Systematic Review and Meta-Analysis

Medicine®

OPEN

# Applying Gini coefficient to evaluate the author research domains associated with the ordering of author names

## A bibliometric study

Tsair-Wei Chien, MBA[a,*], Julie Chi Chow, MD[b], Yu Chang, MD[c], Willy Chou, MD, MBA[d,e,*]

**Abstract**

**Background:** Team science research includes the number of coauthors in publications. Many papers have discussed the ordering of author names and the contributions of authors to a paper. However, no paper addresses the relation between authors' research domains and personal impact factors (PIF) with the ordering of author names. We aimed to apply Gini coefficient (GC) to evaluate the author research domains associated with the PIF and the ordering of author names on academic papers.

**Methods:** By searching the PubMed database (Pubmed.com), we used the keyword "medicine" [journal] and downloaded 10,854 articles published from 1969 to 2018. A total number of 7502 articles labeled with complete author's countries/areas were included in data analysis. We also proposed a PIF index and jointly applied social network analysis (SNA), the GC, and Google Maps to report the following data with visual representations: the trend of author collaboration in *Medicine*; the dominant nations and keywords in *Medicine*; and the author research domains in *Medicine* associated with the PIF and the ordering of author names on academic papers.

**Results:** The trend of author collaboration in *Medicine* is slightly declining (= −0.06) based on the number of authors per article. The mean number of individuals listed as authors in articles is 7.5. Most first authors are from China (3649, 48.64%) and Taiwan (847, 11.29%). The median of GC (0.32) and PIF (0.74) for the middle authors are obviously less than those for the first (0.53, 2.19) and the last authors (0.42, 2.61). A perfect positive linear relation with a large effect exists between GC and PIF because the correlation coefficient is 0.68 (>0.50, $t$=2.48, n=9).

*Editor: Minh Van Hoang.*

*All data were downloaded from MEDLINE database at pbumed.com.*

*All data used in this study is available in Additional files.*

*The authors have no conflicts of interest to disclose.*

*Additional files*

*Additional file 1:*

*MP4: the process of the study to build Google Maps.*

*http://www.healthup.org.tw/marketing/course/marketing/mapingGooglemapformedicine.mp4*

*Additional file 2: Pajek control file and dataset*

*Pajek_aa.net*

*Additional file 3:*

*MP4: The calculation of person impact factor for the author of the Journal of Medicine at https://youtu.be/CJJ-uV8fYls*

*Additional file 4:*

*MP4: Identifying the unique author name.*

*http://www.healthup.org.tw/marketing/course/marketing/uniqueauthordetect.mp4*

*Additional file 5:*

*PDF: How to screen duplicate authors in a network*

*Additional 6*

*The calculation of Gini coefficient in a MS Excel format at www.healthup.org.tw/ GiniA_2015.zip*

*[a] Medical Research Department, [b] Department of Paediatrics, Chi Mei Medical Center, [c] National Taiwan University School of Medicine, [d] Department of Sports Management, College of Leisure and Recreation Management, Chia Nan University of Pharmacy and Science, [e] Ncphrology Department, Chi-Mei Medical Center, Tainan, Taiwan.*

*[*] Correspondence: Willy Chou, Chi-Mei Medical Center, 901 Chung Hwa Road, Yung Kung District, Tainan 710, Taiwan (e-mail: ufan0101@ms22.hinet.net); Tsair-Wei Chien, Medical Research Department, Chi-Mei Medical Center, Tainan, Taiwan (e-mail: smile@mail.chimei.org.tw).*



*Medicine (2018) 97:39(e12418)*

*Received: 23 June 2018 / Accepted: 22 August 2018*

*http://dx.doi.org/10.1097/MD.0000000000012418*

**Conclusion:** Results suggest that the corresponding author is submitting the manuscript to the target journal with a core author's academic background and the personal impact factor related to the research domain and the journal scope in the future. As such, peer reviewers can quickly determine whether the manuscript is a potentially citable research paper.

**Abbreviations:** AD = author domain, CA = core author, CI = confidence interval, CV = curriculum vitae, GC = Gini coefficient, HTML = Hyper Text Mark-up Language, JCR = journal citation report, MESH = medical subject heading, PIF = personal impact factor, SNA = social network analysis, VBA = visual basic for application.

**Keywords:** authorship collaboration, Gini coefficient, Google Maps, personal impact factor, PubMed library, social network analysis

## Key Points

- Google Maps is used to show geographical presentations. This approach is rarely seen in previous papers.
- The personal research domain with a GC and the personal IF are displayed on Google Maps to interpret the author academic scope.
- The way to examine duplicate authors with identical names in the library database was illustrated in this study using social network analysis such as Betweenness Centrality algorithm, which was never mentioned in bibliometric analyses of published papers.

## 1. Introduction

The number of research papers with multiple authors[1] has increased considerably because knowledge discovery no longer occurs merely in local university departments but now occurs across international academic institutes.[2] Accordingly, given increasing academic pressures and the spotlight on individuals with prolific publications, many researchers seek to claim authorship of a paper.[1] Articles were cited more frequently only when the first authors (i.e., primary authors) had higher prestige (i.e., h-indice) and studies were funded,[3] However, rarely do all coauthors contribute to a paper equally[4] and the corresponding authors (i.e., supervisory authors usually responsible for the funding and placed at the last) also contributed a substantial credit to the respective article[5] Whether the ordering of author names is related to the author prestige based on the in-depth research domain[6] remains unclear.

Author placement in the article byline has significant implications for accountability and allocation of credit.[7] Many papers[8–10] have assessed authorship order and the type of contribution to the article. However, a good research paper is achieved through honorable authorship, vigilant editors, robust peer review, and a discerning readership.[7] Whether the core author contributes the most appropriate role to the manuscript should be determined to enable readers or reviewers to quickly see whether the paper exemplifies proper research. The reason is that the editor usually requires authors to provide evidence of statistical (or research) consultation (or at least expertise) by including a statistician/epidemiologist either among the authors or in the acknowledgments. A biostatistician (or, say, core author) may review such manuscripts during the review process.

### 1.1. Research domain

Authors with the depth of research domain have enough prestige to publish papers.[6] An author domain (AD) is similar to a curriculum vitae, which provides a summary of one's experience, skills, and academic background, but it does not include teaching experience, degrees, research, awards, presentations, and other achievements aside from publications. The AD merely focuses on the academic paper features mainly by providing a visual representation to draw the keywords of the PubMed medical subject heading (MESH) terms along with the Gini coefficient (GC) to display the depth or breadth in academic specialization.[6,11] Many bibliometric studies[12–17] have used coword (or coauthor) analysis to visualize their study specialization. However, no studies display their outline in a way that can be further examined by a zoom-in and zoom-out functionality like on Google Maps[4,18,19] or illustrated by a technique to screen out the duplicated authors with identical names in academic databases.

### 1.2. The target journal and a topic we concern

We concerned the journal Impact Factor (IF) based on journal citation report (JCR) annually released by Thomson Reuters and found that the journal of *Medicine* (Baltimore) in 2015 suffered a rapid decrease (=4.75=5.723/1.206) in JCR IF. The topic of whether the ordering of author names is related to the author domain and can predict the journal IF through the personal impact factors (PIF) is needed to investigate.

The person who contributed the most significant share of the actual research is the first author. The one who provided strategic thinking throughout the project is the last author. All others are the middle authors who performed more hands-on work (earlier in the list) or more advisory roles (later). The former 2 usually function as the core authors.[5] Whether research domains are different among the 3 (primary, middle, and supervisory) authors still remains unclear.

### 1.3. The objectives of this study

The objectives of this study are to report the trend of author collaboration in *Medicine*, the dominant nations and keywords in *Medicine*; and the author research domains in *Medicine* associated with the PIF as well as the ordering of author names on an article.

