

# 複雑ネットワーク：統計物理学の視点

羽田野直道

## 概要

この講義では、複雑ネットワークを概観し、統計物理学の視点からの解析について議論します。

まずネットワークの基礎用語について説明し、次に複雑ネットワークのよく知られた性質である「スケールフリー性」と「スモールワールド性」について説明します。近年、そのような性質を持つネットワークがどのように現れるのか、また、そのようなネットワーク上でのモデルが、格子上とどのように異なる振る舞いをするのかが分野横断的に盛んに研究されています。これらを初歩から説き起こしてお話します。

同時に、複雑ネットワークがどのような点で統計物理学の対象となるかを、私見を交えて説明します。統計物理学の根本的な概念として「普遍性」があります。この概念はそもそも物理の根本に関わるものですが、統計物理学では特に相転移の繰り込み群による研究の中で培われてきました。そこで、複雑ネットワークを「普遍性」の観点から見ることを強調しつつお話します。

最後に、我々の最近の研究の中から、ネットワークグリーン関数によるコミュニティ検出と、複雑ネットワーク上のゼロ固有値局在について紹介します。

## 1 グラフとネットワーク

この節では、ネットワークに関する基本的な事項を列挙します。**ネットワーク**とは点を線で結んだものですが、用語は数学と物理で異なります。「ネットワーク」は物理の用語で、数学では「グラフ」と呼ばれて古くから研究されてきました。点は数学では頂点(vertex)と呼びますが、物理では**ノード**と呼び、線は数学では辺または枝(edge)と呼びますが、物理では**リンク**と呼びます。以下では物理の用語を統一して用います。

理論上は、あるノードから出発してすぐに同じノードに戻ってくる「自己リンク」(または「ループ」、図1(a))や、2つのノードを複数のリンクが結ぶ「多重リンク」(図1(b))も考えられますが、以下ではいずれも考慮しません。また、どこにもリンクされていない孤立したノードが存在しない場合(「単連結」と呼びます)を基本的には考えます。リンクに向きがついていない場合を**無向ネットワーク**(図1(c))、ついている場合を**有向ネットワーク**(図1(d))と言います。また、個々のリンクに重み(多くの場合は正の実数)が対応づけられている場合を**重み付きネットワーク**と言います。この講義では、簡単のため重み無しの無向ネットワークを主に説明します。

古くから研究されてきたネットワークは**ランダムネットワーク**です。これは、与えられたノードからランダムに2つ選んでリンクで結ぶという操作を繰り返して生成されるネットワークです(図2)。これに対して、世の中に実在するネットワークの多くが全く異なる

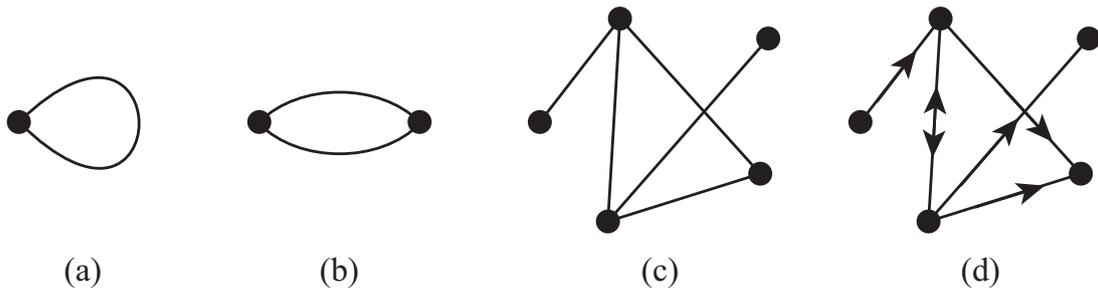


図 1: (a) 自己リンクや (b) 多重リンクのような状況は考えず、(c) 無向ネットワークや (d) 有向ネットワークのように、2つのノードが結ばれているか結ばれていないかのどちらかであるという状況のみを考えます。

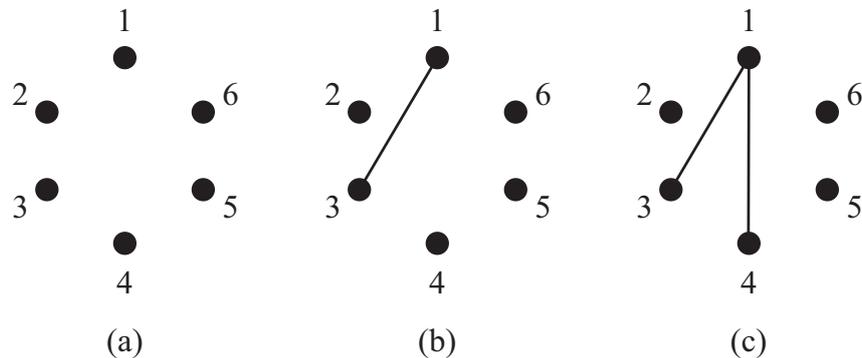


図 2: ランダムネットワークの作り方。(a) 全ノードの中からランダムに2つのノードを選びます。(b) 仮に1と3が選ばれたとします。選ばれたノードをリンクで結びます。さらに2つのノードを選びます。(c) 仮に1と4が次に選ばれたとします。それらをリンクで結びます。なお、多重リンク (図1(b)) を避けるため、既にリンクで結ばれた2つのノードは選ばないことにします。

る統計的性質を持っていることが明らかになってきました。それらを総称して**複雑ネットワーク**と呼び、分野横断的に盛んに研究されています [1-10]。例えば表1のようなものが調べられています。

以下で、ネットワークを研究する上でよく使われる量を紹介します。まず**次数** (degree) と**次数分布**について説明します。あるノードに結ばれているリンクの本数を、そのノードの次数と呼びます。次数  $k$  のノードがいくつ存在するかを示すヒストグラムを**次数分布**と呼び、以下で  $n(k)$  と書くことにします。ランダムネットワークの次数分布はポアソン分布になり、有限の平均と分散が存在します。これに対して、後述するように多くの複雑ネットワークでは**次数分布が<sup>べき</sup>冪分布**になり、平均や分散が存在しない場合があります。冪分布になることを**スケールフリー性**と呼びます。

次に**ノード間距離**について説明します。ネットワーク上のリンクをたどる道筋を**ウォーク** (walk) と呼びます。そのうち、同じリンクは二度と通らない道筋を**パス** (path) と呼びます。ネットワーク上の2ノード間の距離とは、その2つのノードを結ぶ最短のパスの長さです。ノードのペア全て (ただし結ぶパスのないペアを除く) について2ノード間

表 1: 複雑ネットワークの例。

ネットワーク	ノード	リンク
Zachary の空手クラブ [11]	クラブメンバー	友人関係にあるメンバー間
論文の共著者 [12]	研究者	共著論文のある研究者間
映画俳優の共演 [13]	俳優	共演映画のある俳優間
World-Wide Web [14]	ウェブサイト	リンクが存在するサイト間
空港の定期便関係 [15]	空港	定期便がある空港間
線虫の神経回路網 [16]	ニューロン	シナプス結合のあるニューロン間
タンパク質の反応関係 [17]	タンパク質	生化学反応するタンパク質間

