Mapping the collaboration networks of biomedical research in Southeast Asia

Vincent Schubert R. Malbas, MD

University of the Philippines Manila - Philippine General Hospital, Ermita, Manila, Philippines

Corresponding author: **Vincent Schubert R. Malbas, MD**

Postal address: College of Medicine, University of the Philippines, 547 Pedro Gil St, Ermita, Manila, Philippines

Tel. (+632) 5264170

Fax (+632) 5361368

Email: schubertmalbas@post.upm.edu.ph

ABSTRACT

Collaboration forms an integral aspect of global research endeavors, where co-authorship derived from bibliographic records provides the building block for mapping research collaboration networks. Bibliometric techniques and social network analysis tools were applied to measure the scope and depth of collaboration in biomedical research in Southeast Asia during the period 2005-2009. In particular, centrality scores and draw network maps were calculated for both country and institutional levels of aggregation. In the field of biomedical research, Thailand and Singapore are the most productive and collaborative countries in Southeast Asia during the period studied. Using network analysis, there was strong correlation of research productivity by a country or institution with the number of collaboration and its group influence, and weak correlation with maximal data flow within the research network. There were specific clusters of connected institutions in subnetworks for neoplasm, diabetes, and tuberculosis research. Given the observed frequency of regional collaboration in Southeast Asia, in comparison to foreign collaboration, it is argued that increasing the number of collaborations within Southeast Asia will help advance the region's efforts on domestic and regional health issues.

Keywords: Southeast Asia; co-authorship networks; bibliometrics; social network analysis; network centrality

INTRODUCTION

Author collaborations and institutional collaborations are a common aspect in scientific research. It exists in various forms, e.g., as interdisciplinary research work that is shared by groups of different disciplines and/ or as parallel research where individual groups apply a similar protocol in separate study sites. Current trends of scientific collaboration show increasingly global connections, which mirror the transformation of individual research systems into increasingly global efforts. The continuing growth in quantity, size, and diversity of research projects are afforded by the availability of new communication channels and information facilities.

As a field of scholarly interest, scientific research collaboration has been already extensively investigated using bibliometric co-authorships. Scholarly studies on international co-authorships and citation impact of various countries (Glanzel, 2001), domestic and foreign collaborative relations of institutions (Jarneving, 2010; Nwaka et al., 2010), comparison of co-author networks generated from different bibliographic database sources (Newman, 2001), and longitudinal characterization of scientific collaborations over time (Barabasi et al., 2002) have been thoroughly described. However, detailed maps of research collaboration networks comprising institutions in Southeast Asia are lacking in literature.

The stage and scope of scientific research in Southeast Asia is as diverse as the socio-cultural profiles of its component countries. In a recent report by UNESCO (2010), total research output in 2008 published by the region as a whole was largely comprised of clinical medicine (30%) and biomedical science (13%), and each component country presented a particular core competency. Singapore, for example, was a prolific producer in engineering and technology (28.7%) and clinical medicine (20.9%), while the Philippines was mostly productive in biology (36.4%) and clinical medicine (26.3%). This study attempts to describe and quantify this proliferation of biomedical science research in Southeast Asia at both country and institutional levels of aggregation.

Measuring the degree of collaboration in co-authorship networks have been done successfully over the past decade through the use of social network analysis (SNA), which is a field of sociology largely derived from network theory and graph theory. It has lent itself to a variety of disciplines, .e.g., computer science, economics, and biology. Some interesting findings on collaboration networks include the formation of core groups in science networks and the repercussions it may imply in research policy (Leydesdorff & Wagner, 2008) and the method of identifying specific institutions that hold central roles in regional, national, and transnational connectivity (Jarneving, 2010).

In terms of co-authorships by Southeast Asian scientists, UNESCO (2010) reported that across all fields of study, the United States was the most frequent source of co-authors in the period 1998-2008, while Japan, China, and European Union also comprised significant shares. During the same period, Indonesia and Vietnam favored Japanese collaboration, while Malaysia was more strongly connected with China. This study therefore aims to compare these observed patterns to the co-authorship data in biomedical research, and to provide quantifiable indicators of collaboration at the country level and institutional level of aggregation.

