Communications of the Korean Statistical Society
2011, Vol. 18, No. 5, 595–601
DOI: http://dx.doi.org/10.5351/CKSS.2011.18.5.595

Random Generation of the Social Network with Several Communities

Myung-Hoe Huh[1,a], Yonggoo Lee[b]

[a]Department of Statistics, Korea University; [b]Department of Statistics, ChungAng University

Abstract

A community of the social network refers to the subset of nodes linked more densely among them than to others. In this study, we propose a Monte-Carlo method for generating random social unipartite and bipartite networks with two or more communities. Proposed random networks can be used to verify the small world phenomenon of the social networks with several communities.

Keywords: Social network analysis, community, Monte-Carlo generation, bipartite network.

1. Background and Aim

A social network consists of a number of nodes (points) and a number of links (lines) that connect pairs of nodes. Individuals and social entities are denoted by nodes, while the relationship and/or interdependencies are represented by links.

Unipartite social network refers to the network in which all the nodes are individuals or social entities of one level. For instance, the `faux.magnolia.high` network that represents the friendship relations among 1461 students of a high school is unipartite. The dataset is available at the `R`'s `ergm` package. In a bipartite social network, nodes are composed of individuals and social entities, linked by lines from individuals to social entities or from social entities to individuals (Wasserman and Faust, 1994). For instance, the `scotland.net` which represents the membership of 136 directors to 108 Scotland corporate companies is bipartite. The dataset is available at the homepage of `Pajek`, an open source software for large network analysis.

If the links have directions, the network is called directed. If not, the network is undirected. Communities of a network consist of several nodes that are linked more densely among them than to others (Newman, 2004).

In the literature, there appeared a number of algorithms for the Monte-Carlo generation of social networks (Watts and Strogatz, 1998; Barabasi and Albert, 1999; Leskovec *et al.*, 2005). They are useful for studying the general nature of social networks such as the "small world" phenomenon (Travers and Milgram, 1969). As far as the authors know, there are no previous papers for randomly generating social networks composed of communities.

In this study, we propose a Monte-Carlo method for generating undirected random social unipartite and bipartite networks with two or more communities which are linked with each other, extending the rewiring process of Watts and Strogatz (1998). Proposed random networks can be used to verify the small world phenomenon.

[1] Corresponding author: Professor, Department of Statistics, Korea University, Anam-Dong, Sungbuk-Gu, Seoul 136-701, Korea. E-mail: stat420@korea.ac.kr

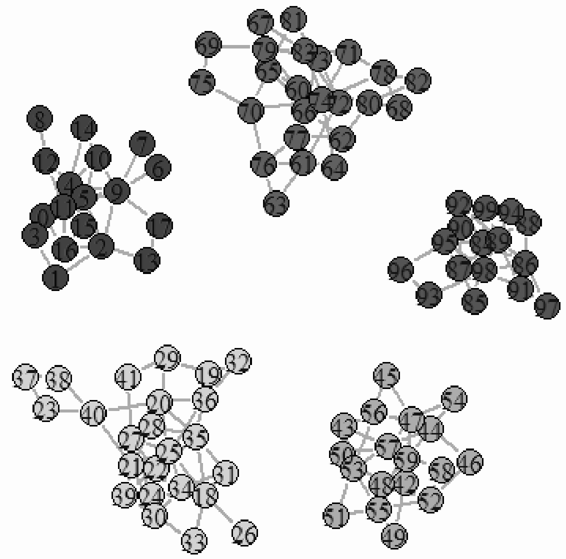

Figure 1: *A random unipartite network with five communities completely split*

2. Random Unipartite Networks with Several Communities

We will generate a random network of N nodes and L links with k communities ($L \gg N$). First, we decompose N and L into k numbers, $n_1, \ldots, n_k$ and $l_1, \ldots, l_k$ from the multinominal distributions $M(N, \mathbf{u}_k)$ and $M(L, \mathbf{u}_k)$ where $\mathbf{u}_k = 1/k\,(1, \ldots, 1)$. We assume that $l_1 \geq n_1, \ldots, l_k \geq n_k$.