の距離を平均した値（様々に呼ばれますが、ここでは**平均ノード間距離**と呼ぶことにします）は、ネットワークの性質を示す重要な指標の一つです。

**クラスター係数**も指標としてよく使われます。これは、あるノードの隣接ノード間にリンクが張られているかどうかを示す指標です。あるノード  $i$  の次数を  $k_i$  とすると、その隣接ノード  $k_i$  個から 2 個選ぶペア  $k_i(k_i - 1)/2$  のうち、 $q_i$  ペアがリンクされていたとします。つまり、ノード  $i$  を頂点とする三角形が  $q_i$  個存在するとします。このとき

$$C_i = \frac{q_i}{k_i(k_i - 1)/2} \quad (1)$$

がノード  $i$  のクラスター係数です。例えば図 3 のネットワークで、ノード 1 は次数が  $k_1 = 3$  なので、隣接ノード 2, 3, 4 から 2 個選ぶペアは (2, 3), (3, 4), (4, 2) の 3 通りあります。そ

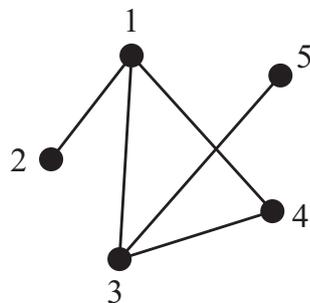


図 3: クラスター係数と隣接行列の説明のためのネットワーク。

のうち、ノード 3 と 4 はリンクで結ばれていますが、他のペアは結ばれていません。したがって  $q_1 = 1$  であり、ノード 1 のクラスター係数は  $C_1 = 1/3$  です。より身近な例で言えば、クラスター係数とは、ある人の 2 人の友人が互いに友人同士である確率です。これを全てのノードについて平均した値 (**平均クラスター係数**) も、ネットワークの性質を示す重要な指標の一つです。

最後に**隣接行列**を説明します。ノード  $N$  個の無向ネットワークを考えましょう。ノードに適当に番号  $i = 1, 2, 3, \dots, N$  をつけておきます。このネットワークに対応する  $N \times N$  行列  $A$  を以下のように構成します。行列要素  $A_{ij}$  を、ノード  $i$  と  $j$  がリンクで結ばれている場合は 1、そうでなければ 0 とします。例えば、図 3 のネットワークの隣接行列は

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (2)$$

となります。つまり無向ネットワーク (図 1(c)) の隣接行列は対称行列です。これに対して有向ネットワーク (図 1(d)) の隣接行列は、ノード  $i$  から  $j$  に向かってリンクがあるときだけ  $A_{ij} = 1$  とするので、一般に非対称行列です。ここでは自己リンク (図 1(a)) が存在しないとしているので、隣接行列の対角要素は必ずゼロです。重み付きネットワークでは、隣接行列の要素 1 をリンクの重みで置き換えます。

## 2 スケールフリー性とスモールワールド性

ここでは、第 1 節でも少し述べた**スケールフリー性**をまず説明します。これは多くの実在するネットワークにおいて次数分布が冪分布になることを指します。冪分布の特殊性は「特徴的な大きさ (スケール)」が存在しないことです。それを指して「スケールフリー」と呼びます。

ランダムネットワークと比較しながら説明を進めましょう。図 4(a) は第 1 節で述べた手順で作ったランダムネットワーク、図 4(b) は第 4 節で述べる手順で作ったスケールフリーネットワークです。より大きいネットワークについて次数分布を示したのが図 5 です。ランダムネットワークの次数分布は平均  $\mu \simeq 200$ 、偏差  $\sigma \simeq 15$  程度のガウス分布

$$n(k) \sim e^{-(k-\mu)^2/(2\sigma^2)} \quad (3)$$

に近い形をしていることがわかります。(正確にはポアソン分布  $n(k) = \lambda^k e^{-\lambda}/k!$  です。パラメータ  $\lambda$  が大きいと、平均  $\mu = \lambda$ 、偏差  $\sigma = \sqrt{\lambda}$  のガウス分布に近づきます。) 一方、スケールフリーネットワークの次数分布は平たくなっています。この程度のサンプル数ではまだよくわからないのですが、より大きいネットワークで計算すると、次数分布は大きい  $k$  に対して冪分布

$$n(k) \sim k^{-\gamma} \quad (4)$$

に近づきます。

式 (3) と (4) の決定的な違いは、式 (4) が自己相似的になっている点です。冪関数 (4) は横軸を 2 倍すると同時に縦軸を  $2^{-\gamma}$  倍すると元の関数に重なります。これを**自己相似的**と呼びます。一方、ガウス分布 (3) やポアソン分布では、そのような操作は存在しません。それは偏差  $\sigma$  という典型的長さ (スケール) があるからです。冪関数にはそのような典型的な長さが無いので、式 (4) をスケールフリーな分布と呼ぶのです。

実在するネットワークの多くで冪分布をする次数分布が観測されています。スケールフリーネットワークを生成するモデルとしては第 4 節で説明する Barabási-Albert 模型 [13] が有名です。

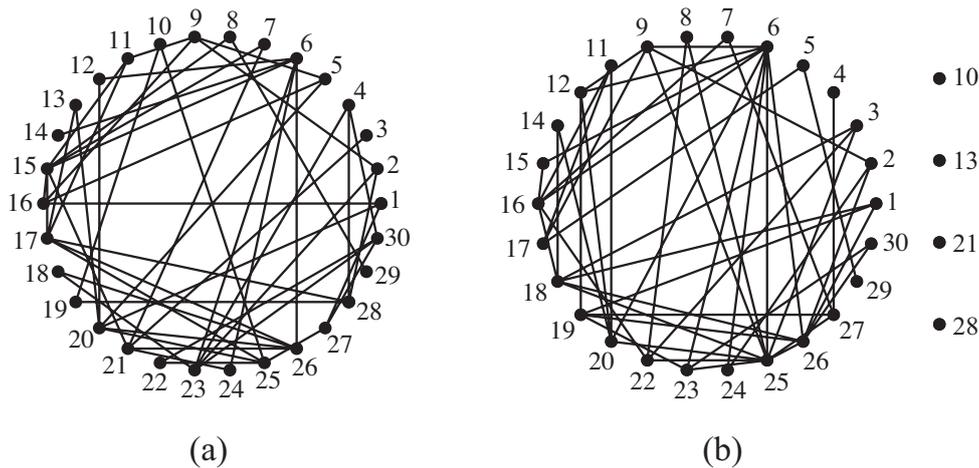


図 4: ノード数 30、リンク数 50 のネットワーク：(a) ランダムネットワーク；(b) スケールフリーネットワーク。

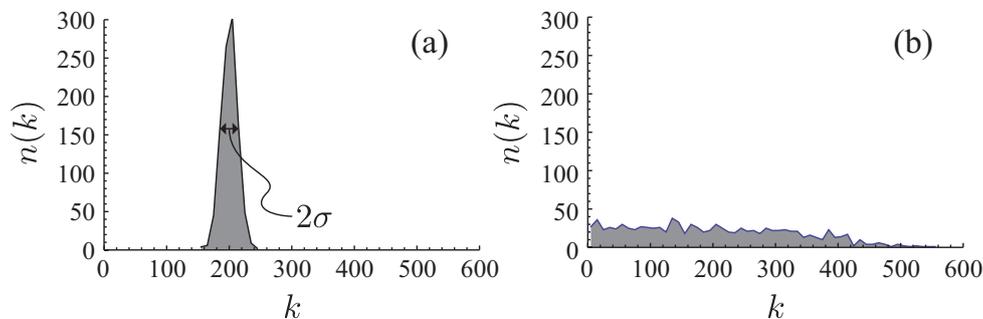


図 5: ノード数 1000、リンク数 200 000 のネットワークの次数分布：(a) ランダムネットワーク；(b) スケールフリーネットワーク。

次に**スモールワールド性**を説明します。よく引用されるのがミルグラム（米国の社会心理学者）の実験です。（ただし実験の確実性を疑う意見はあります。）ある町の多くの住人に手紙を渡し、マサチューセッツ州に住む特定の個人にその手紙が届くように転送することを依頼しました。その際、必ず知人へ転送することを条件としました。これによって、米国内の任意に選んだ 2 名が、何人の知人関係を介してつながっているかを知ろうとしたわけです。結果としては、手紙が届いた場合に限定すると平均 6 人を介していたと報告されています。

