

ネットワーク上の現象に対する新しい分析法

湯浅 友幸^{†1} 白山 晋^{†1}

ネットワーク上の現象は大局的構造から説明されることが多く、局所的な見地から分析されることは少ない。本稿ではノードの性質に注目した新たな分析法を提案する。提案手法では自己組織化マップ(SOM)を用いたノード分類に基づくシミュレーションの可視化と分析を行う。そして、2つのシミュレーションモデルを用いた実験により提案手法の有効性を示す。

New Analysis Method for Phenomena on Networks

TOMOYUKI YUASA^{†1} and SUSUMU SHIRAYAMA^{†1}

Most researches concerning the influence of network structure on phenomena on the network are carried out based on relationships between global statistics of the network structure and characteristic properties of those phenomena, even though the local structure has a significant effect on dynamics of some phenomenon. In this paper, we propose a new analysis method for some phenomena on networks based on the categorization of nodes. First, local statistics such as the average path length, the clustering coefficient for a node are calculated and are given to each node. Secondly, the nodes are categorized by Self-Organizing Map (SOM) and are grouped. Characteristic properties of some phenomenon are visualized into the grouped nodes. From some numerical results using two simulation models, the validity of our method is examined.

1. はじめに

感染症の流行や噂の伝播、集団の合意形成などの世の中の諸現象を背後に存在するネットワーク構造に基づき分析する研究が広くなされている¹⁾。これらの研究の多くでは行動主体

をノード、行動主体間の関係性をリンクで模したネットワーク上で現象をモデル化したシミュレーションの分析を行う。

従来のネットワーク上の現象を分析する方法、あるいは方法論としては、平均頂点間距離やクラスター係数といったネットワーク構造の特徴を表す統計的指標との関係性を調べるものや、次数や媒介中心性などが最も高いノードに着目し、現象との関係性を分析するものが多い。前者はネットワークの大局的構造から、後者は重要なノードから現象を説明するものである。しかしながら、これらの手法だけではネットワーク上の現象の詳細を知ることは難しい。前者では局所的な見方が欠け、後者では重要ノードのみに焦点が当てられるので全体との関係性を見出すことが難しいためである。

これらの欠点を解決する手法として、大局と局所を関係付ける中間的な構造としてのコミュニティと現象の関連性を調べるものがある。しかし、コミュニティの構造はコミュニティ抽出法に依存し、現象との関連性を調べる前にコミュニティ自体の意味付けが問われることが多い。

また、可視化による分析も行われるが、グラフィアウトの任意性に起因する解釈の難しさという問題がある。例えば、格子のように規則的なレイアウトができるネットワークに対する可視化結果の解釈は容易であるが、大規模かつ複雑なネットワークでは文献2)のようにレイアウト自体が煩雑になるため、その上での結果の分析が困難になる。

以上から、ネットワーク上の現象を局所的な見地を踏まえた上でより詳細に分析するためには新たな分析法が必要になる。

本稿では、ネットワーク内のノードをその特徴量に基づいて分類し、分類されたグループに対して可視化を行うという手法を提案する。コミュニティに属するノードをノード分類の結果と考えると、コミュニティを用いる方法でも行われるものといえる。しかし、多くのコミュニティ抽出法がネットワーク構造の1つの特徴量(例えば、モジュラリティ Q)に基づくため、ノード分類はネットワーク構造の一側面を反映したものにすぎないと考えられる。そこで、提案手法では、ノードの属性をいくつかの局所的な特徴量を成分とする多変量変数とし、多変量変数に対してデータマイニングを行うことでノードを分類し、多角的な分析を試みる。また、2つのシミュレーションモデルを用いたシミュレーション結果に提案手法を適用し、手法の有効性を示す。

^{†1} 東京大学大学院工学系研究科
School of Engineering, the University of Tokyo

2. 提案手法

2.1 ノードの分類

2.1.1 ノードの属性

平均頂点間距離やクラスター係数などはネットワーク全体の構造の特徴を表す統計的指標であるが、多くはノード毎に算出できるものである。本研究では、その点に着目し、ノード v_i に属性を与える。具体的には、次数 k_i 、隣接ノードの平均次数 k_{nn} 、媒介中心性 b_i 、平均頂点間距離 L_i 、クラスター係数 C_i を成分とする多変量変数 \mathbf{n}_i をノード v_i の属性とする。

