

Using Visual Representations to Present the Pattern of International Co-Author Collaboration in the Field of Molecular and Genetic Research

Chien TW^{1*}, Chang Y², Chow JC³ and Chou W^{4,5*}

¹Department of Research, Chi-Mei Medical Center, Tainan, Taiwan

²National Taiwan University, School of Medicine, Taiwan

³Department of Paediatrics, Chi Mei Medical Center, Tainan, Taiwan

⁴Department of Physical Medicine and Rehabilitation, Chi Mei Medical Center, Tainan, Taiwan

⁵Department of Recreation and Health-Care Management and Institute of Recreation Industry Management, Chia Nan University of Pharmacy, Tainan, Taiwan

Abstract

Objective: The pattern of international co-author collaboration in molecular and genetic research remains unclear. We collected data from Medline and report the results with graphical presentations using Google maps and social network analysis (SNA).

Methods: Downloading 6,732 abstracts on December 13, 2017 from the Medline library with keywords of Molecular (Title) AND Genetic (Title), we reported following features: (1) nation and journal distribution; (2) main keywords frequently presented in papers; (3) the eminent author and key indicators in SNA. We programmed Microsoft Excel VBA to organize data. Google Maps and SNA Pajek were used for displaying results in molecular and genetic research.

Results: We found that (1) the most number of nations are from U.S. (1622,31.88%), China (361, 7.10%), and Japan (356, 7.00%); (2) the most number of journals is Genetika (103, 1.53%); (3) two clusters of RT-PCR and genetic association earn the highest cluster coefficient; (4) the eminent with the highest cluster coefficient is J Barhanin from Italy.

Conclusion: Social network analysis provides wide and deep insight with the relationships among entities of interest. The results drawn by Google maps can be offered to readers for future submission to journals.

Keywords: Authorship collaboration; Social network analysis; Google maps; Medline

Introduction

Many papers have been saved in Medline library. International collaboration in science has increased rapidly in recent decades [1]. The dominant nations in Science come from U.S. and Europe [2,3]. Whether the field of molecular and genetic research is also dominated by U.S. and Europe remains unclear.

The co-author relation is similar to the comorbid co-occurred with one another in medicine. Molecular and genetic terms are often combined together in literature. Many studies have made efforts to explore the association of two or more relevant entities such as molecular epidemiological analysis used for revealing a high level of genetic diversity [4] and molecular and genetic determinants together existed in glioma cell invasion [5]. The pattern of international co-author collaboration in molecular and genetic research is worthwhile to study. It is rather hard using traditional statistics to observe the association of two or more symptoms co-occurred at one moment till social network analysis (SNA) [6] has been launched to explore the pattern of elements in a system.

Google maps offers a global view of geospatial visualization for our interesting objects dispersed on a map [7,8]. However, only four papers were found in Medline library using keyword google map (Title) on December 12, 2017. Many papers [9-11] have researched on co-author collaboration in academics, but failed to incorporate SNA with Google maps in their studies. We are thus interested in investigating whether it is possible to show the pattern of author collaboration in molecular and genetic research using Google maps.

Our aims are to investigate patterns of international co-author collaborations in molecular and genetic research by collecting data from Medline and to visualize results with following representations:

(1) nation and journal distribution; (2) main keywords frequently presented in papers; (3) the eminent author in the field of molecular and genetic research.

Methods

Data sources

We programed Microsoft Excel VBA (visual basic for applications) modules for extracting abstracts and their corresponding coauthor names as well as keywords on December 12, 2017 from Medline library. Only those abstracts entitled with molecular and genetic topics and labelled with Journal Article were included. Others like those labelled with Published Erratum, Editorial or those without author nation were excluded from this study. A total of 6,732 eligible abstracts were obtained from Medline since 1983. Only 5,088 papers are labeled with 1st author nation in Medline.