## 2. Methods

### 2.1. Data source

By searching the PubMed database (Pubmed.com) maintained by the US National Library of Medicine, we used the keyword "medicine" [Journal] on March 7, 2018, sorted the results according to the most recent works, and downloaded 10,854 articles that were published from 1969 to 2018. An author-made Microsoft Excel Visual Basic for Application (VBA) module was used to analyze and present the research results. All downloaded abstracts (n = 10,854) meeting the requirement for the type of journal article; see Additional File 1 and Fig. 1. Ethical approval was not necessary for this study because all the data come from the PubMed library, which is available on the Internet.

### 2.2. Social network analysis and Pajek software

Social network analysis (SNA)[20] was developed to explore the pattern of entities in a system. Pajek[21] is one of the most popular SNA software in the literature;[6] see Additional File 2. In keeping with the guidelines of Pajek, we defined an author (or paper keyword of the MESH term) as a node (or an actor) connected to other nodes through the edge (or, say, the relation). Usually, the weight between 2 nodes is the number of connections.

Centrality is a vital index for analyzing the network. Any individual or keyword that lies in the center of the social network will determine its influence on the system and its speed to gain information.[22] We applied the betweenness centrality to emphasize the pivotal role (i.e., bridge) in the network and illustrated Additional File 1 to describe the details of the graphical process by using SNA and Google Maps.

### 2.3. Reporting the research results

**2.3.1. Trend of author collaboration in Medicine.** We selected *Medicine* as the target journal. Two cross tables (i.e., columns for publication years and rows for the first author countries/areas and the most productive authors) show the distribution of nations and authors in *Medicine* across years. The top 3 authors who published the most number of papers in *Medicine* are in the 3 study groups: the first, the middle, and the last authors. The statistics of mean, median, minimum, and a maximum number of authors per article in *Medicine* across years were calculated by using descriptive statistics.

The results are shown on Google Maps. The large bubble denotes the number of published papers for certain countries/areas (authors). A wide line corresponds to a strong relation between the 2 entities (i.e., the nation or the author). Clusters separated by the algorithm of the partitioned communities are denoted by bubbles in different colors.

**2.3.2. Dominant nations and Mesh terms in Medicine.** According to the table about the coauthors and their collaboration nations, we can plot the distribution of the article counts across countries on Google Maps and examine the trend of the papers in *Medicine* for each continent and county/area. The top 10 Mesh terms over the years for the Journal of *Medicine* were also displayed in a table.

**2.3.3. Difference in patterns of research domain across the author order.** The MESH terms with an asterisk were extracted from each paper by using the MS Excel VBA function of split (MESH,"/"). SNA cluster analysis using Pajek was performed to obtain the maximal betweenness centrality (i.e., the most number of connections with others) from each MESH term in a cluster. We applied the top 5 maximal degree centralities to calculate the GC in the network because the GC measures the inequality for the data elements, such as the frequency distribution (for example, levels of income for a nation) of the 5 sequential data for a nation's income. The threshold is set at 0.4[23,24] to differentiate the income inequality for a specific country or area.

We define the research domain by the GC formula $q/(q-1) \times \sum_{i=1}^{n} |D_i|/2 \times n \div \sum_{i=1}^{n} X_i$, where D is the absolute difference of each pair data element, X is the data element value, n is the number of clusters, and q/(q − 1) is the adjustment for the data element number to reach 1.0 (i.e., an extreme inequality) because the GC is dependent on the number of calculated frequencies (or bins).

The process includes selecting the maximal number of connections for a node (or an entity) of each cluster; selecting the top 5 of the data elements in clusters obtained by performing first step; and using the abovementioned Gini formula to determine the depth (≥0.4) or the breadth (<0.4) of a specific author research domain. Interested readers are suggested to click the hyperlink for calculating the Gini coefficient on their own at the reference.[25]

**2.3.4. Personal research impact factors (PIF) based on PubMed.** The h-index[26] is an author-level metric that measures both the productivity and citation impact of the publications of a scientist or scholar. Comparing researchers' achievements using h-index is problematic like many indices due to citation patterns substantially different among scientific disciplines.[27] Other shortcomings include a gender effect,[28] age and career factors,[29] and the assumption of equal credits across all coauthors in an article byline.[4] We thus mimicked the JCR IF and calculated PIF [30] with additional co-author weights (i.e., contributions or credits) using the formula $(= \frac{\sum Cited.papers.based.on.SCI.IF_i \times W_j.in.the.recent.3.years}{Citable.papers \times W_j.in.the.given.2.years})$, whereas $W_j \cdot (= \frac{\exp(\gamma_j)}{\sum_{j=0}^{m}\exp(\gamma_j)} = \frac{2.72^\wedge\gamma_j}{\sum_{j=0}^{m} 2.72^\wedge\gamma_j}$ denotes the weights based on the order j in the article byline, $\gamma_j$ is the author contribution with an integer number from m to 0 in descending order, and m+1 = author number. Accordingly, the summation of all authors' weights per paper equals 1.0. Primary authors gain 2.72 times of credit to supervisory authors and another 2.72 times to the third, etc. If $\gamma_j$ is less than 1.0, the credits for each author will be inversed in accending order. If all $\gamma_j$ are monotonously increasing with a mean of zero, the middle authors will gain the most part of weights.

The credit formula can be further extended to a general model $(= \; = \frac{Base_k{}^\wedge\gamma_j}{\sum_{j=0}^{m} Base_k{}^\wedge\gamma_j})$ used in other situations. For instance, all $\gamma_j$ assigned by an instant zero imply equal contributions to all authors. The parameters (the Base and the Power $\gamma_j$ ) could enable the ratio of credits between any 2 sequential authors to be greater or less than 2.72 used in this study. The PIF is also affected by the JCR IF of cited journals, see the formula above. If a journal IF is less than 1.0 or the journal is not indexed by the JCR, the weighted score is assigned as 1.0. Otherwise, it is endorsed by the JCR IF.

In 2015 and 2016, 5293 citable papers in *Medicine* were cited by 4627 times from journals on PubMed in 2017, and 649 PIFs (>1.0) for *Medicine* authors are generated in comparison, see Additional file 3.

### 2.4. Statistics and requiring unique author name

Analysis of variance was performed to examine the difference in paper quantity. The medians of GC and PIF among the 3 study groups (i.e., the first, the middle, and the last author) are compared. The Kendall coefficient (W)[31] was used for evaluating the concordance of density parameters in a network. We expect that the larger network (i.e., with the more number of members) has, the lower density coefficient will be.