Completely split communities

For the community $j\,(= 1, \ldots, k)$, the node numbers from 1 to n_j are arranged to form a sequence of length l_j that contains n_j numbers at least once with extra $l_j - n_j$ numbers by randomly sampling from 1 to n_j with replacement. Form the second sequence of length l_j by randomly permuting the first sequence; then, these two sequences are bound by columns that are used as the links within the community. It is assumed that the network is undirected. The network may contain the loops and multiple links. In such cases, we simplify the network by discarding such redundant links.

Figure 1 shows a random network with $5\,(= k)$ communities which accommodate $100\,(= N)$ nodes (split into 18, 24, 18, 24, 16 nodes) and $250\,(= L)$ links (split into 45, 65, 40, 54, 46 links). The network when simplified contains 202 links, reduced from 250 links.

Inter-connected communities

We first generate a unipartite network with k communities completely split. Then for each link $l\,(= 1, \ldots, L)$ the one end node is switched to another with probability p. The replacement node is chosen uniformly from all nodes of the network. Consequently, k communities of the network are likely to be connected with each other. The parameter p influences the strength of connectedness. When simplified, the network may contain a slightly different number of links from the originated network, after removing the loops and multiple links.

Figure 2 shows a descendant network of Figure 1 with $p = 0.1$. The simplified network has 211 links. We see that all the communities are connected.

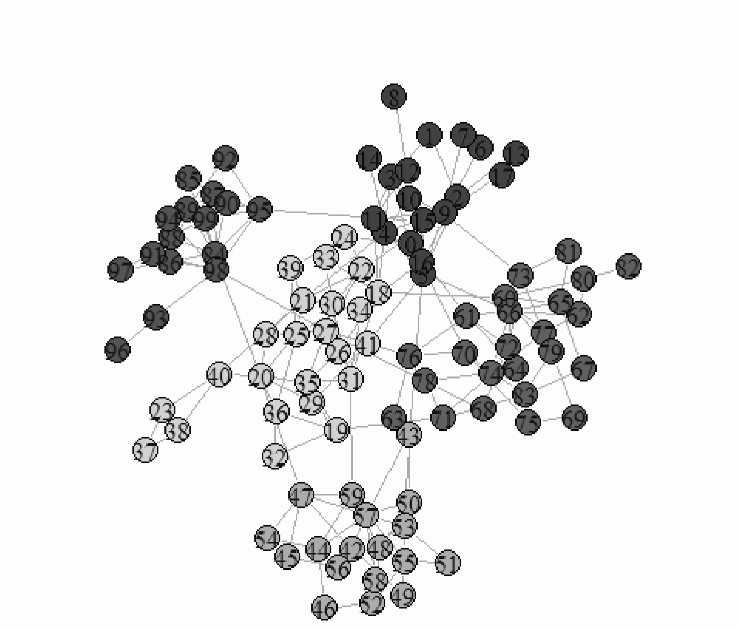

Figure 2: *A random unipartite network with five communities inter-connected*

3. Random Bipartite Networks with Several Communities

The aim is to generate a random bipartite network of A social entities and B individuals, which are connected by L links ($L \gg A$ and $L \gg B$), forming k communities. First, we randomly divide A, B and L into k components for community elements using the multinominal distribution with equal probabilities. Turnout numbers are denoted by a_j, b_j, l_j ($j = 1, \ldots, k$) and we will assume that $l_j \geq a_j$ and $l_j \geq b_j$ for $j = 1, \ldots, k$, so that each individual is connected to at least one social entity and vice versa.

Completely split communities

Within each community $j\,(= 1, \ldots, k)$, the social entities are numbered from 1 to a_j and the individuals are numbered by 1 to b_j. The social entities are arranged to form a random sequence of length l_j that contains a_j numbers at least once ($l_j \geq a_j$). Similarly, a random sequence of length l_j for b_j individuals ($l_j \geq b_j$) is formed. Then two sequences are bound by columns and used for the links within the community. We simplify multiple links; therefore, the effective number of links may not equal to L.

Figure 3 shows a case in which $A = 50$ (of sizes 13, 10, 4, 10, 13), $B = 100$ (of sizes 15, 22, 22, 22, 19), $L = 250$ (of sizes 47, 49, 55, 52, 47) and $k = 5$. The network when simplified contains 220 links. In the figure, circles represent individuals and squares represent social entities.

