新ネットワーク思考 (A.L.barabasi)
- スモールワールド・ネットワークー世界を知るための新科学的思考法 (ダンカン・ワッツ)
- 人脈作りの科学 (安田雪)
- 複雑な世界, 単純な法則 ネットワーク科学の最前線 (マーク・ブキャナン)
- 複雑ネットワークの科学 (増田直紀, 今野紀雄)

などが役に立つのではないかとおもう。

2 Network in the real world

2.1 Type of networks

ネットワークには有向グラフ (directed graph) と無向グラフ (undirected graph) とがある。有向グラフで表わされるものとしては、個人間の電話連絡や、E-mail のやり取りなど、ある方から他方へという流れを持つ行為から成り立つ関係性を表現できる。また、有向グラフには閉路を含む循環路を持つ可能性がある。また、食物連鎖のように循環路を持たない構造もある。それ以外の特徴あるネットワークとしては、異

⁴Statistical mechanics of complex networks, Rev.Mod.Phys, 74, pp 47–97

⁵Evolution of networks, Advances in Physics 51, pp.1079–1187

なる2種類のノードとその種類が異なるノード間でしかリンクが発生しない、構成される二項グラフなどがある。affiliation networks の様に個人とグループという二つのノード間に成り立つ関係性など。ネットワークの研究は、決して完全な科学ではまだなく、まだ深く探求できる多くの可能性を残している。

とりあえず、ネットワークの研究に必要な最低限の専門用語は

Vertex: 時には site(physics), node(computer science), actor(sociology) などと呼ばれる
 Edge: 時には bond(physics), link(computer science), tie(sociology) と呼ばれる
 Directed/undirected:
 Degrees(次数): ノードから生えているエッジの数
 Component: 連結しているネットワーク
 Geodistic path: 任意のノード間を連結する最短経路。必ずしも一つとは限らない。
 Diameter: ネットワークの直径

等がある。

2.2 Social networks

社会ネットワークとは、人間もしくはグループにおける関係性を表現したものである。例えば、友人関係や婚姻関係などは古くから研究されてきた。社会ネットワークの研究は、社会学の分野で長い歴史がある。過去は、小さな活動グループや子供の友人関係などを対象とし、Anatol Rapoport による数学的なモデルでは、初めて次数分布の重要性が語られている。最近では、ビジネスコミュニティや Sexual Contact などその対象となって言っている。

それ以外に重要なものとして Milgram による “Small-World の実験” がある。これらの実験が現実のものとして再現されてはいないが、少くともネットワークの構造というものを教えてくれる実験であった。また、この実験は、“six degrees of separation” として世の中に知られている。

これらの社会ネットワークの実験での問題点は、取得できるデータが不確実であり不正確であることにある。そのため、対象とするデータはより正確にとれる、WWW, Email, instant messaging 情報へと変遷した。

2.3 Information network

Information networks は、別名 knowledge network と呼ぶ。対象となるものは citation のネットワークになる。

- WWW
- the networks of relations between word classes.
- preference networks

2.4 Technological network

- electric power grid
- networks of roads

- railways
- vascular networks

2.5 Biological network

- the network of metabolic path ways
- the network of mechemistic physical interactions between proteins
- neural networks

3 Properties of networks

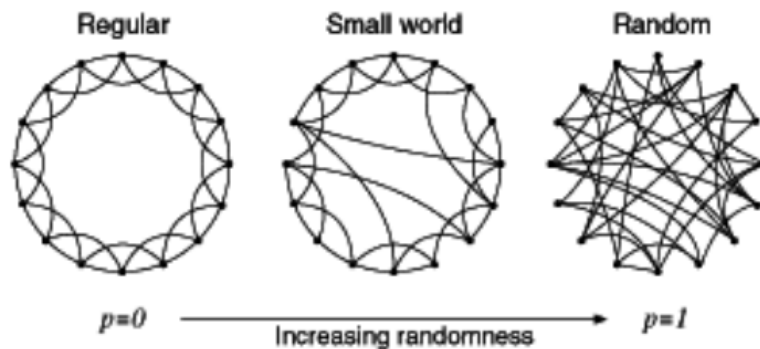


図 3: レギュラーグラフとスモールワールドとランダムネットワーク

現実のネットワークは決してランダムなネットワークではなく、何かしらのガイドに沿うメカニズムを持っており、ある目的を達成するためのネットワーク構造を発見に役立つ方法を提案する。

