Applying clustering coefficient to the pattern of international author collaboration in the topic of addiction and clinical research

Julie Chi Chow[1], Tsair-Wei Chien[2], Chien-Ho Lin[3], Willy Chou[4,5*]

OBJECTIVE: To apply cluster coefficient (CC) to the pattern of international author collaborations on the topic of addiction and clinical research using data from Medline and to visualize the results using Google maps and social network analysis (SNA).

METHOD: We obtained 647 abstracts on December 22, 2017, from Medline based on the keywords of addiction and clinical research since 1989. The author names, countries, and keywords were recorded. We also made a note of the following features: (1) nation distribution for 1st author's and most popular journals; (2) eminent authors in addiction and clinical research, (3) notable keywords representing addiction and clinical research, and (4) cluster coefficients similar or different between author and keyword networks. We programmed Microsoft Excel VBA routines to extract data from Medline. Google Maps and SNA Pajek software were performed to display the graphical representations with an easy-to-read feature for readers.

RESULTS: We found that (1) the most number of papers on the topic of addiction and clinical research are from the U.S. (225, 37.19%) and Australia (86,14.21%); (2) the productive authors with the highest cluster coefficient in addiction and clinical research are Abe, Yoshinari(Canada) and Agid, Ofer(Japan); (3) the most linked keywords are clinical research and drug addiction; (4) both author and keyword networks present higher CC in their networks.

CONCLUSION: Social network analysis provides wide and deep insight into the relationships among nations and co-authors. The results can provide readers with knowledge and concept diagram for future submission to a journal in addiction and clinical research.

Key Words: *Authorship collaboration; Google Maps; Social network analysis; Medline*

Addiction is a brain disorder characterized by compulsive engagement in rewarding stimuli, despite adverse consequences (1,2). Although the involvement of some psychosocial factors is relevant and a biological process is induced by repeated exposure to an addictive stimulus, the core pathology drives the development and maintenance of addiction remain unclear (2). Also, the two properties that characterize all addictive stimuli are that they are reinforcing and intrinsically rewarding all the time. This intoxication culture among Danish adolescents is alarming as adolescents are particularly vulnerable to substance use.

As of December 20, 2017, more than 57,795 papers were found by the keyword addiction and 12,105 by addiction (title) in Medline library. There are four topics that intrigue us, including (1) which countries/regions contribute most in the paper publication of addiction? (2) which journals published most of the addiction and clinical research? (3) which keywords are more frequently found in addiction and clinical research? (4) which authors are the most eminent on the topic of addiction and clinical research.

It is hard to find the relationship between multiple entities. We often can only get a sense of their correlation. Just as, when many customers purchase their goods by placing them in a shopping cart. What are the correlations of these goods. An apocryphal story was often told to tell us the concept of co-occurrence that is about beer and diaper sales which usually goes along with a strong correlation on Friday (4-6). Many data scientists have developed ways to discover new knowledge from the vast quantities of increasingly available information (7), especially applying social network analysis (SNA) (8-10) to big data analysis.

Authorship collaboration using SNA is an example illustrated by many authors in recent years (8) because co-authors among researchers form a type of social network. Whether the keyword network in addiction and clinical research earns an equally important impact is interesting to explore. We are thus interested in using SNA to explore the features in addiction and clinical research from published papers we observed in Medline library.

Google maps have provided users to gain an overall geospatial visualization

(11,12). However, few were found in Medline library when searching the keyword google map (Title) on December 22, 2017. Many papers (9,10) have conducted studies on co-author collaboration. However, none display these results combined with Google maps and social network analysis.

Our aims are to apply clustering coefficient (12) to the pattern of international author collaboration in addiction and clinical research on the following topics: (i) nation and journal distribution across years; (ii) the most eminent authors in addiction and clinical research; (iii) the recent research domains defined by authors; (iv) the cluster coefficients in different networks.

METHODS

Data sources

We programed Microsoft Excel visual basic for applications (VBA) modules to extract abstracts and their corresponding coauthor names as well as author-defined keywords for each article on December 22, 2017, from Medicine National Institutes of Health (Medline) since 1989. Only those abstracts published with the keyword in (addiction and "clinical research") (title) and labeled with Journal Article were included. Others like those marked with Published Erratum, Editorial or without author nation name were excluded from this study. A total of 674 eligible abstracts were obtained from Medline in which 605 with nation name for the 1st author were found.