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り、次式で求められる。他の任意の2ノードを、始点ノード v_{i_s} 、終点ノード v_{i_t} とすれば、

$$b_i = \frac{\sum_{i_s=1; i_s \neq i}^N \sum_{i_t=1; i_t \neq i}^{i_s-1} \frac{g_i^{(i_s i_t)}}{N_{i_s i_t}}}{(N-1)(N-2)/2} \quad (1)$$

である。ここで、 $g_i^{(i_s i_t)}$ は始点 v_{i_s} から終点 v_{i_t} への最短路の中で v_i を通るものの数、 $N_{i_s i_t}$ は v_{i_s} から v_{i_t} への最短路の総数である。また、 N はノード数であり、分母は規格化定数である。

平均頂点間距離 L_i はノード v_i から他の任意のノードへの最短路の平均であり、ノード v_i とノード v_j の最短路の距離を $d(v_i, v_j)$ とすれば次式となる。

$$L_i = \sum_{i \neq j} \frac{d(v_i, v_j)}{N-1} \quad (2)$$

クラスター係数 C_i は v_i とリンクするノード間に存在するリンク数を E_i とすれば、その次数 k_i を用いて次式で表される。

$$C_i = \frac{E_i}{k_i(k_i-1)/2} \quad (3)$$

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。

2.1.2 自己組織化マップによる分類と可視化

本稿では、ノードを分類する手法として自己組織化マップ(SOM)を用いる。また、SOMの結果の表現としては2次元格子を用いることにした。図1に、 5×5 格子の場合を示す。

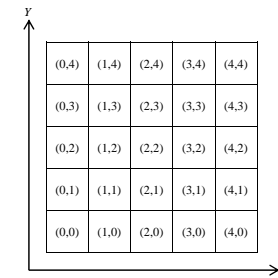


図1 SOM内のカテゴリへのナンバリング
 Fig.1 Categories' Number in SOM

ここで、一つ一つの格子をセルと呼び、セルを図1の座標軸に従って (X, Y) 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 $M \times M$ のいずれかのセル内に配置される。類似の特徴を持つデータが近い位置に配置されるのがSOMの特徴である。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、さらにグループに分けるが、本稿ではセル自身をグループとして扱う。また、セルが一つのカテゴリに相当するものとし、カテゴリ自体を (X, Y) で表す(図1)。

SOMの分類傾向については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、確認を行う。

2.2 シミュレーション結果の可視化

本稿のネットワーク上のシミュレーションはノードに対してカテゴリカルな状態を与えられるものとする。SOMでは近い性質を持つデータが近くに配置されるため、近傍のデータをまとめるような可視化によってシミュレーション結果の傾向が把握しやすくなる。そこで、セル内の各状態のノードの割合を示す円グラフによる可視化を行う。また、状態の変化を時系列で表現する。図2に一例を示す。

SOMの分類傾向を示すヒートマップとSOM上のシミュレーション傾向を比較することによりネットワーク上の現象をノードの特徴量に基づいて分析する。

3. 実験と考察

3.1 シミュレーションモデル

3.1.1 ネットワーク生成

本稿では、後述するHKモデル³⁾とCNNモデル⁴⁾の2種類のネットワークモデルによっ

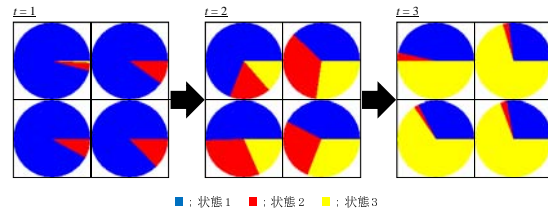


図2 シミュレーションの可視化例
 Fig.2 Example of Visualization

ネットワークを生成する．ノード数は、 $N = 10000$ ，平均次数は $\langle k \rangle \approx 8$ とする．

HK モデルは Holme らによって提案されたネットワークモデルである．ネットワークの成長と優先的選択，クラスター形成の3つのプロセスによって生成される．ノードの次数分布がべき乗分布に従うスケールフリー性を有する．また，ノードのクラスター係数が高い．

CNN モデルは Vázquez によって提案されたネットワークモデルである．ネットワークの成長と潜在リンクのリンク化のプロセスにより生成される．スケールフリー性を有し，クラスター係数が高い．また，次数の近いノードどうしがリンクしやすい同質局在性を持つ．