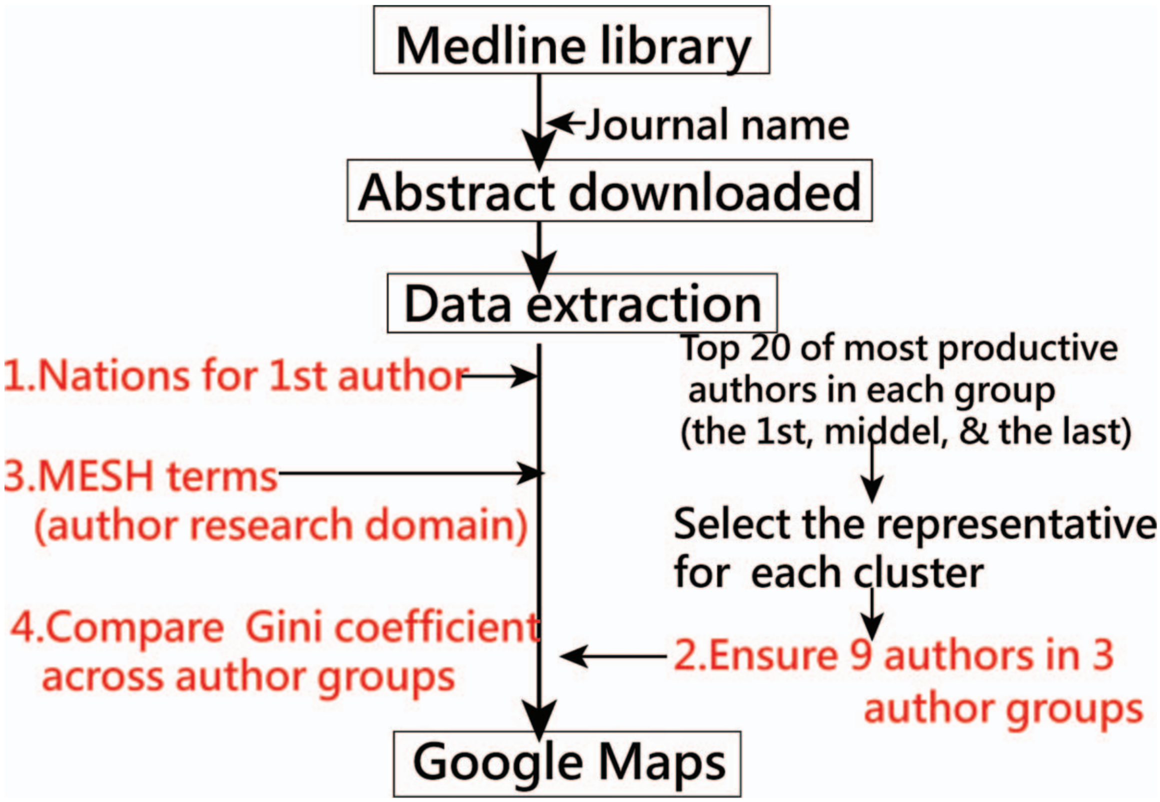


**Figure 1.** The flowchart and concept of this study.

There are many duplicate authors in PubMed library. We applied SNA (using partition-component method) to confirm the author name is unique. The reason is that the productive author must have many coauthors jointly constructing a big network and any 2 authors with identical names are rarely listed in a science research team (or, say, network). Applying the Between Centrality algorithm to search the Bridge role in a network, we can select the possible authors whose names might be identical but are different persons. Another way is to exclude the target author and examine any network isolated from the main network. If it has happened, we then carefully inspect the affiliated institute and make sure the author is finally identical. Interested readers are recommended to see Additional files 4 and 5.

## 3. Results

### 3.1. Trend of author collaboration in Medicine

A total of 7502 (=69.11%) abstracts were included due to their author countries/areas exactly listed in PubMed library (Table 1). The trend of author collaboration in *Medicine* is slightly declining (= −0.06) based on the number of authors per article, as shown at the bottom of Table 1. The mean number of individuals listed as authors in articles is around 7.5 for *Medicine*.

### 3.2. Change in dominant nations in Medicine

A total of 7502 eligible papers with entire author nations since 1969 are shown in Table 1. Most authors are from China (3649, 48.64%) and Taiwan (847, 11.29%). The trend in the number of publications for countries is shown in the column of growth in Table 1. The 3 continents of Asia, Europe, and North America have the highest growth rate (≥0.96). The diagram (shown by SNA and Google Maps) in Fig. 2 displays the author collaborations among nations in pivotal roles (i.e., bridges). The top 4 are those from the United States, Germany, China, and Brazil. The top 2 continents with the most number of papers in *Medicine* are Asia (71.3%) and Europe (16.26%). Any nation that collaborated with other nations is shown with a blue line. Interested authors can click the bubble of interest to see details on a website at reference.[32]

### 3.3. The most frequency of Mesh terms in Medicine

The top 2 in counts of PubMed Mesh terms in *Medicine* are method and surgery which were rarely occurred before 2014. Other 8 terms are similarly with a highly increasing trend due to the publication outputs that are increasing since 2015, see the last column in Table 2.

### 3.4. Three groups of productive authors selected in Medicine

The representatives in the 3 groups are as follows:

1. Marginean, Cristina Oana (Romania), Jang, Sung Ho (Republic of Korea), and Lee, Jacky W Y (China);
2. Chen, Tzeng-Ji (Taiwan), Yu, Chang Sik (Republic of Korea), and Jeng, Jen-Eing (Taiwan);
3. Kao, Chia-Hung (Taiwan), Gonzalez-Gay, Miguel A (Italy), and Bouza, Emilio (Spain); see Table 3.

**Table 1**

**The number of published papers in Medicine across years and nations.**

| Continent | 1969–07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | Total | % | Growth |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Africa | | | | | | | | 1 | 16 | 24 | 20 | | 61 | 0.81 | 0.88 |
| Egypt | | | | | | | | | 10 | 15 | 7 | | 32 | 0.43 | 0.96 |
| Others | | | | | | | | 1 | 6 | 9 | 13 | | 29 | 0.39 | 0.82 |
| Asia | | 4 | 3 | 8 | 10 | 5 | 3 | 123 | 1049 | 1741 | 2296 | 71 | 5349 | 71.3 | 0.96 |
| China | | 1 | 1 | 1 | | | | 69 | 630 | 1093 | 1803 | 51 | 3649 | 48.64 | 0.99 |
| Taiwan | | 1 | 2 | 3 | 3 | 1 | | 23 | 250 | 369 | 180 | 8 | 847 | 11.29 | 0.77 |
| Japan | | | | | 5 | 1 | 2 | 14 | 71 | 101 | 161 | 5 | 361 | 4.81 | 0.8 |
| South Korea | | | | | | | | 1 | 23 | 37 | 45 | 2 | 108 | 1.44 | 0.99 |
| Turkey | | | | 1 | | 2 | | 6 | 19 | 33 | 24 | 1 | 89 | 1.19 | 0.95 |
| Hong Kong | | | | | | | 1 | 2 | 16 | 25 | 10 | 1 | 58 | 0.77 | 0.95 |
| Others | | 2 | 0 | 3 | 2 | 1 | | 9 | 56 | 107 | 93 | 3 | 298 | 3.97 | 0.92 |
| Europe | | 16 | 22 | 21 | 23 | 20 | 10 | 51 | 218 | 330 | 289 | 9 | 1220 | 16.26 | 1.00 |
| France | | 6 | 8 | 12 | 11 | 11 | 7 | 13 | 35 | 71 | 44 | | 303 | 4.04 | 0.97 |
| Italy | | | 2 | 1 | 1 | | | 9 | 50 | 78 | 56 | | 208 | 2.77 | 0.98 |
| Spain | | 8 | 7 | 6 | 9 | 6 | 3 | 11 | 26 | 22 | 31 | | 194 | 2.59 | 0.9 |
| Germany | | | 1 | | | | | 4 | 21 | 32 | 27 | | 89 | 1.19 | 0.92 |
| Switzerland | | | 1 | 2 | | | | 2 | 15 | 18 | 23 | 1 | 71 | 0.95 | 0.97 |
| Poland | | | | | | | | 1 | 16 | 20 | 20 | 3 | 60 | 0.8 | 0.99 |
| Others | | 2 | 3 | 0 | 2 | 3 | 0 | 11 | 55 | 89 | 88 | 5 | 295 | 3.93 | 0.99 |
| North America | | 15 | 17 | 11 | 13 | 10 | 15 | 22 | 61 | 145 | 116 | 4 | 746 | 9.94 | 0.98 |
| USA | | 14 | 17 | 10 | 12 | 10 | 15 | 17 | 54 | 129 | 99 | 3 | 672 | 8.96 | 1.00 |
| Canada | | 1 | | 1 | 1 | | | 4 | 1 | 14 | 14 | | 54 | 0.72 | 0.92 |
| Others | | | | | | | | 1 | 5 | 2 | 3 | 1 | 19 | 0.25 | 0.45 |
| Oceania | 36 | | | | 1 | | | 2 | 14 | 16 | 8 | | 44 | 0.59 | 0.85 |
| Australia | | | | | 1 | | | 1 | 12 | 11 | 8 | | 36 | 0.48 | 0.98 |
| New Zealand | 3 | | | | | | | 1 | 1 | 4 | | | 6 | 0.08 | 0.66 |
| Others | 2 | | | | | | | | 1 | 1 | | | 2 | 0.03 | 0.79 |
| South America | | 1 | | | | | | 6 | 23 | 19 | 24 | 1 | 82 | 1.09 | 0.7 |
| Brazil | 19 | 1 | | | | | | 5 | 18 | 15 | 19 | 1 | 64 | 0.85 | 1 |
| Peru | | | | | | | | 1 | 2 | 1 | 4 | | 8 | 0.11 | 0.89 |
| Others | 1 | | | | | | | | 3 | 3 | 1 | | 10 | 0.13 | 0.5 |
| Total | | 36 | 42 | 40 | 47 | 35 | 28 | 205 | 1381 | 2275 | 2753 | 85 | 7502* | 100 | 1 |
| The trend of author collaboration in *Medicine* | | | | | | | | | | | | | | | |
| # per article | | | | | | | | | | | | | | | |
| Mean | 4.7 | 7.9 | 7.5 | 8.7 | 8.0 | 9.0 | 9.5 | 7.8 | 7.6 | 7.5 | 6.5 | 6.1 | 7.5 | | −0.03 |
| Median | 4 | 6 | 7 | 7 | 7 | 7 | 8 | 7 | 7 | 7 | 6 | 6 | | | 0.29 |
| Minimal | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | | | −0.46 |
| Maxmal | 23 | 22 | 16 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 15 | | | −0.13 |