Data arrangement to fit SNA requirement

Before visualizing our results using SNA, we organized data in compliance with the format and guidelines defined by Pajek software (13). Microsoft Excel VBA routines were used to deal with data fitting to the SNA requirement.

Graphical representations to report

(1) **Author nations and their relations:** Two tables (i.e. columns for publication years and rows for the 1st author nations and journals) were made to represent the distribution of nations and journals in addiction and clinical research. The bigger bubble means, the number of the nodes (i.e.,

[1]Department of Paediatrics, Chi Mei Medical Center, Tainan, Taiwan, [2]Research Department, Chi-Mei Medical Center, Tainan, Taiwan, [3]Department of Psychiatry, Chi-Mei Medical Center, Taiwan, [4]Department of Physical Medicine and Rehabilitation, Chi Mei Medical Center, Tainan, Taiwan, [5]Department of Recreation and Health-Care Management and Institute of Recreation Industry Management, Chia Nan University of Pharmacy, Tainan, Taiwan

Correspondence: Dr. Willy Chou, Department of Recreation and Health-Care Management and Institute of Recreation Industry Management, Chia Nan University of Pharmacy, Tainan, Taiwan. Tel: 886-6-2812811, email: ufan0101@ms22.hinet.net

nations, authors or keywords in this study). The wider line indicates, the stronger relations between the two ties. Community clusters with a closer relation are filled with different colors in the bubbles.

(2)   **Keywords to present the research domain:** If our keywords represent the mainstream topics in addiction and clinical research, the stronger relations between the two keywords can be highlighted and linked by SNA. Like the concept of co-occurrence about beer and diaper sales. The presentation for the bubble and line is interpreted similarly to the previous section.

Statistical tools and data analyses

Google Maps (14) and SNA Pajek software (13) were used to visualize the eminent authors with the keywords of addiction and clinical research (Additional file 1). Author-made Excel VBA modules were applied to organize data. Cluster coefficient (12) represents the density of a network as below:

=, whereas n=the number of nodes in a network and m=the number of other connected nodes with a specific ego node. A significant lever (>1.96) is defined by t-value as the formula (=cc*√((n-2)/(1-cc2)).

In contrast, E-I index is defined by the formula, where EL= the number of external friendship links and IL= the number of internal friendship links (15,16). The negative E-I index means a coherence cluster in existence. Similarly, the higher CC indicates many members are other linked members' friends. Density is defined as the ratio of the linked members over all possible linked members.

RESULTS

Author nations and their relations

A total of 605 eligible papers with complete author nations based on journal article since 1989 are shown in Table 1. We can see that the most number of papers are from the U.S. (225, 37.19%) and Australia (86,14.21%). The trend in the number of publications is present in the column of growth in Table 1. All continents present a positive increase in a paper publication.

The diagram shown by SNA and Google Maps in Figure 1 displays the author's collaboration among nations based on addiction and clinical research. Overall, the highest productive nations are from the U.S. and Europe (Figure 1). Any nation collaborated with other nations are shown with a blue line. Interested authors are recommended to click the bubble of interest to see details on a website at reference (17).
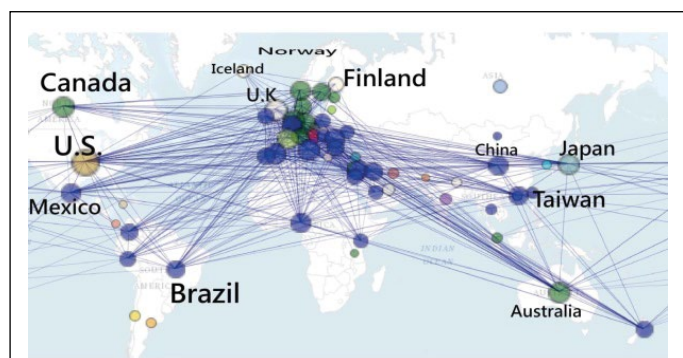

Figure 1) *Google Maps on the topic of author collaboration in addiction and clinical research (CC=0.51)*

Journals and the trend

A total of 674 eligible abstracts were included in the current study of addiction and clinical research. The most numbers of journals in production are Addiction (60, 8,90%) and Drug Alcohol Depend (11, 1.63%). The trend for a journal is shown in the column of correlation in Table 2. We can see the numbers of articles in Journal of Subst Abuse Treat is increasing.