Data arrangement to fit SNA requirement

We analyzed all eligible papers with complete data consisting of author countries and journal names. Prior to visualize representations

***Corresponding author:** Dr. Willy Chou, Department of Physical Medicine and Rehabilitation, Chi Mei Medical Center, Tainan, Taiwan, Tel: 90 474 225 11 67; E-mail: ufan0101@ms22.hinet.net

Received December 20, 2017; **Accepted** January 18, 2018; **Published** January 23, 2018

Citation: Chien TW, Chang Y, Chow JC, Chou W (2018) Using Visual Representations to Present the Pattern of International Co-Author Collaboration in the Field of Molecular and Genetic Research. J Mol Genet Med 12: 319 doi:10.4172/1747-0862.1000319

Copyright: © 2018 Chien TW, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

using SNA, we organized data in compliance with the SNA format and guidelines of Pajek software [12]. Microsoft Excel VBA was used to deal with data fitting to the SNA requirement.

Graphical Representations to Report

Author nations and their relations

Two cross tables (i.e. columns for publication years and rows for the 1st author nations as well as journals) were generated for showing the distribution of nations and the most number of journals publishing papers of molecular and genetic research. The bigger bubble means the more number of the nodes (i.e., nations, or authors). The wider line indicates the stronger relations between two nodes. Community clusters are filled with different colors in bubbles.

Keywords and authors to present the feature of molecular and genetic research

Keywords in abstract are defined by authors. Research domain can be highlighted by the relation between any pair of two keywords using SNA. The representation for the bubble and line is interpreted similar to the previous section.

Statistical Tools and Data Analyses

Google Maps [13] and SNA Pajek software [12] were used to display visualized representations for papers published in the field of molecular and genetic research. Author-made Excel VBA modules were used to organize research data.

Cluster coefficient represents the density of a network as below=

$$\frac{\text{number of closed triplets}}{\text{number of possible connected triplets of nodes}} = \frac{\sum_{i=1}^n \text{other_two_nodes_connected}}{\sum_{i=1}^n (m_i * (m_i - 1) / 2)}, \text{ whereas}$$

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 of $\frac{EL - IL}{EL + IL}$, where EL=hte number of external friendship links and IL= the number of internal friendship links [14]. 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.

Nation	1983-2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total	%	Growth
Africa	7	3	3	1	2	5	4	2	3	3	5	38	0.75	0.39
Tunisia	2	-	1	1	1	2	-	-	1	2	-	10	0.20	0.09
Nigeria	1	-	-	-	-	2	3	-	-	-	-	6	0.12	-
Egypt	1	-	-	-	1	-	-	-	-	-	3	5	0.10	-
Kenya	1	1	-	-	-	-	-	2	-	-	1	5	0.10	0.46
Others	2	2	2	0	0	1	1	0	2	1	1	12	0.24	0.37
Asia	440	38	43	63	71	75	78	65	85	94	91	1143	22.46	0.91
China	90	16	13	19	28	25	32	33	37	31	37	361	7.10	0.17
Japan	222	7	12	21	19	17	9	7	10	18	14	356	7.00	0.06
India	28	4	7	8	11	17	15	10	12	17	11	140	2.75	0.91
Taiwan	24	1	3	7	1	6	4	3	2	3	5	59	1.16	-0.03
Others	76	10	8	8	12	10	18	12	24	25	24	227	4.46	5.40
EUROPE	1123	61	66	78	69	73	64	64	63	97	82	1844	36.24	0.50
U.K	232	10	10	9	13	2	6	5	10	16	7	320	6.29	-0.17
Germany	213	10	9	12	7	11	6	8	9	7	9	301	5.92	0.14
France	130	7	6	7	7	6	4	10	4	10	6	197	3.87	0.41
Italy	96	2	5	11	8	14	8	11	5	7	11	180	3.54	-0.42
Netherlands	79	2	7	3	5	6	8	5	5	8	4	132	2.59	-0.11
Others	373	30	29	36	29	34	32	25	30	49	45	714	14.03	1.98
N. AMERICA	1245	54	63	41	53	61	56	62	59	51	46	1792	35.22	-0.11
U.S	1135	51	59	38	46	54	49	52	52	46	40	1622	31.88	0.32
Canada	97	3	3	3	5	6	6	5	6	2	5	142	2.79	0.52
Mexico	7	-	1	-	2	-	1	4	1	3	-	19	0.37	-
Cuba	3	-	-	-	-	-	-	-	-	-	-	3	0.06	-
Others	3	-	-	-	-	1	-	1	-	-	1	6	0.12	0.03
OCEANIA	75	3	5	6	8	3	2	7	7	7	5	128	2.52	0.29
Australia	64	3	4	6	7	2	2	5	5	6	4	108	2.12	0.13
New Zealand	11	-	-	-	1	1	-	-	2	-	-	15	0.29	0.17
Palestine	0	-	1	-	-	-	-	1	-	1	1	4	0.08	0.16
Myanmar	0	-	-	-	-	-	-	1	-	-	-	1	0.02	0.43
S. AMERICA	49	6	4	12	8	6	10	11	15	14	8	143	2.81	0.59
Brazil	20	6	2	4	3	4	8	7	9	9	7	79	1.55	0.17
Argentina	8	-	2	3	1	-	2	2	2	2	1	23	0.45	0.72
Chile	11	-	-	1	2	-	-	1	4	1	-	20	0.39	0.27
Colombia	5	-	-	3	-	1	-	-	-	-	-	9	0.18	-0.30
Others	5	-	-	1	2	1	-	1	-	2	-	12	0.24	0.32
Total	2939	165	184	201	211	223	214	211	232	266	237	5088	100.00	0.89

