2 次元しきい値モデルの構造解析

システム情報数理学
上野裕也

2011 年 2 月 10 日

1　序文

世界には様々なネットワークが存在している. 例えば, インターネット通信網や交通網といった物理的なネットワーク, 人間関係や一族の家系図, 伝染病の流行といった社会的・認知的なネットワークなど, 多様な関係性から生ずるネットワークが存在している. これらのネットワークは複雑ながらもいくつかの規則性があると考えられ研究されてきた.

本研究では, しきい値モデルに着目する. しきい値モデルは, ランダムグラフや他のモデルでは導入されていない頂点重みを用いることにより, 人と人, 物と物がつながるためのプロセスを考えたネットワークの構成を再現している点で興味深い.

本研究の目的は, 重みとなる確率変数列を 2 つ用意することで, より多様なしきい値モデルを提案し, しきい値モデルの特性量と比較することである.

2　ネットワークの特性量

V を空でない集合 (頂点集合), E を V の 2 点集合の部分集合 (辺集合) とするとき, $G = (V, E)$ をグラフという. また, 確率的なルールによって生成されたグラフをランダムグラフといい, 両方を称してネットワークという.

以下のグラフ $G = (V, E)$ では, 頂点総数を n とする.

定義 2.1. 辺の密度

頂点 v の次数 $d(v)$ は, v から出ている辺の個数とする. グラフ $G = (V, E)$ の辺の密度を,

$$D = \frac{1}{n(n-1)} \sum_{v \in V} d(v)$$

と定義する.

定義 2.2. 平均頂点間距離

v, w を 2 つの頂点とするとき, v から w へ行くために通らなくてはならない辺の個数の最小値を $d(v, w)$ と書いて距離という. グラフ $G = (V, E)$ の平均頂点間距離を,

$$L = \binom{n}{2}^{-1} \sum_{\text{異なる頂点対}} d(v, w)$$

と定義する.

定義 2.3. クラスター係数

頂点 v のクラスター係数 $C(v)$ を,

$$C(v) = \binom{d(v)}{2}^{-1} (\text{頂点 } v \text{ を含む三角形の個数})$$

と定義する. ただし, $d(v) = 0$ または 1 の場合には $C(v) = 0$ とする. グラフ $G = (V, E)$ のクラスター係数を,

$$C = \frac{1}{n} \sum_{v \in V} C(v)$$

と定義する.

3　しきい値グラフ

定義 3.1. グラフ $G = (V, E)$ に対して, 非負実数 $w_v (v \in V)$ と t が存在して, 頂点 u と v が,

$$w_u + w_v > t \Longleftrightarrow uv \in E$$

を満たす時, G をしきい値グラフという.

しきい値グラフには以下のような特徴がある.

(1) 4 頂点以上のしきい値グラフは, P_4, C_4, $2K_2$ を含まない.

(2) しきい値グラフは, 頂点 1 個からなるグラフに, 孤立的な頂点 (頂点を加える前にあった頂点に対し辺を結ばない) かスター的な頂点 (頂点を加える前にあった全ての頂点に対し辺を結ぶ) を繰り返し加えていくことで構成することができる.

4　2 値しきい値モデル

頂点総数を n とする. $\{X_i\}_{i=1,\ldots,n}$ が $P(X_1 = 1) = p$, $P(X_1 = 0) = 1 - p$ を満たす独立同分布な確率変数列とする. しきい値 θ を $0 \leq \theta < 1$ とし, $X_i + X_j > \theta$ のとき, 頂点 i と頂点 j を辺 $\langle i, j \rangle$ で結ぶ (図 1). 辺の密度, 平均頂点間距離, クラスター係数に関する極限を求める.

4.1　辺の密度に関する極限

辺の総数 D_n とする. 辺の密度に関する平均は,

$$\lim_{n \to \infty} \binom{n}{2}^{-1} E[D_n] = p(2 - p)$$

である. 辺の密度に関する極限は,

$$\lim_{n\to\infty}\binom{n}{2}^{-1}D_n = p(2-p),\ \ a.s.$$

さらに, 辺の密度に関する分散の極限は,

$$V\Bigg[\binom{n}{2}^{-1}D_n\Bigg]\sim\frac{p(1-p)^3}{n}+\cdots$$

となる.

