

複雑ネットワークに対するノード分類法

Categorization Method for Nodes in Complex Networks

湯浅 友幸*¹

Tomoyuki Yuasa

白山 晋*¹

Susumu Shirayama

*¹東京大学大学院工学系研究科

School of Engineering, the University of Tokyo

Nodes are often categorized using some centrality or similarity measures in order to analyze the structure of complex networks. Sometimes a community structure is used for the node categorization. However, there are few studies that the nodes are categorized based on multiple characteristic properties which can be defined at each node such as the degree, local clustering coefficient, local geodesic distance, etc. In this study, we propose a new categorization method for nodes in complex networks. First, we calculate several local characteristic properties at each node, and define the attribute vector of the node which each component corresponds to such properties. Second, the nodes are categorized by clustering multivariate data, i.e. the attribute vectors. SOM-based simple clustering method is used in this paper. Finally, several examples are demonstrated to show how the proposed method works well.

1. はじめに

従来の社会ネットワーク分析 [金光 03] と 90 年代後半に生まれた複雑ネットワーク分析 [増田 10] の大きな違いの 1 つがネットワークの規模にあるとされる。前者では数百程度のノードに対して、個々のノードの属性やリンクの形成過程などの詳細な分析が行われる。後者では、属性や形成過程よりも、数千、数万、あるいは数千万のノードによって構成されるネットワークの位相構造を統計的に分析するという手段がとられることが多い。ただし、前者では対象とする規模が増し、後者では大局的な構造を代表する平均値に基づく統計的指標から内在する局所的な構造へと分析対象が変化することで、分析手法からは両者の区別は難しくなっている [安田 09]。

いずれにしてもネットワーク上で生じる現象を理解するために構造を分析することは重要であり、様々な手法が提案されている。特に、社会ネットワーク分析で用いられてきたいくつかの中心性指標や構造同値といった位相構造の中でノードの役割を見出す方法は両者に共通するものとして発展している。しかしながら、特徴あるノードと多くのノードとの関係性を直接的に調べることは、大規模なネットワークでは未だに難しい。このため、ノードをクラスタに分け、クラスタ内およびクラスタ間の関係性の分析によって、ネットワーク構造を理解するという方法が用いられている。この代表として Newman らのコミュニティ抽出法との併用がある (複雑ネットワーク分析でのコミュニティは、位相構造に基づくノードのクラスタを意味することが多い。本稿ではそれに倣う)。彼らの方法は、ノードが密にリンクする部分を抽出し、ネットワークをコミュニティに分割するというものである [Newman 04]。この方法によってノードはコミュニティによって分類され、同じような属性のノードが同じコミュニティ内に存在するとされる。実際に、多くのネットワークにおいてそれが示されている。

一方、次数や中心性指標などのノード毎に算出できる特徴量によってノードをグループに分類し、グループ毎、あるいはグループ間の関係性を調べるという方法もある。また、Leicht らは、構造同値の考え方を発展させる形で、類似するノードの

連絡先: 湯浅友幸, 東京大学大学院工学系研究科, 〒113-8656
東京都文京区本郷 7-3-1, yuasa@nakl.t.u-tokyo.ac.jp

周囲のリンク構造は類似するという仮定の下でノードの類似度指標を提案し、ノードの分類を試みている [Leicht 06]。

しかしながら、これらの手法による分類では、局所構造が十分に反映されないという課題がある。Newman らのコミュニティ抽出では、モジュラリティ Q と呼ばれるリンクの粗密から得られる指標によってコミュニティが抽出される。したがって、リンクの粗密というネットワーク構造の側面のみを反映したノード分類となる。ノード毎に算出できる特徴量に基づく分類では、1 つの特徴量によるものが多く、この場合も一側面しか反映されない。構造同値、あるいは Leicht らのノードの類似度指標を用いたものでは、ある程度の広範囲のリンク構造が類似したノードが抽出される。しかしながら、異なる性質のノードどうしの類似性、あるいは相違性は 1 つの類似度で表される。ノード毎に算出できる 1 つの特徴量に基づく分類と同様に一側面しか反映されない。ネットワークにはノード毎に算出できる独立した様々な指標が存在する。このため、一側面のみからの分類では局所構造が十分に反映されたものとはいえない。