この実験は知人間のネットワークでの平均ノード間距離が小さいことを示しています。クラスター係数については、この実験から詳細は得られませんが、平均ノード間距離が小さく、同時にクラスター係数が大きい状況を指してスモールワールド性と呼びます。ランダムネットワークでは平均ノード間距離は小さいのですが、同時にクラスター係数も小さい値を取ります。一方、規則的なネットワーク（例えば格子）では、クラスター係数は大きいのですが、同時に平均ノード間距離も大きくなります。

ところが、実在するネットワークの多くでは、平均ノード間距離が小さく、同時にクラスター係数が大きくなっています。スモールワールドネットワークを生成するモデルとして

は、これも第4節で説明する Watts-Strogatz 模型 [18] が有名です。

### 3 統計物理学における普遍性

統計物理学が複雑ネットワークを扱うにあたって最も重要視するのは**普遍性**であると言えるでしょう。究極の目標は、複雑ネットワークを**普遍性クラス**に分類することであると言っても過言ではありません。第4節と第5節で複雑ネットワークのモデルを導入する際も、普遍性という考え方が重要になります。そこでこの節では、統計物理学における普遍性という考え方を復習しておきましょう。

そもそも物理学における普遍性 (universality) とは、様々な系を貫いて統一的に支配する法則が存在することを言います。物理学において普遍性が重要視されるのは、物理学が本来は神の存在を証明するための学問であったからです。万有引力 (universal gravity) の法則が偉大なのも、星の運動とリンゴの運動が全く同じ法則で説明でき、それが神の存在を示唆しているように思えるからです。

その普遍性が統計物理学の中で驚異的な形で現れるのが**臨界現象**です。例えば、強磁性体を高温から徐々に冷却すると磁化率が次第に大きくなり、**臨界点**  $T_c$  に近づくと磁化率は冪関数  $(T - T_c)^{-\gamma}$  の形で発散します。その冪関数の指数  $\gamma$  を**臨界指数**と呼びます。臨界点を過ぎると自発磁化が、やはり冪関数  $(T_c - T)^\beta$  の形で発生します。この指数  $\beta$  もまた臨界指数です。また、液体気体相転移の臨界点においては、圧縮率が  $(T - T_c)^{-\gamma}$  の形で発散します。臨界点より下では、密度が  $(T_c - T)^\beta$  の形で変化します。

驚くことに、イジング的な異方性を持つ磁性体の臨界指数と、液体気体相転移の臨界点における臨界指数が全く同じなのです。これは、両者の自由度が共に  $Z_2$  という対称性を持っているからです。つまりイジング的な磁性体では、各原子 (または分子) の磁気モーメントが上向きか下向きかどちらかという対称性を持っています。一方、液体気体相転移では、流体分子がある場所に存在するかしないかという対称性を持っています。このように系の自由度に2つの状態がある対称性が  $Z_2$  です。こうした対称性 (と空間次元) さえ同じであれば臨界指数は同じになるのです。臨界点においては系全体の無数の自由度が複雑に絡み合い、系はいわば大混乱の状態にあります。そのような混乱の極みの中から、臨界指数が同じという普遍性が立ち現れてくるのは実に驚嘆すべきことです。

臨界現象の普遍性は**繰り込み群**によって理解されます。臨界点では系の相関長が冪関数の形で発散するので大域的な相関が発生し、それが系の振る舞いを支配します。逆に、系の局所的な自由度は系の振る舞いに影響しなくなります。そこで系の細かな自由度から順に次々に消去して**粗視化**を段階的に進めることによって、系の大域的な振る舞いがどのように臨界点に向かって変化していくかを知らうとするのが繰り込み群です。

繰り込み群では、系の粗視化を繰り返すと、いずれ系は**固定点**に到達すると考えます。固定点とは、それ以上粗視化しても系がもはや変化しない状態のことです。例えば強磁性イジング模型では、高温相で粗視化を繰り返すと高温極限に到達し、低温相で粗視化を繰り返すと低温極限に到達します。一方、臨界点直上で粗視化を繰り返しても臨界点直上から動きません。つまり、強磁性イジング模型には高温極限と低温極限という安定な固定点が2つと、臨界点直上という不安定な固定点が1つ存在します。

固定点では粗視化しても系が変化しないので、特徴的な長さが系から消滅していることがわかります。これは、第2節で述べたように冪分布に典型的な大きさが無いことと対応しています。このため臨界点直上の固定点では様々な物理量が冪関数の振る舞いを示すのです。磁性体と液体気体相転移のように一見まったく異なる系も、粗視化を繰り返した後で同じ固定点に行き着くのであれば、それらは同じ臨界指数の異常性を示すでしょう。このようにして臨界現象の普遍性が立ち現れると考えられます。

さらに、固定点の種類が有限個、あるいは加算無限個であれば、それらのうちのどの固定点に行き着くかによって、あらゆる系の相転移現象を分類することができると予想されます。これが普遍性クラスという考え方です。実際に、空間2次元の古典模型の相転移は**共形場理論**によって、ほぼ完全な形で普遍性クラスに分類されています。

さて、複雑ネットワークに普遍性クラスの考え方を適用するには、もう一つ、**自己組織化臨界現象**という概念を説明する必要があります。これまでに述べた臨界現象では、系の温度や圧力などのパラメーターを外から調節することによって、系を臨界点に誘導する必要がありますがありました。ところが、系が自然に臨界的になる場合が存在するという説があり、そのような現象を自己組織化臨界現象と呼びます。

砂山模型を例に話を進めましょう。上から砂粒をポツポツと同じ場所に落として砂山を築くことを考えます。砂粒が中央にたまってある程度の塔になると、雪崩が起こって塔は崩壊します。これが繰り返されてある一定の傾斜角の砂山が築かれます。このとき、雪崩の大きさの頻度分布が冪分布であると考えられています。

つまり、小さな雪崩も非常に多く起こるが、非常に大きい雪崩も必ず少しは起こることです。これはあたかも、パラメータを調節することなしに、系が自然に臨界的な固定点に到達したかのように見えます。そこでこのような現象を自己組織化臨界現象と呼びます。

他の例としては、地震の規模の頻度分布が冪分布であることがグーテンベルク・リヒター則として知られています。また、株価や外国為替の変動幅の頻度分布も冪分布であると考えられています。同じように、実在する多くのスケールフリーネットワークも自己組織化臨界現象としてとらえることができるでしょう。

これまでに、自己組織化臨界現象を普遍性クラスに明確に分類したという研究は、私の知る限りまだありませんが、現在の統計物理学の大きな課題の一つであると言えるでしょう。その観点から、スケールフリーネットワークを普遍性クラスに分類しようというのが、複雑ネットワークにおける統計物理学の野望です。

現実には野望の達成はまだ遠いかも知れませんが、ただ普遍性という考え方は、統計物理学における模型の設定にも強い影響を及ぼしていることに注意が必要です。現実には複雑な詳細が存在する場合でも、そのような詳細は臨界的な振る舞いには影響しないだろうと期待されます。同じ普遍性クラスに属する最も簡単な模型の振る舞いさえ調べれば、臨界的振る舞いを知るには十分であると考えられるのです。第4節と第5節で紹介する模型も、見方によっては非常に単純な模型に見えますが、上述のような考え方に基づいて研究が進められています。

