

# ある特殊形状の ソーシャルネットワークの生成の研究

武者佑哉<sup>†</sup> 原田紀夫<sup>††</sup>

ソーシャルネットワーク (SN) において、大きなサイクルを持つ巨大な連結成分の存在が観測されている。この形状は STD などの感染拡大に影響する。ここではこの形状が生成されるメカニズムを研究する。

## Research of generation of social network with a peculiar shape

YUYA MUSHA<sup>†</sup> NORIO HARADA<sup>††</sup>

In social network (SN), the existence of a huge connected component with a big cycle is observed. This shape influences the infection spread such as STD. The mechanism with which this shape is generated is researched here.

## 1. はじめに

私達が生活する社会には、友人関係や恋人関係など個人として、会社と会社などのグループとしてのソーシャルネットワーク (以降 SN と略す) などが存在する。また近年では、mixi や facebook などのようにインターネット上での SN も存在するように、多種多様な SN が存在する。その中でも興味深いものが、論文 [2] で紹介されている Jefferson のある高校 (以降 J 校と略す) での SN である。この SN をグラフにより可視化すると図 1 のようになる。この形は、自己ループを持たず、節点数と枝数が等しい連結グラフで、1つのサイクルを持つという特徴を持つ調和グラフ [5] に近い形をしている。この形状の類似ということから興味を持った。本研究ではこのようなグラフになる SN 生成のメカニズムがどのようなものかを研究する。また、シミュレーションを行い、J 校のデータ、[1] の論文のシミュレーションのデータとの比較を行う。

## 2. [1] の論文の概要

### 2.1 要約

STDs (Sexual Transmitted Diseases : 淋病・ヘルペス・HIV など ; 以降ここでは STDs と略す) がどのように感染を広げているのかを調べている論文である。最も性的関係が頻繁に入れ替わる 16 歳から 18 歳 (つまり青春期) に着目し、データをとっている。その中で、J 校における性的関係の SN が図 1 である。一般的には、サイクルを持たない SN になると考えられ、感染の可能性が低い。しかし、J 校においては、いくつかのサイクルが存在しているため、病気に感染する可能性が高くなると考えられる。このことから、J 校の SN になるシミュレーションを行っている。そのため、J 校の SN に近づくことを目指して、シミュレーションにおいて条件を設けている。

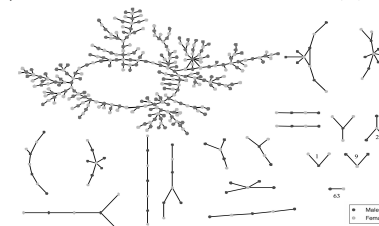


図 1 J 校の性的関係の SN

Figure1 SN of sexual relation of J school

\*<sup>†</sup> 拓殖大学大学院工学研究科  
Takushoku University graduate school engineering research course

<sup>††</sup> 拓殖大学工学部情報工学科  
Information Engineering Section of Department of Engineering at Takushoku University.

表1 J校のグラフの次数の分布

Table1 distribution of degree in graph of J school

degree	度数	degree	度数	degree	度数
1	338	4	30	7	0
2	140	5	6	8	0
3	57	6	1	9	1

● J校と同じ次数の分布, 人数 (表1)

● 長さ4のサイクルを除く

また, 特性データの計算上においても条件を設けている.

● 二者関係を除いて計算する

この3つの条件のみ明記されているが, 実際には他にも条件があると思われる. そして, この条件でシミュレーションを1,000回行い, そのデータとJ校のデータと比較をした結果, 近いデータを出した.

## 2.2 SNの特性データ

STD感染拡大に関わるSNの特性データが次の9つのデータである. この9つのデータを用いて比較を行う.

### (1) 最大コンポーネントの大きさ

1つのSNに携わっている最大の人数のこと. 病気に感染する可能性がある人数を表している.

### (2) コンポーネントの数

現れるSNの数. 数が少なければ1つのSNに集中するので, STDに感染する可能性がある人数が多くなる.

### (3) 到達可能性

他の人にSTDを感染させる可能性を表している.

### (4) 閉塞率の平均

STDに感染する可能性の閉塞率を表している.

### (5) 最大コンポーネントにおける閉塞率

(4)のうち最大コンポーネントのみにおける閉塞率である.

### (6) 直径

最短経路を通る最大の距離. これが大きいとより遠い人にまでSTDに感染する可能性がある. また, (7)と組み合わせると, (7)の値が大きいとより早くSTDに感染する可能性がある.

### (7) サイクルの数

深さ優先探索したときのサイクルの数である. 図2のようなネットワークにおいて, 実際は「A-B-C-A」, 「A-C-D-A」, 「A-B-C-D-A」の3つのサイクル数になるが, 深さ優先で探索しているため, 「A-B-C-A」, 「A-B-C-D-A」の2つのサイクル数になる. サイクルがあることにより, 病気感染の可能性が高くなる.