3.1 The small-world effect

スモールワールド効果は、Stanley Milgram によって 1960 年代に行なわれた実験によれば、およそ 6 人でいどだった。これは、人間関係が短いパスでつながれていることを示す初めての実験である。

スモールワールドでは、任意の二点間の距離の平均 L と クラスタリング係数 C の値に注目する。

$$l = 2/n(n+1) \sum_{i \leq j} d_{ij}$$

この時、 d_{ij} は 頂点 i から 頂点 j への距離、各最短距離の総和を、接続可能なリンクの総数 $2/n(n+1)$ で割ることで距離の平均がだせる。ただし、正規化のために l^{-1} が一般的に用いられる。

3.2 Transitivity or clustering

クラスタリングとは、ネットワーク内にどの程度偏りがあるのかを図るものである。具体的には、あるノードと隣接関係にある二つのノード同士が隣接関係にあるときにはそのノードのクラスタリング係数は1となる。つまり、すべてのノードの近接ノード集合のネットワークの密度を計算することでクラスタリング係数が算出できる。

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}$$

$$C = \frac{1}{n} \sum_i C_i$$

3.3 Degree distributions

次数分布は、ある次数 k がネットワークに閉める割合 P_k をプロットすることで示せる。

限られた大規模なグラフでは binomial もしくは Poisson 分布を示す。現実のネットワークにおいてランダムなグラフのもつ次数分布を見ることが出来る。

3.3.1 Scale-free networks

Scale-free ネットワークでは、次数の分布がベキ分布を示す。そして、このベキ分布の発現がネットワークの規模にとらわれない。このスケールフリーに関する研究は、Price が行なった論文参照ネットワークが始まりであり、最近になって barabási により改めて発見され注目を浴びるようになった。

この次数分布が power-law を示すネットワークとしては、citation network, WWW, Internet, metabolic networks, telephone call graphs, the network of human sexual contacts.

cutoff がある場合のネットワークは、power grid, railway networks movie actors, collaboration networks

3.4 Network resilience

ネットワークの弾性力(耐久能力) どのくらいのノードやエッジを消したときに、ネットワークはバラバラにならないで済むのか

3.5 Mixing patterns

ノードに特徴をつけてその関係性をグラフとしてプロットすることで分析する。

3.6 Degree correlations(次数相関性)

対象とするネットワークが、高い次数が高い次数をもつノードと、もしくは低い次数のノード同士と連結しているのかを調べるための指数である。この次数相関性の数量化には幾つかの手法がある、

- [Maslov et al.] 両端にあるノードの次数を二次元ヒストグラムにプロットする。protein interaction network と Internet のデータを結果にしめした。

- [Pastor-Satorras et al.] よりコンパクトに状況を示せる手法である．ある次数 k のノードの関数として、近隣ノードの次数平均を計算する．この手法では、もしネットワークが適度に複雑であれば、 k にしたがって増加する曲線となる．
- [Newman] エッジの両端の次数の Pearson 相関係数の計算により、基準を一つの数値に落とした．これにより、正の相関では複雑なネットワークであり、負の相関ではそうではないものとえられる．論文に添付されている TableII をみると分かるが、social network ではすべて正の相関を示しているが、それ以外のネットワーク (information network, technological networks, biological networks) では負の相関が現れている．この結果に関する説明はまだ明らかではない．