これらのネットワークを用いて，以下に示すネットワーク上の感染症伝播と空間囚人のジレンマのシミュレーションを行う．

3.1.2 ネットワーク上の感染症伝播

感染症伝播を表すモデルとして SIR モデルを用いる．各ノードは，S (感染状態; Susceptible)，I (感染状態; Infectious)，R (隔離状態; Removed) の3状態をとり，隣接するノード状態を加味した状態遷移を行う．

病原の感染率，治癒率をそれぞれ λ ， μ ，隣接する状態 I のノード数を $n(I)$ とすれば，以下の (a) ~ (c) のプロセスで微小時間 dt が進行する．

- (a) ランダムにノードを一つ選択する．
- (b1) 選択ノードの状態が S であれば，確率 $\lambda n(I)dt$ で選択ノードの状態を I に変更する．
- (b2) 状態が I であれば，確率 μdt で選択ノードの状態を R に変更する．
- (b3) 状態が R であれば，特に処理を行わない．
- (c) (a)，(b) をノード数の回数分繰り返す．

本稿では，はじめに状態 I のノードをネットワーク内にそれぞれ 10 ノード配置し，残りを状態 S とする．そして，系から状態 I のノードが消滅するまで上記のプロセスを繰り返す．

表 1 SPD の利得行列

Table 1 Payoff Matrix of SPD

	C (協力)	D (裏切)
C (協力)	1,1	0,T
D (裏切)	T,0	0,0

す．また，感染率 $\lambda = 0.2$ ，治癒率 $\mu = 1$ ，離散時間 $dt = 0.01$ とした．

3.1.3 空間囚人のジレンマ

空間囚人のジレンマ (SPD) は社会的ジレンマ構造における集団の意思決定を模したモデルである．各ノードは C (協力戦略; Cooperator)，D (裏切戦略; Defector) の2戦略をとり，隣接するノード状態を加味した戦略変更を行う．本稿では，以下の (a)，(b) を繰り返して1ステップが進行し，全てのノードが同期的に戦略を変更するモデルを用いる．

- (a) 各ノードは隣接する全てのノードと対戦し，各々の戦略と表 1 に基づいた利得を得る．
- (b) 自分と周囲のノードの中で最も利得の大きいノードの戦略を採用する．ただし，協力ノードと裏切ノードの最大利得が等しい場合は戦略を変更しない．

本稿では，ネットワーク内に協力，裏切のノードを半数ずつランダムに配置し，両戦略のノード数の変動がなくなるまで上記のプロセスを繰り返す．また，表 1 の T は 1.5 とした．

3.2 ノード分類の結果

はじめに HK モデル，CNN モデルによって生成されたネットワークに対して，ノードの属性 n_i を算出し，SOM の 5×5 のセルを利用してノードを分類する．次に，各カテゴリ (セル) 内のノードに対して， k ， k_{nn} ， b ， L ， C の平均値を計算し，平均値に基づくヒートマップを表示すると，図 3 と 4 になる．図は，それぞれ HK モデル，CNN モデルのものである．また，各セル内のノード数に対するヒートマップを図の左上に示してある．

この分類傾向をまとめると HK モデルから作られたネットワークに対しては以下の性質がある．

- k に関しては， X 軸正方向ほど値が大きい．例外はあるが， Y 軸正方向ほど値が大きい．特に (4, 4) で最大値をとる．
- k_{nn} に関しては， Y 軸正方向ほど値が大きい． X 軸方向は中央ほど値が大きいケースが多い．
- b に関しては， X 軸， Y 軸それぞれ正方向ほど値が大きい．最大値をとる (4, 4) で他のカテゴリより非常に大きな値になっている．
- L に関しては， X 軸， Y 軸それぞれ正方向ほど値が小さい． b の場合と異なり，カテゴ