### (8) 情報中心性の平均

最短経路の距離を用いて計算を行う. この値が高いと, より中心性が高い.

### (9) 情報中心性の標準偏差

(8)の標準偏差である.

[1]のJ校のデータと1,000回行ったシミュレーションのデータを後の比較のために表2, J校のコンポーネントの分布を表3として引用する.

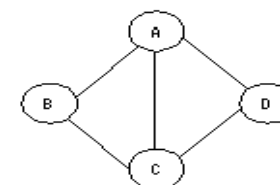


図2 サイクルを持つネットワークの例

Figure2 example of network with cycles

表2 J校とシミュレーションの特性データ[1]

Table2 the characteristic data of J school and of simulation [1]

Measure	Observed	Mean	Min	Max	Std Dev	IQ Range
(1)	288	289.83	108	344	28.5	30
(2)	101	103.7	96	113	2.31	3
(3)	0.254	0.262	0.073	0.36	0.042	0.05
(4)	0.017	0.012	0.01	0.029	0.003	0.004
(5)	0.064	0.072	0.045	0.123	0.008	0.011
(6)	37	36.5	18	63	6.04	8
(7)	5	8.24	1	17	2.24	3
(8)	0.003	0.004	0.002	0.009	0.0005	0.0004
(9)	0.0007	0.00076	0.00047	0.0019	0.00012	0.00012

表3 J校のコンポーネントの分布[1]

Table3 distribution of component of J school [1]

男女計	comp 数	男女計	comp 数	男女計	comp 数	男女計	comp 数
288	1	6	1	5	1	4	2
10	2	7	1	5	1	3	12
8	1	7	1	4	2	3	9
8	1	5	2	4	1	2	63

表4 長さ4を除かないときの特性データ

Table4 the characteristic data when four in length is not excluded

Measure	Mean	Min	Max	Std Dev	IQ Range
(1)	299.76	186	344	18.7	24
(2)	110.1	100	124	3.65	4
(3)	0.276	0.138	0.36	0.032	0.04
(4)	0.017	0.01	0.032	0.001	0.001
(5)	0.088	0.050	0.116	0.009	0.012
(6)	30.65	21	52	4.63	6
(7)	14.15	4	28	3.65	4
(8)	0.003	0.003	0.003	0.0001	0.0001
(9)	0.00055	0.00044	0.0007	0.00004	0.00005

### 3. シミュレーション

日常生活で性的関係が繋がっていくのは、まず任意の2人の関係がつながり、また他の任意の2人がつながる。もしくは、その2人のつながりにさらに1人がつながり3人になる。さらに、4人、5人・・・と関係のつながりが増えていくのが一般的だと考えた。そこでここでは、SNの生成は次のアルゴリズムとする。

#### 3.1 SN生成アルゴリズム

- ① 人数・次数の分布をJ校と同じに設定
- ② ランダムに2人を選択
- ③ 選択した2人がすでにつながっていたら②へ
- ④ 選択した2人をつなげて次数の分布を超えるなら②へ  
 ※次数7: 次数8が0かつ次数9が0かつ次数7が1を超えないなら⑤へ  
 ※次数8: 次数7が0かつ次数9が0かつ次数8が1を超えないなら⑤へ
- ⑤ 選択した2人をつなげる
- ⑥ 次数の分布がすべて一致していないなら②へ
- ⑦ 計算処理へ

このアルゴリズムを用いて、[1]と同じ条件で1,000回シミュレーションを行った。[1]の条件において、「長さ4のサイクルを除く」とあるが、J校のSNには、長さ4のサイクルは存在している。また、日常生活でも現れる可能性がある。しかし、[1]のシミュレーションでは何故か除いている。そのため、長さ4のサイクルを除いてシミュレーションをした結果、除かない場合とかわらない結果となったので、ここでは長さ4のサイクルを除かないでシミュレーションを行う。このことより、長さ4のサイクルを除かないシミュレーションの結果の特性データが表4となり、コンポーネントの分布を1,000回のシミュレーションでの合計として表したものが表5となる。

表5 長さ4を除かないときのコンポーネントの分布 (1,000回の合計)

Table5 distribution of component when four in length is not excluded (1,000 total)