そこで、本稿では複数の指標を利用したノードの分類法を提案する。はじめに、局所構造を反映する次数、クラスター係数などをノード毎に算出し、これらをノードの属性とする。次に、ノード毎に算出できる平均頂点間距離、媒介中心性などを属性に加える。結果として、ノードの属性は多変量変数として表される。その多変量変数に対してデータマイニング手法を適用しノードを分類する。本稿では、自己組織化マップを利用する。提案手法を 2 つのネットワークに適用し、コミュニティ構造に基づく分類と比較し、その違いについて考察する。また、ネットワーク上のシミュレーションの可視化分析への応用について述べる。

2. 提案手法

2.1 ノードの属性

はじめに、ノード v_i に、ネットワークの局所構造を反映すると考えられる複数の属性値を与える。それらの属性値を成分とする多変量変数 n_i を定義する。

属性には、次数のような直接的に局所構造を反映するものと、平均頂点間距離のようなネットワーク全体の影響を受ける

C ₂₁	C ₂₂	C ₂₃	C ₂₄	C ₂₅
C ₁₆	C ₁₇	C ₁₈	C ₁₉	C ₂₀
C ₁₁	C ₁₂	C ₁₃	C ₁₄	C ₁₅
C ₆	C ₇	C ₈	C ₉	C ₁₀
C ₁	C ₂	C ₃	C ₄	C ₅

図 1: SOM 内のカテゴリへのナンバリング

が局所的に（ノード毎に）定義できるものの 2 種類が考えられる．本稿では，前者として，次数 k ，隣接ノードの平均次数 k_{nn} ，クラスター係数 C を，後者として，媒介中心性 b ，平均頂点間距離 L を扱い，両者を区別せず \mathbf{n}_i の成分とする．

ここで，隣接ノードの平均次数 k_{nn}^i はノード v_i とリンクするノードの平均次数である．また，ノード v_i の媒介中心性 b_i は v_i が他の任意の 2 ノード間をつなぐ最短路上に存在する割合を示す特徴量である．

以上を成分とする多変量変数 $\mathbf{n}_i = (k_i, k_{nn}^i, b_i, L_i, C_i)$ を用いてノードを分類する．

2.2 自己組織化マップ (SOM) によるノード分類

ノード属性を示す多変量変数 \mathbf{n}_i に対してデータマイニング手法を用いてノードを分類する．

被説明変数のない多変量変数を分類する方法としていくつかのクラスタリング手法が提案されている．類似度（あるいは非類似度）の定義の正当性，分類結果に対し成分の寄与を判別できること等によって，手法の選択が行われることが多い．本稿では，自己組織化マップ (SOM) を用いた．その理由は，ノード属性に対する明確な類似度の定義が不要であること，SOM の結果をそのまま分類に利用できること，決定木によって成分の寄与を判別できることである．また，主成分分析との比較ができる点 [北 10] にも注目した．

SOM は 2 次元マップとし，その表現として $M \times M$ の 2 次元格子を用いる．本稿で用いる 5×5 の 2 次元格子を図 1 に示す．ここで，1 つ 1 つの格子をセルと呼び，セルを図 1 に従って $C_X (X = 1, 2, \dots, 25)$ で識別することにする．SOM のアルゴリズムによって， N 個のノードがいずれかのセル内に配置される．一般的には，配置されたものを凝集法などのクラスタリング手法によってグループに分ける．本稿では，類似の特徴を持つデータが近い位置に配置されるという SOM の特徴を考慮し，セル自身をグループとして扱う．セルの数や配置後のクラスタリングなどは今後の課題である．

また，1 つのカテゴリが 1 つのセルに相当するものとし，カテゴリ自体を C_X で表す（図 1）．例えば， C_1 のセルに配置されたノードは，カテゴリ C_1 に属するノードとして分類される．

