

Using Structure Holes for Determining Key Factors: An Illustration of Reporting Eradication of Amoebiasis

Tsair-Wei Chien^{1, 2}, Shih-Bin Su^{3, *}

¹Research Departments, Chi-Mei Medical Center, Tainan, Taiwan

²Department of Hospital and Health Care Administration, Chia-Nan University of Pharmacy and Science, Tainan, Taiwan

³Department of Occupation Medicine, Chi-Mei Medical Center, Tainan, Taiwan

Email address:

smile@mail.chimei.org.tw (Tsair-Wei Chien), shihbin.su@msa.hinet.net (Shih-Bin Su)

*Corresponding author

To cite this article:

Tsair-Wei Chien, Shih-Bin Su. Using Structure Holes for Determining Key Factors: An Illustration of Reporting Eradication of Amoebiasis.

Applied and Computational Mathematics. Special Issue: Some Novel Algorithms for Global Optimization and Relevant Subjects.

Vol. 6, No. 4-1, 2017, pp. 55-63. doi: 10.11648/j.acm.s.2017060401.15

Received: December 20, 2016; **Accepted:** January 9, 2017; **Published:** January 24, 2017

Abstract: *Background:* Many researches aim to determine key factors affecting their concerns of interest using traditional statistical techniques, such as logistical or linear regressions. Social network analysis (SNA) is a newly novel way determining key roles through the use of network and graph theories recently. An example of commonly visualized through SNA is the disease transmission path of Middle East respiratory syndrome (MERS). *Purpose:* To determine key roles using structure holes of SNA for further improvement, and to show the SNA advantage over traditional classic test theory. *Methods:* Data were records regarding 443 adult mentally retarded residents who were infected with amoebiasis and distributed in 10 houses in past 10 years. A series of intensive mass screenings and treatment interventions were conducted. Structure holes were applied to verify the efficacy of determining key roles and strong associations for the domains of interest in a network and compared with the result obtained from the traditional Chi-square statistics. *Results:* The classification of key roles in a network (e.g., with which year the residency room with amoebiasis cases has strongly association) can be effectively discriminated through the structure holes of SNA. Though the result is similar to the traditional Chi-square statistics, the structure holes can release much more useful and valuable information for further investigation. *Conclusions:* Because of advances in computer technology, the number of healthcare studies for the group classification and association assertion continues to increase and benefit comparisons of data if structure holes of SNA are applied.

Keywords: Social Network Analysis, Structure Holes, Chi-Square Statistics, Middle East Respiratory Syndrome, Amoebiasis

1. Background

A great deal of work has been devoted to finding key factors and correct grouping classifications in the past years [1, 2]. Logistical regression, multiply linear regressions, cluster analysis, and exploration factor analysis are often used for determining them. A series of cross (contingency) tables reporting demographic data and prevalence of the study population using Chi-square are frequently seen [3, 4]. A further report regarding the key factor (in row) significantly associated with the predicted group (in column) using the criterion of the standardized residual value(>1.96) is required to display once an overall independence feature is rejected ($p < 0.05$) with Chi-square test. Unfortunately,

few studies were found displaying such further more information to readers in past published papers.

Similarly, South Korea was experiencing the largest outbreak of Middle East respiratory syndrome (MERS) coronavirus infections outside the Arabian Peninsula in 2015 [5]. We saw a simplified transmission diagram that merely descriptively and graphically was illustrating the spreading events associated with Cases 1, 14, 16 of MERS-CoV without an objective and statistical base for disclosing the three above-mentioned cases or more key role cases, like the demonstration in Figure 1 we downloaded data [6] and made it using Social network analysis (SNA).