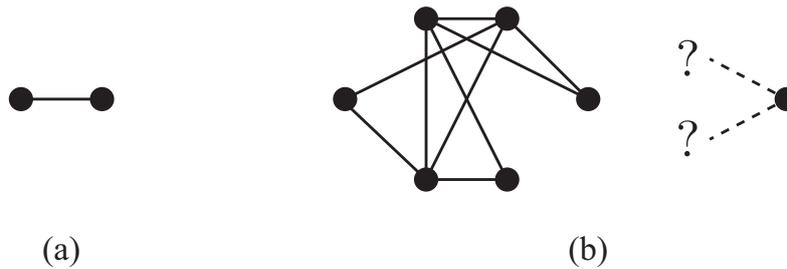


図 6: Barabási-Albert 模型の構成法。ここでは  $m = 2$  とします。(a) 最初の  $m = 2$  個のノード。(b) 後から追加するノードを、既存のノードのうちの  $m = 2$  個のノードとリンクで結びます。このとき、選択する確率は (5) とします。

## 4 複雑ネットワークの成長模型

この節では複雑ネットワークの構造に関する模型、次節では複雑ネットワーク上のダイナミクスを示す模型を紹介します。前者として具体的には、スケールフリー性を再現する模型として良く知られている **Barabási-Albert 模型** [13] と、スモールワールド性を再現する模型として有名な **Watts-Strogatz 模型** [18] を紹介します。

### 4.1 Barabási-Albert 模型

Barabási-Albert 模型 [13] は以下のようにしてネットワークを構築する模型です。

- (i) 最初に  $m$  個のノードを用意し、全てのノード間をリンクで結びます (図 6(a))。整数  $m$  は 1,2,3 程度の小さい整数です。
- (ii) 以下の手順を繰り返して、ネットワークを成長させます。
  - (a) 新しいノードを用意し、以下の手続きに従って既存のノード  $n$  個のうちの  $m$  個のノードにリンクをつなぎます。(したがって、最初にノードを付け加える時は、既存の  $m$  個のノード全てに自動的にリンクをつなぎます。) 既存のノード  $i = 1, 2, \dots, N$  それぞれの次数 (リンクの本数) を  $k_i$  とします。このとき、 $N$  個のノードから  $m$  個のノードを確率

$$p_i = \frac{k_i}{\sum_{i=1}^N k_i} \quad (5)$$

にしたがって選択します。

- (b) 新しく付け加わったノードも既存のノードの個数に含めて、 $N$  を 1 だけ増加させます。

ここでポイントとなるのは、確率 (5) です。この確率に従うと、既にリンクを多く持っているノードほど、より高い確率で新しいリンクを獲得します。これを**優先的接続** (preferential attachment) と呼びます。このために、非常に多くのリンクを持つノードがランダムネットワークよりも優先的に生成される一方、リンクの少ないノードはいつまで経ってもリン

クが少ないままということになります。これが図 5(a) と (b) の違いを生みます。中でも周囲のノードよりも特にリンク数の多いノードを**ハブ** (hub) と呼ぶこともあります。

この手順に従ってネットワークを成長させると、その次数分布は大きい  $k$  に対して

$$n(k) \sim k^{-3} \quad (6)$$

となることが示せます。つまり式 (4) において  $\gamma = 3$  であるスケールフリーネットワークを構成できたことになります。この場合、 $k$  の平均も分散も発散することに注意しましょう。

なお、式 (5) を拡張して、例えば [19]

$$p_i = \frac{k_i + a}{\sum_{i=j}^N (k_j + a)} \quad (7)$$

とパラメーター  $a$  を導入すると、式 (6) は

$$n(k) \propto k^{3+a/m} \quad (8)$$

と変更され、指数  $\gamma$  が 3 ではない解を与えます。現実のスケールフリーネットワークでは、 $2 \leq \gamma \leq 3$  となっている場合が多く、これは  $-m \leq a \leq 0$  とおくと説明できることになります。(ただし「普遍性クラス」は曖昧になると言わざるを得ません。)

以下で式 (6) を導いてみましょう。まず時刻変数  $t$  を導入し、最初に  $m$  個のノードを用意した時刻を  $t = 0$ 、新しいノードを 1 個追加する度に時刻が 1 経過するとします。すると、時刻  $t$  において全ノード数は  $N(t) = m + t$  です。一方、全リンク数は最初に  $m(m-1)/2$  本、新しいノードを追加する度に  $m$  本増えるので、 $E(t) = m(m-1)/2 + mt$  です。各ノード  $i = 1, 2, \dots, N(t)$  の次数  $k_i(t)$  の合計はリンク数の 2 倍なので、時刻  $t$  が十分に経過したときには

$$\sum_{j=1}^{N(t)} k_j(t) = 2 \left[ \frac{m(m-1)}{2} + mt \right] \simeq 2mt \quad (9)$$

となります。したがって、式 (5) は

$$p_i(t) \simeq \frac{k_i(t)}{2mt} \quad (10)$$

と近似できます。

次に、あるノード  $i$  が時刻  $t+1$  において追加されたノードから受け取るリンクの期待値を求めましょう。本来は、新しいノードから同時に 2 本以上のリンクを受けないという制約があります。しかし、もしこの制約を課していなかったとしても、時間が十分に経過して全ノード数が  $m$  に比べて非常に大きくなれば、新しく追加されたノードから同じノードにリンクが結ばれることはほとんど起こらないでしょう。そこで以下ではこの制約を無視することにします。すると、受け取るリンク数の期待値は  $mp_i$  となります。これが単位時間当たりに増加する  $k_i$  の値だと考えられます。

そこで、時間を連続変数に近似して微分方程式を導入すると

$$\frac{dk_i}{dt} = mp_i = \frac{k_i}{2t} \quad (11)$$

となり、その解は

$$k_i(t) = m \left( \frac{t}{t_i} \right)^{1/2} \quad (12)$$

で与えられます。なお、 $t_i$  はノード  $i$  が追加された時刻を表し、その際にノード  $i$  は  $m$  本のリンクを結ぶので、初期条件を  $k_i(t_i) = m$  としました。式 (12) は、時間の経過と共に  $\sqrt{t}$  で次数  $k_i$  が増えていくこと、そして、後から追加されたノードほど次数が小さいことを表しています。

次数分布  $n(k)$  を求めるには、まず  $k$  本より少ないリンクを持っているノードの数を求めます。式 (12) によれば、これは時刻  $t_i = t(m/k)^2$  より後に、現時点  $t$  までの間に追加されたノードの数です。(ただし、 $k$  は  $m$  よりも十分大きいとします。) すなわち

$$N(t) - N(t_i) = t - t_i = t \left[ 1 - \left( \frac{m}{k} \right)^2 \right] \quad (13)$$

です。次数分布  $n(k)$  は、式 (13) を  $k$  で微分して

$$n(k) = \frac{2m^2t}{k^3} \quad (14)$$

となります。こうして式 (6) が導かれました。

なお、式 (5) を拡張して、例えば

$$p_i = \frac{k_i + a}{\sum_{i=j}^N (k_j + a)} \quad (15)$$

とパラメータ  $a$  を導入すると、式 (14) は

$$n(k) \propto k^{3+a/m} \quad (16)$$

と変更され、指数  $\gamma$  が 3 ではない解を与えます。現実のスケールフリーネットワークでは、 $2 \leq \gamma \leq 3$  となっている場合が多く、これは  $-m \leq a \leq 0$  とおくと説明できることとなります。