Research Questions

The primary objective of this study is to provide detailed maps of regional and foreign research collaborations by Southeast Asian countries and institutions in the field of biomedical research during the period 2005-2009. Two levels of investigation are sought: country analysis pertains to data for Brunei, Cambodia, Lao PDR, Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam, while institution analysis characterizes and ranks universities, medical centers, and other research institutions identified in bibliographic records of Sciverse Scopus (Elsevier, 2010). Specific subsets of biomedical science were further explored, particularly for disease fields of cancers and neoplasms, diabetes, infectious diseases, malaria, and tuberculosis. Specifically, the following research questions were investigated:

- Do co-authorship networks of countries and institutions in Southeast Asia in 2005-2009 demonstrate an efficient structure as a research network?

- Are there notable clusters of countries or institutions in each network?
- Which countries and institutions occupy central roles in each network?

To address these research questions, each country network and institution network was described according to the number of nodes, edges, density, cohesion, and extent of clustering. Institutions that occupy central roles in the disease study network, particularly by calculating degree, closeness and eigenvector centrality scores were also identified (Bonacich, 1972; Borgatti, Everett, & Freeman, 2002; Freeman, 1979).

Scope and Limitations

Bibliographic co-authorship data have been exported from Sciverse Scopus on July 3, 2010, and comprise biomedical publications made between January 2005 and December 2009. As a bibliographic database, Scopus presents several advantages and disadvantages compared to Thomson Reuters Web of Science, NCBI Pubmed, and Google Scholar (Kulkarni, Aziz, Shams, & Busse, 2009; Falagas, Pitsouni, Malietzis, & Pappas, 2008; Jacso, 2005).

Collaboration networks were examined in terms of weighted undirected edges among countries or institutions, where edges or links are assigned with similarity values corresponding to the number of co-authorships. Co-authorship is defined as the co-occurrence of two or more affiliations in a bibliographic record, and specifically prefer whole author counting for assigning the weight of edges between co-authors (Gauffriau, Larsen, Maye, Roulin-Perriard, & von Ins, 2007). Note that there are other author counting methods, i.e., fractional author counting and first author counting, which will yield different weight assignments. Specifically, fractional author counting implies that each of the A authors receives a score equal to 1/A, while first author counting assigns one full credit to the first author while other authors do not receive any credit (Gauffriau et al., 2007).

METHODS

Data Sourcing

Bibliographic records, including biomedical articles, reviews, letters, conference papers, and proceedings, are queried in the Sciverse Scopus on July 3, 2010. Particularly, these records have been published between January 2005 and December 2009, and have at least one author affiliated to a Southeast Asian address, i.e., Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Myanmar, Malaysia, Philippines, Singapore, Thailand, or Vietnam. To further restrict the dataset to the biomedical field, only related Scopus subject categories are selected. Ten (10) sets of ASCII-format text files are then exported from Scopus, corresponding to 10 Southeast Asian countries.

Data Processing

All text files exported from Scopus are merged into a single master text file, and duplicate bibliographic records are then carefully deleted. Publication dates are screened to restrict the dataset within 2005-2009. The master text file is converted into DBase files using the MS-DOS executable Scopus.exe (Leydesdorff, 2008), and then translated into a single Microsoft Access database comprising three core tables that contain article data (34,125 records), institutional affiliations (112,956 records), and author information (178,185 records). Excluding all errata notes published, the articles datatable contains 34,059 records. Since Scopus provides variable names and addresses of local and foreign institutions based directly on data given by publishers, the country names, institutional names and addresses were thoroughly checked, corrected, and standardized.