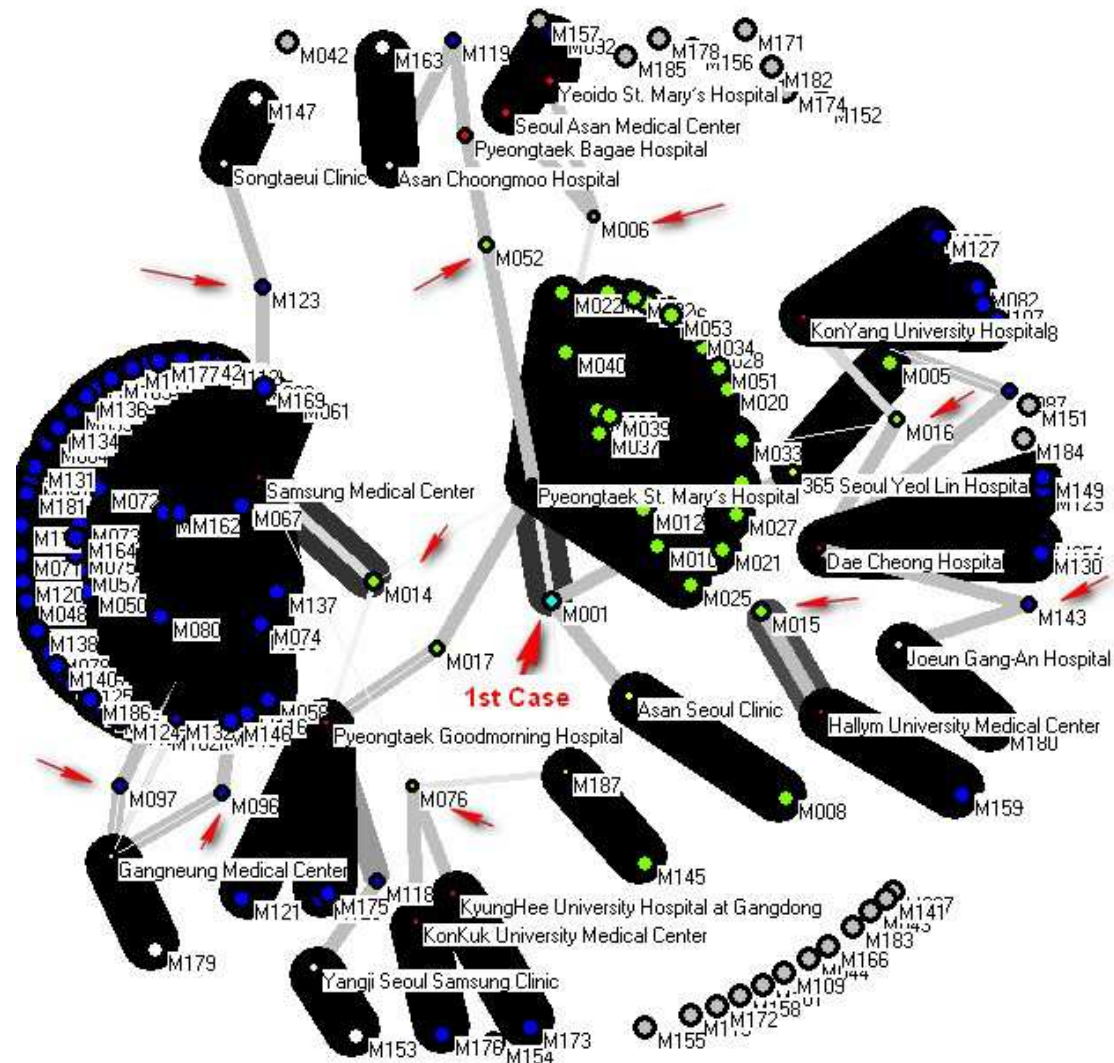


Figure 1. Objectively and statistically displaying the three above-mentioned cases of MERS-CoV in 2015.

An individual (or group/organization) who acts as a mediator (or bridge/gatekeeper) between two or more closely connected groups of people could gain important comparative advantages more than others. Through structure holes of SNA, as a gap between two individuals who have complementary sources to information, individuals (or groups/organizations) hold certain positional advantages/disadvantages from where they are embedded in neighborhoods of a social structure [7-9]. A detailed introduction of structure holes will be described in Methods.

A published paper [3] reported a successful experience in eradicating amoebiasis through a series of intensive mass screenings and treatment interventions in a large institute for mentally retarded in Taiwan. A total of 443 adult mentally retarded residents who were infected with amoebiasis and distributed in 10 houses in past 10 years were included in that study. The prevalence of amoebiasis at the beginning was 14.7%, and then gradually reduced to 12.9%, 10.8%, 6.3%, 3.6%, 2.7%, 3.4%, and 2.2%. Finally, no more positive cases found during the last survey in 2004. All factors (including

age, period of institutionalization, period of institutionalization, and motor activity) but gender ($p=0.002$) are independent of infected subject number. Not only has the association between rooms and years not reported in that article, but also the individual infected occurrence rate for further investigation. It is due to limitation in traditional statistical technique.

We are interested in illustrating structure holes of SNA for finding more information, and redrawing conclusions for the amoebiasis occurrence on gender.

2. Methods

2.1. Data Source

With permission, we obtained the original study dataset from authors of amoebiasis eradication in a large institution for mentally retarded in Taiwan [3], where accommodated around 450 persons who are 18 years old and over and suffer from severe or profound mental retardation or multi-disability with partial retardation in possession of a directory for the

mentally and physically disabled.

The institute in southern Taiwan had 182 employees and 448 residents, 349 male (77.9%) and 99 female (22.1%), who were distributed in 10 houses. In which, the tenth house had 14 residents living outside but near the institute campus. Each of the remaining 9 houses had an average of 47 residents. The house 3 and 4 were for female only. The 9 houses had the same layout, including four bedrooms, two consultant rooms, one restroom, one bathroom, and one central living room.

2.2. Time Points of Screenings and Treatment Interventions

Through consecutive intensive mass screenings and treatment interventions for all the residents, including 2001/Mar; 2001/Aug., 2001/Nov., 2002/Mar., 2002/Aug., 2003/Jan., and 2004/May, 7 surveys totally. Infected cases were treated with standard protocol by CDC Taiwan. The surveys in 1995 and 1997 performed by CDC another program were reported with prevalence rate of 14.7% and 12.9%, respectively.

2.3. Structure Holes of Social Network Analysis

For determining the key factor (in row) associated with a domain (in column), we choose the largest structure hole as the determinant. The formula of structure holes for each component cell is shown in Equation (1), whereas Cij stands for the structure hole of a component cell, others are referred to Table 1 and detail in the link [10].

$$C_{ij} = (P_{ij} + \sum_q P_{iq} P_{qj})^2 \tag{1}$$

Figure 2 shows five nodes from 1 to 5. Each value near to the node means the out-degree structure hole, like values from left to right in Table 1. In contrast, the in-degree structure holes are read from top to bottom. The largest value in column means contribution most to the domain of interest. That is existing the closest row relation to the column. For instance, node 1 contributes most to node 3(1.0 greater than others in column 1), node 5 is attributable to node 2 and node 4, node 1 to node 3, and node 5 can be explained by node 2 and node 4.