* Authors with 69.11% (=7502/10,855) have countries/areas in MEDLINE.

Nine clusters are extracted by using principal component analysis (Fig. 3). The cluster features are shown in Table 4, showing which relevant parameters and coefficients are appropriate for describing the characteristics of the respective cluster. For instance, all clusters are independent of E-I index (i.e., all equal to $-1$, implying all members disconnected to outside clusters), and the author Kao, Chia-Hung has the greater number of members (=232). Seven clusters have a significant network density (i.e., $t$ value $> 2.0$ on the cluster coefficient). As expected, the larger network (at the right side in Table 4) has the lower density coefficient (at the left side in Table 4) with a significant higher Kendall W ($P < .05$). Interested authors can view Reference[33] to see the details of the author clusters and their collaborations in academic studies. The calculation of Kendall coefficient (W) is referred to the Reference[34] and Additional File 4.

### 3.5. Research domain in depth and breadth for an author

GC is applied to the research domain to interpret its depth and breadth. We illustrate an example of the author Tzeng-Ji Chen (a middle author) in Fig. 4. When his papers were downloaded from *Medicine* by searching the keyword of the author name, a total of 21 abstracts were obtained. The top 4 domains are epidemiology, economics, statistics and numerical data, and complications with an overall GC of 0.52, indicating his research domains are in-depth (>0.40, with an unequal size for the top 5 bubbles in the network). Interested readers may view the link in Reference[35] to see the details. All papers of the author Tzeng-Ji Chen in *Medicine* are shown on the web of the link [Publications] clicked at the upper left portion of the map. All relevant GCs at the bottom in Table 4 computed by the formula are given in Methods 2.2.3 and demonstrated at the Reference.[25] A demonstration regarding GC calculation in an MS Excel format is provided in Additional file 6.

### 3.6. Patterns of research domain in the ordering of author names

No difference was found in the number of papers [$F(2, 6) = 1.01$, $P = .42$] among author groups. Figure 5 shows a comparison of research domains and PIFs across the ordering of author names

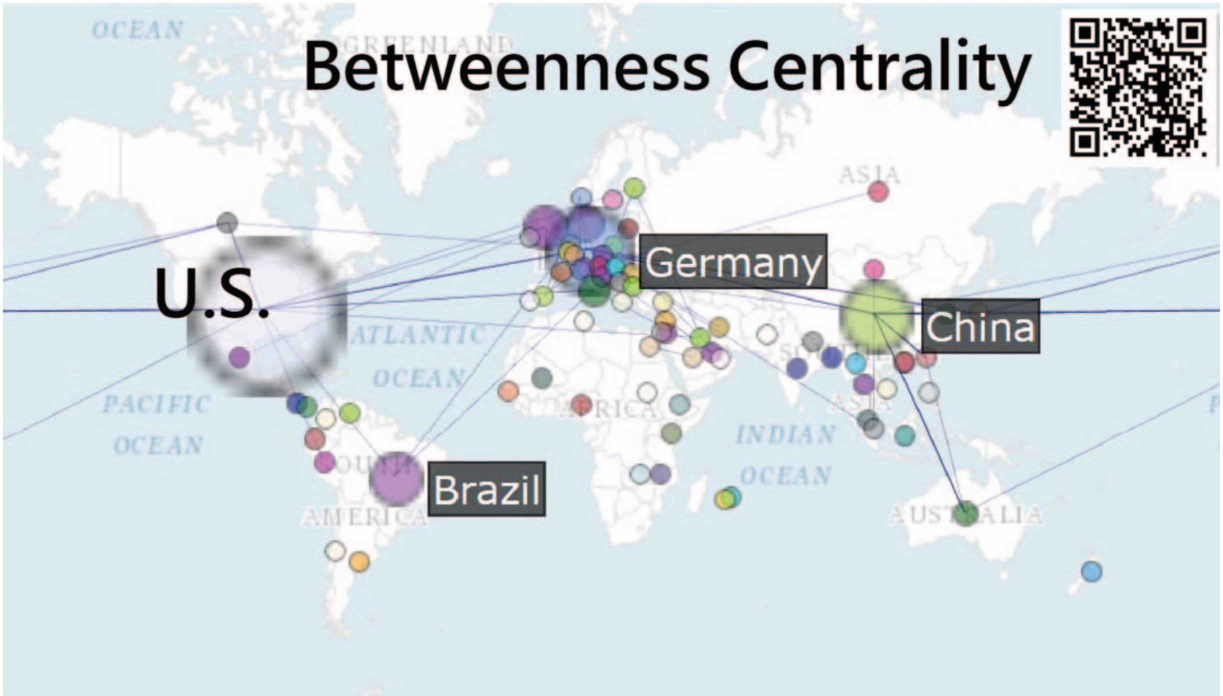


**Figure 2.** Author nationals distributed on the Google Maps.

by using the GC cutting point at 0.4 and PIF at 1.0. We can see that the median of GC (0.32) and PIF (0.74) for the middle authors are significantly less than them for the 1st (0.53, 2.19) and last authors (0.42, 2.61), respectively. A perfect positive linear relation with a large effect[36] exists between GC and PIF because the correlation coefficient is 0.68 (>0.50, $t=2.48$, n=9).

Among 649 authors with PIFs greater than 1.0 in Fig. 6, both Hyun Chul Lee (31.8) and Eun Young Lee (30.4) had the highest PIFs in 2017, and both Chia-Hung Kao (62.12) and Sung Ho Jang (14.21) earned the most highly cited scores contributed to the Journal of *Medicine*.

## 4. Discussion

The trend of author collaborations in *Medicine* is slightly declining (=−0.06), as shown by the number of authors per article. The mean number of authors in articles is 7.5 for *Medicine*; this figure was 6.1 in 2018. This number is higher than that in PubMed from 1.9 (1975) to 5.67 (2016),[37] but less than that in the 3 leading general medicine journals (*JAMA*, *The Lancet*, and *New England Journal of Medicine*) (from a range of 8–11 in 2005 to 11–18 in 2015) in 2005, 2010, and 2015.[38]

### 4.1. What this knowledge adds to what we already knew

Most authors of papers published in *Medicine* are from China (3649, 48.64%) and Taiwan (847, 11.29%), indicating that the dominant countries in scientific research have shifted from the United States and Europe[39,40] to Asia, such as China and Japan.