Data Analysis

Two-mode crosstab matrices are obtained from the Microsoft Access database according to the level of aggregation (i.e., country level or institutional level) and particular subset (i.e., all biomedical research, cancers and neoplasms, diabetes, infectious diseases, malaria, or tuberculosis). These 2-mode matrices comprise bibliometric records versus their corresponding co-authors (country or institution), with data containing binary values.

In counting the number of collaborations among countries and institutions, the method of whole author counting was used, where each author (country or institution) receives one full score. In assigning bibliometric records to specific subset (e.g., cancer and neoplasm research), specific keywords were used for data mining within the list of titles and abstracts of the Microsoft Access database.

To work around the maximum column limit of crosstab queries in Microsoft Access, these crosstab matrices are further merged in Microsoft Excel and IBM SPSS into 13 unique 2-mode matrices. Each 2-mode matrix is then transformed into a weighted, undirected 1-mode network, i.e., a similarity matrix based on the number of co-authorships, using Ucinet v6.286 by Analytic Technologies (Borgatti et al., 2002).

Social network measures are explored using automated software functions in Ucinet. Degree centrality, closeness centrality, and eigenvector centrality (Bonacich, 1972; Borgatti et al., 2002; Freeman, 1979) are calculated for each country or institution in the network. Descriptive statistics for each network, e.g., number of nodes, network density, and average centrality scores are also calculated using Ucinet. Pajek (Batagelj & Mrvar, 1998) was used as parallel software to confirm the number of nodes and edges. For each network analyzed, countries or institutions are ranked according to centrality scores.

Network mapping is done using Pajek and Netdraw (Borgatti, 2002). The strength of the link or edge between any pair of institution or country is defined as the number of bibliographic records wherein the pair cooccurs in the affiliation field in 2005-2009. Clustering methods are also automatedly derived (Girvan & Newman, 2002; Borgatti, 2002).

For statistical analysis, tabulated data for each network are subjected to Spearman rank correlation and Pearson product moment correlation, where 0.72 is the critical r & ρ value for 2-variable comparison at the country level of aggregation at N=10, at 0.050 significance p, and 0.21 is the critical r & ρ value for 2-variable comparison at the institutional level of aggregation at N>200, at 0.050 significance p. These coefficients are calculated in order to determine if, for example, the number of collaborations is correlatable with the amount of research output and centrality scores.

RESULTS

Social network analysis yields a comprehensive array of statistics to characterize scientific collaboration networks. At the country level of aggregation, infectious disease research and neoplasm research comprise 12.98% and 8.19% of 34,059 biomedical publications in 2005-2009, respectively (Table 1). Tuberculosis research, diabetes research, and malaria research also represent a significant share of the entire dataset (1.40%, 2.41%, and 2.78%, respectively).

Table 1: Descriptive statistics on country collaborations by type of biomedical research (2005-2009).

NETWORK SUMMARY	**RESEARCH NETWORKS AMONG COUNTRIES (2005-2009)**					
	all biomedical	**neoplasm**	**diabetes**	**infectious diseases**	**malaria**	**tuberculosis**
no. of articles (sources)	34,059	2,789	822	4,420	947	477
no. of nodes (countries)	163	101	68	120	85	57
no. of nodes within SEA	10	9	10	10	9	10
no. of nodes outside SEA	153	92	58	110	76	47
no. of edges (links)	4,245	1,657	634	1,751	687	338
ave. node degree	659.288	154.436	41.941	158.45	58.635	27.333
ave. node closeness	106.405	65.036	41.118	72.944	48.588	32.673
network density	4.0697	1.5444	0.626	1.3315	0.698	0.4881
ave. distance	1.322	1.328	1.278	1.245	1.192	1.212
distance-based cohesion	0.839	0.836	0.861	0.877	0.904	0.894
weighted overall graph clustering coefficient	14.403	4.407	1.821	5.466	3.225	1.682

A total of 163 countries are identified in the entire dataset (Table 1). Note that no bibliographic records have been exported from Scopus for Brunei under malaria research and Lao PDR under neoplasm research for this 5-year period. The infectious disease research network comprise 120 countries (nodes) or 73.62% of the total countries identified. The number of links is also greatest in infectious disease research, comprising 1,751 edges or 41.25% of the total 4,245 links among countries. Note that the number of edges in Table 1 simply describes whether two countries are linked or not.