Table 1: Nation distribution based on the 1st author in molecular and genetic research.

Results

Author nations and their relations

A total of 5,088 eligible papers with complete author nations since 1983 are shown in Table 1. We can see that the most number of nations are from U.S. (1622,31.88%), China (361, 7.10%), and Japan (356, 7.00%). The trend in the number of publications for countries is present in the column of growth in Table 1. We can see Asia with the highest growth rate (0.91) The diagram (shown by SNA and Google maps) in Figure 1 displays author collaboration among nations. The highest productive nations are from Europe (36.24%) and North America (35.22). Any nation

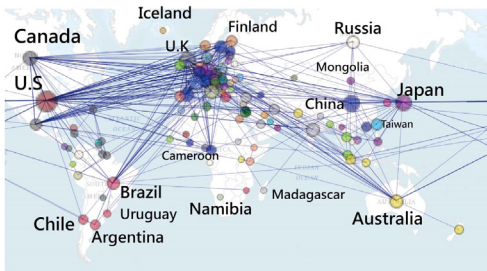


Figure 1: International coauthor collaboration in molecular and genetic research.

collaborated with other nations are shown with a blue line. Interested authors are recommend clicking the bubble of interest to see details on a website at reference (Table 1 and Figure 1) [15].

Journals and the trend

A total of 6,732 eligible abstracts were analyzed regarding title with either molecular or genetic keyword. The most number of journals in production is Genetika (103, 1.53%). The trend for a journal is shown in the column of correlation in Table 2. BoTh journals of PLoS One (0.84) and Genet Mol Res (0.85) earn the highest growth in past years. We can see other journals are increasing or decreasing in papers regarding molecular or genetic research (Table 2).

Keywords to present the feature of research domain

Two clusters of RT-PCR and genetic association earn the highest cluster coefficient (Table 3) [16]. We can see that the two bigger bubbles are of RT-PCR and genetic association in respective clusters (Table 3).

Eminent authors selected by SNA

The eminent with the highest cluster coefficient is J Barhanin form Italy shown in Table 3 (bottom) or click it on the reference [17]. We can see that the top 10 with a higher cluster coefficient are present in (Figure 2).