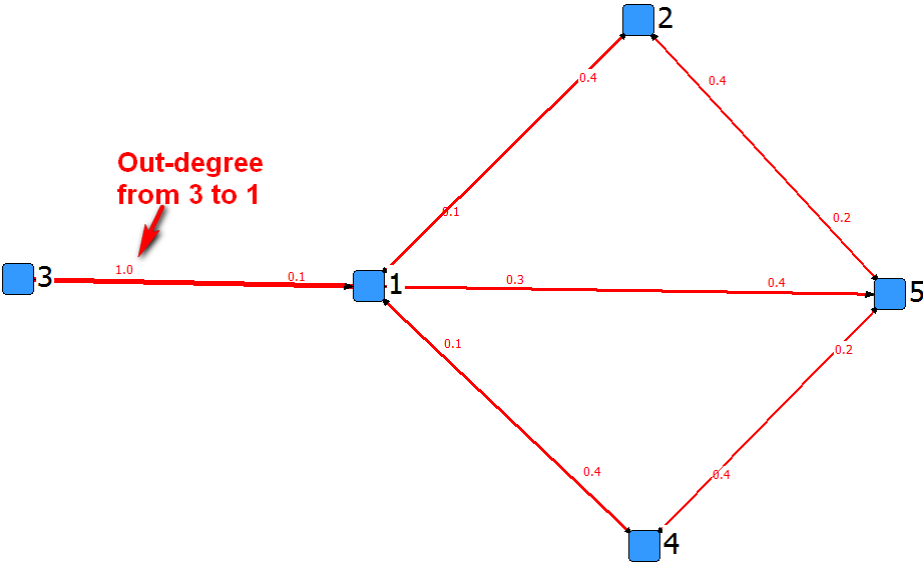


Figure 2. Calculation of structure holes.

Table 1. Calculation of structure holes.

The 1-mode dataset					
	1	2	3	4	5
1	0	1	1	1	1
2	1	0	0	0	1
3	1	0	0	0	0
4	1	0	0	0	1
5	1	1	0	1	0

Proportion for each row in a cell(Pij)					
	1	2	3	4	5
1	0	0.25	0.25	0.25	0.25
2	0.5	0	0	0	0.5
3	1	0	0	0	0
4	0.5	0	0	0	0.5
5	0.33	0.33	0	0.33	0

Piq*Pqj obtained by MMult(Pij,Pij)					
	1	2	3	4	5
1	0.58	0.08	0.00	0.08	0.25
2	0.17	0.29	0.13	0.29	0.13
3	0.00	0.25	0.25	0.25	0.25
4	0.17	0.29	0.13	0.29	0.13
5	0.33	0.08	0.08	0.08	0.41

Pij +Piq*Pqj(=(2)+(3))					
	1	2	3	4	5
1	0.58	0.33	0.25	0.33	0.50
2	0.67	0.29	0.13	0.29	0.63
3	1.00	0.25	0.25	0.25	0.25
4	0.67	0.29	0.13	0.29	0.63
5	0.66	0.41	0.08	0.41	0.41

Cij (=sqrt(4)) if cells in (1) >0						
	1	2	3	4	5	C _i
1	0	0.1	0.1	0.1	0.3	0.5
2	0.4	0	0	0	0.4	0.8
3	1	0	0	0	0	1.0
4	0.4	0	0	0	0.5	0.9
5	0.4	0.2	0	0.2	0	0.8

Note:

To find key factor to the domain: select the largest in column, eg, node 3 related to node 1, node 4 to node 5, and node 1 to node 3.

If not one values exist in a row, choose the largest as the only factor to represent the specific domain

2.4. Organizing Data in SNA Software

To investigate which time point (i.e., eight screening points in column) that can be explained most by a specific house (i.e., 10 accommodation area in row), we designed a pajek control file (directing house to time points with a weight value 1) [11, 12] to execute social network analysis and plot graphical representation with UCINET[13]. A total of 17 nodes (=9 houses + 8 time points, excluding House 10 without any infected case in past 10 years) was set and 250 command codes were programmed, e.g., the code 9 17 1 indicates house 9 had an infected case(weighted with 1) in Year 2003(assigned with 17). The whole codes are included in the link [10] and Additionally supplemental file.

We executed Pajek software through following steps: (1) File>open the control file, (2) Network>Create vector>structure holes, (3) Network>Create partition>k-Neighbors>Input, (4) Draw>Network+first partition +first vector, (5) Layout>Energy>Kamada Kawai>Separate component, (6) Layout>Energy>Fruchterman Reingold>2D. The structure holes for each time point (i.e., C_j) can be obtained, but without the detail calculation results like we show in the link [10] and Additionally supplemental file.

2.5. Chi-Square Test for Determining Key Factors

Standardized Z-scores(=(observed – expected)/√expected) for each cell can be yielded and used to choose the largest value in each column for determining which row factor that can explain the specific domain (in column) most. The detailed module used for executing the selection of key factors is in the link [10] and Additionally supplemental file.