4.2 平均頂点間距離に関する極限

平均頂点間距離に関する極限は,

$$\lim_{n\to\infty}L_n = 2-p(2-p),\ \ a.s.$$

となる.

4.3 クラスター係数に関する極限

クラスター係数 C_n に関する極限は,

$$\lim_{n\to\infty}C_n = 1-(1-p)^2p,\ \ a.s.$$

となる.

5 2 値 2 次元しきい値モデル

頂点総数を n, しきい値 θ を $0 \le \theta < 1$ と固定する. $\{X_i\}_{i=1,2,\ldots,n}$, $\{Y_i\}_{i=1,2,\ldots,n}$ は,

$$P(X_i=1)=p_1 \qquad P(X_i=0)=1-p_1$$
$$P(Y_i=1)=p_2 \qquad P(Y_i=0)=1-p_2$$

を満たす独立同分布な確率変数列とする. $X_i+X_j > \theta$, $Y_i+Y_j > \theta$ のときのみ頂点 i と頂点 j を辺 $\langle i,j\rangle$ で結ぶ. 各頂点の重みを (X_i, Y_i) と表記する. 辺の密度, 平均頂点間距離, クラスター係数に関する極限を求める.

例 5.1. 頂点の重みが $(1,0),(0,1),(1,0),(0,0)$ で与えられたとき, 右図 1 のようになる.

図 1 : 2 値 2 次元しきい値モデル

5.1 辺の密度に関する極限

定理 5.1. 辺の密度に関する平均は,

$$\lim_{n\to\infty}\binom{n}{2}^{-1}E[D_n]$$

である. 辺の密度に関する極限は,

$$\lim_{n\to\infty}\binom{n}{2}^{-1}D_n = p_1p_2(2-p_1)(2-p_2),\ \ a.s.$$

さらに, 分散は,

$$\begin{aligned}V\Bigg[\binom{n}{2}^{-1}D_n\Bigg]\sim\frac{p_1p_2}{n}(&-p_1^3p_2^3-15p_1^2p_2^2-15p_1p_2+1\\&+4p_1^3p_2^2+4p_1^2p_2^3-4p_1^3p_2-4p_1p_2^3+15p_1^2p_2\\&+15p_1p_2^2-p_1^2-p_2^2+p_1+p_2)+\cdots\end{aligned}$$

となる.

5.2 平均頂点間距離に関する極限

定理 5.2. 平均頂点間距離に関する極限は,

$$\lim_{n\to\infty}L_n = 2-p_1p_2(2-p_1)(2-p_2)\ \ a.s.$$

5.3 クラスター係数に関する極限

定理 5.3. クラスター係数に関する極限は,

$$\begin{aligned}\lim_{n\to\infty}C_n &= p_1^2p_2^2(2-p_1)(2-p_2)+p_1^2(2-p_1)(1-p_2)\\&\quad+p_2^2(2-p_2)(1-p_1)+(1-p_1)(1-p_2),\ \ a.s.\end{aligned}$$

6 結論

しきい値モデルで実現しないグラフを 2 次元しきい値モデルでは多くのグラフを実現することがわかった. このことから, 2 次元しきい値モデルはしきい値モデルよりも多様な構造を持つことがわかった. 一般の 2 次元しきい値モデルに対して, 3 つの特性量の極限定理を得るまでに至らなかったが, 確率変数がベルヌイ試行に従う場合を扱い, 3 つの特性量を具体的に計算した. しきい値モデル, 2 次元しきい値モデルについて 3 つの特性量を取り扱ったが, 高次元化されたしきい値モデルの構造, 特性量についても興味深いものがあると考えられる.

参考文献

[1] Y. Ide, N. Konno and N. Masuda: *Statistical properties of a generalized threshold network model,* Methodol. Comput. Appl. Probab 12 (2010), 361–377.

[2] N. V. R. Mahadev and U. N. Peled: *Threshold graphs and related topics,* North-Holland, 1995.

[3] R. J. Serfling: *Approximation theorems of mathematical statistics,* John Wiley Sons, Inc. 1980.

[4] 今野紀雄 :「複雑ネットワーク入門」講談社, 2008.

[5] 藤澤洋徳:「確率と統計」朝倉書店, 2006.