各カテゴリに分類されたノードの特徴量の平均からヒートマップを作成し，SOM の分類傾向について調べる．

2.3 コミュニティ抽出によるノード分類との比較

提案手法との比較のために，コミュニティ抽出に基づくノード分類を行う．コミュニティ抽出法としては，Newman らのもの [Newman 04] を用いる．本稿では，Clauset らのアルゴリズム [Clauset 04] により，コミュニティ抽出を行い，属するコミュニティの違いによってノードの分類を行う．

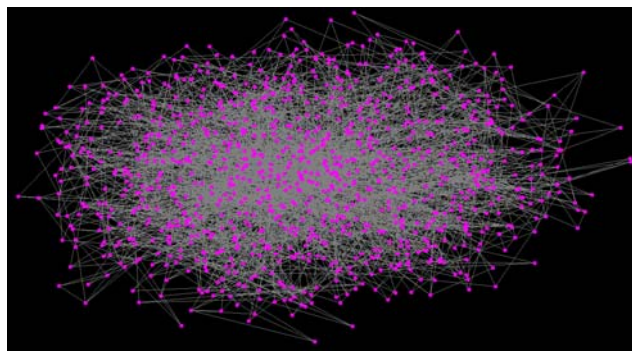


図 2: HK モデルで生成したネットワーク

3. 実験と考察

3.1 ネットワークモデル

ネットワークモデルによって生成したネットワークに対してノード分類を行う．本稿では，HK モデル [Holme 02] と KE モデル [Klemm 02] を用いた．いずれのネットワークでもノード数は 1000，平均次数は 6 とした．図 2 に HK モデルで生成したネットワークを示す．ノードの配置は，Pajek [Batagelj 03] における Kamada-Kawai アルゴリズム [Kamada 89] により決定した．

3.2 ノード分類の結果

はじめに，図 4 と図 5 に HK モデルと KE モデルによって生成されたネットワークに対する SOM による分類結果を示す．図では，ノードの属性である，次数，隣接ノードの平均次数，クラスター係数，媒介中心性，平均頂点間距離に対して，各カテゴリ内の平均値（セル内の数値）に基づいてセルが色付けされている．なお，各図の下段右に各カテゴリに属するノード数を示している．また，図 2 のノードに対してカテゴリ毎に色付けしたものを図 3 に示す．配色はカテゴリ番号に対応させ，赤橙黄緑青藍紫の順に変化させている．番号が小さなカテゴリほど紫に近い色，大きなカテゴリほど赤に近い色である．

図 4 と図 3 を比べると，図 4 では，各属性値の大小に基づくノードの分類が直感的にとらえられることがわかる．例えば，図 4 ではマップのより左側のカテゴリの方が k の値が大きく，より下側の方が k_{nn} が大きな傾向にある．また，こうした傾向は属性値ごとに表れている．一方，図 3 からは，隣接関係にあるノードが必ずしも同じカテゴリにはないことがわかる．両者を併用することでネットワーク構造のより深い理解が可能になるものと期待できる．

次に，同じネットワークに対して，コミュニティ抽出による分類を行った結果を示す．抽出されたコミュニティを $G_X (X = 1, 2, \dots)$ で表す．表 1 と表 2 は，それぞれ HK モデルと KE モデルによって生成されたネットワークに対する分類結果である．各コミュニティ内のノード数と各属性値の平均を示している．

2 つの表からコミュニティに基づく分類では，ノードの属性値の大小が反映されていないことがわかる．例えば，図 4 の分類と表 1 の分類を比べると大きな違いがある．図 4 からは，カテゴリ間での各属性値の差違が明確であることがわかるが，表 1 からコミュニティ間の差違を読み取ることは難しい．また，図 5 の分類と表 2 の分類の比較からも同様の傾向が得られる．このように，コミュニティに基づく分類では，位相構造から求められるノードの属性値から各コミュニティの識別は難しいことがわかる．