2.6. Data Processing and Analysis

After assigning weight value of 2 (as a weak tie to structure holes in number of tie line in SNA compared to 1 for occurrence once, 0.5 for twice, and 0.33 for trice, etc.) to cases without amoebiasis infection in past10 years, we compared the gender-, House-, and age-specific structure hole and evaluated the differences by one-way ANOVA. All the statistical analyses were conducted using Statistical Package for the MedCalc for Windows 9.5.0.0 MedCalc Software, Mariakerke, Belgium, and all the statistical tests were performed at the two-tailed significance level of 0.05.

3. Results

3.1. Structure Holes and Prevalence Rates of Amoebiasis in Year

Both structure holes and prevalence rates of amoebiasis in year are drawn in Figure 2. We can see the R-square is 0.918. That is, the correlation coefficient reaches 0.96(=√0.918), indicating structure holes hold highly consistent with prevalence rates using traditional statistical technique.

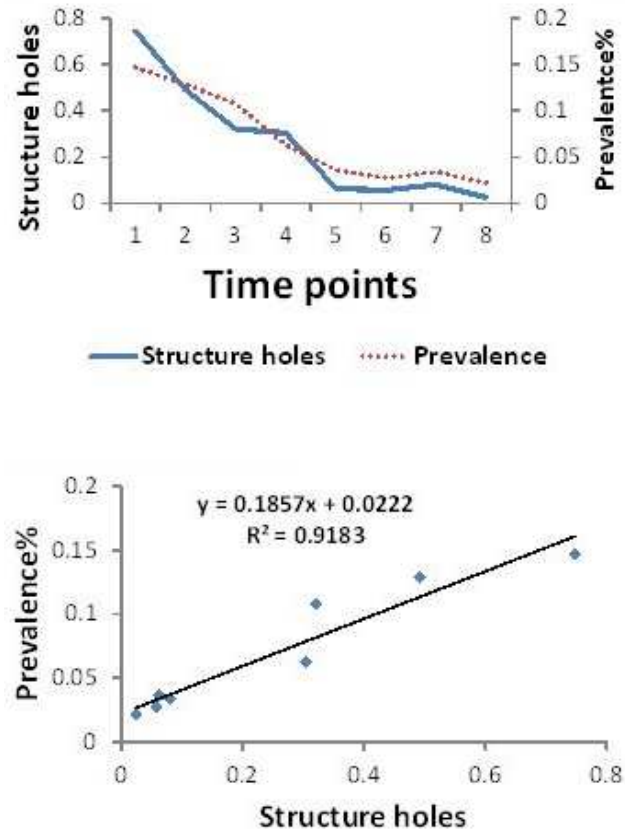
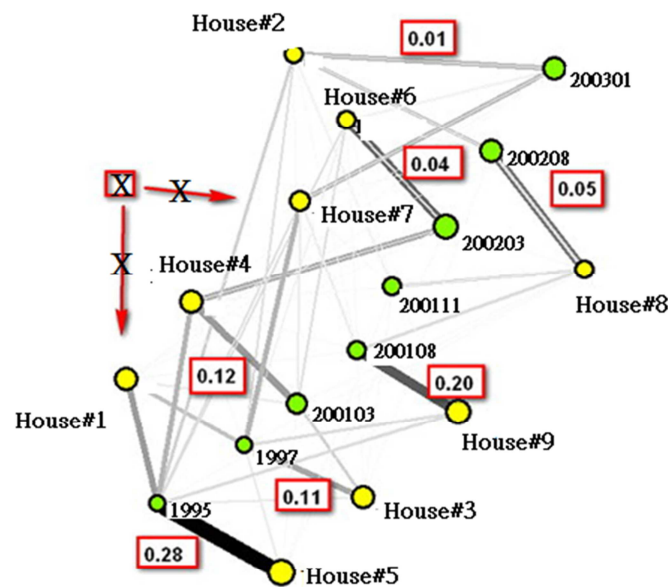
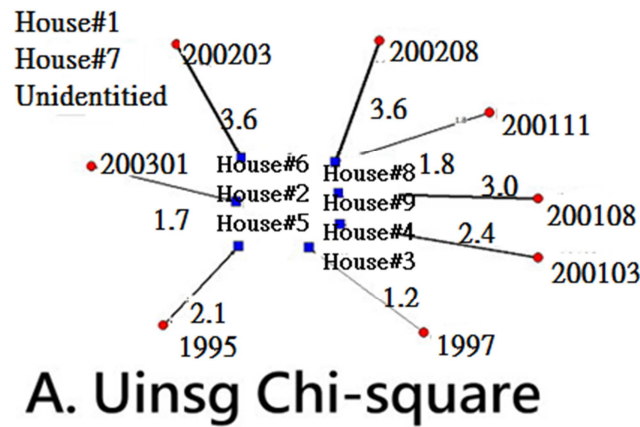


Figure 3. relations between structure holes and prevalence rates in year.

3.2. Comparison Between Structure Holes and Chi-Square Test

The result yielded by both structure holes and Chi-square test are similar in classifying houses mostly related to the time points (i.e., House 2 related to Time 9, House3 to Time 2, House 4 to Time 3, House 5 to Time 1, house 6 to Time 6, House 8 to time 8, House 9 to Time 4), see Table 2, Table 3, and Figure 3. Chi-square value with 99.29 (p<0.001) is needed to further investigate which row is closely related to a specific column. We found that House 1 and House 7 were not associated to any time points, indicating the number of infected cases in both houses were not so high enough to explain one of the time points.