In contrast, node degree is the number of ties incident upon a node (country or institution), and average node degree calculates the mean of all degree scores normalized against the number of nodes involved in the network. Degree score was described as numerically equivalent to the number of collaborations for that particular node. It is also equivalent to the sum of all weights of edges attached to that node. Thus, the network on infectious disease research has the most number of collaboration per country under the field of biomedical research.

Average node closeness, on the other hand, tells us how fast a unit of information can arrive on a given node of a network. A closer node therefore allows faster data exchange. Based on this network indicator, countries involved in the infectious disease research network are closer to each other compared to other disease subsets of biomedical research.

For a valued network, density computes the total of all values divided by the number of possible ties (Borgatti, 2002). Therefore, network density values greater than 1 represent an estimate that any pair of nodes (country or institution) collaborated multiple times during 2005-2009. The network on cancer and neoplasm research is denser compared to all other subsets. However, neoplasm research collaborations among countries tend to be slightly less cohesive based on average distance, which is described as the average number of steps it takes to travel from a given node to any other node. All networks in this study, nevertheless, show that the average number of steps between nodes is less than two, which is significantly lower than the popular concept of "six degrees of separation", or theoretically, the "small world" concept first demonstrated by Travers and Milgram (1969).

The weighted overall clustering coefficient, by definition, is the weighted mean of the clustering coefficient of all the nodes each one weighted by its degree (Borgatti et al., 2002). In other words, it tells us the probability that any one country belongs to a cluster or subset of nodes having multiple redundant connections formed with each other. Data suggests that clustering is most likely to occur in neoplasm research.

The same array of network statistics is calculated for the institutional level of aggregation (Table 2). Institutional networks on neoplasm research and malaria research have more nodes outside Southeast Asia (i.e., foreign institutions), while a larger percentage of nodes are within the region for tuberculosis research. Infectious disease research and diabetes research networks are more or less equally distributed between regional and foreign nodes. Comparison of average node degrees and network density shows that neoplasm research has generated more collaboration per institution. Infectious disease research, on the other hand, is more cohesive and has afforded the closest interaction among institutions.

Table 2: Network statistics on institutional collaborations by type of biomedical research (2005-2009).

NETWORK SUMMARY	RESEARCH NETWORKS AMONG INSTITUTIONS (2005-2009)				
	neoplasm	**diabetes**	**infectious diseases**	**malaria**	**Tuberculosis**
no. of articles (source)	2,789	822	4,420	947	477
no. of nodes (institutions)	913	449	1,330	527	369
no. of nodes within SEA	336	211	658	222	218
no. of nodes outside SEA	577	238	672	305	151
no. of edges (links)	9,001	1,668	9,609	2,249	909
ave. node degree	35.139	12.695	24.463	14.129	7.182
ave. node closeness	310.558	117.407	454.077	174.912	83.536
network density	0.0385	0.0283	0.0184	0.0269	0.0195
ave. distance	1.022	1.017	1.011	1.016	1.013
distance-based cohesion	0.989	0.992	0.995	0.992	0.993
weighted overall graph clustering coefficient	1.194	0.801	0.566	0.548	0.641

Based on the biomedical research network for 2005-2009, Singapore is the most collaborative country in Southeast Asia during the period, with 9,995 instances of collaboration, while Thailand was the most productive nation with 8,384 co-authorships with other countries (Table 3). Centrality measures are then calculated in Ucinet for each Southeast Asian country.

Table 3: Network collaborations and centrality indicators based on biomedical articles (2005-2009).