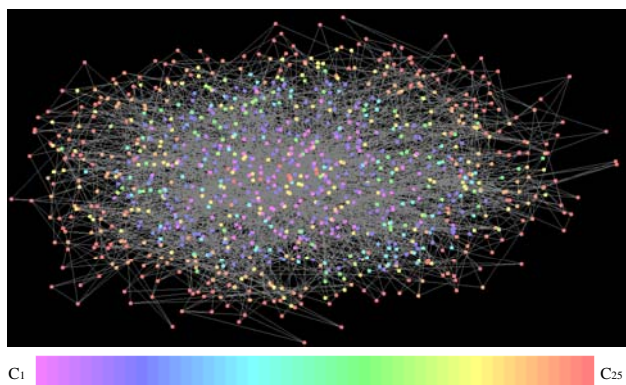


図 3: 提案手法による分類の可視化結果 (HK モデル)

また, SOM による分類とコミュニティ構造による分類との間でモジュラリティ Q の値を比較すると, HK モデルでは, SOM による分類で -0.016, コミュニティ抽出による分類で 0.651 となった. また, KE モデルでは SOM による分類で -0.021, コミュニティ抽出による分類で 0.753 となった. 上述のノード属性に対する SOM による分類でのカテゴリに属するノードとモジュラリティ Q に基づくコミュニティに属するものとは異なるものと考えられる.

これらはネットワークの可視化からも示される結果である. 図 3 は, HK モデルに対して SOM によるノード分類の結果を各カテゴリごとに色分けして示した図である. 図 6 は, コミュニティに基づくノード分類の結果を各コミュニティごとに色分けして示した図である. ノードの配色は, カテゴリ番号, あるいはコミュニティ番号に対応させ, 赤橙黄緑青藍紫の順に変化させている. また, ノードの配置は, Pajek[Batagelj 03] における Kamada-Kawai アルゴリズム [Kamada 89] により求められている. なお, このアルゴリズムでは, 次数の高いノードほど中央近くに配置されやすい傾向にある.

はじめに, HK モデルの可視化結果を比較すると図 3 においては同じ色のノードがおおよそ同心円状に配置される傾向にあることがわかる. 特に, 最も外側には赤色のノードが存在する. これはカテゴリ 25 にあたり, 最も次数 k の小さなカテゴリである. 一方, 中央付近に散在する黄色のノードは最大の k を示すカテゴリ 21 に相当する. このように SOM による分類は, ノードの属性値を反映することが可視化からも確認される. 一方, 図 6 のコミュニティに基づく分類では同色のノードが固まって配置される傾向にある. これは位相的に距離の近いノードどうしが同じコミュニティに属しやすいという Newman らのコミュニティ抽出法の性質を表している. これらの傾向は, KE モデルの場合により顕著に表れている.

以上のように, 提案した SOM によるノードの分類法は, 従来のものとは異なる分類法であることがわかる.

3.3 シミュレーションの可視化分析への応用

従来のネットワーク上のシミュレーションの可視化は, 図 2 のようにレイアウトされたネットワーク上で行われるものが多い. 位相構造とシミュレーション結果の把握を同時に行わなければならない, 現象の詳細が理解しにくいという問題があった. また, グラフレイアウトの手法によって結果の解釈が変わるといった問題がある. 提案手法の SOM によるノード分類上でシミュレーション結果を可視化することにより, ノードの属性値である構造の特徴量と現象の関係性をより明確に示すことがで

表 1: コミュニティ構造に基づく分類結果 (HK モデル)

	k	k_{nn}	b	L	C	ノード数
G ₁	6.33	20.07	7.9e-3	3.66	0.57	165
G ₂	6.24	17.67	6.8e-3	3.70	0.53	210
G ₃	6.32	15.96	6.1e-3	3.85	0.55	178
G ₄	5.72	16.21	4.2e-3	3.79	0.56	109
G ₅	5.84	14.01	5.2e-3	4.03	0.52	106
G ₆	5.65	14.47	4.2e-3	3.94	0.55	99
G ₇	5.40	12.67	2.9e-3	3.93	0.45	35
G ₈	5.53	14.12	3.7e-3	3.98	0.51	70
G ₉	4.92	10.37	1.8e-3	4.35	0.58	12
G ₁₀	4.56	16.81	2.0e-3	3.84	0.68	9
G ₁₁	4.57	13.14	1.5e-3	3.83	0.40	7