B.Using structure holes

Figure 4. Using two ways to show the room associated with the year Referring to the tie line between notes (the stronger, the thicker, in Table 3(6)).

Table 2. (1) Subjects infected recorded across years and rooms.

House	1995	1997	200103	200108	200111	200203	200208	200301	Sum
1	10	8	5	2	0	1	2	1	63
2	7	6	3	3	3	0	4	3	55
3	5	9	7	3	2	0	1	0	48
4	15	8	17	3	2	4	0	1	24
5	8	3	2	0	1	1	0	0	15
6	5	4	5	1	1	5	1	2	12
7	9	13	8	6	2	0	1	3	15
8	4	4	1	6	4	1	6	0	10
9	2	2	0	4	1	0	0	0	242
Sum	58	58	54	100	30	48	84	52	18
(2)Expected scores									
1	7.51	6.59	5.55	3.24	1.85	1.39	1.73	1.16	
2	7.51	6.59	5.55	3.24	1.85	1.39	1.73	1.16	
3	6.99	6.13	5.16	3.01	1.72	1.29	1.61	1.08	
4	12.95	11.35	9.56	5.58	3.19	2.39	2.99	1.99	
5	3.88	3.41	2.87	1.67	0.96	0.72	0.90	0.60	

House	1995	1997	200103	200108	200111	200203	200208	200301	Sum
6	6.22	5.45	4.59	2.68	1.53	1.15	1.43	0.96	
7	10.88	9.54	8.03	4.69	2.68	2.01	2.51	1.67	
8	6.73	5.90	4.97	2.90	1.66	1.24	1.55	1.04	
9	2.33	2.04	1.72	1.00	0.57	0.43	0.54	0.36	
p=	0.0003								
Chi-sq=	99.29								
(3)Standardized Z-scores									
1									
2								1.72	
3		1.16							
4			2.41						
5	2.09								
6						3.60			
7									
8							3.57		
9				2.99					

Table 3. (1) Using original data to calculate structure holes.

Notes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Sum
House1	0	0	0	0	0	0	0	0	0	10	11	5	2	0	1	2	1	29
House 2	0	0	0	0	0	0	0	0	0	7	6	3	3	3	0	4	3	29
House 3	0	0	0	0	0	0	0	0	0	5	9	7	3	2	0	1	0	27
House 4	0	0	0	0	0	0	0	0	0	15	8	17	3	2	4	0	1	50
House 5	0	0	0	0	0	0	0	0	0	8	3	2	0	1	1	0	0	15
House 6	0	0	0	0	0	0	0	0	0	5	4	5	1	1	5	1	2	24
House 7	0	0	0	0	0	0	0	0	0	9	13	8	6	2	0	1	3	42
House 8	0	0	0	0	0	0	0	0	0	4	4	1	6	4	1	6	0	26
House 9	0	0	0	0	0	0	0	0	0	2	2	0	4	1	0	0	0	9
1995	10	7	5	15	8	5	9	4	2	0	0	0	0	0	0	0	0	65
1997	8	6	9	8	3	4	13	4	2	0	0	0	0	0	0	0	0	57
200103	5	3	7	17	2	5	8	1	0	0	0	0	0	0	0	0	0	48
200108	2	3	3	3	0	1	6	6	4	0	0	0	0	0	0	0	0	28
200111	0	3	2	2	1	1	2	4	1	0	0	0	0	0	0	0	0	16
200203	1	0	0	4	1	5	0	1	0	0	0	0	0	0	0	0	0	12
200208	2	4	1	0	0	1	1	6	0	0	0	0	0	0	0	0	0	15
200301	1	3	0	1	0	2	3	0	0	0	0	0	0	0	0	0	0	10
(2)Obtaining Pij(=proportion for each row, e.g., Room1 to column 17=1/29=0.03)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
House1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.34	0.28	0.17	0.07	0.00	0.03	0.07	0.03	
House 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.21	0.10	0.10	0.10	0.00	0.14	0.10	
House 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.33	0.26	0.11	0.07	0.00	0.04	0.00	
House 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.16	0.34	0.06	0.04	0.08	0.00	0.02	
House 5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.20	0.13	0.00	0.07	0.07	0.00	0.00	
House 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.17	0.21	0.04	0.04	0.21	0.04	0.08	
House 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.31	0.19	0.14	0.05	0.00	0.02	0.07	
House 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.15	0.04	0.23	0.15	0.04	0.23	0.00	
House 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.22	0.00	0.44	0.11	0.00	0.00	0.00	
1995	0.15	0.11	0.08	0.23	0.12	0.08	0.14	0.06	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1997	0.14	0.11	0.16	0.14	0.05	0.07	0.23	0.07	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
200103	0.10	0.06	0.15	0.35	0.04	0.10	0.17	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
200108	0.07	0.11	0.11	0.11	0.00	0.04	0.21	0.21	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
200111	0.00	0.19	0.13	0.13	0.06	0.06	0.13	0.25	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
200203	0.08	0.00	0.00	0.33	0.08	0.42	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
200208	0.13	0.27	0.07	0.00	0.00	0.07	0.07	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
200301	0.10	0.30	0.00	0.10	0.00	0.20	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

Table 3. (3) Using Excel MMULT function to calculate $P_{iq} * P_{qj}$ of above metric.