一方、例えば  $p_i$  を定数

$$p_i = \frac{1}{N} \simeq \frac{1}{t} \quad (17)$$

にすると、微分方程式 (11) は

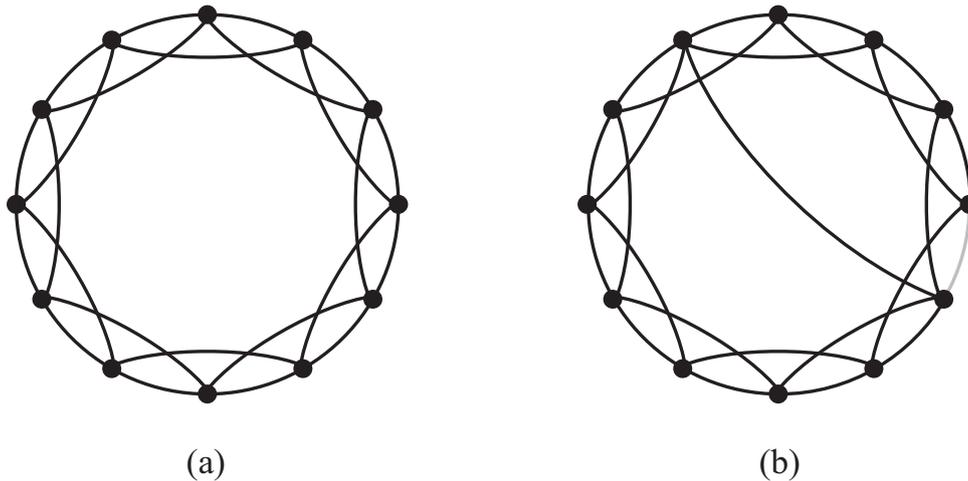


図 7: Watts-Strogatz 模型の構成法。(a) 最初に次数  $m$  のノード  $N$  個が規則的にリンクされている状況を考えます。(b) ランダムにリンク (ここでは右の灰色のリンク) を切断し、別のリンク (ここでは対角の長いリンク) につなぎ替えます。

$$\frac{dk_i}{dt} = \frac{m}{t} \quad (18)$$

と変更され、微分方程式の解は

$$k_i = m + m \log \frac{t}{t_i} \quad (19)$$

$k_i = m[1 + \log(t/t_i)]$  と変更され、次数分布は

$$n(k) \propto \exp(-k/m) \quad (20)$$

となるので、ポアソン分布の裾野に近い形を与えます。

## 4.2 Watts-Strogatz 模型

次に、Watts-Strogatz 模型は以下のようにしてネットワークを構成する模型です。まず、最初に次数  $m$  のノード  $N$  をリング状に連結した状況を考えます (図 7(a))。つまり  $Nm/2$  本のリンクが存在します。このなかからランダムに  $pNm/2$  本を選び、図 7(b) の要領でランダムにつなぎ替えます。割合が  $p = 0$  のときは規則的なネットワークのまま、割合が  $p = 1$  のときはランダムネットワークが得られます。

まず  $p = 0$  の場合に平均ノード間距離と平均クラスター係数を求めてみましょう。この場合には全てのノードが同じ立場にあるので、ノード間距離を求めるにあたって、片方のノードを固定して計算できます。図 7(a) において、あるノードから円環に沿って  $m/2$  ステップの範囲にあるノードとの間の距離は 1 です。円環に沿って  $x$  ステップのところのノードとの間の距離は、およそ  $x/(m/2)$  です。円環の反対側のノードが最も遠いノード

ですが、そこまでは円環に沿って  $N/2$  なので、距離はおよそ  $N/(m/2)$  です。以上から、平均ノード間距離は

$$L \simeq \frac{1}{N/2} \sum_{x=1}^N \frac{x}{m/2} \simeq \frac{2N}{m} \quad (21)$$

となります。一方、クラスター係数は、これも特定のノードに固定して計算できます。あるノードには  $m$  個のノードが隣接しており、隣接ノードのペアは  $m(m-1)/2$  ペア存在します。この内、円環に沿って1ステップ離れたノードでつながっているのが  $m-2$  ペア、円環に沿って2ステップ離れたノードでつながっているのが  $m-3$  ペアです。最長では、円環に沿って  $m/2$  ステップ離れたノードでつながっているのが  $m/2-1$  ペアです。したがって、隣接ノード間でつながっているペアの総数は

$$\frac{1}{2} \times \left[ (m-2) + \left( \frac{m}{2} - 1 \right) \right] \times \frac{m}{2} = \frac{3}{8} m(m-2) \quad (22)$$

です。したがって、平均クラスター係数は

$$C = \frac{3(m-2)}{4(m-1)} \quad (23)$$

となります。

次に  $p=1$  のランダムネットワークの場合には、平均ノード間距離と平均クラスター係数はどのように計算できるでしょうか。計算に入る前に、Watts-Strogatz 模型におけるリンクのつなぎ替えの操作中、リンクの総数は保存されていることに注意しましょう。したがって、 $N$  個のノードのペア  $N(N-1)/2$  のうちで、 $Nm/2$  ペアがリンクで結ばれていることとなります。つまり、このランダムネットワークにおいては、全体のペアの内、 $m/(N-1)$  の割合でランダムにリンクが結ばれていることがわかります。これは、あるノードの隣接ノード間のペアについても全く同じです。したがって、ランダムネットワークのクラスター係数が直ちに

$$C = \frac{m}{N-1} \quad (24)$$

であることがわかります。

このクラスター係数は、ネットワークが大きくなるとゼロに近づきます。つまりランダムネットワークでは、三角形がほとんど構成されていないことがわかります。そこで、全くループが存在しないネットワーク (図8) で平均ノード間距離を考えてみましょう。ある一つのノードに注目し、そこから距離  $l$  の位置にあるノードの総数  $N(l)$  を概算すると、平均次数が  $m$  なので、およそ

$$N(l) \simeq m(m-1)^{l-1} \sim m^l \quad (25)$$

です。



図 8: 樹のネットワーク。

実は、樹のネットワークの特徴は、 $N(l)$  が  $N(l-1)$  よりも  $m$  倍も大きい点です。したがって、ノード間の平均距離を計算する際には遠いノードの寄与が支配的です。そのため、全ノード数  $N$  と平均ノード間距離  $L$  の間にも、およそ

$$N \sim m^L \quad (26)$$

という関係があります。以上のことから

$$L \sim \ln N \quad (27)$$

という概算ができます。

以上をまとめると、平均ノード間距離は  $p=0$  で  $O(N^1)$  なのに対して、 $p=1$  で  $O(N^0)$  に減ります。一方、平均クラスター係数も  $p=0$  で定数なのに対して、 $p=1$  で  $O(N^{-1})$  に減ります。より詳細に議論すると、 $p$  が増えるにつれて平均ノード間距離はすぐに  $O(N^0)$  に減るのに対して、平均クラスター係数はなかなか  $O(N^{-1})$  に減りません。つまり、中間領域でクラスター係数は比較的大きく、一方でノード間距離は小さいというスモールワールドが実現しているのです。

## 5 複雑ネットワーク上の模型

第4節では、複雑ネットワークの構造を説明するための模型を説明しました。それに対して、複雑ネットワークは与えられたものとして、その上で何らかの統計物理学的模型の振る舞いを調べるといった研究も多く存在します。

すぐに考えつくのは、格子上の模型と複雑ネットワーク上の模型では相転移の臨界性がどのように異なるかということです [20]。例えば強磁性イジング模型は複雑ネットワーク上で相転移を起こすのでしょうか。起こすとしたら、その臨界指数はどのような普遍性クラスに属するのでしょうか。