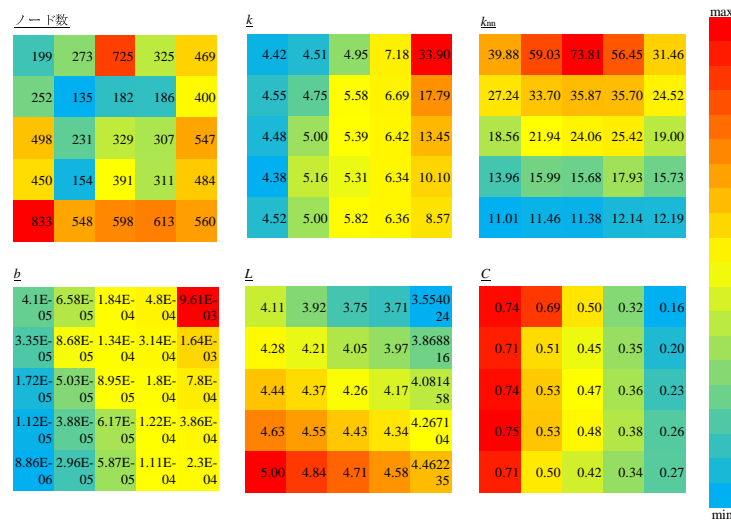


図3 HKモデルの分類結果に対するヒートマップ
Fig.3 Heat Map for Categorization of nodes in HK Model

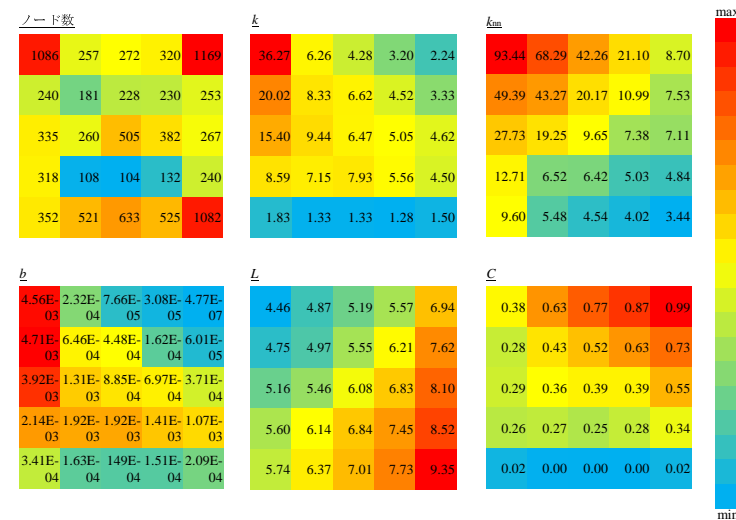


図4 CNNモデルの分類結果に対するヒートマップ
Fig.4 Heat Map for Categorization of nodes in CNN Model

りごとの値は緩やかに変動する。

- ・ C に関しては、 X 軸負方向ほど値が大きい。 Y 軸方向に関しては正負方向と値の大小に大きな相関はない。

CNNモデルから作られたネットワークに対しては以下である。

- ・ k に関しては、 X 軸負方向ほど値が大きい。 Y 軸方向に関しては正負方向と値の大小に大きな相関はない。
- ・ k_{nn} に関しては、 X 軸に関しては負方向ほど、 Y 軸に関しては正方向ほど値が大きい。
- ・ b に関しては、 X 軸負方向ほど大きな値をとる。 Y 軸に関しては $(0, 4)$ 、 $(0, 3)$ のケース以外は中央付近で大きな値をとる。
- ・ L に関しては、 X 軸に関しては負方向ほど、 Y 軸に関しては正方向ほど値が小さい。
- ・ C に関しては、 X 軸、 Y 軸それぞれ正方向ほど大きな値をとる。

3.3 ネットワーク上の感染症伝播の分析

代表的な時間ステップの結果を提案手法に基づいて可視化したものを図5と6に示す。図5は、HKモデルによって生成されたネットワーク上のシミュレーション結果の可視化であ

り、図6はCNNモデルに対するものである。

図5 (HKモデル)からは、はじめに $t = 3.0$ において $(4, 4)$ を中心に感染が拡大し始めることがわかる。これは図3における k の大きな領域に一致し、ハブから感染が始まることを示している。次に、 $t = 4.0$ において X 軸、 Y 軸ともに正方向に値が大きい $(0, 4) \sim (3, 4)$ 、 $(4, 1) \sim (4, 3)$ といったカテゴリ(セル)を中心に感染が拡大する。これらのカテゴリは図3における k または k_{nn} の大きなものであり、 L の小さなものに一致する。以降 $t = 5.0$ 、 $t = 7.5$ において X 軸、 Y 軸ともに負方向のカテゴリへと感染が波及していく。このため、 k や k_{nn} 、 L の小さなノード群には感染が広がりにくいことがわかる。最後に、 $t = 16.78$ の終状態を見ると、黄色で示される総感染者数の割合は図3の b の大小の傾向とほぼ一致する。HKモデルで生成されたネットワークでは、最短路に位置しやすいノードは感染の伝播経路にもなりやすいことを示している。