Journal	1983-2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total	%	Growth
Genetika	54	9	6	3	8	5	7	3	5	1	-	103	1.53	-0.74
Genetics	89	-	1	3	-	-	-	-	-	-	1	94	1.40	-0.25
PLoS One	0	1	2	7	5	9	13	15	14	12	11	89	1.32	0.84
Theor Appl Genet	48	1	3	2	3	2	1	3	2	3	4	72	1.07	0.49
J Bacteriol	56	-	-	-	-	1	-	-	-	-	-	64	0.95	-0.06
Proc Natl Acad Sci U S A	35	2	1	1	3	2	-	1	-	3	-	60	0.89	-0.27
Genet Mol Res	1	-	2	4	3	6	3	5	10	9	7	50	0.74	0.85
Am J Surg Pathol	22	7	5	1	2	1	1	3	2	2	2	48	0.71	-0.54
Infect Genet Evol	3	-	2	3	4	7	7	-	5	9	3	43	0.64	0.45
Mol Ecol	26	2	5	2	1	1	1	-	-	-	1	39	0.58	-0.70
Others	334	22	27	26	29	34	33	30	38	39	29	662	9.83	0.72
Total	3933	226	222	254	264	281	264	301	290	305	265	6732	100.00	0.78

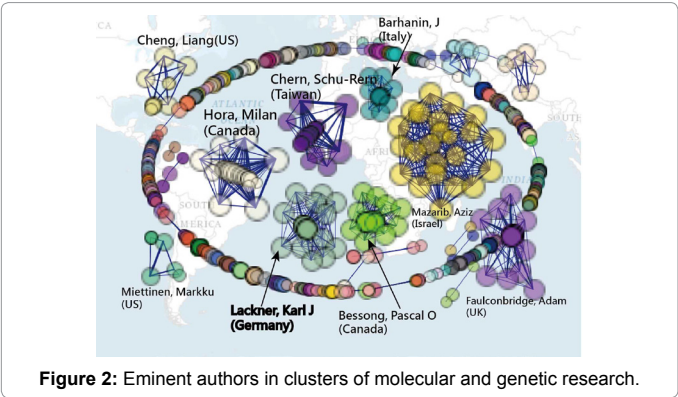
Table 2: Journal distribution for papers in molecular and genetic research.

Name	CC	Density	Weighted	EI	Count of Nodes	degree	Weighted	t-value
RT-PCR	0.89	0.02	0.04	-1.00	60	44	68	14.87
Genetic association	0.88	0.04	0.05	-0.91	23	11	12	8.49
Microsatellites	0.83	0.09	0.09	-0.71	27	30	30	7.44
Fish	0.81	0.01	0.01	-0.57	49	9	12	9.47
Genetics	0.80	0.15	0.30	-1.00	37	97	201	7.89
Genetic analysis	0.76	0.03	0.03	-0.70	44	26	27	7.58
Genotyping	0.71	0.05	0.05	-0.62	60	83	83	7.68
Mutations	0.70	0.05	0.05	-0.99	58	83	87	7.34
Mucopolysaccharidosis type iva	0.68	0.03	0.04	-0.61	48	33	43	6.29
Mutation	0.65	0.05	0.06	-0.86	64	101	113	6.73
Barhanin, J (Italy)	0.91	1.00	3	-1	12	66	198	6.94
Cao, Shinuo (Japan)	0.78	1.00	3	-1	9	36	108	3.3
Aghokeng, Avelin F (Estonia)	0.57	1.00	3	-1	21	210	630	3.02
Blankenberg, Stefan (Netherlands)	0.49	1.00	3	-1	22	231	693	2.51
Ivanova-Smolenskaia, I A (UK)	0.48	0.53	2.24	-1	10	24	101	1.55
Blatt, Ilan (Israel)	0.47	0.93	3.77	-1	23	236	955	2.44

Najera, Rafael (Cuba)	0.46	1.00	3.02	-1	10	45	136	1.47
Chen, Chih-Ping (Taiwan)	0.34	0.45	2.18	-1	18	69	333	1.45
Cheng, Liang (US)	0.33	0.51	2.45	-1	11	28	135	1.05
Michal, Michal (Czech Republic)	0.32	0.39	1.66	-1	28	149	626	1.72

CC: Cluster Coefficient; Density Centrality: The Number of Connected Relation/All Possible Connected Relation ($=n(n-1)/2$, where n: the Number of Nodes (or Entities); Weighted Centrality includes Repeated Connection Relations; Count of Nodes: the Number of Members in The Network; Degree: the Number of Connected Relation; Weighted: all Repeated Connection Relations; t-value: Student t Statistics.