また、臨界指数が求められたとして、それをどのように説明したらよいのでしょうか。第3節でも述べたように、格子上の模型における臨界現象の理解では、相関長が発散して系全体の自由度が結合することによって相転移が起こります。臨界指数の値は、空間次元、

相関長の発散の指数、臨界点直上における系のフラクタル性などから説明できます。しかし、複雑ネットワーク上ではユークリッド幾何学的な「長さ」を簡単には定義できません。最短パスで定義するノード間距離はピタゴラスの定理を満たさないからです。したがって、そもそも「相関長」や「フラクタル性」をどのように定義したらいいかという問題が存在します。

その他の問題意識としては、複雑ネットワーク上で考える方が自然であるモデルが存在します。例えば、経済現象や世論形成を説明するモデルは、多くの場合、格子上で考えられてきましたが、現実の企業間ネットワークや知人間ネットワークが格子ではなくて複雑ネットワークであることを考えると、複雑ネットワーク上でどのようなダイナミクスを示すかを考える方が自然です [21-25]。

上述の問題は、複雑ネットワークの構造が複雑ネットワーク上のモデルの振る舞いにどのような影響を与えるかを知ろうとするものでした。逆に、複雑ネットワーク上のモデルの振る舞いから複雑ネットワークの構造を知ろうとする研究もあります。例えばスピンモデルのダイナミクスからネットワークのコミュニティを検出する手法が提案されています [26-29]。コミュニティとは、ネットワーク上でひとかたまりのグループを構成する部分ですが、統一的な定義は実は存在しません。幾つかのテストケースにおいてもっともらしい答えを与える定義という観点から「よい定義」と「悪い定義」、「よい検出方法」と「悪い検出方法」が判断されています。以下で、統計力学的計算を利用したコミュニティの検出方法を紹介しましょう。

ここで紹介する方法 [26] では、コミュニティの定義として

- (i) コミュニティ内部のノードはリンクで強く結びついている、
- (ii) コミュニティ内部と外部の結びつきは弱い、

という2つの点を強調しています。この2つの特徴に対応して、 $q$  状態ポッツ模型のハミルトニアン

$$H = - \sum_{i \neq j} (A_{ij} - \gamma p_{ij}) \delta(\sigma_i, \sigma_j) \quad (28)$$

を考えます。ポッツ模型とはスピンの  $q$  個の離散的状態をとるもので、相互作用し合う2つのスピンが同じ状態をとるとエネルギーが発生するようなモデルです。(2状態ポッツ模型はイジング模型と等価です。イジング模型では、スピンの反対向きの場合をエネルギーの原点とすれば、スピンの同じ向きを向いたときにエネルギーが発生するからです。) 式 (28) において  $\sigma_i$  は1から  $q$  までの整数の値を取るスピン変数で、 $\delta(\sigma_i, \sigma_j)$  は  $\sigma_i = \sigma_j$  のときだけ値が1になり、それ以外はゼロであるクロネッカーのデルタです。また  $A_{ij}$  は隣接行列  $A$  の要素、 $p_{ij}$  はノード  $i$  と  $j$  がリンクする確率、 $\gamma$  は適当なパラメータです。(詳細は原論文 [26] を参照してください。) ハミルトニアン (28) は、 $A_{ij} > \gamma p_{ij}$  の場合には、ノード  $i$  と  $j$  の間にリンクが存在すれば両者のスピン変数間に強磁性的相互作用が働き、リンクが存在しなければ、反強磁性的相互作用が働くように設定されています。このハミルトニアンの最低エネルギー状態を求めると、上の2つの特徴を備えたコミュニティがある程度の精度で検出されます [26]。

以上の研究では、ネットワーク上のモデルのダイナミクスを考える際には、ネットワークそのものは固定されています。しかし最近では、第4節で解説した複雑ネットワークの成長と、この節で解説した複雑ネットワーク上のモデルのダイナミクスを絡ませた研究も行われ始めています。実際、ネットワーク上のモデルのダイナミクスが複雑ネットワークの成長に影響を及ぼす場合も考えられます。そのため、ネットワークの成長の時間スケールと、ネットワーク上のモデルのダイナミクスの時間スケールが同程度の場合、どのようなことが起こるかにも興味を持たれています。

## 6 我々の研究

最後に、最近の我々の研究を2つ紹介しておきます。ノード間の結びつきの強さを表す *communicability* という指標の導入 [30,31] と、隣接行列の固有値スペクトルについての研究です。(いずれも詳細は論文を参照して下さい。)

### 6.1 Communicability によるコミュニティ検出

第1節で紹介したように、ノード間の距離は最短パスの長さで定義されます。しかし、ノード間の結びつきの強さは、必ずしも最短パスの長さだけで表されるものではありません。例えば図9(a)では左端と右端のノード間距離は2、図9(b)では3です。距離だけ考えると図9(a)の左端と右端のノード間の方が近いのですが、図9(b)の左端と右端のノードの方が結びつきが強いと考えるのが自然です。このように、最短パス以外のウォーク(定義は第1節を参照)も考慮するために Estrada 氏が導入したのが *communicability* [30,31]

$$G_{ij} = \sum_{n=0}^{\infty} \frac{1}{n!} (A^n)_{ij} \quad (29)$$

です。ここで  $A$  は第1節で導入した隣接行列です。実は  $(A^n)_{ij}$  は、ノード  $i$  と  $j$  の間を  $n$  ステップで結ぶウォークの総数を与えます。なぜなら

$$(A^n)_{ij} = \sum_{i_1, i_2, \dots, i_{n-1}} A_{ii_1} A_{i_1 i_2} A_{i_2 i_3} \cdots A_{i_{n-1} j} \quad (30)$$

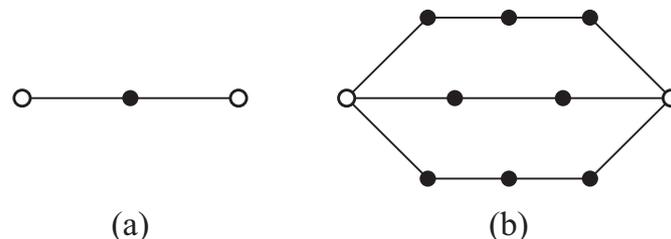


図 9: (a) 左端と右端のノード間距離は 2、(b) こちらは 3。

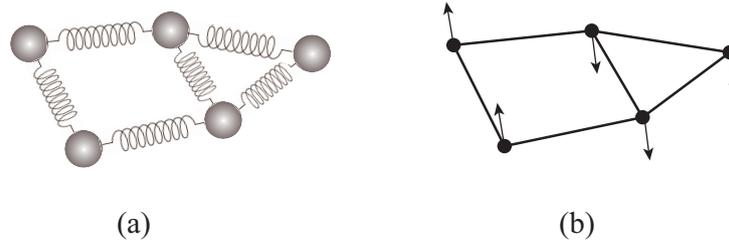


図 10: (a) ボールがバネで繋がれている系の概念図。(b) その系の振動の概念図。振動の向きは任意です。左の2つのノードと右の3つのノードの間に「節」があります。