3.7 Community structure

多くの social network は、community structure と仮定される．このネットワークの例としては、グループ内でのエッジの高い密度が、グループ間でのエッジの密度は低い．これは、人々が、関心事や、就職した年代などによってグループに分かれるのは経験的なできごととして理解できるだろう．このような構造のタイプを時に階層化 (stratified) と呼ぶ．

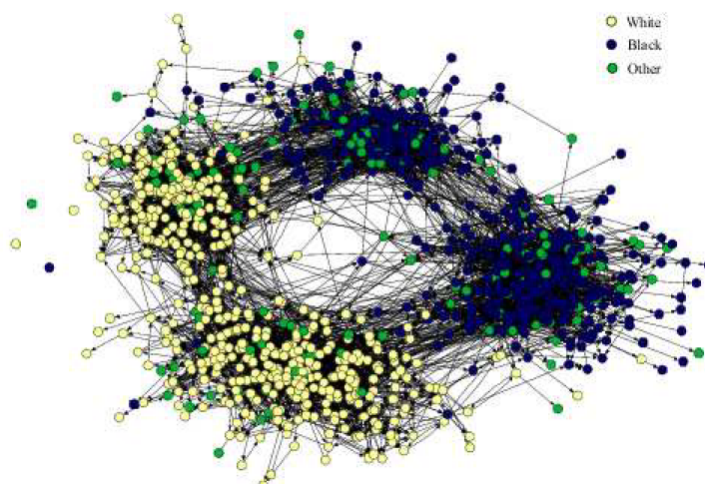


図 4: Friendship network of children in a US school. 友人関係は参加者に尋ねることで決定、したがって有向グラフ、ノードの色は人種で分けている．図から人種により分ける線があることは明らか、このグラフはバネモデルを用いて描かれている

コミュニティは発見する伝統的な手法として cluster analysis(hierarchical clustering) がある．この手法では、ノード間に何かしらの“関係の強さ (connection strength)” を割り当てる．また、一般的に、すでにあるエッジだけでなく n 個のノードがあるならすべてノード間 $\frac{1}{2}n(n-1)$ に与える．その一方で、すべてのペアに strength を与えない場合もあり、残りのペアの値は zero と仮定する．その際、 n 個のエッジを持たないノードからスタートする、頂点間のつながりの強さを解消するエッジを追加する．任意の時点でこの手続きをやめてその状態でのコンポーネントを調べる．この時できたコンポーネントがクラスタである．

クラスターは、connection strength の定義の違いによってクラスタリングは可能となる．最近、良く取り上げられるものとしては“edge betweenness” がある．この edge betweenness とは、対象とするノード間に存在する geodistic path が幾つあるのかを調べる．結果は、少なくとも social network と biological

network では、コミュニティの構造は共通のネットワークの特徴を持つ。食物連鎖はいかなる簡単な方法でもコミュニティは壊れないことが分かっているにも関わらず。

ネットワーククラスタリングは、データクラスタリングの技術と混同しないように、両者の手法には共通点がおおく、互いに使えたりするけど.....

社会ネットワーク分野では、ネットワーククラスタリングは、ネットワークを何かしらの基準でコミュニティに切り分ける block models として長い間議論されている。また、社会学者は特にネットワークの構造的同値 (structural equivalence) に強い関心を持っている。構造的同値とは、例えば二つのノードがあるとき、それらが同様の近隣を持っているならば同値であると言える。

Flake は、コミュニティの構造とは異なる問題であるが、ここで関係するものとして、もし既知のネットワークから得られたノードが、どのコミュニティに属しているのかを明らかにする。この問題の解を得るアルゴリズムにより、WWW と citation networks の様なネットワークでは特別な値が明らかになっている。Flake の論文では、少なくとも Web のコンテンツに関して、maximum flow method に基づくアルゴリズムが明らかにされている。