[illegible]

$P_{iq} \cdot P_{qj}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
House 6	0.11	0.10	0.09	0.23	0.06	0.16	0.14	0.08	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
House 7	0.12	0.12	0.12	0.19	0.05	0.08	0.19	0.09	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
House 8	0.10	0.15	0.10	0.13	0.04	0.08	0.15	0.20	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
House 9	0.10	0.12	0.11	0.14	0.05	0.06	0.19	0.15	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1995	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.22	0.20	0.10	0.06	0.05	0.05	0.04
1997	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.25	0.19	0.11	0.06	0.04	0.05	0.04
200103	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.23	0.24	0.08	0.05	0.06	0.03	0.04
200108	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.23	0.14	0.18	0.08	0.03	0.08	0.03
200111	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.21	0.15	0.15	0.09	0.04	0.09	0.04
200203	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.18	0.23	0.06	0.05	0.13	0.04	0.04
200208	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.21	0.11	0.15	0.10	0.03	0.15	0.04
200301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.23	0.18	0.10	0.06	0.05	0.06	0.07
(4) Calculation of $P_{ij} + P_{iq} \cdot P_{qj}$																	
$P_{ij} + P_{iq} \cdot P_{qj}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
House1	0.00	0.11	0.11	0.20	0.07	0.09	0.17	0.09	0.03	0.34	0.28	0.17	0.07	0.00	0.03	0.07	0.03
House 2	0.11	0.15	0.10	0.16	0.05	0.08	0.17	0.13	0.04	0.24	0.21	0.10	0.10	0.10	0.00	0.14	0.10
House 3	0.12	0.11	0.13	0.20	0.06	0.08	0.18	0.10	0.04	0.19	0.33	0.26	0.11	0.07	0.00	0.04	0.00
House 4	0.12	0.09	0.11	0.25	0.07	0.11	0.16	0.07	0.03	0.30	0.16	0.34	0.06	0.04	0.08	0.00	0.02
House 5	0.13	0.10	0.10	0.23	0.09	0.10	0.15	0.07	0.03	0.53	0.20	0.13	0.00	0.07	0.07	0.00	0.00
House 6	0.11	0.10	0.09	0.23	0.06	0.16	0.14	0.08	0.02	0.21	0.17	0.21	0.04	0.04	0.21	0.04	0.08
House 7	0.12	0.12	0.12	0.19	0.05	0.08	0.19	0.09	0.04	0.21	0.31	0.19	0.14	0.05	0.00	0.02	0.07
House 8	0.10	0.15	0.10	0.13	0.04	0.08	0.15	0.20	0.05	0.15	0.15	0.04	0.23	0.15	0.04	0.23	0.00
House 9	0.10	0.12	0.11	0.14	0.05	0.06	0.19	0.15	0.09	0.22	0.22	0.00	0.44	0.11	0.00	0.00	0.00
1995	0.15	0.11	0.08	0.23	0.12	0.08	0.14	0.06	0.03	0.29	0.22	0.20	0.10	0.06	0.05	0.05	0.04
1997	0.14	0.11	0.16	0.14	0.05	0.07	0.23	0.07	0.04	0.26	0.25	0.19	0.11	0.06	0.04	0.05	0.04
200103	0.10	0.06	0.15	0.35	0.04	0.10	0.17	0.02	0.00	0.27	0.23	0.24	0.08	0.05	0.06	0.03	0.04
200108	0.07	0.11	0.11	0.11	0.00	0.04	0.21	0.21	0.14	0.22	0.23	0.14	0.18	0.08	0.03	0.08	0.03
200111	0.00	0.19	0.13	0.13	0.06	0.06	0.13	0.25	0.06	0.23	0.21	0.15	0.15	0.09	0.04	0.09	0.04
200203	0.08	0.00	0.00	0.33	0.08	0.42	0.00	0.08	0.00	0.27	0.18	0.23	0.06	0.05	0.13	0.04	0.04
200208	0.13	0.27	0.07	0.00	0.00	0.07	0.07	0.40	0.00	0.21	0.21	0.11	0.15	0.10	0.03	0.15	0.04
200301	0.10	0.30	0.00	0.10	0.00	0.20	0.30	0.00	0.00	0.24	0.23	0.18	0.10	0.06	0.05	0.06	0.07

Table 3. (5) Getting structure holes for each cell with $P_{ij} + P_{iq} \cdot P_{qj})^*$ $P_{ij} + P_{iq} \cdot P_{qj})$ if the value in each cell of the original data is greater than zero.