Inter-connected communities

We first generate a bipartite network with k communities completely split. Then, for each link $l\,(= 1, \ldots, L)$, the social entity node is switched to another social entity with probability p, while the individual node is kept intact. The replacement social entity is chosen uniformly from all social entity nodes of the network. Consequently, k communities of the individuals and the social entities are likely to be connected with each other. The parameter p influences the strength of connectedness.

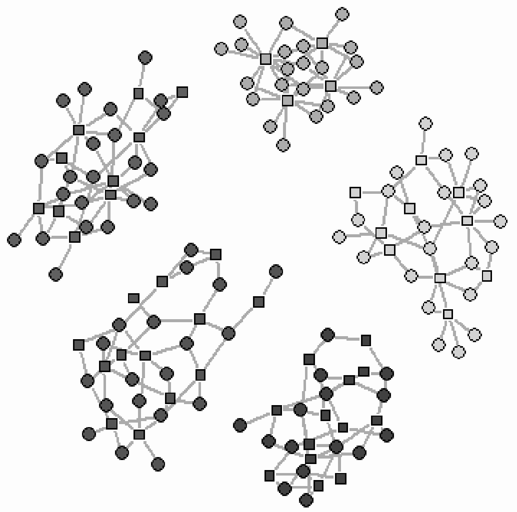

Figure 3: *A random bipartite network with five communities completely split*

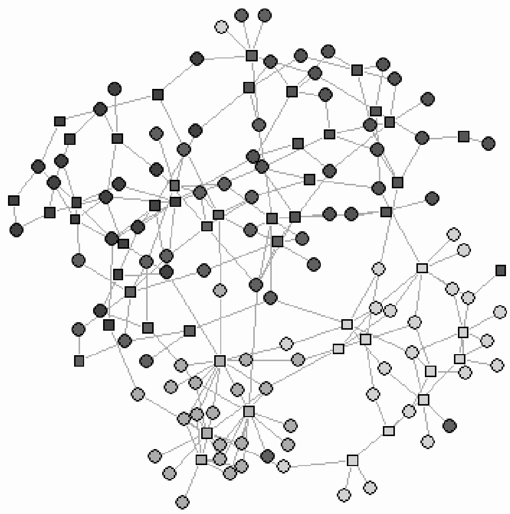

Figure 4: *A random bipartite network with five communities inter-connected*

Figure 4 shows a descendant network of Figure 3 with $p = 0.1$. The network has the same number of social entity nodes and individual nodes, but may have different number of links from the designated number L when multiple links are simplified. Simplified network contains 232 links connecting social entities and individuals. We see that the communities are connected.

4. Applications of Random Networks with Several Communities

We will suggest an application of random unipartite networks with k communities.

Table 1: Small world phenomenon in the unipartite networks

(1) N = 100 nodes, L = 250 links, k = 5 communities

p	Q1	Q2	Q3	5% Trimmed Mean
0.0	∞	∞	∞	∞
0.05	3	5	6	4.77
0.1	3	4	5	3.93
0.2	3	4	5	3.86
0.3	3	3	4	3.44
0.4	2	3	4	3.05
0.5	2	3	4	3.12

(2) N = 1000 nodes, L = 2500 links, k = 50 communities

p	Q1	Q2	Q3	5% Trimmed Mean
0.0	∞	∞	∞	∞
0.05	7	9	11	∞
0.1	5	7	8	∞
0.2	5	6	6	5.55
0.3	4	5	6	5.00
0.4	4	5	5	4.69
0.5	4	4	5	4.52

"Small world" phenomenon is well known in the literature on social networks and is demonstrated only for the case of unipartite networks without community structure.

For unipartite random networks of $N = 100$ nodes, $L = 250$ links, $k = 5$ communities and $p = 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$, the quartiles including median (the second quartile) and 5% trimmed mean of inter-node distances are computed, as shown in Table 1 (1).

For $p = 0$, five communities are disconnected, so that three quartiles are infinite. For $p = 0.05$ in which only 5% of links are expected to change, the median of inter-distances drops to 5 from ∞. The median decreases to 4 (to 3) as p increases to 0.1 (to 0.3). Then the median does not decrease further.

For unipartite random networks of $N = 1000$ nodes, $L = 2500$ links, $k = 50$ communities and $p = 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$, the quartiles including median and 5% trimmed mean of inter-node distances are reported in Table 1 (2).