男女計	comp 数	男女計	comp 数	男女計	comp 数	男女計	comp 数
- 5	102,836	- 95	0	- 185	0	- 275	26
- 10	5,036	- 100	1	- 190	1	- 280	46
- 15	818	- 105	0	- 195	0	- 285	58
- 20	264	- 110	0	- 200	0	- 290	74
- 25	85	- 115	0	- 205	0	- 295	96
- 30	60	- 120	0	- 210	0	- 300	108
- 35	22	- 125	0	- 215	1	- 305	105
- 40	11	- 130	0	- 220	0	- 310	116
- 45	7	- 135	0	- 225	0	- 315	101
- 50	2	- 140	0	- 230	0	- 320	80
- 55	2	- 145	0	- 235	2	- 325	57
- 60	1	- 150	0	- 240	2	- 330	46
- 65	1	- 155	0	- 245	2	- 335	11
- 70	0	- 160	0	- 250	7	- 340	13
- 75	0	- 165	0	- 255	3	- 345	6
- 80	2	- 170	0	- 260	5	- 350	0
- 85	0	- 175	0	- 265	20	- 355	0
- 90	0	- 180	0	- 270	14	- 360	0

### 3.1.1 この SN 生成アルゴリズムの考察

[1]のJ校のデータ及びシミュレーションのデータ表2と比較する。(1)について平均は10大きく、最小は90程度大きく、標準偏差が10小さいため、分布の広がりが少ない生成となった。そして、(2)についても平均が10程度大きい。これより、今回用いた3.1の生成アルゴリズムでは、多くの人がつながる1つの大きなコンポーネントができ、少ない人数がつながるコンポーネントが多く生成される傾向があることがわかる。(3)についても0.01大きい。このようなSNは、よりSTDに感染する可能性がある人数が多いSNになっている。さらに、(6)は6小さい。これは(7)が6大きいためだと考えられる。(7)が大きく、(6)が小さいため、STDに感染するスピードが速いと考えられる。

このことより、SN生成アルゴリズムでは、ランダムに2人を選択するときに、すでにネットワークにつながっている人を選ぶ確率が高いのではと考えられる。さらに、日常生活において3.1のアルゴリズムのように誰とでもつながるといことは考えづらい。このことより、次のような好感度を用いたSN生成アルゴリズムに改訂した。

### 3.2 好感度を考慮した SN 生成アルゴリズム

- ① 人数・次数の分布をJ校と同じに設定
- ② 個々の好感度に1~10のステータスをつける
- ③ ランダムに2人を選択
- ④ 選択した2人がすでにつながっていたら②へ
- ⑤ 2人の好感度を足してX(3~10・15)未満なら個々に1を足して②へ
- ⑥ 選択した2人をつなげて次数の分布を超えるなら、2人の好感度のステータスをつけなおして②へ  
 ※次数7: 次数8が0かつ次数9が0かつ次数7が1を超えないなら⑤へ  
 ※次数8: 次数7が0かつ次数9が0かつ次数8が1を超えないなら⑤へ
- ⑦ 選択した2人をつなげ、その他の好感度は1を引く
- ⑧ 次数の分布がすべて一致していないなら②へ
- ⑨ 計算処理へ

このアルゴリズムを用いて、[1]と同じ条件で1,000回シミュレーションをX(=3~10と15)について行った。その結果、最もJ校に近い値になったX=5のときの特性データが表6となり、コンポーネントの分布が表7となる。最もJ校に遠い値になったX=15のときの特性データが表8となり、コンポーネントの分布が表9となる。

表6 X=5のときの特性データ

Table6 the characteristic data when X is 5

Measure	Mean	Min	Max	Std Dev	IQ Range
(1)	285.07	120	341	28.2	32
(2)	105.4	97	116	3.23	5
(3)	0.253	0.077	0.35	0.042	0.05
(4)	0.017	0.02	0.033	0.001	0.001
(5)	0.080	0.045	0.110	0.010	0.013
(6)	33.71	22	66	5.95	8
(7)	9.36	1	20	3.23	5
(8)	0.003	0.003	0.003	0.0001	0.0001
(9)	0.00060	0.00047	0.0007	0.00004	0.00005

表7 X=5のときのコンポーネントの分布(1,000回の合計)

Table7 distribution of component when X is 5 (1,000 total)

男女計	comp 数	男女計	comp 数	男女計	comp 数	男女計	comp 数
- 5	95,884	- 95	1	- 185	0	- 275	54
- 10	6,347	- 100	4	- 190	0	- 280	72
- 15	1,257	- 105	0	- 195	1	- 285	59
- 20	429	- 110	0	- 200	3	- 290	75
- 25	200	- 115	0	- 205	1	- 295	98
- 30	84	- 120	0	- 210	4	- 300	87
- 35	48	- 125	2	- 215	5	- 305	93
- 40	35	- 130	1	- 220	5	- 310	84
- 45	20	- 135	0	- 225	5	- 315	52
- 50	12	- 140	1	- 230	4	- 320	35
- 55	6	- 145	2	- 235	9	- 325	30
- 60	10	- 150	1	- 240	17	- 330	22
- 65	5	- 155	1	- 245	9	- 335	8
- 70	5	- 160	1	- 250	15	- 340	3
- 75	7	- 165	1	- 255	25	- 345	1
- 80	3	- 170	1	- 260	30	- 350	0
- 85	0	- 175	0	- 265	30	- 355	0
- 90	3	- 180	3	- 270	53	- 360	0