表 2: コミュニティ構造に基づく分類結果 (KE モデル)

	k	k_{nn}	b	L	C	ノード数
G ₁	6.34	19.91	7.3e-3	4.44	0.48	161
G ₂	6.28	14.02	8.0e-3	4.53	0.48	64
G ₃	6.09	21.72	7.9e-3	4.45	0.57	204
G ₄	5.83	16.57	7.4e-3	4.69	0.58	177
G ₅	5.94	11.30	6.7e-3	4.77	0.63	36
G ₆	5.99	10.32	7.2e-3	4.90	0.56	152
G ₇	5.71	11.81	6.9e-3	5.01	0.54	93
G ₈	5.63	16.85	6.9e-3	4.85	0.56	113

きる. その詳細については講演で示す.

4. 結言

複雑ネットワークのノードを, 局所構造を反映した複数の特徴量に基づき分類するという新たな手法を提案した. 本稿では, ノード毎に算出された 5 つの特徴量を多変量変数として扱い, 自己組織化マップによってノードを分類した.

提案手法を 2 つのネットワークに対して適用した結果, 各特徴量の大小, および相関関係を直感的に把握できるノードの分類が可能であることがわかった. また, コミュニティ抽出に基づくノード分類との差異を示した.

提案手法とグラフレイアウトに基づくネットワーク可視化を併用することでネットワーク構造のより深い理解が可能になるものと期待される.

参考文献

- [Batagelj 03] Batagelj, V. and Mrvar, A.: Pajek - analysis and visualization of large networks, *Lecture Notes in Computer Science*, Vol. 2265, pp. 77-103 (2003)
- [Clauset 04] Clauset, A., Newman, M., and Moore, C.: Finding community structure in very large networks, *Physical Review E*, Vol. 70, No. 6, p. 066111 (2004)
- [Holme 02] Holme, P. and Kim, B.: Growing scale-free networks with tunable clustering, *Physical Review E*, Vol. 65, No. 2, p. 26107 (2002)

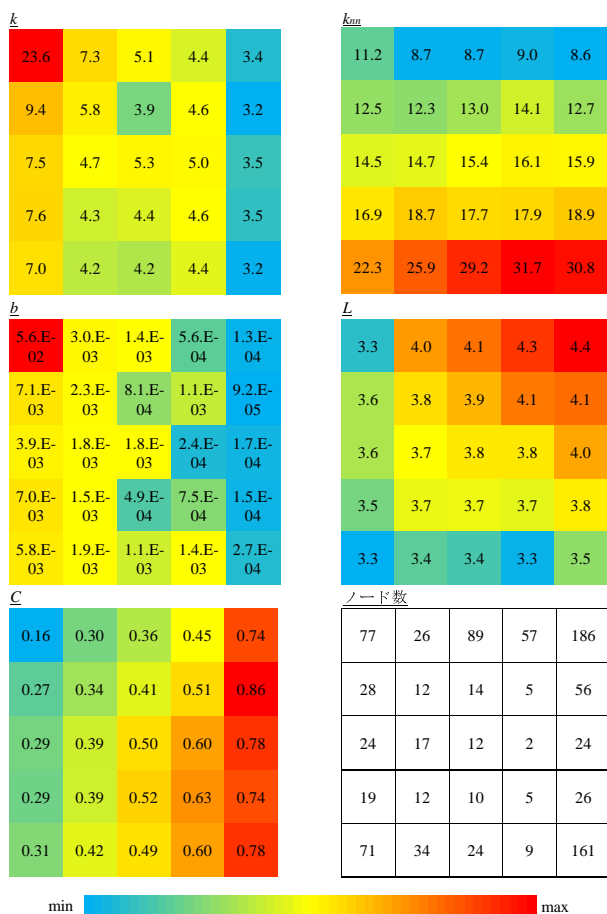


図 4: 提案手法による分類結果 (HK モデル)