$P_{ij} + P_{iq} \cdot P_{qj})^{\wedge 2}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	C.j
House1										0.12	0.08	0.03	0.00	0.00	0.00	0.00	0.00	0.24
House 2										0.06	0.04	0.01	0.01	0.01	0.00	0.02	0.01	0.16
House 3										0.03	0.11	0.07	0.01	0.01	0.00	0.00	0.00	0.23
House 4										0.09	0.03	0.12	0.00	0.00	0.01	0.00	0.00	0.24
House 5										0.28	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.35
House 6										0.04	0.03	0.04	0.00	0.00	0.04	0.00	0.01	0.17
House 7										0.05	0.10	0.04	0.02	0.00	0.00	0.00	0.01	0.21
House 8										0.02	0.02	0.00	0.05	0.02	0.00	0.05	0.00	0.18
House 9										0.05	0.05	0.00	0.20	0.01	0.00	0.00	0.00	0.31
1995	0.02	0.01	0.01	0.05	0.02	0.01	0.02	0.00	0.00									0.14
1997	0.02	0.01	0.02	0.02	0.00	0.00	0.05	0.00	0.00									0.14
200103	0.01	0.00	0.02	0.13	0.00	0.01	0.03	0.00	0.00									0.20
200108	0.01	0.01	0.01	0.01	0.00	0.00	0.05	0.05	0.02									0.15
200111	0.00	0.04	0.02	0.02	0.00	0.00	0.02	0.06	0.00									0.16
200203	0.01	0.00	0.00	0.11	0.01	0.17	0.00	0.01	0.00									0.31
200208	0.02	0.07	0.00	0.00	0.00	0.00	0.00	0.16	0.00									0.26
200301	0.01	0.09	0.00	0.01	0.00	0.04	0.09	0.00	0.00									0.24
Ci.	0.09	0.23	0.08	0.35	0.03	0.24	0.25	0.28	0.03	0.75	0.49	0.32	0.30	0.06	0.06	0.08	0.02	
Note. Using Excel function(=IF(original cells=0,0, previous cells^2) to get the result																		
(6)Selecting largest value in column being the bridge toward the specific domain																		
	1	2	3	4	5	6	7	8	9	T1	T2	T3	T4	T5	T6	T7	T8	
House1																		
House 2																	0.01	
House 3											0.11							
House 4												0.12						
House 5										0.28								
House 6															0.04			
House 7																		
House 8														0.02		0.05		
House 9													0.20					

Note. Using Excel function (=IF(previous cell =max(the column), max(the column), "")) to get the largest value for each column

3.3. ANOVA for Testing Group Mean Difference in Structure Hole

We compared the differences of gender-, House-, and age-specific structure hole and obtained results of gender ($F=13.30$, $p<0.001$), house ($F=5.771$, $p<0.001$), and age ($F=0.364$, $p=0.636$). Female with average structure hole of 1.36 less than male 1.57, indicating amoebiasis occurrence on female easier than male. The lowest structure hole ($=1.195$) is in House 4 for female only and results in different structure hole in house type. The highest average structure hole ($=2.0$, without any amoebiasis case in all time points) is house 10.

Referring House 4 for female only in Table 2, we obtained number of amoebiasis occurrence as {15, 8, 17, 3, 2, 4, 0, 1} in time points and found that the high number of 17 in third time point (Year 200103) might mislead to draw conclusions: (1) female easier to suffer amoebiasis than male, (2) different structure hole in house type, and (3) time point 3 (Year 200103) was fully explained by House 4.

4. Discussion

Principal Findings

Our most important findings were that (1) structure holes have closely high consistent to prevalence rates, (2) the structure holes have similar effects examining mostly strong association between rows and columns using Chi-square test, and (3) individual structure holes yielded by SNA can be as predicted parameters used for ANOVA and linear regression.

Implications and Future Considerations

Social networks have been successfully applied to many diverse fields [14]. These include, broadly, the social sciences [15], human disease [16, 17], scientific collaboration [18, 19], social contagion [20], and many others.

Comparing to the non-parameter Chi-square test used for a cross (contingency) table ($p=0.002$) [3], the statistical power is less than the parameter one (e.g., the structure hole used in this study, $p<0.001$). This is the advantage using the additionally yielded individual structure holes over the traditional statistics without such personal parameters used for further analyses.

We used structure holes of SNA to gain more information than the traditional statistics and drew a conclusion that the observed number of amoebiasis for House 4 in Year 200103 is significantly beyond our expectation with a standardized Z-score of 2.41 (see Table 2(3)) and might mislead to conclude: (1) female easier to suffer amoebiasis than male, (2) different structure hole in house type, and (3) time point 3 (Year 200103) was fully explained by House 4. Further investigation such as a typo in digit is required to reconfirm.

Strengths of this study

Due to the previous published paper [3] without reporting the house association to the amoebiasis, we were interested in investigating whether house factor is associated with the amoebiasis occurrence using SNA and found that female in House 4 at time point 3 (Year 200103) should be further investigated under scrutiny. Why the three values (15, 17 in

House 4 at the 1st and third time point and 8 in House 5 at the 1st time point) were significantly beyond our expectation.

Similarly, House 4 was subject to accommodating Female only that caused amoebiasis occurrence higher than other houses. After further investigating House 4 (with a standardized Z-score of 2.09, see Table 2(3)) strongly associated with Year 1995 (time point), see Table 2 and Table 3, we can draw another conclusion that there might be no any difference in group of gender and house when discarding data of the first aberrant time point. This is why no any further investigation in the previous published paper [3] using traditional statistics. In contrast, social structures commonly visualized through social network analysis using structure holes can yield similar result in comparison with the traditional statistics.

We also demonstrated the calculation process of structure holes in SNA (see the link [10] and Additionally supplemental file) which is superior to other SNA studies [21-23] without disclosing detailed and useful information of calculation process to readers.

Limitations

This study has at least three limitations. First, structure holes in Table 3(6) are not like Chi-square method showing standardized Z-score in Table 2(3) for distinguishing its statistical significance of effect. Second, SNA visualization to present data should be cautiously interpreted with implied knowledge such as a pivot role with the highest structure hole in counterparts. For instance, important cases were found with a high value of structure hole in a thicker tie line. Third, we have not verified whether those numbers of female amoebiasis occurrence at the 1st and third time points are attributable to typos or other reasons such as clinical features of asymptomatic, chronic and long incubation time. Our conclusion should be challenged regarding houses independent of time points in amoebiasis occurrence at random when all conditions of housing equipment and environment are equal. Further studies are recommended and encouraged to compare both methods of Chi-square test and structure holes in consistence for determining key factors using other healthcare data.