表 8 X=15 のときの特性データ  
Table8 the characteristic data when X is 15

Measure	Mean	Min	Max	Std Dev	IQ Range
(1)	209.02	65	308	48.8	67
(2)	99.9	96	107	2.16	3
(3)	0.151	0.039	0.29	0.051	0.08
(4)	0.016	0.01	0.030	0.002	0.001
(5)	0.087	0.049	0.176	0.016	0.019
(6)	31.79	17	60	6.45	9
(7)	3.89	0	11	2.16	3
(8)	0.003	0.003	0.003	0.0001	0.0001
(9)	0.00068	0.00058	0.0008	0.00003	0.00004

表 9 X=15 のときのコンポーネントの分布 (1,000 回の合計)  
Table9 distribution of component when X is 15 (1,000 total)

男女計	comp 数	男女計	comp 数	男女計	comp 数	男女計	comp 数
- 5	82,641	- 95	17	- 185	29	- 275	24
- 10	11,589	- 100	13	- 190	30	- 280	17
- 15	2,402	- 105	17	- 195	20	- 285	13
- 20	914	- 110	17	- 200	35	- 290	9
- 25	441	- 115	17	- 205	31	- 295	5
- 30	234	- 120	10	- 210	42	- 300	5
- 35	153	- 125	4	- 215	45	- 305	5
- 40	108	- 130	12	- 220	39	- 310	2
- 45	93	- 135	20	- 225	42	- 315	0
- 50	63	- 140	16	- 230	47	- 320	0
- 55	33	- 145	15	- 235	42	- 325	0
- 60	41	- 150	14	- 240	44	- 330	0
- 65	32	- 155	8	- 245	39	- 335	0
- 70	32	- 160	29	- 250	42	- 340	0
- 75	19	- 165	19	- 255	38	- 345	0
- 80	26	- 170	19	- 260	48	- 350	0
- 85	24	- 175	19	- 265	28	- 355	0
- 90	11	- 180	23	- 270	26	- 360	0

### 3.2.1 好感度を考慮した SN 生成アルゴリズムの考察

X の値が大きくなるにつれ、最大コンポーネントの大きさが小さくなることが分かった。これについて X=5 である表 7 と X=15 である表 9 のコンポーネントの分布を比較してみると、X=5 のときは、中間の大きさのコンポーネントはあまり生成されていないが、X=15 のときは、中間の大きさのコンポーネントが多くできている。これにより、X の値が大きくなれば、すでに SN につながっている人を選ぶ確率が低くなり、X の値が小さくなれば、SN につながっている人を選ぶ確率が高くなるということがわかった。

### 3.3 2つのアルゴリズムに関する考察

3.1. の SN 生成アルゴリズムと 3.2. の好感度を考慮したものとを比較してみると、3.2. のほうが X=5 の近くでは、最大コンポーネントの大きさと到達可能性が J 校に近い結果となった。しかし、コンポーネントの分布において、J 校では中間の大きさのコンポーネントは生成されていないのに対し、3.2. の SN 生成アルゴリズムでは生成されている。シミュレーションの結果は 1,000 回の和を表しているの確率の面では小さいが、好感度を入れると発生の割合が大きくなることは興味深い。

## 4. おわりに

今回行ったシミュレーションでは、枝の次数を合わせるだけで他はランダムとした 3.1. のアルゴリズムと好感度を考慮したアルゴリズムについて行った。一連のシミュレーションによって好感度の和 X=5 の辺りに J 校のデータに近いものが得られた。この理由については、まだ明確なものは得られていない。今後これらの分析とともにアルゴリズムのより妥当性が得られるものを考えていきたい。

## 参考文献

- 1) Peter S. Bearman, James Moody, Katherine Stovel, "CHAINS OF AFFECTION: THE STRUCTURE OF ADOLESCENT ROMANTIC AND SEXUAL NETWORKS", ISERP WORKING PAPER 02-04
- 2) David Krackhardt, "A Plung into Networks", SCIENCE VOL 326 20 OCTOBER 2009 pp.47-48 AAAS
- 3) Matthew O. Jackson, "Social and Economic Networks", Princeton university press pp.6-9
- 4) Stanley Wasserman and Katherine Faust, "Social Network Analysis Methods and Applications", Cambridge university press
- 5) 原田紀夫, "システムの超調和理論", 拓殖大学研究叢書 1998 年 3 月 p.345
- 6) "Pajek", <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>