The eminent authors in addiction and clinical research

The most eminent authors who published the most number of paper in addiction and clinical research are Yoshinari Abe (Canada) and Ofer Agid (Japan) (Figure 2). The link on website was referred to reference (18).

Keywords to present the research domain of addiction and clinical research

The most linked keywords are clinical research and drug addiction (Figure 3)

or click it on the reference (19). We can see that the keywords consist of many clusters with different cluster coefficients.

Cluster coefficients in a network

Each cluster has its cluster coefficient to represent the density of a network. We found that author clusters earn an approximately equivalent CC of keyword clusters (Table 3). Cluster coefficient has a significant effect in comparison with a significant t-value (>2.0), indicating both author and keyword networks have the feature of high density in a network (Table 3).

DISCUSSION

This study found that 1) the most number of papers in addiction and clinical research are from the U.S.(225, 37.19%) and Australia (86, 14.21%); (2) the productive authors with the highest cluster coefficient in addiction and clinical research are Abe, Yoshinari (Canada) and Agid, Ofer (Japan); (3) the most linked keywords are clinical research and drug addiction; (4) both author and keyword networks present higher CC in their networks.

What this adds to what was known

Many previous types of research (8-10) have inspected coauthor collaboration using social network analysis. The results (the most number of articles in addiction and clinical research are from the U.S. and Europe) are similar to the findings that dominant nations in science come from the U.S. and Europe (20,21). We showed a novel method incorporating SNA with Google maps to explore the data of publication in addiction and clinical research. It can be seen that visual representations rendered to the reader are rare in literature. Traditionally, it is very hard to observe the association of two or more symptoms or ties together appeared in a network at the moment.

Journal authorship collaboration can be compared with each other using Google Maps. We can see that many links are connecting two nations. Such a network can be defined as a collaboration pattern which results are similar to the previous study (5). Accordingly, the researchers have a high level of international author collaboration in anesthesiology, which is consistent with the previous studies on investigating scientific collaboration of Iranian Psychology and Psychiatry Researchers (22,23).

There are 1,084 papers with the keyword social network analysis in the paper title when searching Medline on December 21, 2017, in which two papers (24,25) incorporated MeSH into SNA to disclose relevant knowledge to readers. However, no such papers have incorporated Google maps link as we did in the current study. The CCs we illustrated (17-19) are called overall CCs. The highest value represents the closer networks, as shown in figures; 0.56, 0.90 and 0.96, respectively. The global CCs and individual CC are shown on the map with each cluster and each node and have their different meaning and implication on the density of an existing network. Evidence suggests that in most real-world networks, nodes tend to create a tightly knit groups characterized by a relatively high density of ties; this likelihood tends to be greater than the average probability of a tie randomly established between two nodes (14,26).

What it implies and what should be changed?

Scientific publication is one of the objective measurements to evaluate the achievements of a medical specialty or discipline (27,28). It is worth combining SNA and Google Maps to disclose knowledge and information to the readers for reference in the future.

Many algorithms and measures (or indicators) have been developed using SNA to graphically explore data (8). This kind of data can identify the interested target journal for an author with paper of selected topics; It means that the core subject can be analyzed using the centrality measure (10,23) yielded by SNA.

Strengths of this study

The way we incorporated SNA with Google Maps is unique in comparison with others (8-10). The network density can be replaced by CC used in this study. Another strength and feature for this study are that Google Maps are sophisticatedly used and linked in references (17-19) for each interested topic, the reader can manipulate the link by their ways as a dashboard to understand the feature of the author collaboration or the keyword relationship. The nation distribution in Figure 1 is an easily understood feature of international author collaborations on addiction and clinical research. One picture is worth ten thousand words. We hope following studies can report other types of information using Google API to readers in the future.