BIOMEDICAL RESEARCH		DIRECT LINKAGES				DATA FLOW		GROUP INFLUENCE	
country	**no. articles**	**total collabs**	**no. collabs within SEA**	**degree centrality**	**rank**	**closeness centrality**	**rank**	**eigenvector centrality**	**rank**
Thailand	12,568	8,384	487	8,384	**2**	150.5	**1**	0.636	**2**
Singapore	12,405	9,995	421	9,995	**1**	138.5	**2**	0.659	**1**
Malaysia	7,071	2,912	401	2,912	**3**	136	**3**	0.05	**3**
Indonesia	1,324	1,880	232	1,880	**4**	133	**4**	0.017	**4**
Philippines	574	1,189	199	1,189	**5**	130.5	**5**	0.012	**5**
Cambodia	318	620	115	620	**6**	119	**6**	0.006	**6**
Lao PDR	168	348	92	348	**7**	104	**8**	0.005	**7**
Brunei	163	97	30	97	**10**	91.833	**10**	0.001	**9**
Myanmar	103	226	46	226	**8**	107	**7**	0.002	**8**
Vietnam	86	133	29	133	**9**	102	**9**	0.001	**10**

In particular, each measure of centrality assumes a different set of characteristics regarding information flow in the network of nodes, particularly in the way how information is diffused or exchanged within the network, and in which path or trajectory the information is passed from one node to another (Borgatti, 2005). Specifically, the mechanism of transmission from node to node in collaborative research networks is likely through parallel duplication, where information is copied from the original node and simultaneously performed in several receiving nodes. This is similar to a radio broadcast network where information is recorded in the radio station and broadcasted to all radio listeners. Borgatti (2005) has argued that only specific centrality measures can be applied to certain network types. In the case of an R&D network, the most appropriate centrality measures are Freeman closeness, Freeman degree, and Bonacich eigenvector (Bonacich, 1972; Freeman, 1979). Highly collaborative countries are greatly prolific in output ($r=0.968$, $p=0.000$; $\rho=0.952$, $p=0.000$) and highly influential based on eigenvector ($r=0.977$, $p=0.000$; $\rho=0.997$, $p=0.000$). However, high eigenvector scores cannot be correlated with high closeness scores ($r=0.663$, $p=0.036$; $\rho=0.973$, $p=0.000$). These correlation values support the notion that centrality scores have been formulated to evaluate distinct aspects of a network.

The network map of countries on biomedical research co-authored by Southeast Asia during 2005-2009 shows an elegant depiction of centrality of Singapore and Thailand (Figure 1). Note the relatively central positions of the United States, Australia, Japan, and United Kingdom in this 2-dimensional model. Five other countries in Southeast Asia abound the periphery, i.e., Malaysia, Indonesia, Philippines, Cambodia, and Lao PDR indicating lower degree scores. Brunei, Myanmar, and Vietnam are not shown.