Table 3: Indicators of SNA for the keyword clusters in a global network.



Discussion

This study used SNA techniques to report that (1) the most number of nations are from U.S. (1622,31.88%), China (361, 7.10%), and Japan (356, 7.00%);(2) the most number of journals is Genetika (103, 1.53%); (3) two clusters of RT-PCR and genetic association earn the highest cluster coefficient; (4) the eminent with the highest cluster coefficient is J Barhanin from Italy.

What this adds to what was known

An apocryphal story is often told to discover the co-occurrence about beer and diaper sales [18-20]. It is hard to see all possible pairs of our observed entities at one short moment. In literature, no such examples but studies [9,10] were illustrated to inspect co-author collaboration using SNA. We demonstrated SNA incorporated with Google maps to display valuable information to readers, which is rare seen in previous papers.

Clusters can be compared with each other using Google maps. We can see that many links connecting nations, indicating a collaboration pattern to the previous study [11]. The results in this study show a huge international co-author collaboration in molecular and genetic research which is consistent with the previous studies that investigated scientific collaboration of Iranian Psychology and Psychiatry Researchers [21,22].

Two papers [23,24] incorporated MeSH (Medical subject heading) with social network analysis to explore knowledge in journal topics. However, no any incorporated SNA with Google maps to show research results like we did in the current study. The way we illustrated here in Figures is novel and promising in academics, especially in the field of molecular and genetic research.

What it implies and what should be changed?

Scientific publication is one of the objective measurements to evaluate the achievements of a medical research [25]. Using SNA and Google Maps is appropriate to report journal features or author research domains in future. Several algorithms have been developed

in computer science and have applied SNA to researches. If we further investigate whether author domains or paper keywords are most fitting the scope of a journal, the centrality measures [9] is recommend to readers. It means that the core research domain can be analyzed using the centrality measure [11,23] produced in social network analysis.

Strengths of this study

The way we used with SNA and Google Maps is unique, which is rare seen in previous papers. Another strength (or feature) is regarding Google Maps provided to interested readers who can practice it on their own ways by clicking the links in references [15-17]. The nation distribution in Figure 1 is easy to know the feature of molecular and genetic research. One picture is worth ten thousand words. We expect following studies that can report more information using SNA and Google Maps to readers.

Limitations and future study

The interpretation and generalization this study should be cautious. First, the data were downloaded from Pubmed. Any attempt to generalize the findings should be subject to the similar background or the journal with similar topic and scope.

Second, data were extracted from Pubmed. We also put a lot of efforts on every linkage, the original downloaded data including some errors in symbols such as period, comma or others in author address that might result in some bias.

Third, there are many computer algorithms in social network analysis. We only applied on way to show data. Any changes made in algorithm used for exploration will display different layout of pattern.

Fourth, the social network analysis is not limited in Pajeck software we used in this study, Others such as Ucinet [26] and Gephi [27] are suggested to readers for use in future.

Conclusion

Social network analysis provides wide and deep insight with the relationships among entities of interest. The results drawn by Google maps can be offered to readers for future submission to journals.

References

1. National Science Board. Science and Engineering Indicators. Washington DC: National Science Foundation. 2017/11/3 available at <http://www.nsf.gov/statistics/seind12/>.
2. Leydesdorff L, Wagner C, Park HW, Adams J (2013) International collaboration in science: The global map and the network. El Profesional de la Informacion 22: 87-95.
3. Glänzel W, Schlemmer B (2007) National research profiles in a changing Europe (1983–2003) An exploratory study of sectoral characteristics in the Triple Helix. Scientometrics 70: 267-275.
4. Guo W, Jelocnik M, Li J, Sachse K, Polkinghorne A, et al. (2017) From genomes to genotypes: Molecular epidemiological analysis of *Chlamydia gallinacea* reveals a high level of genetic diversity for this newly emerging chlamydial pathogen. BMC Genomics 18: 949.