Limitations and future study

The interpretation and generalization of the conclusions should be cautious.
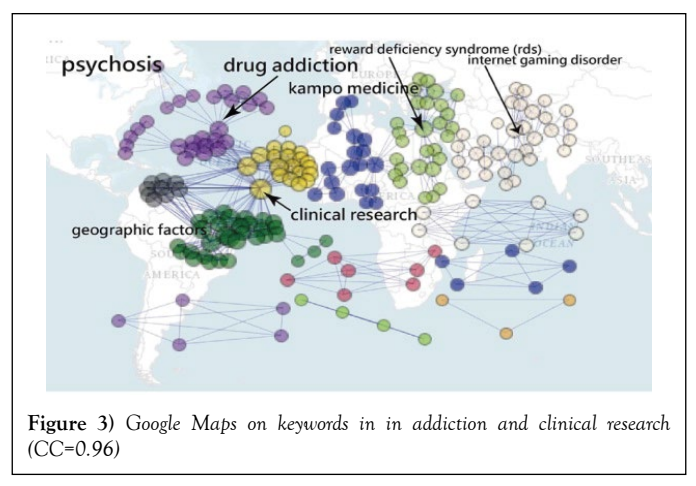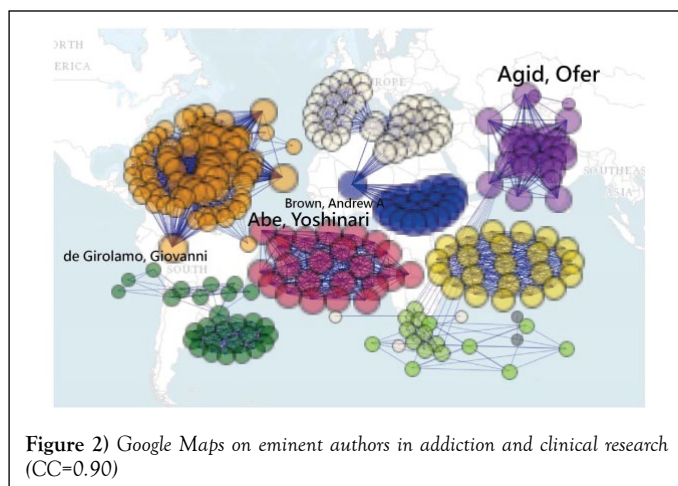
TABLE 1

Nation distribution based on the 1st author for papers published in addiction and clinical research.

Continent	1989-2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	total	%	Growth
Africa	**0**	-	-	-	-	-	-	-	-	-	**3**	**3**	**0.50**	**0.52**
Nigeria	0	-	-	-	-	-	-	-	-	-	2	2	0.33	0.52
Tanzania	0	-	-	-	-	-	-	-	-	-	1	1	0.17	0.52
Asia	**8**	-	**1**	**1**	-	**1**	**2**	**8**	**3**	**11**	**19**	**54**	**8.93**	**0.81**
China	1	-	-	-	-	-	-	1	2	6	9	19	3.14	0.79
Japan	3	-	-	-	-	-	1	4	-	1	4	13	2.15	0.63
Iran	0	-	1	-	-	-	-	1	1	2	2	7	1.16	0.74
Others	4	0	0	1	0	1	1	2	0	2	4	15	2.48	0.73
Europe	**13**	**6**	**1**	**2**	**4**	**5**	**3**	**15**	**27**	**34**	**38**	**148**	**24.46**	**0.87**
U.K	1	-	-	-	-	-	1	5	2	5	7	21	3.47	0.85
Norway	0	-	-	-	1	-	-	1	3	9	4	18	2.98	0.72
France	1	-	-	-	1	-	1	2	3	5	4	17	2.81	0.90
Switzerland	5	-	-	-	1	-	-	-	6	3	2	17	2.81	0.61
Germany	2	5	1	1	-	1	-	2	-	2	2	16	2.64	-0.24
Others	4	1	0	1	1	4	1	5	13	10	19	59	9.75	0.86
N. America	**84**	**6**	**13**	**12**	**13**	**9**	**16**	**16**	**38**	**33**	**58**	**298**	**49.26**	**0.84**
U.S	56	1	9	9	10	4	14	7	35	29	51	225	37.19	0.82
Canada	28	5	4	3	3	5	2	9	3	4	5	71	11.74	0.14
Mexico	0	-	-	-	-	-	-	-	-	-	2	2	0.33	0.52
Oceania	**4**	-	**5**	**2**	**5**	**10**	**2**	**8**	**26**	**16**	**20**	**98**	**16.20**	**0.80**
Australia	4	-	2	2	2	8	2	8	25	15	18	86	14.21	0.82
New Zealand	0	-	3	-	3	2	-	-	1	1	2	12	1.98	-0.03
S. America	**1**	-	-	-	-	**1**	-	**1**	-	**1**	-	**4**	**0.66**	**0.34**
Brazil	1	-	-	-	-	1	-	1	-	1	-	4	0.66	0.34
Total	110	12	20	17	22	26	23	48	94	95	138	605	100.00	0.89