3.8 Network navigation

Stanley Milgram の有名な small-world の実験は、手紙を知人関係を利用して目的の人物に受け渡すことで、そこにある最短パスを利用しようとした。ただし、この実験から Milgram も気が付かなかった異なる結論を Kleinberg が 2000 年に指摘した。Milgram の結果は既存のネットワークのつながりを明らかにしたが、普通の人々が適切なパスを探せることもデモンストレートしている。彼の実験の参加者は目標の人物に特別な知見もないのだから。それにも関わらず、短いステップでつながっていたということは、そのネットワークに特別な構造があったことが予測される。Kleinberg が指摘したように、変わりにランダムネットワークであったならば短いパスを指摘するということは非情に難しい事である。もし、人工的にネットワークを組むことができるならば、社会ネットワークに現れるのと同様の方法でナビゲーションが容易になる。効率的なデータベース構造やより良い peer-to-peer コンピューティングネットワークなどが。

3.9 Other network properties

- the size of the largest component (“giant component”)

This is a measure of the effectiveness of networks at doing its job

- the distribution of the “betweenness centrality”

4 Random graphs

4.1 Poisson random graph

4.2 Generalized random graph

5 Exponential random graphs and Markov graphs

6 The small-world model

Watts と Strogatz により提案された、small-world model は、それほど洗練されていない扱いが容易なネットワークのモデルである。スモールワールドモデルは、低い次元のレギュラーグラフから作ることが

ら始め、ショートカットとなるエッジを追加もしくは張り替えることで作られる。Small-World はどのような lattice もしくは dimension or topology から作ることはできるが、一番良い方法は 1 次元の lattice に、近隣の k 個のノードにリンクを接続して k 次数の lattice をもつネットワークを作る。そして、エッジを確率 p で自己ループやすでに存在するエッジに重ならないように適度に組み替える。

レギュラーグラフの時の平均パス長は L が大きければ $\frac{L}{4k}$ となる。そして、クラスター係数は $C = (3k-3)/(4k-2)$ であり、 k が大きければ $\frac{3}{4}$ になる。そして、ランダムグラフのときには、最大 $\log L / \log k$ であるが、クラスタリング係数は C は $2k/L$ と非常に小さい。

Watts と Strogatz により提案されたオリジナルな SmallWorld モデルはどこか堅苦しい?。例えば、エッジのどちらかの両端のみを書き換える。そのため自己ループに陥ることはなく、すでに存在するところにはエッジを追加しない。

これら以外の方法としては Monasson や Newman と Watts によって提案されたものなどがある。これらでは、エッジを書き換えるのではなく、新しいペアにエッジを追加する方式である。これを行うことで、ネットワークのコンポーネントが分断される危険性がなくなる。

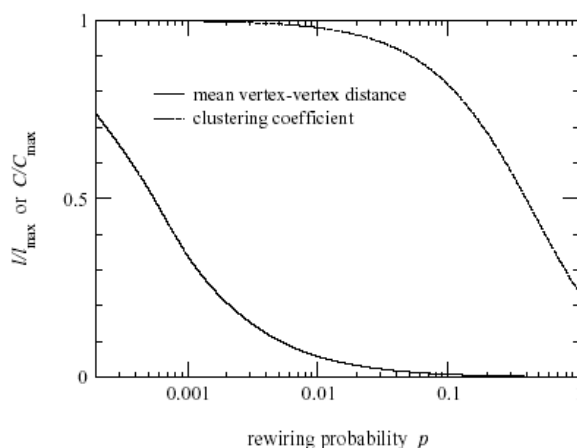


図 5: Watts と Strogatz の提案したスモールワールドネットワークにおけるクラスタリング係数 C と頂点間距離 l

7 Models of network growth

これまでのネットワークのモデルでは、まず第一に、どのようにそれらがもつ特徴からネットワークが作られるのかの理解の助けにはならない。この章では、それを目標とするネットワークの特長を試みたモデルを紹介する。つまり、それらのネットワークでは、基本的にノードやエッジを何かしらの法則に従って加えることで成長を実現する。これらの成長は特徴あるネットワークを作り出す。

7.1 Price's Model

Price のモデルは、論文の citation network

7.2 BA Model

8 Processes taking place on networks