一方、図6 (CNNモデル)では、はじめに $t = 0.5$ において $(0, 4)$ で爆発的に感染が広がる。これは図4における k 、 k_{nn} ともに最大値をとるカテゴリであることから、ハブの相互感染によって急激に感染が広がるものと考えられる。次に、 $t = 1.0$ 、 $t = 1.5$ において

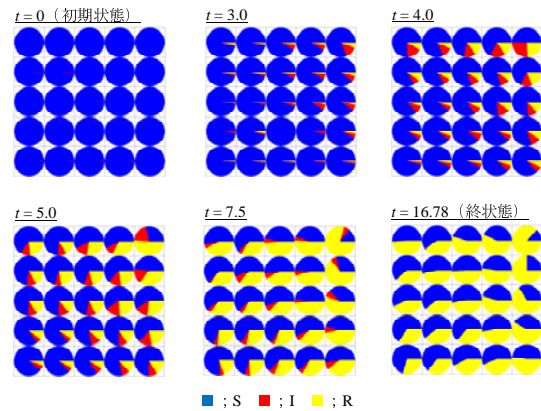


図5 HKモデルにおけるSIRモデルのシミュレーションの可視化
Fig.5 Visualization of Results by SIR Model on HK network

(0, 4) 付近のカテゴリで感染が広がる．これらのカテゴリは k_{nn} が大きい， k は必ずしも大きくない．さらに， $t = 2.0$ ， $t = 3.0$ ， $t = 5.0$ にわたって X 軸負方向， Y 軸正方向のカテゴリを中心に感染が広がる．一方， $(0, 0) \sim (0, 4)$ ， $(0, 4) \sim (4, 4)$ といったカテゴリではほとんど感染が広がらない．これらのカテゴリは k ， k_{nn} が小さく， L が大きなカテゴリにあたり，ハブを含む他ノードとのリンクが少なく，他のノードとの距離が大きなノードが感染しにくい傾向が HK モデルの場合よりも顕著に現れる．最後に， $t = 12.6$ の終状態では，その総感染者数の割合は図 4 における k_{nn} が大きな順， L が小さな順にほぼ一致する．HK モデルの場合と異なり，CNN モデルでは b が有意な影響を及ぼさないことがわかる．

3.4 空間囚人のジレンマの分析

SPD のシミュレーションに対して，代表的な時間ステップの結果を提案手法に基づいて可視化したものを図 7，8 に示す．なお，SPD においては協力，裏切のいずれもが支配的になりうるが，本節では協力が支配的になる場合を示す．

図 7 (HK モデル) からは，はじめに $t = 1$ ， $t = 2$ ， $t = 3$ ， $t = 6$ において，協力が一時的に増加するカテゴリがあるものの最終的には裏切がネットワーク内に広がることがわかる．一時的な協力の拡大が顕著なのは $(2, 4)$ ， $(3, 4)$ ， $(4, 4)$ である．これらは図 3 における k_{nn} が最大のカテゴリ， k_{nn} が 3 番目に大きなカテゴリ， k が最大のカテゴリである．次に， $t = 8$ では協力が拡大するが， $t = 13$ では裏切が再度支配的になる．この際も協力の拡大の

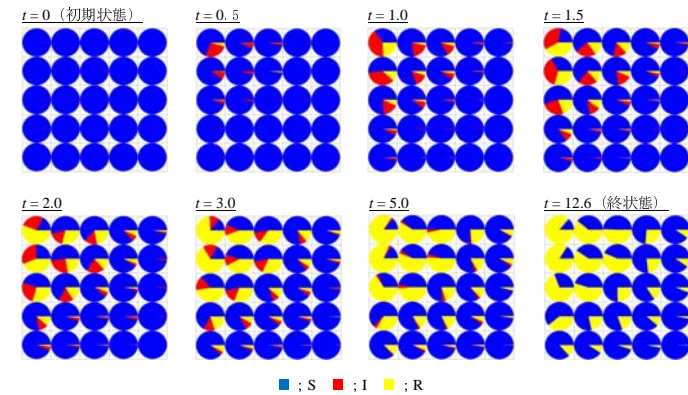


図6 CNNモデルにおけるSIRモデルのシミュレーションの可視化
Fig.6 Visualization of Results by SIR Model on CNN network