Usually, the person who provided strategic thinking throughout the project is the last author (this person is also typically responsible for the funding). The last author often receives as much credit as the first author, because he or she is assumed to be the intellectual and financial driving force behind the research.[41] We found that the medians of GC and PIF in groups of the first and the last authors display higher than the counterpart of the middle authors. The primary and supervisory should be shared with more credits contributed to a given article.[4,5]

**Table 2**

**The top 10 Mesh terms over the years for the journal of *Medicine*.**

| Mesh term | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | Total | Growth |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Methods | | | | 7 | 1 | | 81 | 429 | 809 | 730 | 2057 | 0.95 |
| Surgery | | | | | 1 | | 52 | 408 | 673 | 865 | 1999 | 0.82 |
| Complications | 11 | 19 | 12 | 13 | 12 | 7 | 65 | 335 | 627 | 622 | 1723 | 0.83 |
| Genetics | 2 | 1 | 3 | 3 | 7 | 7 | 86 | 398 | 632 | 540 | 1679 | 0.77 |
| Epidemiology | 3 | 13 | 16 | 26 | 15 | 12 | 52 | 438 | 683 | 385 | 1643 | 0.86 |
| Diagnosis | 13 | 19 | 14 | 15 | 9 | | 79 | 368 | 609 | 491 | 1617 | 0.83 |
| Pathology | 8 | 9 | 3 | 4 | 13 | 1 | 64 | 397 | 546 | 557 | 1602 | 0.86 |
| Therapeutic use | 7 | 3 | 10 | 26 | 9 | | 53 | 351 | 570 | 531 | 1560 | 0.80 |
| Blood | 2 | 7 | 8 | 4 | 3 | 4 | 34 | 315 | 542 | 396 | 1315 | 0.82 |
| Drug therapy | 6 | 3 | 8 | 17 | 4 | 1 | 46 | 258 | 475 | 456 | 1274 | 0.82 |

**Table 3**

**The distribution of paper publication for 3 types of author in this study.**

| Author | 1969–07 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | Total | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| The 1st author | | | | | | | | | | | | | | |
| Marginean, Cristina Oana (Romania) | | | | | | | | | | 8 | 7 | 2 | 17 | 6.77 |
| Jang, Sung Ho (Republic of Korea) | | | | | | | | | 2 | 15 | 17 | 6 | 40 | 15.94 |
| Lee, Jacky W Y (China) | | | | | | | | 3 | 7 | 2 | | | 12 | 4.78 |
| The middle author | | | | | | | | | | | | | | |
| Chen, Tzeng-Ji (Taiwan) | | | | | | | | 1 | 9 | 9 | 2 | | 21 | 8.37 |
| Yu, Chang Sik (Republic of Korea) | | | | | | | | | 2 | 7 | 3 | | 12 | 4.78 |
| Jeng, Jen-Eing (Taiwan) | 2 | 1 | 1 | | | | | | | | | | 4 | 1.59 |
| The last author | | | | | | | | | | | | | | |
| Kao, Chia-Hung (Taiwan) | | | | | | | | | 73 | 34 | 2 | | 109 | 43.43 |
| Gonzalez-Gay, Miguel A (Italy) | 7 | 2 | 2 | | 3 | | 1 | 2 | 1 | 1 | | | 19 | 7.57 |
| Bouza, Emilio (Spain) | 2 | 1 | 2 | 2 | | 1 | | 1 | 7 | 1 | | | 17 | 6.77 |
| Total | 11 | 4 | 5 | 2 | 3 | 1 | 1 | 7 | 101 | 77 | 31 | 8 | 251 | 100 |

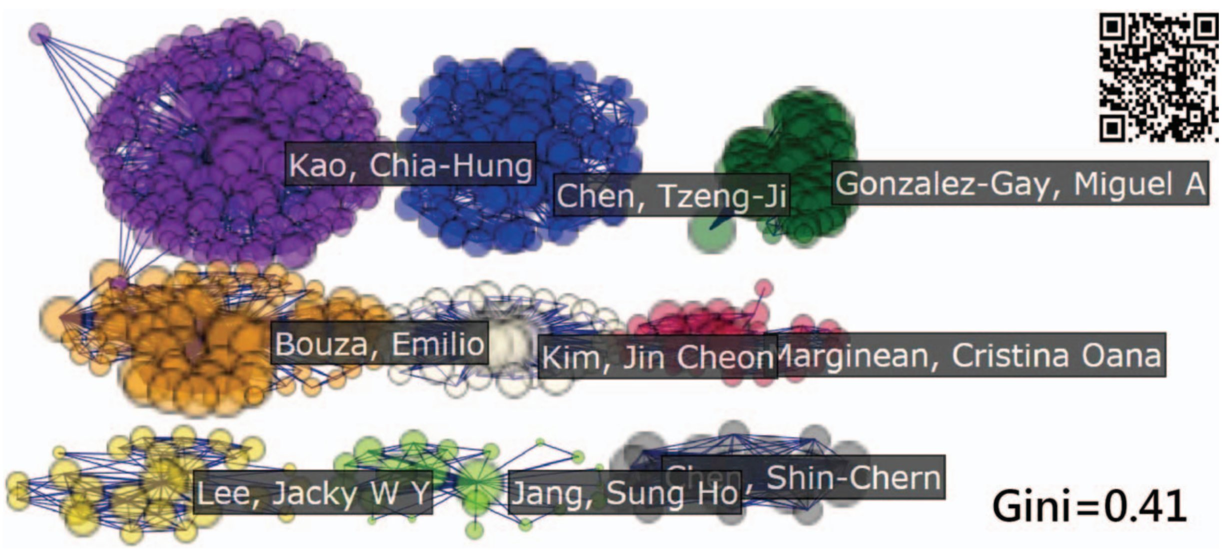


**Figure 3.** Author distribution on the Google map.

**Table 4**

**Density parameters for each cluster.**

| Author (PIF) | CC | *t* value | Density | D_Weight | E-I index | n | Link | L_Weight | Gini |
|---|---|---|---|---|---|---|---|---|---|
| The 1st author | | | | | | | | | |
| Marginean, Cristina Oana (2.194) | 0.44 | 2.68* | 0.26 | 0.42 | −1 | 32 | 128 | 209 | 0.64 |
| Jang, Sung Ho (1.214) | 0.40 | 1.90 | 0.25 | 0.52 | −1 | 21 | 52 | 109 | 0.53 |
| Lee, Jacky W Y (2.648) | 0.56 | 3.24* | 0.37 | 0.56 | −1 | 25 | 112 | 167 | 0.49 |
| The middle author | | | | | | | | | |
| Chen, Tzeng-Ji (1.679) | 0.46 | 5.26* | 0.12 | 0.15 | −1 | 105 | 672 | 809 | 0.52 |
| Yu, Chang Sik (0.747) | 0.34 | 2.01* | 0.45 | 0.93 | −1 | 33 | 235 | 490 | 0.32 |
| Jeng, Jen-Eing | 0.73 | 4.14* | 0.90 | 2.03 | −1 | 17 | 122 | 276 | 0.20 |
| The last author | | | | | | | | | |
| Kao, Chia-Hung (2.612) | 0.11 | 1.68 | 0.05 | 0.07 | −1 | 232 | 1233 | 1973 | 0.64 |
| Gonzalez-Gay, Miguel A (1.0) | 0.61 | 6.49* | 0.24 | 0.37 | −1 | 73 | 626 | 961 | 0.33 |
| Bouza, Emilio (3.178) | 0.53 | 5.00* | 0.21 | 0.43 | −1 | 66 | 448 | 918 | 0.42 |
| Gini coefficient | 0.13 | 0.17 | 0.22 | 0.26 | 0 | 0.29 | 0.30 | 0.30 | 0.11 |
| Kendall W=0.68, $\chi^2$ = 16.53, *d. f.* = 8, *P* =.04; | | W=0.94, $\chi^2$ = 22.58, *d. f.* = 8, *P*<.001. | | | | | | | |

CC = cluster coefficient, the more means the higher dens sample size of CC, Density = the ratio between the number without duplicated connections and the maximal possible connections (=n (n-1)/2), D_Weight = density with the number of repeated connections, E-I index = the difference between the external and internal contacts divided by the total number of connections, the less means, the higher convergent centrality, n = the number of members, Link = the number of nonduplicated connections, L_Weight = the number including duplicated connections, Gini = gini coefficient.ity in the network, *t* value = dependent of data elements.