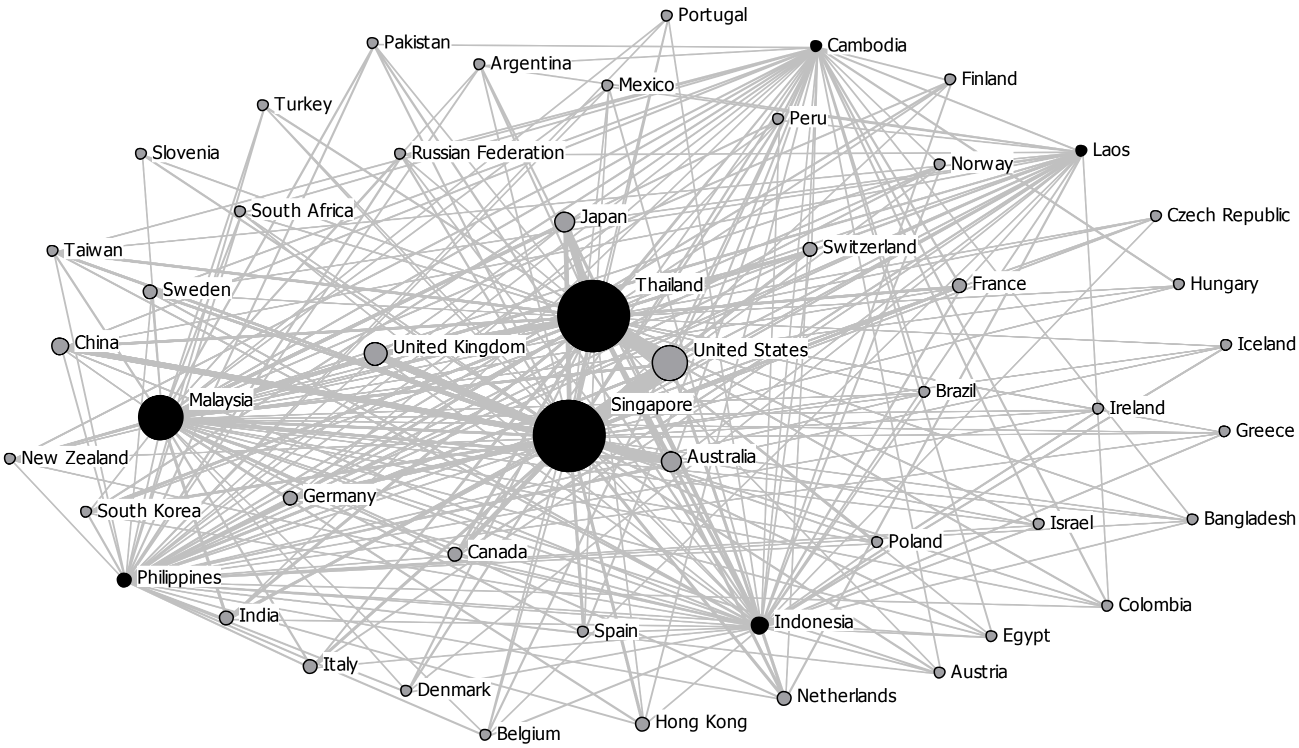

Figure 1: Network map of top collaborative countries on biomedical research (2005-2009). Only edges to or from Southeast Asia are shown. Sizes of nodes are scaled to number of publications for this dataset. Density = 4.0697, Ave Clustering = 50.783.

It is important to note that the entire dataset and its consequent subsets are Southeast Asia-centric. Thus, the node size for the United States in Figure 1, for example, reflects only its publication output within the limits of this Southeast Asian dataset. The United States and other foreign countries are assumed to have produced more biomedical output than reported here in the same 5-year period.

Thailand is the most productive during 2005-2009, and has the highest closeness centrality. In a flow context, closeness was interpreted as an index of the expected time until arrival of data flowing through the R&D network. Therefore, nodes like Thailand with high closeness scores have shorter total distances from other nodes, and they can receive data flows sooner. Thailand is well-positioned to obtain/receive and share/give novel information early in the process.

Singapore, on the other hand, has the highest degree centrality and eigenvector centrality. Eigenvector centrality is defined as the principal eigenvector of the adjacency matrix defining the network. In simple terms, a node has a high eigenvector score when it is adjacent to nodes that are themselves high scorers. The idea is that even if a node influences just one other node, who subsequently influences many other nodes (who themselves influence still more others), then the first node in that chain is highly influential. Thus, Singapore is the most influential country in Southeast Asia in the field of biomedical research.

An important research question for this study is to determine if Southeast Asian nations are cooperating with each other, or does each country pursue diverging motives in biomedical research. For instance, have Singapore and Thailand, being important knowledge hubs in the regional network map, created a viable communication channel in between? The dynamics of regional cooperation in biomedical research can be gleaned through general patterns of co-authorship of countries (Table 4).

Table 4: Top countries for transnational co-authorship on biomedical research (2005-2009).