中心は $(2, 4)$ ， $(3, 4)$ ， $(4, 4)$ である． $t = 14$ ， $t = 15$ では，協力の拡大が見られるが裏切が再度支配的になることはない．以降の $t = 16$ ， $t = 17$ では k_{nn} の大きなカテゴリから，特に k の大きなものから協力が拡大していく．終状態の $t = 21$ において裏切戦略をとるものは k_{nn} の小さな $(0, 0)$ ， $(0, 1)$ などである．以上の結果から HK モデルによって生成されたネットワーク上の SPD では一旦裏切が支配的になるが，ハブの協力戦略が波及することで協力が支配的になるといえる． k_{nn} が高いノードから協力が拡大するのもこれらのノードがハブとリンクするためである．

一方，図 8 (CNN モデル) では，はじめに $t = 1$ ， $t = 2$ ， $t = 4$ において，全体的に裏切がネットワーク内に広がる．このとき，裏切の拡大は $(0, 4)$ を中心とした k ， k_{nn} の大きなカテゴリから始まる．次に， $t = 6$ では協力の拡大が $(0, 4)$ を中心に始まる．以降 $t = 7 \sim t = 12$ に至るまで協力の拡大が続く．これは図 4 における k ， k_{nn} が大きなカテゴリ，特に k_{nn} の大きなカテゴリを中心としたものである．また，これらのカテゴリでは L が小さくなる．また，時間ステップが進行しても X 軸正方向かつ Y 軸負方向に存在するカテゴリでは協力が波及しない．終状態である $t = 16$ においても協力が支配的になる中でこれらのカテゴリでは裏切が残っている．一方，図 4 からは，これらのカテゴリは $(1, 0) \sim (4, 0)$ のような k の非常に小さなカテゴリ，または， $(4, 1) \sim (4, 4)$ のような k と k_{nn} が小さい中でも特に C が大きな値をとるカテゴリであることがわかる．他のノードとのリンクの少なさや

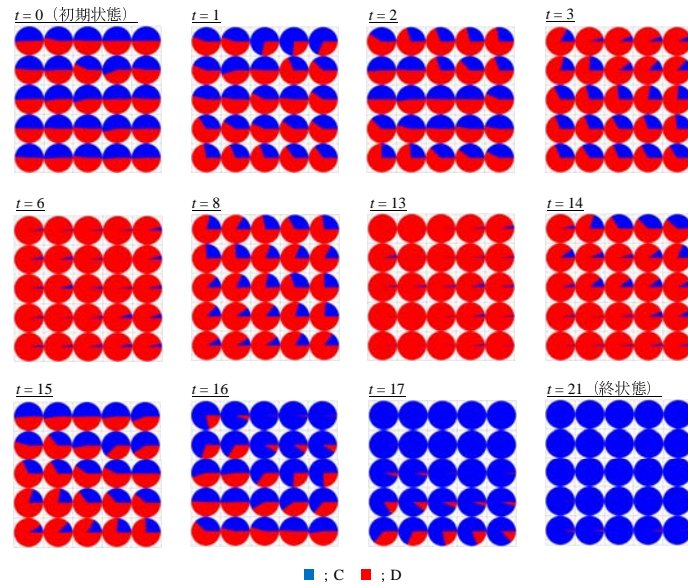


図7 HKモデルにおけるSPDシミュレーションの可視化
Fig.7 Visualization of Results by SPD on HK network

C の大きなノード周辺のリンクが密な特殊な構造が裏切の生き残りを促すものと考えられる。以上の結果から CNN モデルにより生成されたネットワーク上の SPD では、HK モデルでのものと同様に、はじめに裏切が支配的になるが、 k の大きなハブの協力が k_{nn} の大きなノードを中心に波及することがわかる。一方で、HK モデルの場合と比べると、 k の小さなノード、 C の大きなノードにおいて裏切が生き残りやすくなる。

3.5 考察

2つのネットワークモデルによって生成されたネットワーク上での2種類のシミュレーション結果からは、特に SIR モデルを用いた感染伝播シミュレーションでは、 k の大きなハブからの感染の波及が、空間囚人ジレンマからは同様にハブからの協力の波及が見られた。これらの知見は従来研究^{5),6)}でも指摘されていたものだが、提案手法による分析では、 k が小さくとも k_{nn} の大きなノードが波及を促すという結果も得られた。また、感染症や協力の波及や、終状態における各状態のノードの割合に対しては、 k 、 k_{nn} だけでなく b 、 L 、 C