For small values of p such as 0.05 and 0.1, the inter-individual distances almost double as the scale of the network becomes 10-fold; however, for moderate values of p between 0.2 and 0.5, the inter-individual distances increase 30% to 50% as the scale of the network becomes 10-fold.

For bipartite random networks of $A = 50$ social entities, $B = 100$ individuals, $L = 250$ links, $k = 5$ communities and $p = 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$, the quartiles including median and 5% trimmed mean of inter-individual and inter-social entity distances are computed, as shown in Table 2 (1).

For $p = 0$, five communities of individuals and social entities are not connected, so three quartiles are infinite. For p as small as 0.05, the median of inter-individual and inter-social entity distances drops to 2.5 and 2.2. For $p \geq 0.1$, the median inter-individual distances and the median inter-social entity distances stay stable at 2.

For bipartite random networks of $A = 500$ social entities, $B = 1000$ individuals, $L = 2500$ links, $k = 50$ communities and $p = 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$, the quartiles including median and 5% trimmed mean of inter-individual and inter-social entity distances are reported in Table 2 (2). For various values of p, the median inter-individual distances increases moderately even though the scale of the network becomes 10-fold, showing "small-world effect".

Table 2: Small world phenomenon in the bipartite networks

(1) A = 50 social entities, B = 100 individuals, L = 250 links, k = 5 communities

Individuals				
p	Q1	Q2	Q3	5% Trimmed Mean
0.0	∞	∞	∞	∞
0.05	2.0	2.5	3.0	2.58
0.1	2.0	2.0	3.0	2.32
0.2	2.0	2.0	3.0	2.28
0.3	2.0	2.0	2.5	2.24
0.4	2.0	2.0	2.5	2.12
0.5	1.8	2.0	2.5	2.09

Social entities				
p	Q1	Q2	Q3	5% Trimmed Mean
0.0	∞	∞	∞	∞
0.05	2.0	2.2	3.0	2.37
0.1	1.5	2.0	2.5	2.09
0.2	1.5	2.0	2.5	2.00
0.3	1.5	2.0	2.0	1.95
0.4	1.5	2.0	2.0	1.87
0.5	1.3	2.0	2.0	1.81

(2) A = 500 social entities, B = 1000 individuals, L = 2500 links, k = 50 communities

Individuals				
p	Q1	Q2	Q3	5% Trimmed Mean
0.0	∞	∞	∞	∞
0.05	4.0	5.0	6.0	∞
0.1	3.0	4.0	5.0	3.99
0.2	3.0	4.0	4.0	3.48
0.3	3.0	3.0	4.0	3.27
0.4	3.0	3.0	4.0	3.22
0.5	3.0	3.0	4.0	3.20

Social entities				
p	Q1	Q2	Q3	5% Trimmed Mean
0.0	∞	∞	∞	∞
0.05	4.0	4.0	6.0	∞
0.1	3.0	4.0	4.0	3.55
0.2	2.0	3.0	4.0	3.11
0.3	2.0	3.0	4.0	2.97
0.4	2.0	3.0	3.0	2.95
0.5	2.0	3.0	3.0	2.91

5. Remarks

In this study, random or Erdos and Renyi's (1959) networks were used as components for k separated communities and then end nodes of links were switched by another with specified probability p, so that communities get connected with each other.

There is no reason for component networks to be restricted to the class of Erdos-Renyi networks. Other classes such as regular lattice type and Barabasi-Albert type networks can be adopted instead.

For the detection of communities within the network, there are several algorithms available such as the walktrap communities, the fast greedy communities and/or the edge-betweenness communities. Monte-Carlo random networks could be valuable in the comparative assessment of community detection methods; however, this study does not cover the topic.

References

Barabasi, A.-L. and Albert, R. (1999). Emergence of scaling in random networks, *Science*, **286**, 509–512.

Erdos, P. and Renyi, A. (1959). On random graphs I, *Publicationes Mathematicae Debrecen*, **6**, 290–297.

Leskovec, J., Kleinberg, J. and Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations, *KDD '05: Proceeding of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 177–187.

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks, *Physical Review*, **E 69**, 066133.

Travers, J. and Milgram, S. (1969). An experimental study of the small world problem, *Sociometry*, **32**, 425–443.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small world networks, *Nature*, **393**, 440–442.

Received June 2011; Accepted June 2011