と書いたときに、和記号の中で  $A$  の要素の積が 1 になる場合は  $i \rightarrow i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_{n-1} \rightarrow j$  というウォークが可能な場合のみだからです。(ここで、ウォークは行きつ戻りつする経路も含むことに注意して下さい。) 式 (29) では、ノード  $i$  と  $j$  を結ぶ全てのウォークを考慮していますが、係数  $1/n!$  があるので、長いウォークほど影響がそれなりに小さくなっています。係数の選び方には  $1/n!$  以外にも方法が考えられますが、式 (29) のよい点は

$$G_{ij} = (e^A)_{ij} \quad (31)$$

のように、統計力学の分配関数に似た形にまとめられることです。実際、新たにパラメータを導入して

$$G_{ij}(\beta) = \sum_n \frac{\beta^n}{n!} (A^n)_{ij} = (e^{\beta A})_{ij} \quad (32)$$

とすれば、 $\beta$  を変化させることによって長いウォークの寄与を大きくしたり小さくしたり調整できます。隣接行列  $A$  は、ある近似のもとでボールがバネで繋がれている系 (図 10(a)) のハミルトニアン  $H$  の符号を反転したものと見なすことができます。したがって式 (32) は、その系の温度グリーン関数と見なせます。

この communicability を使って、以下のようなコミュニティ検出の方法を提案しました [30,31]。まず、隣接行列の固有値と固有ベクトルを  $\lambda_\mu$  と  $\vec{\psi}_\mu$  と書くことにします。ノード数を  $N$  とすると隣接行列は  $N \times N$  行列なので、 $\mu = 1, 2, \dots, N$  です。このとき、式 (31) はスペクトル分解して

$$G_{ij} = \sum_{\mu=1}^N \psi_\mu(i) \psi_\mu(j) e^{\lambda_\mu} \quad (33)$$

と書けます。ここで  $\psi_\mu(i)$  は固有ベクトル  $\vec{\psi}_\mu$  の第  $i$  成分です。

上述したように隣接行列  $A$  の符号を反転したものを、ボールがバネで繋がれている系のハミルトニアン  $H = -A$  と見なすと、隣接行列の最大固有値  $\lambda_1$  はハミルトニアン基底エネルギー  $-\lambda_1$  に対応します。ペロン・フロベニウスの定理から、最大固有値に縮退はなく、その固有ベクトルは全成分が同符号です。次の固有値  $\lambda_2$  は第一励起準位  $-\lambda_2$  に対応します。その固有ベクトルには「節」が一つ存在するはず (図 10(b))。その節が 2

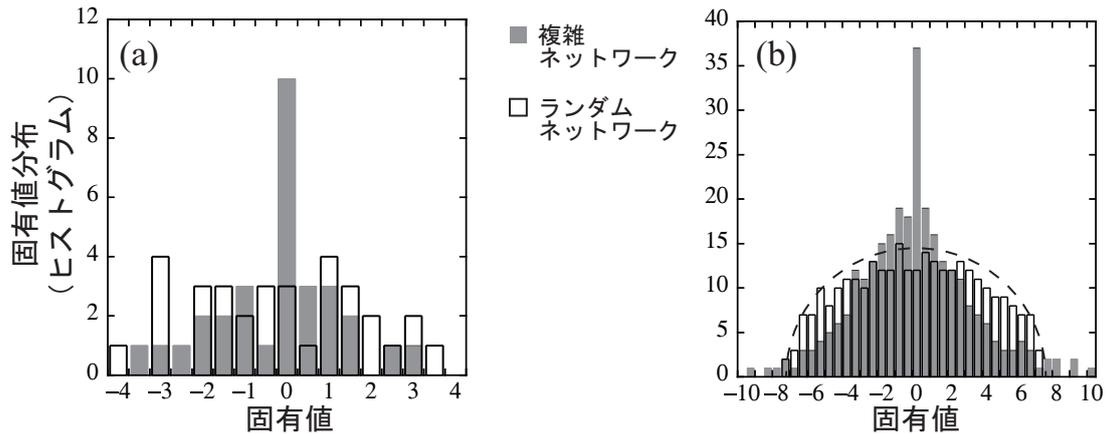


図 11: (a)Zachary の空手クラブの友人ネットワーク [11] と (b) 線虫の神経回路網 [16] の実データに基づく隣接行列の固有値分布 (灰色の固有値分布)。白抜きの固有値分布は、同じノード数・同じリンク数を持つランダムネットワーク (計算機で発生させた) の隣接行列の固有値分布。

つのコミュニティを分割する線に対応していると考えます。するとノード  $i$  と  $j$  が同じコミュニティに属していれば  $\psi_2(i)$  と  $\psi_2(j)$  は同符号、異なるコミュニティに属していれば  $\psi_3(i)$  と  $\psi_3(j)$  が異符号になります。同様に、第 2 励起状態  $\lambda_3$  は節が二つ存在するので、コミュニティが三つ存在する場合には 2 つの節が分割する線になると期待できます。この場合も、ノード  $i$  と  $j$  が同じコミュニティに属していれば  $\psi_2(i)$  と  $\psi_2(j)$  は同符号、隣り合うコミュニティに属していれば  $\psi_3(i)$  と  $\psi_3(j)$  が異符号になります。

以上のことから、式 (33) から基底状態の成分を除いた残りを、 $\psi_\mu(i)$  と  $\psi_\mu(j)$  が同符号の項と異符号の項に分けます：

$$\begin{aligned} \Delta G_{ij} &= G_{ij} - \psi_1(i)\psi_1(j)e^{\lambda_1} \\ &= \sum_{\mu}^{++/--} |\psi_\mu(i)\psi_\mu(j)|e^{\lambda_\mu} - \sum_{\mu}^{+/-+} |\psi_\mu(i)\psi_\mu(j)|e^{\lambda_\mu}. \end{aligned} \quad (34)$$

右辺の第 1 項が同符号の項、第 2 項が異符号の項です。第 1 項が大きいほど同じコミュニティに属している可能性が高いと考えられます。そこで、あらゆる  $i$  と  $j$  の組み合わせのうちで  $\Delta G_{ij}$  が正のものだけを結んだ新しいネットワークをつくり、それを communicability network と呼ぶことにします。この新しいネットワークからコミュニティの存在を判定します。

## 6.2 隣接行列のゼロ固有値と局在

ここでは、隣接行列の固有値分布についての我々の研究を紹介します [32]。例として、2 つの実データについて隣接行列の固有値分布を図 11 に示しました。いずれの分布でも中央に鋭いピークがあります。それに比べ、同じノード数・同じリンク数を持つランダムネットワークの隣接行列の固有値分布はなだらかです。(ランダムネットワークの隣接行

表 2: 4つのネットワークの隣接行列のゼロ固有値数をランダムネットワークと比較しました。ゼロ固有値数の上段が複雑ネットワークに対するもの、下段が同じノード数・リンク数のランダムネットワークに対するもの。

	空手クラブ [11]	神経回路網 [16]	タンパク質 [17]	論文共著者 [12]
ノード数 / リンク数	34/78	297/2148	2114/2203	40421/175692
ゼロ固有値数	10	15	888	2678
(- 孤立ノード数)	(-0)	(-0)	(-268)	(-844)
ゼロ固有値率	29.4%	5.1%	29.3%	4.5%
ゼロ固有値数	0	0	412	5
(- 孤立ノード数)	(-0)	(-0)	(-265)	(-5)
ゼロ固有値率	0%	0%	6.9%	0%