* A significant level at 0.5.

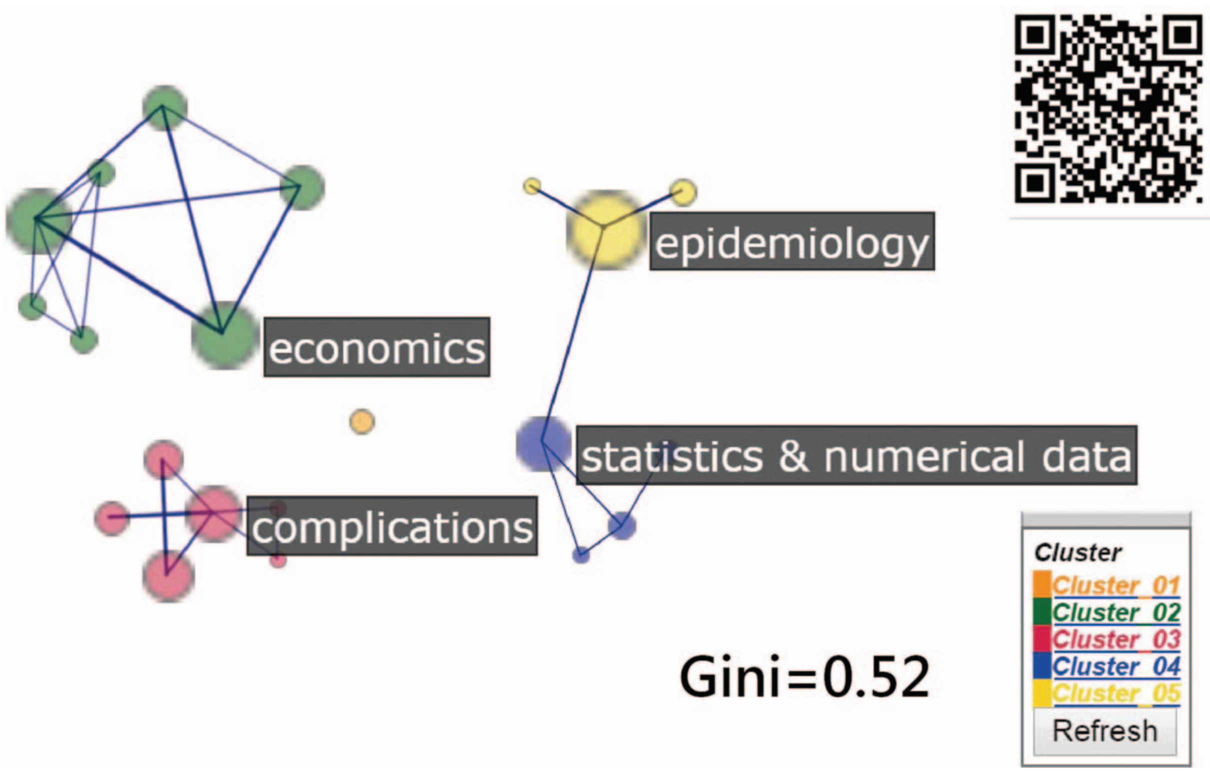


**Figure 4.** The research domain of the author Tzeng-Ji Chen.

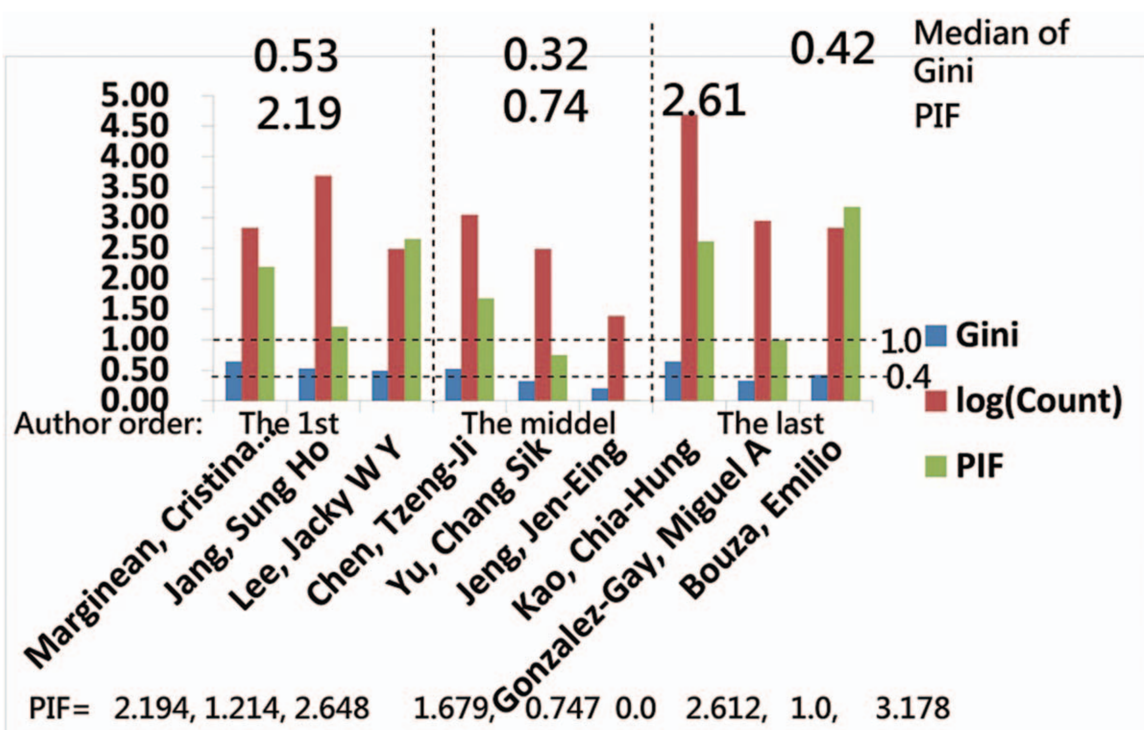


**Figure 5.** Comparison of research domains across the ordering of author names using Gini coefficients.

We applied the Gini coefficient to present the extent of in-depth author domains like the World Bank, UN, did for evaluating the inequality of a nation's wealth and many healthcare settings assessing the distribution of hospital beds,[42] the inefficiencies in esophageal adenocarcinoma screening,[43] the equality of medical health resource allocation on geographical areas in Chian,[44] and the child mortality in Taiwan compared with industrialized countries.[45]