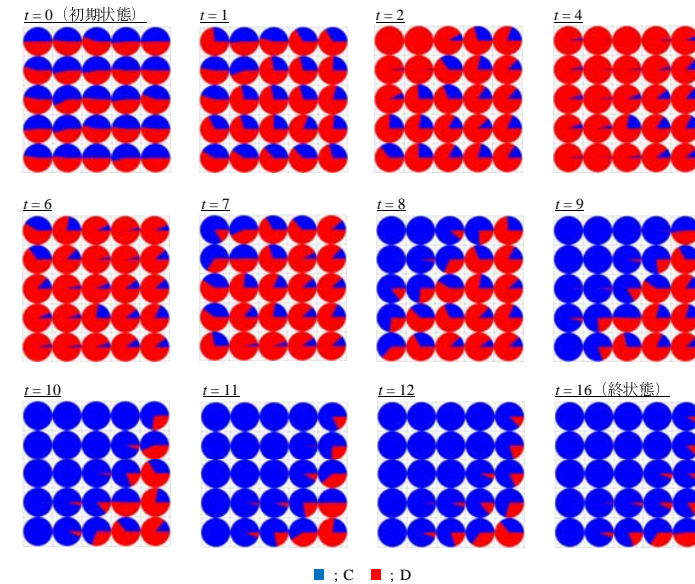


図8 CNNモデルにおけるSPDシミュレーションの可視化
Fig.8 Visualization of Results by SPD on CNN network

が影響を与えることがわかった。また、提案手法による可視化では、カテゴリの状態が隣接カテゴリに波及していく様子が示される。SOM でのカテゴリ間の距離の近さは必ずしもグラフ距離の近さを示すものではない。そのため、SOM によって分類されたノードに対して状態の局在化が生じ、それが近い性質のカテゴリ（セル内のノード群）に伝播していくものと考えられる。このような知見は従来の分析手法では得られるものではない。しかしながら、SOM によって生じるカテゴリの意味付けは定性的なものであり、定量的な観点からの議論は難しいものと考えられる。今後の課題である。

4. 結 言

本稿では、ネットワーク上の現象を局所的な見地から分析する新たな手法を提案した。はじめに、ノード毎に、次数、隣接ノードの平均次数、媒介中心性、平均頂点間距離、クラスター係数を算出し、それらを成分とする多変量変数をノードの属性とし、この多変量変数に対して自己組織化マップによってノードを分類する。次に、ネットワーク上のシミュ

レーションを行い、分類されたノード群のノードの状態変化を可視化するというものである。

提案手法を、2つのネットワークモデルによって生成されたネットワーク上の感染症伝播と空間囚人ジレンマのシミュレーションの結果の分析に適用した。従来手法による現象とネットワーク全体の統計的指標との関係よりも詳細な局所的な構造との関係性が顕在化できることがわかった。また、従来のネットワーク可視化を用いた分析はグラフィックアウトに大きく依存するが、提案手法では、近い性質を持つノードが近くに存在するようなレイアウトになるため、直感的な現象の理解が可能であることがわかった。

今後の展望としては、ネットワークモデルから生成されたネットワーク上でのシミュレーション全般に対する提案手法の有効性を示すことである。また、実ネットワークを用いたシミュレーションや実観測データの分析においても提案手法の適用を考えていきたい。

参 考 文 献

- 1) 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).
- 2) 斉藤和巳：ネットワークの可視化技術 大規模情報から意味情報の抽出，ネットワーク科学への招待 世界の”つながり”を知る科学と思考 (青山秀明，相馬 亘，藤原義久，編)，サイエンス社，pp.166-171 (2008).
- 3) Holme, P. and Kim, B.: Growing scale-free networks with tunable clustering, *Physical Review E*, Vol.65, No.2, p.26107 (2002).
- 4) Vázquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Physical Review E*, Vol.67, No.5, p.56104 (2003).
- 5) Moreno, Y., Pastor-Satorras, R. and Vespignani, A.: Epidemic outbreaks in complex heterogeneous networks, *The European Physical Journal B-Condensed Matter and Complex Systems*, Vol.26, No.4, pp.521-529 (2002).
- 6) Tsukamoto, E. and Shirayama, S.: Influence of the variance of degree distributions on the evolution of cooperation in complex networks, *Physica A: Statistical Mechanics and its Applications*, Vol.389, No.3, pp.577 - 586 (2010).