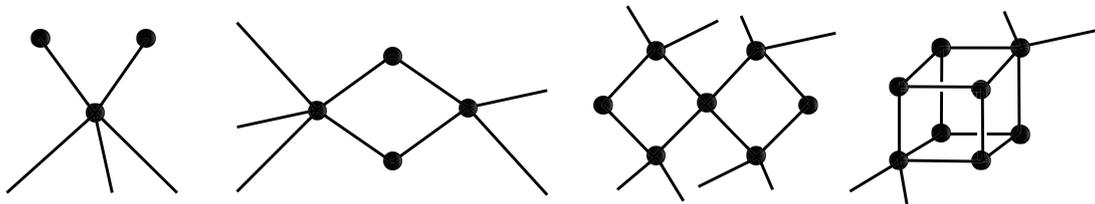


図 12: ゼロ固有値を与える局所的構造の例。

列はいわゆるランダム行列です。ランダム行列の固有値分布は行列サイズが大きくなると半円に収束することが知られていますが、確かに図 11(b) のランダムネットワークの固有値分布は半円（破線）に近い形をとっています。）

他の 2 つの例も併せて詳細を表 2 に示しました。ランダムネットワークでもゼロ固有値が全くないというわけではありません。特にリンク数が少ない（疎な）ネットワークでは、ランダムネットワークの隣接行列でもゼロ固有値が発生することが、行列の「ランク落ち」として知られていました。しかし、表 2 の 4 つの例のいずれにおいても、複雑ネットワークのゼロ固有値がランダムネットワークよりもかなり多いことがわかります [32]。

ネットワークに例えば図 12 のような局所構造がある時にゼロ固有値が発生することを示せます。このような構造によってゼロ固有値が発生することは、格子状の模型において既に知られていましたが、複雑ネットワークにおいて同じことが大規模に起こっていることがここでのポイントです。このような構造はランダムネットワークでは偶然にしか起こりえず、リンク間の相関が存在することによって初めて出現しやすくなります。つまり、複雑ネットワークの成長過程において図 12 のような局所構造が出現しやすいようなリンク間相関が存在することを示唆しています。

次に、ゼロ固有値に対応する固有ベクトルを議論します。複雑ネットワークの実データにおいて、大量に縮退するゼロ固有値の固有ベクトルを適切な形で求めるのは、実は容易ではありません。大量に縮退しているため、ゼロ固有値の固有ベクトルの任意の線形結合が全てゼロ固有値を与える固有ベクトルだからです。

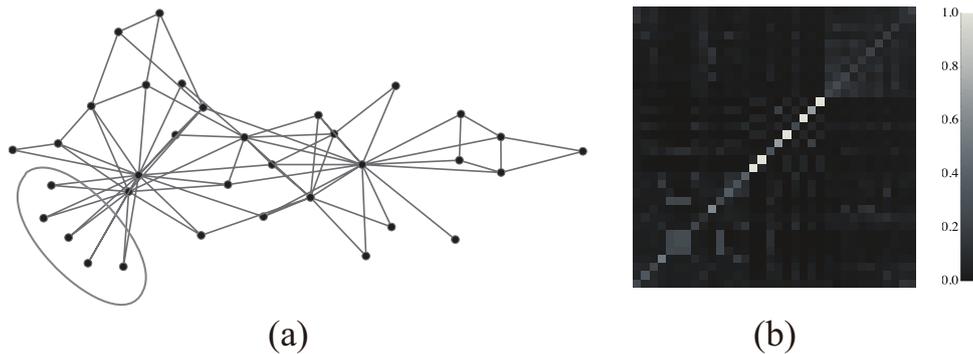


図 13: (a)Zachary の空手クラブのネットワーク上で、灰色の曲線に囲まれた5つのノードが図 12 の局所構造を持っています。(b) そこに量子ウォークが局在しています。

そこで我々は、ネットワーク上での連続時間量子ウォークを考えることによって、ゼロ固有値に対応する固有ベクトルが図 12 のような局所構造に強く局在していることを示しました (図 13) [32]。この局在は、ネットワーク上のランダムネス (例えばランダムポテンシャル) によって引き起こされるアンダーソン局在とは全く異なる局在です。アンダーソン局在は、波が干渉によってポテンシャル内にトラップされる現象ですが、一般に固有値スペクトルの両端の固有状態に起こります。それに対して、ここで議論する局在は、固有値スペクトルのほぼ中央に位置するゼロ固有値の固有状態に起こります。そこで、この局在を「ゼロ固有値局在」と呼んで、特にアンダーソン局在と区別することにしています。

## 7 まとめ

この講義では、まず複雑ネットワークの基礎を説明しました。特にスケールフリー性に着目し、それを再現するモデルを説明しました。また、ネットワーク上のダイナミクスによるモデルも説明しました。最後に、最近の我々の研究を紹介しました。今後、統計物理学の視点からの複雑ネットワークの「普遍性」の研究が進展することを期待します。

## 参考文献

- [1] 増田直紀・今野紀雄「複雑ネットワークの科学」(産業図書, 2005)
- [2] 増田直紀・今野紀雄「複雑ネットワーク 基礎から応用まで」(近代科学社, 2010)
- [3] A.-L. Barabási “Linked: The New Science of Networks” (Perseus Publishing, 2002)
- [4] M. Newman, A.-L. Barabási, D.J. Watts “The Structure and Dynamics of Networks” (Princeton University Press, 2006)
- [5] M. Newman “Networks: An Introduction” (Oxford University Press, 2010)
- [6] R. Cohen, S. Havlin “Complex Networks: Structure, Robustness and Function” (Cambridge University Press, 2010)
- [7] E. Estrada “The Structure of Complex Networks: Theory and Applications” (Oxford University Press, 2012)

- [8] R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74** (2002) 47
- [9] S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. **51** (2002) 1079
- [10] M.E.J. Newman, SIAM review **45** (2003) 167
- [11] W.W. Zachary, J. Anthropol. Res. **33** (1977) 452
- [12] M.E.J. Newman, Proc. Natl. Acad. Sci. USA **98** (2001) 404
- [13] A.-L. Barabási, R. Albert, Science **286** (1999) 509
- [14] R. Albert, H. Jeong, A.-L. Barabási, Nature **401** (1999) 130
- [15] V. Batagelj, A. Mrvar, <http://vlado.fmf.uni-lj.si/pub/networks/data/>
- [16] D.J. Watts and S.H. Strogatz, Nature **393** (1998) 440
- [17] H. Jeong, S. Mason, A.-L. Barabási, Z.N. Oltvai, Nature **411** (2001) 411
- [18] D.J. Watts, S.H. Strogatz, Nature **393** (1998) 440
- [19] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. Lett. **85** (2000) 4633
- [20] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, Rev. Mod. Phys. **80** (2008) 1275
- [21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Phys. Rep. **424** (2006) 175
- [22] G. Szabó, G. Fáth, Phys. Rep. **446** (2007) 97
- [23] K. Szhajd-Weron, Acta Physica Polonica B **36** (2005) 2537
- [24] A. Grabowski, R.A. Kosiński, Physica A **361** (2006) 651
- [25] Q. Li, L.A. Braunstein, H. Wang, J. Shao, H.E. Stanley, S. Havlin, J. Stat. Phys. **151** (2013) 92
- [26] J. Reichardt, S. Bornholdt, Phys. Rev. Lett. **93** (2004) 218701; Phys. Rev. E **74** (2006) 016110
- [27] I. Ispolatov, I. Mazo, A. Yuryev, J. Stat. Mech. (2006) P09014
- [28] S.W. Son, H. Jeong, J.D. Noh, Eur. Phys. J. B **50** (2006) 431
- [29] P. Ronhovde, Z. Nussinov, Phys. Rev. E **81** (2010) 046114
- [30] E. Estrada, N. Hatano, Phys. Rev. E **77** (2008) 036111; *ibid.* **78** (2008) 026102; Appl. Math. Comp. **214** (2009) 500
- [31] E. Estrada, N. Hatano, M. Benzi, Phys. Rep. **514** (2012) 89
- [32] B. Ruben, N. Hatano, in preparation