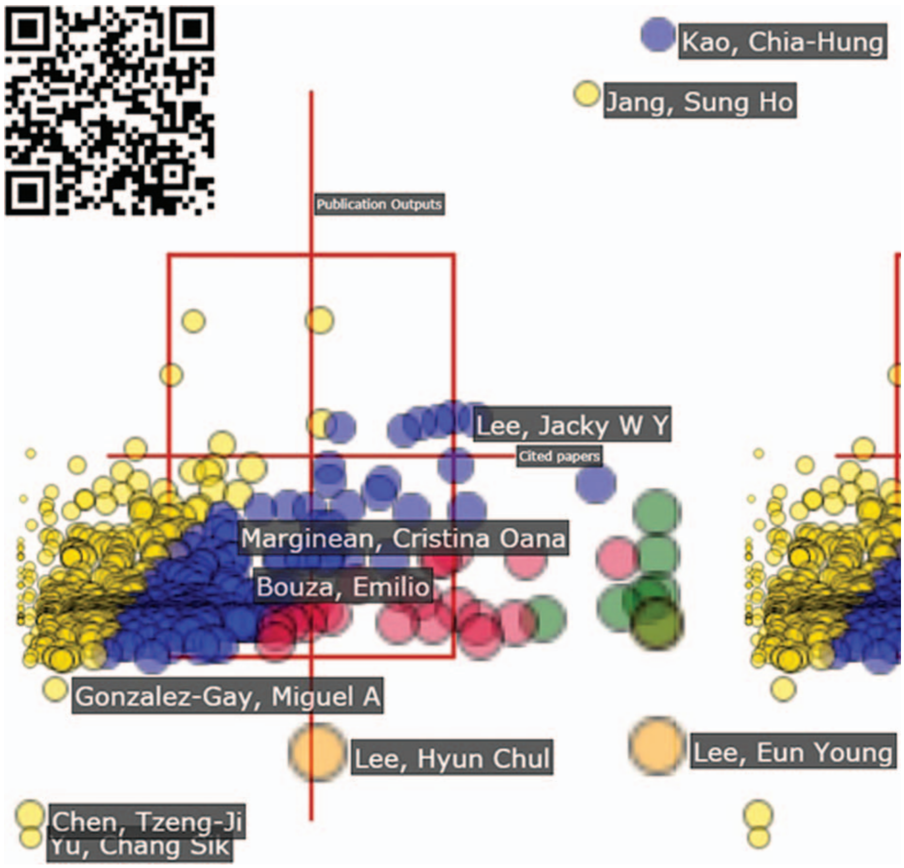


**Figure 6.** Personal impact factors for Medicine authors on Googel Maps.

Through the visual representations in Figures, we see that any 2 groups can be significantly separated from each other if the E-I index equals −1.0. Our use of data mining and the SNA algorithm in this study is similar to the apocryphal story that found that beer and diaper products in a supermarket have a strong correlation[46,47] and a study that evaluated research topic evolution in psychiatry by using coword analysis.[48,49]

### 4.2. What the findings imply and what should be changed?

We verified that the association between GCs of author domain and PIFs is positively related. The journal IF can be predicted (or improved) by high-quality citable papers published by authors whose PIFs are high. The core author we address in this study is required to disclose his or her expertise to readers (or reviewers) in the future.

We particularly incorporated SNA with Google Maps, as similarly used in a previous study[6] that discovered the most productive authors by using the National Health Insurance Database. The dashboards with the hyperlink[32,33] can be manipulated by readers to view details about the entity relationship in general. The strength of the study is that it provides readers with MP4 videos in Additional files. As such, the research process involves automatically building Hyper Text Mark-up Language for use on Google Maps, which was never applied in previous papers that employed SNA to visualize author collaboration characteristics.[50] Another feature is the way to screen out the duplicate authors with identical names shown in Additional File 4 which is also rarely seen and frequently reported as limitations in the literature.[50]

Future studies are encouraged to incorporate SNA with Google Maps to present knowledge concept maps.[51,52] The author domains are combining SNA and GC on Google Maps as a dashboard that is required to be shown in the cover letter of the submission paper to indicate that the manuscript is relevant to the author expertise and the PIF.

### 4.3. Strengths of this study

We used SNA to analyze coauthor collaboration and research domain for a particular author, such as Tzeng-Ji Chen in Fig. 4, which is different from the traditional approaches used for displaying knowledge concept maps in previous studies.[53,54] We suggested all submission papers reporting author domains with the GC and PIF in the cover latter to display the author academic domains to the journal editors and reviewers in the future.

Notably, we transformed the coordinates from Pajek into Google Maps so that all nodes (authors or keywords) can be precisely located on Google Maps and all clusters can be gathered in appropriate colors and sizes on Google Maps, which were rarely seen using social network analysis to show the results with a dashboard on Google Maps. For more details, see Additional Files.

Each year in June, millions of scholars pay close attentions to the JCR locating journal impact factors. No such PIF was applied to individual scholars for a scientific discipline or institute as JCR annually updating the IFS for the indexed journals. We demonstrated a scheme of quantifying coauthor contribution in an article byline and developed a formula based on PubMed publications for calculating PIFs. The PIF can be applied to the productivity and impact of a scholarly journal as well as a group of scientists, such as a department or university or country. All coauthor contributions are automatically sized by 1.0 for each published article through the probability theory based on the Rasch rating scale model.[55] The process of PIF calculation can be referred to as Additional file 3.

### 4.4. Limitations and future study

Although findings are based on the above analysis, there are still several potential limitations that may encourage further research efforts. First, all data were extracted from the PubMed database. Some papers that have incomplete information (or some authors have an identical name to other authors though we have made efforts to detect them, see Additional File 4) might affect the results of this study. For instance, only 69.11% of authors in the country/area might result in a selection bias. If those missing data were randomized, the inference made in Table 1 would be accepted. Studies are recommended to verify the results (i.e., the dominant countries in *Medicine* have shifted from the United States and Europe to Asia) in the future.

Second, many algorithms have been used for SNA. We merely applied the algorithm of degree centrality in the figures. Any changes in the algorithm used in this study might present a different pattern and judgment to the results.

Third, the data extracted from *Medicine* cannot be generalized to other journals, such as the declining trend of author collaboration (= − 0.06) based on the number of authors per article and the shift in nations that dominate science research from the United States and Europe[39,40] to Asia. More journals should be included in studies on a similar topic in the future.

Fourth, author academic domains and cited papers are determined by the paper selections on PubMed. Whether the positive linear relation between GCs (i.e., author domain across the ordering of authors in an article byline) and the corresponding PIFs still exists in other journals is required to further inspect in the future.

Fifth, we demonstrated a model for quantifying coauthor contributions. The parameters were arbitrarily set for calculating author PIFs in an exponentially descending order. Whether the PIFs can help editors (or readers) know who are the most highly cited authors in a scientific discipline is needed to verify in the future.

Sixth, the assumption of corresponding (or supervisory) authors being the last authors might be challenged, especially in computing PIFs. Any parameters changed in our proposed formula will affect the author contribution weights and the PIFs in results. The parameters set to calculate weights in this study might accommodate the ordering of authors in the biomedical field (i.e., the first authors own the most, the last followed, and the middle third, and the others less and less).[5]

## 5. Conclusion

These findings demonstrate that SNA combined with Google Maps is feasible for the development of the knowledge concept (i.e., the author research domain and the PIF display for authors). Results suggest that the corresponding author is submitting the manuscript to the target journal with a core author's academic background and the personal IF related to the research domain and the journal scope in the future. As such, peer reviewers can quickly determine whether the manuscript is a potentially citable research paper.

## Acknowledgments

The authors thank Enago (www.enago.tw) for the English language review of this manuscript.

## Author contributions

T-WC conceived and designed the study, JCC and YC interpreted the data, and WC monitored the process and the manuscript. T-WC drafted the manuscript. All authors read the manuscript and approved the final manuscript.

**Conceptualization:** Tsair-Wei Chien.
**Data curation:** Julie Chi Chow, Yu Chang.
**Formal analysis:** Tsair-Wei Chien, Yu Chang.
**Methodology:** Julie Chi Chow, Yu Chang.
**Project administration:** Julie Chi Chow, Yu Chang, Willy Chou.
**Resources:** Tsair-Wei Chien.
**Software:** Tsair-Wei Chien, Yu Chang.
**Supervision:** Willy Chou.
**Writing – original draft:** Tsair-Wei Chien.