5. Masui K, Kato Y, Sawada T, Mischel PS, Shibata N (2017) Molecular and genetic determinants of glioma cell invasion. *Int J Mol Sci* 18: 1-2.
6. Bright CF, Haynes EE, Patterson D, Pisu M (2017) The value of social network analysis for evaluating academic-community partnerships and collaborations for social determinants of health research. *Ethn Dis* 27: 337-346.
7. Dasgupta S, Vaughan AS, Kramer MR, Sanchez TH, Sullivan PS (2014) 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* 3: e24.
8. Kobayashi S, Fujioka T, Tanaka Y, Inoue M, Niho Y, et al. (2010) A geographical information system using the Google Map API for guidance to referral hospitals. *J Med Syst* 34: 1157-1160.
9. Sadoughi F, Valinejadi A, Shirazi MS, Khademi R (2016) Social network analysis of Iranian researchers on medical parasitology: A 41-year co-authorship survey. *Iran J Parasitol* 11: 204-212.
10. Osareh F, Khademi R, Rostami MK, Shirazi MS (2014) 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* 8: 263-271.
11. Liu X, Bollen J, Nelson ML, Van de Sompel H (2005) Co-authorship networks in the digital library research community. *Info Process Manag* 41: 1462-1480.
12. de Nooy W, Mrvar A, Batagelj V (2011) *Exploratory social network analysis with pajek: Revised and expanded*, (2nd edn), Cambridge University Press, New York, USA.
13. Phan TG, Beare R, Chen J, Clissold B, Ly J, et al. (2017) Googling service boundaries for endovascular clot retrieval hub hospitals in a metropolitan setting: Proof-of-concept study. *Stroke* 48: 1353-1361.
14. Krackhardt D, Stern RN (1988) Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly* 51: 123-140.
15. <http://www.healthup.org.tw/gps/Molecular.html>
16. <http://www.healthup.org.tw/gps/Molecularkey.html>
17. <http://www.healthup.org.tw/gps/Molecularauthor.html>
18. Domingos P (2012) A few useful things to know about machine learning. *Communications of the ACM* 55: 78-87.
19. Verhoef PC, Kooge E, Walk N (2016) *Creating value with big data analytics: Making smarter marketing decisions*, Routledge, London.
20. <https://goo.gl/1FtGft>
21. Erfanmanesh MA, Rohani VA, Jahromi BR, Gholamhosseinzadeh Z (2013) Investigating Scientific collaboration of Iranian psychology and psychiatry researchers. *J Info Process Manag* 29: 137-163.
22. Osareh F, Chakoli NA, Keshvari M (2010) 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* 25: 573-595.
23. Hu SK, Huang J, Hong WD, Du XJ, Jin R, et al. (2017) The 50 Most-cited articles in gastroenterology and hepatology from Mainland China. *Pak J Med Sci* 33: 215-220.
24. Coelho DH, Edelmayer LW, Fenton JE (2014) A century of citation classics in otolaryngology-head and neck surgery journals revisited. *Laryngoscope* 124: 1358-1362.
25. Chang HT, Lin MH, Hwang IH, Chen TJ, Lin HC, et al. (2017) Scientific publications in gastroenterology and hepatology in Taiwan: An analysis of Web of Science from 1993 to 2013. *J Chin Med Assoc* 80: 80-85.
26. Borgatti SP, Everett MG, Freeman LC (2002) *Ucinet for Windows: Software for Social Network Analysis*. MA: Analytic Technologies, Harvard, USA.
27. Bastian M, Heymann S, Jacomy M (2009) Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media* 2009.