TABLE 2

Journal distribution for papers in addiction and clinical research

Journal	1989-2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total	%	Growth
Addiction	21	1	3	2	3	8	3	2	6	4	7	60	8.90	0.57
Drug Alcohol Depend	2	-	1	-	-	-	-	2	1	1	4	11	1.63	0.67
PLoS One	0	-	-	-	-	-	2	2	2	3	1	10	1.48	0.76
Psychiatry Res	0	-	-	-	-	-	-	-	4	-	5	10	1.48	0.56
Nicotine Tob Res	3	-	-	-	-	-	1	-	1	-	4	9	1.34	0.61
Psychol Med	0	1	-	-	-	-	-	2	-	4	2	9	1.34	0.58
Addict Behav	4	-	1	-	-	-	1	1	-	1	-	8	1.19	0.14
Int J Drug Policy	0	-	-	-	-	-	-	-	4	3	1	8	1.19	0.62
J Subst Abuse Treat	4	-	-	2	-	-	-	-	-	1	-	7	1.04	-0.08
Alcohol Clin Exp Res	1	-	-	-	-	-	-	-	2	1	2	6	0.89	0.76
Others	87	11	16	13	21	19	18	40	68	73	91	457	67.80	0.90
Total	122	13	22	17	24	27	25	52	95	97	180	674	100.00	0.85

Figure 2) *Google Maps on eminent authors in addiction and clinical research (CC=0.90)*

Figure 3) *Google Maps on keywords in in addiction and clinical research (CC=0.96)*

First, the data were extracted from Medline. It is worth noting that any generalization should be made in the similar fields of paper contents.

Second, although the data were extracted from Medline and were carefully dealt with in every linkage as correctly as possible, the originally downloaded contexts including some errors in symbols which might affect the resulting reports in this study may be present.

Third, there are many algorithms used for SNA. We merely applied community cluster and density with weighted degrees in Figures. Any

TABLE 3

Association between SSP (Subjective Social Position) and frequent cannabis smoking.

Name	CC	E-I	node	degree	t-value
			Keyword		
*mobile phone problematic use	0.96	-1.00	8	28	8.40*
*schizophrenia	0.81	-0.73	27	106	6.91*
*ecological momentary assessment	0.81	-1.00	7	19	3.09*
*condom	0.75	-0.23	9	36	3.00*
*clinical research	0.74	-0.51	20	112	4.67*
*acetium(r) lozenge	0.67	-1.00	3	3	0.90
dopamine	0.64	-0.87	19	50	3.43*
reward deficiency syndrome (rds)	0.59	-0.92	28	82	3.73*
internet gaming disorder	0.58	-1.00	31	88	3.83*
Asia	0.46	-0.80	32	222	2.84*
			Author		
de Girolamo, Giovanni(Italy)	0.59	-1.00	77	184	6.33*
Brown, Andrew A(Norway)	0.58	-1.00	23	231	3.26*
Abe, Yoshinari(Japan)	0.90	-1.00	23	233	9.46*
Agid, Ofer(Japan)	0.80	-1.00	26	231	6.53*

Note. The higher cluster coefficient(CC) and the lower external-internal index (E-I)indicate closer and denser cluster in existence

changes made along with algorithm will present different pattern and inference making.

Fourth, the social network analysis is not subject to the Pajeck software we used in this study, Others such as Ucinet (28) and Gephi (29) are suggested to readers for use in the future study.

CONCLUSION

Social network analysis provides wide and deep insight into the relationships among nations and coauthors. The results can provide readers with a concept of diagram for future submission to a journal in addiction and clinical research.

Additional file 1:

How to obtain data from PubMed and to build the Google maps with social networks with an mp4 video at

http://www.healthup.org.tw/marketing/course/marketing/AdditionGooglemap.mp4

REFERENCES

1. Singer BF, Fadanelli M, Kawa AB, et al. Are cocaine-seeking "habits" necessary for the development of addiction-like behavior in rats? J Neurosci. 2017;20:2458-17.

2. American Society for Addiction Medicine. Definition of Addiction. 2017/12/21 available at https://www.asam.org/for-the-public/definition-of-addiction

3. Fonteneau ML, Scott R, Duka T. Attentional responses to stimuli associated with a reward can occur in the absence of knowledge of their predictive values. Behavioural Brain Res. 2018;341:26-36.