Tsair-Wei Chien orcid: 0000-0003-1329-0679

## References

[1] Avula J, Avula H. Authors, authorship order, the moving finger writes. J Indian Soc Periodontol 2015;19:258–62.
[2] Gibbons M, Limoges C, Nowotny H, et al. The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies. Sage, London:1994.
[3] Tanner-Smith EE, Polanin JR. Brief alcohol intervention trials conducted by higher prestige authors and published in higher impact factor journals are cited more frequently. J Clin Epidemiol 2016;75:119–25.
[4] Sekercioglu CH. Quantifying coauthor contributions. Science 2008;322: 371.
[5] Mongeon P, Smith E, Joyal B, et al. The rise of the middle author: investigating collaboration and division of labor in biomedical research using partial alphabetical authorship. PLoS One 2017;12:e0184601.
[6] Chien TW, Chang Y, Wang HY. Understanding the Productive Author who Published Papers in Medicine Using National Health Insurance Database: a systematic review and meta-analysis. Medicine (Baltimore) 2018;97:e9967.
[7] Baerlocher MO, Newton M, Gautam T, et al. The meaning of author order in medical research. J Investig Med 2007;55:174–80.
[8] Bhandari M, Guyatt GH, Kulkarni AV, et al. Perceptions of authors' contributions are influenced by both byline order and designation of corresponding author. J Clin Epidemiol 2014;67:1049–54.
[9] Igou ER, van Tilburg WA. Ahead of others in the authorship order: names with middle initials appear earlier in author lists of academic articles in psychology. Front Psychol 2015;6:469.
[10] Fontanarosa P, Bauchner H, Flanagin A. Authorship and team science. JAMA 2017;318:2433–7.
[11] Gini C. Concentration and dependency ratios (in Italian). English translation in Rivista di PoliticaEconomica 1909;87:769–89.
[12] Pu QH, Lyu QJ, Liu H, et al. Bibliometric analysis of the top-cited articles on islet transplantation. Medicine (Baltimore) 2017;96:e8247.
[13] Tian J, Li M, Lian F, et al. The hundred most-cited publications in microbiota of diabetes research: a bibliometric analysis. Medicine (Baltimore) 2017;96:e7338.
[14] Miao Y, Liu R, Pu Y, et al. Trends in esophageal and esophagogastric junction cancer research from 2007 to 2016: a bibliometric analysis. Medicine (Baltimore) 2017;96:e6924.
[15] Zhang Y, Huang J, Du L. The top-cited systematic reviews/meta-analyses in tuberculosis research: a PRISMA-compliant systematic literature review and bibliometric analysis. Medicine (Baltimore) 2017;96:e4822.
[16] Liao J, Wang J, Liu Y, et al. Modern researches on Blood Stasis syndrome 1989-2015: a bibliometric analysis. Medicine (Baltimore) 2016;95: e5533.

[17] Li H, Zhao X, Zheng P, et al. Classic citations in main primary health care journals: A PRISMA-compliant systematic literature review and bibliometric analysis. Medicine (Baltimore) 2015;94:e2219.
[18] 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:e24.
[19] 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;34:1157–60.
[20] Bright CF, Haynes EE, Patterson D, et al. The value of social network analysis for evaluating academic-community partnerships and collaborations for social determinants of health research. Ethn Dis 2017;27 (suppl 1):337–46.
[21] de Nooy W, Mrvar A, Batagelj V. Exploratory Social Network Analysis With Pajek: Revised and Expanded. 2nd ed. Cambridge University Press, New York, NY:2011.
[22] 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;48:1353–61.
[23] Tao Y, Wu XJ, Li CS. Rawls' fairness, income distribution and alarming level of Gini coefficient. Economics Discussion Papers 2017; No 2017-67, Kiel Institute for the World Economy. 2018/5/2 retrieved at https://arxiv.org/ftp/arxiv/papers/1409/1409.3979.pdf.
[24] Biancotti C. A polarization of inequality? The distribution of national Gini coefficients 1970–1996. J Econ Inequality 2006;4:1–32.
[25] Chien TW. Calculation of Gini coefficient on Google Maps. August 8, 2018. Available at: http://www.healthup.org.tw/gps/google_gini.asp.
[26] Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005;102:16569–728.
[27] Kokko H, Sutherland WJ. What do impact factors tell us? Trends Ecol Evol 1999;14:382–4.
[28] Sax LJ, Hagedorn LS, Arredondo M, et al. Faculty research productivity: exploring the role of gender and family-related factors. Res Higher Educ 2002;43:423–336.
[29] Kelly CD, Jennions MD. Thh index and career assessment by numbers. Trends Ecol Evol 2006;21:167–70.
[30] Pan RK, Fortunato S. Author Impact Factor: tracking the dynamics of individual scientific impact. Sci Rep 2014;4:4880.
[31] Kendall MG, Babington SB. The problem of m rankings. Ann Math Statist 1939;10:275–87.
[32] Chien TW. Google map on author nation distribution for the Journal of Medicine. February 13, 2018. Available at: http://www.healthup.org.tw/gps/medicinanation.htm.
[33] Chien TW. Google map on author collaboration for the *Journal of Medicine*. February 13, 2018. Available at: http://www.healthup.org.tw/gps/medicineauthor.htm.
[34] Zaiontz C. Kendall's Coefficient of Concordance (W) using MS Excel. June 12, 2018. Available at: http://www.real-statistics.com/reliability/kendalls-w/.
[35] Chien TW. Google map on the author research domain of Tzeng-Ji Chen. June 13, 2018. Available at http://www.healthup.org.tw/gps/Tzeng-Ji.htm.
[36] Cohen J. A power primer“. Psychol Bull 1992;112:155–9.
[37] US National Library of Medicine. Number of authors per MEDLINE/PubMed citation. Available at: http://www.nlm.nih.gov/bsd/authors1.html. Updated May 16, 2017.
[38] Muth CC, Golub RM. Trends in authorship in team science in major medical journals, 2005-2015. Presented at: Eighth International Congress on Peer Review and Scientific Publication; September 11, 2017; Chicago, IL.
[39] Leydesdorff L, Wagner C, Park HW, et al. International collaboration in science: the global map and the network. CoRR abs/1301 0801 2013.
[40] 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;70:267–75.
[41] Tscharntke T, Hochberg ME, Rand TA, et al. Author sequence and credit for contributions in multiauthored publications. PLoS Biol 2007;5: e18.
[42] Asl IM, Abolhallaje M, Raadabadi M, et al. Distribution of hospital beds in Tehran Province based on Gini coefficient and Lorenz curve from 2010 to 2012. Electron Physician 2015;7:1653–7.
[43] Hur C, Zhan T, Thrift AP, et al. Lorenz curves and Gini coefficient analyses indicate inefficiencies in esophageal adenocarcinoma screening. Clin Gastroenterol Hepatol 2018;Epub ahead of print.
[44] Jin J, Wang J, Ma X, et al. Equality of medical health resource allocation in china based on the Gini coefficient method. Iran J Public Health 2015;44:445–57.
[45] Wu JC, Chiang TL. Comparing child mortality in Taiwan and selected industrialized countries. J Formos Med Assoc 2007;106:177–80.
[46] Verhoef PC, Kooge E, Walk N. Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. Routledge, London:2016.
[47] Power DJ. What is the "true story" about data mining, beer and diapers? DSS News. March 20, 2017. Available at: http://dssresources.com/faq/index.php?action=artikel&id=41.
[48] Wu Y, Jin X, Xue Y. Evaluation of research topic evolution in psychiatry using co-word analysis. Medicine (Baltimore) 2017;96:e7349.
[49] González LM, García-Massó X, Pardo-Ibañez A, et al. An author keyword analysis for mapping Sport Sciences. PLoS One 2018;13: e0201435.
[50] Shen L, Xiong B, Li W, et al. Visualizing collaboration characteristics and topic burst on international mobile health research: bibliometric analysis. JMIR Mhealth Uhealth 2018;6:e135.
[51] McAleese R. The Knowledge arena as an extension to the concept map: reflection in action. Interactive Learning Environ 1998;6:251–72.
[52] Birbili M. Mapping knowledge: concept maps in early childhood education. Early Childhood Res Pract 2008;8:1–2.
[53] Stewart SA, Abidi SS. Applying social network analysis to understand the knowledge sharing behavior of practitioners in a clinical online discussion forum. J Med Internet Res 2012;14:e170.
[54] Zhao K, Wang X, Cha S, et al. A multirelational social network analysis of an online health community for smoking cessation. J Med Internet Res 2016;18:e233.
[55] Andrich D. Relationships between the Thurstone and Rasch approaches to item scaling. Appl Psychol Meas 1978;2:449–60.