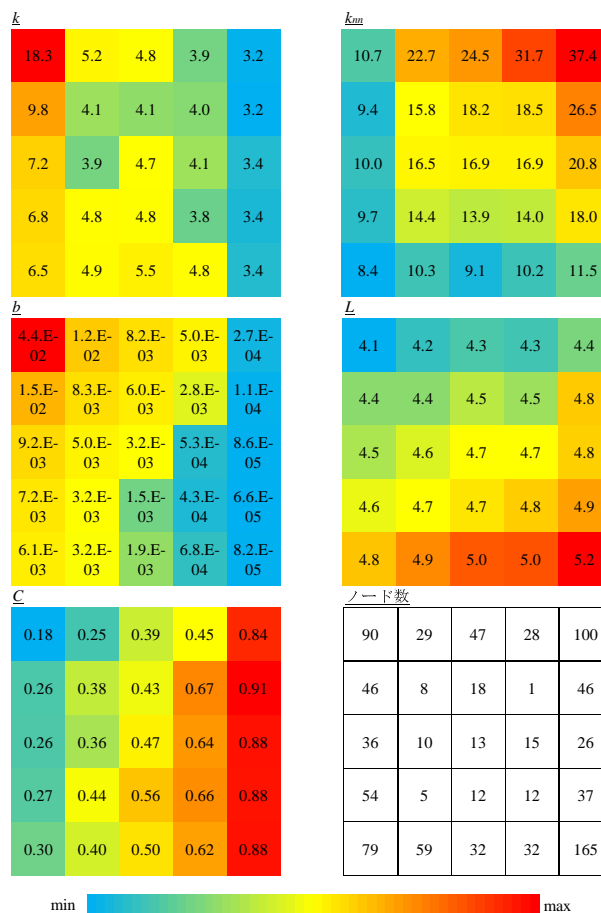


図 5: 提案手法による分類結果 (KE モデル)

- [Kamada 89] Kamada, T. and Kawai, S.: An algorithm for drawing general indirect graphs, *Information processing letters*, Vol. 31, pp. 7–15 (1989)
- [Klemm 02] Klemm, K. and Eguíluz, V.: Growing scale-free networks with small-world behavior, *Physical Review E*, Vol. 65, No. 5, p. 57102 (2002)
- [Leicht 06] Leicht, E., Holme, P., and Newman, M.: Vertex similarity in networks, *Physical Review E*, Vol. 73, No. 2, p. 026120 (2006)
- [Newman 04] Newman, M. and Girvan, M.: Finding and evaluating community structure in networks, *Physical Review E*, Vol. 69, No. 2, p. 026113 (2004)
- [安田 09] 安田雪: Pajek を活用した社会ネットワーク分析, 東京電機大学出版局 (2009)
- [金光 03] 金光淳: 社会ネットワーク分析の基礎 社会関係資本論に向けて (2003)
- [増田 10] 増田 直紀: 複雑ネットワーク 基礎から応用まで, 近代科学社 (2010)
- [北 10] 北 雄介, 門内 輝行: 自己組織化マップを用いた都市の様相の分析, Design シンポジウム 2010 論文集, pp. dss10-00174 (2010)

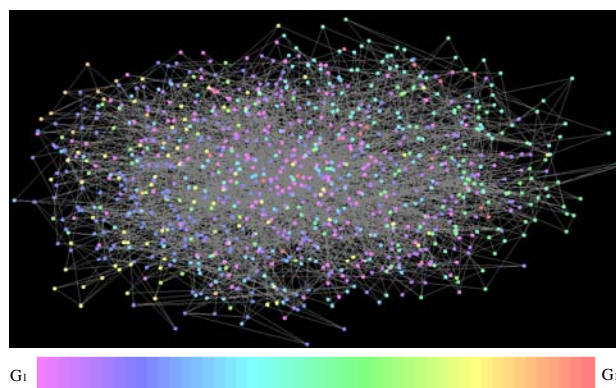


図 6: コミュニティ構造による分類の可視化結果 (HK モデル)