4. Domingos P. A few useful things to know about machine learning. Communications of the ACM 2012;55(10):78-87.

5. Verhoef PC, Kooge E, Walk N. Creating value with big data analytics: Making smarter marketing decisions. London: Routledge, UK. 2016.

6. Power DJ. What is the "true story" about data mining, beer and diapers? DSS News. 2017/3/20 available at https://goo.gl/1FtGft

7. Morris MA, Saboury B, Burkett B, et al. Reinventing Radiology: Big Data and the Future of Medical Imaging. J Thorac Imaging. 2018;33(1):4-16.

8. Sadoughi F, Valinejadi A, Shirazi MS, et al. Social network analysis of Iranian researchers on medical parasitology: A 41 year co-authorship survey. Iran J Parasitol. 2016;11(2):204-212.

9. Osareh F, Khademi R, Rostami MK, et al. Co-authorship network structure analysis of iranian researchers' scientific outputs from 1991 to 2013 based on the social science citation index (SSCI). Collnet J Scientometr Info Manag. 2014;8 (2): 263-271.

10. Liu X, Bollen J, Nelson ML, et al. Co-authorship networks in the digital library research community. Info Process Manag. 2005;41(6):1462-80.

11. Dasgupta S, Vaughan AS, Kramer MR, et al. Use of a google map tool embedded in an internet survey instrument: Is it a valid and reliable alternative to geocoded address data? JMIR Res Protoc. 2014;3(2):e24.

12. Kobayashi S, Fujioka T, Tanaka Y, et al. A geographical information system using the Google Map API for guidance to referral hospitals. J Med Syst. 2010;34(6):1157-60.

13. de Nooy W, Mrvar A, Batagelj V. Exploratory social network analysis with Pajek: revised and expanded, (2nd ed). New York, NY: Cambridge University Press, USA. 2011.

14. Holland PW, Leinhardt S. Transitivity in structural models of small groups. Comparative Group Studies 1971;2:107-124.

15. Phan TG, Beare R, Chen J, et al. Googling service boundaries for endovascular clot retrieval hub hospitals in a metropolitan setting: Proof-of-concept study. Stroke. 2017;48(5):1353-1361

16. Krackhardt D, Stern RN. Informal networks and organizational crises:An experimental simulation. Social Psychology Quarterly 1988;51:123-140.

17. Chien TW. Google Maps on the topic of author collaboration in addiction and clinical research. 2017/12/22 available at http://www.healthup.org.tw/gps/addiction.htm

18. Chien TW. Google Maps on eminent authors in addiction and clinical research. 2017/12/22 available at http://www.healthup.org.tw/gps/addiction author.htm

19. Chien TW. Google Maps on MESH terms in addiction and clinical research. 2017/12/22 available at http://www.healthup.org.tw/gps/addiction key.htm

20. Leydesdorff L, Wagner C, Park HW, et al. International collaboration in science: the global map and the network.CoRR abs/1301.0801 (2013)

21. Glänzel W, Schlemmer B. National research profiles in a changing Europe (1983-2003) An exploratory study of sectoral characteristics in the Triple Helix. Scientometrics 2007;70(2):267-275.

22. Erfanmanesh MA, Rohani VA, Basirian Jahromi R, et al. Investigating scientific collaboration of iranian psychology and psychiatry researchers. J Info Process Manag. 2013;29(1):137-163.

23. Osareh F, Norouzi Chakoli A, et al. Co-authorship of Iranian researchers in science, social science, art and humanities citation indexes in the web of science between 2000 and 2006. J Info Sci Technol. 2010;25(4):573-95.

24. Hu SK, Huang J, Hong WD, et al. The 50 most-cited articles in gastroenterology and hepatology from Mainland China. Pak J Med Sci. 2017;33(1):215-220.

25. Coelho DH, Edelmayer LW, Fenton JE. A century of citation classics in otolaryngology-head and neck surgery journals revisited. Laryngoscope. 2014;124(6):1358–1362.

26. Watts DJ, Srogatz S. Collective dynamics of 'small-world' networks". Nature 1998;6684:440–442.

27. Chang HT, Lin MH, Hwang IH, et al. Scientific publications in gastroenterology and hepatology in Taiwan: An analysis of Web of Science from 1993 to 2013. J Chin Med Assoc. 2017;80(2):80-85.

28. Borgatti SP, Everett MG, Freeman LC. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies 2002.

29. Bastian M, Heymann S, Jacomy M. An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. 2009.