5. Conclusions

Because of advances in computer technology, the number of healthcare studies for the group classification and association assertion continues to increase and benefit comparisons of data if structure holes of SNA can be applied.

Acknowledgements

We thank Frank Bill who provided medical writing services to the manuscript and Chi-Mei Medical Center for offering grand fund to the cost spent of the study.

Funding

There are no sources of funding to be declared.

Availability of Data and Materials

This research is based on a methodology study. All codes and data can be obtained from <http://www.healthup.org.tw/structureHoles2.zip>

Authors' Contributions

TW conceived and designed the study, SB interpreted the data, and Both of authors drafted the manuscript as well as approved the final manuscript.

Authors' Information

TW is an assistant professor at ChiMei Medical Center, Taiwan. He is an expert in computer science and Rasch modelling, mainly in the field of data analysis using statistical technique. SB is a medical doctor with PhD working as an specialist in ChiMei Medical Center, Taiwan.

Competing Interests

The authors declare that they have no competing interests.

Consent for Publication

Not applicable.

Ethics Approval and Consent to Participate

Not applicable.

List of Abbreviations

ANOVA: Analysis of variance
MERS: Middle East respiratory syndrome
SNA: social network analysis
VBA: Visual Basic for Applications

Additional Files

Microsoft Excel-based computer module for calculation process of structure holes

References

- [1] Friedlander S, Silver M. A quantitative study of the determinants of fertility behavior. *Demography*. 1967;4 (1): 30-70.
- [2] Zhai L, Fu S, Zhang C, Liu Y, Wang L, Liu G, Yang M. An efficient classification method based on principal component and sparse representation. *Springerplus*. 2016;5 (1): 832.
- [3] Su SB, Guo HR, Chuang YC, Chen KT, Lin CY. Eradication of amoebiasis in a large institution for mentally retarded in Taiwan. *Infect Control Hosp Epidemiol*. 2007;28 (6): 679-83.
- [4] Lai WP, Chien TW, Lin HJ, Su SB, Chang CH. A screening tool for dengue fever in children. *Pediatr Infect Dis J* 2013;32 (4): 320-4.
- [5] Cowling BJ, Park M, Fang VJ, Wu P, Leung GM, Wu JT. Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015. *Euro Surveill*. 2015;20 (25): 7-13.
- [6] Cyram Inc. Data: MERS-CoV. 2016/07/14 retrieved at <http://www.netminer.com/community/event/event-readList.do>
- [7] Burt RS. *Structural Holes: The Social Structure of Competition*. Cambridge: Harvard University Press, 1995.
- [8] Burt RS. Structural holes and good ideas. *American Journal of Sociology* 2004;110: 349–399.
- [9] Burt RS, Hogarth RM, Michaud C. The social capital of French and American managers. *Organization Science* 2000;11: 123–147.
- [10] Chein. Calculation of structure holes. 2016/07/14 retrieved at <http://www.healthup.org.tw/structureHoles2.zip>
- [11] Batagelj V, Mrvar A. Pajek - Analysis and Visualization of Large Networks. in Jünger, M., Mutzel, P., (Eds.) *Graph Drawing Software*. Springer, Berlin 2003;77-103.
- [12] Batagelj V, Mrvar A. Pajek-program for large network analysis. *Connections*. 1998;21: 47–57.
- [13] Borgatti SP, Everett MG, Freeman LC. *UCINET 6.0*, Version 1.00. Lexington, KY: Analytic Technologies; 1999.
- [14] Grunspan DZ, Wiggins BL, Goodreau SM. Understanding Classrooms through Social Network Analysis: A Primer for Social Network Analysis in Education Research. *CBE Life Sci Educ*. 2014;13 (2): 167–178.
- [15] Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. *Science*. 2009;323: 892–895.
- [16] Morris M. *Network Epidemiology: A Handbook for Survey Design and Data Collection*. Oxford, UK: Oxford University Press; 2004.
- [17] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12: 56–68.
- [18] Newman ME. The structure of scientific collaboration networks. *Proc Natl Acad Sci USA*. 2001;98: 404–409.
- [19] West JD, Bergstrom TC, Bergstrom CT. The Eigenfactor MetricsTM: a network approach to assessing scholarly journals. *Coll Res Libr*. 2010;71: 236–244.
- [20] Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. *Stat Med*. 2013;32: 556–577.
- [21] Wang Y, Tan XD, Zhou C, Zhou W, Peng JS, Ren YS, Ni ZL, Liu B, Yang F, Gao XD. Exploratory social network analysis and gene sequencing in people who inject drugs infected with hepatitis C virus. *Epidemiol Infect*. 2016;13: 1-11.
- [22] Lee IC, Ting TT, Chen DR, Tseng FY, Chen WJ, Chen CY. Peers and social network on alcohol drinking through early adolescence in Taiwan. *Drug Alcohol Depend*. 2015;153: 50-8.
- [23] Zare-Farashbandi F, Geraei E, Siamaki S. Study of co-authorship network of papers in the Journal of Research in Medical Sciences using social network analysis. *J Res Med Sci*. 2014;19(1): 41-6.