BIOMEDICAL RESEARCH	Degree Score	Countries with most co-authorships (no. of collaborations)					
Singapore	9,995	US (2279)	AU (979)	GB (967)	CN (706)	DE (441)	
Thailand	8,384	US (2081)	JP (878)	GB (798)	AU (561)	FR (269)	
Malaysia	2,912	GB (286)	US (265)	AU (249)	SG (177)	IN (171)	
Indonesia	1,880	US (254)	JP (216)	NL (173)	AU (144)	GB (113)	
Philippines	1,189	US (190)	JP (84)	AU (72)	GB (62)	TH (59)	
Cambodia	620	US (88)	FR (74)	TH (51)	AU (38)	GB (32)	
Lao PDR	348	TH (61)	GB (51)	US (33)	FR (33)	CH (25)	JP (25)
Myanmar	226	AU (27)	JP (27)	US (15)	TH (13)	GB (10)	IN (10)
Vietnam	133	JP (15)	US (10)	AU (9)	SE (8)	TH (8)	
Brunei	97	SG (18)	GB (12)	TW (11)	US (11)	MY (9)	

Transnational co-authorship is common among Southeast Asian nations, but largely to foreign scientists. Overall, United States collaborators are the most frequent partners. Malaysia continues its partnership with United Kingdom (GB) in biomedical research, while Indonesia maintains its Dutch (NL) roots. The French (FR) connection in Indochina is still evident in Cambodia and Lao PDR. Elsewhere, Australia (AU) and Japan (JP) command significant influence in Southeast Asia. China (CN), Germany (DE), India (IN), Switzerland (CH), Sweden (SE), and Taiwan (TW) are interesting additions to the scientific domination of the United States in the region. Based on Table 4 and Figure 1, the communication channel between Singapore and Thailand remains weak, and the primary conduit, at least in biomedical research, between the two countries is through the United States.

Neoplasm Research

Bibliometrics has found several applications in exploring the scope of cancer and neoplasm research. Cambrosio, Keating, Mercier, Lewison, and Mogoutov (2006) demonstrated through network analysis the consolidation of basic cancer research and clinical research concepts during 1980-2000. Glynn, Chin, Kerin, & Sweeney (2010) showed that among publications about neoplasms in 2007, the most frequent output pertain to breast, prostate, lung, and intestinal cancer, and leukemia. These illnesses were also positively correlated to their rate of prevalence. It has also been reported that North America and Europe were responsible for more than three-fourths of published oncological articles and abstracts during the period of 2000-2006 (Tas, 2008). It is therefore interesting to explore how smaller regions like Southeast Asia have contributed to the global understanding of cancers and neoplasms.

The subnetwork of Southeast Asian countries on neoplasm research comprise about 8.19% of the total dataset for biomedical research in 2005-2009, and reflects a trend similar to the overall biomedical network. In particular, high levels of output by Thailand and Singapore correspond to high instances of collaboration in the field ($r=0.936$, $p=0.000$; $\rho=0.967$, $p=0.000$). The amount of published output is also positively correlated with closeness ($r=0.934$, $p=0.000$; $\rho=0.983$, $p=0.000$) and eigenvector ($r=0.940$, $p=0.000$; $\rho=0.996$, $p=0.000$).

At the institutional level of aggregation, the cancer research network comprises 336 institutions within Southeast Asia and 577 foreign institutions. Large institutions in Singapore such as the National University, Genome Institute, and the National Cancer Center have the most instances of collaboration. Several other agencies, universities, and medical centers from Singapore, Thailand, and Malaysia comprise the most productive institutions in the region. High publication output is strongly correlated with the frequency of collaboration ($r=0.922$,$p=0.000$; $\rho=0.791$, $p=0.000$), moderately correlated with eigenvector influence ($r=0.768$, $p=0.000$; $\rho=0.630$, $p=0.000$) and weakly correlated with the closeness score of an institution ($r=0.362$, $p=0.000$; $\rho=0.705$, $p=0.000$).

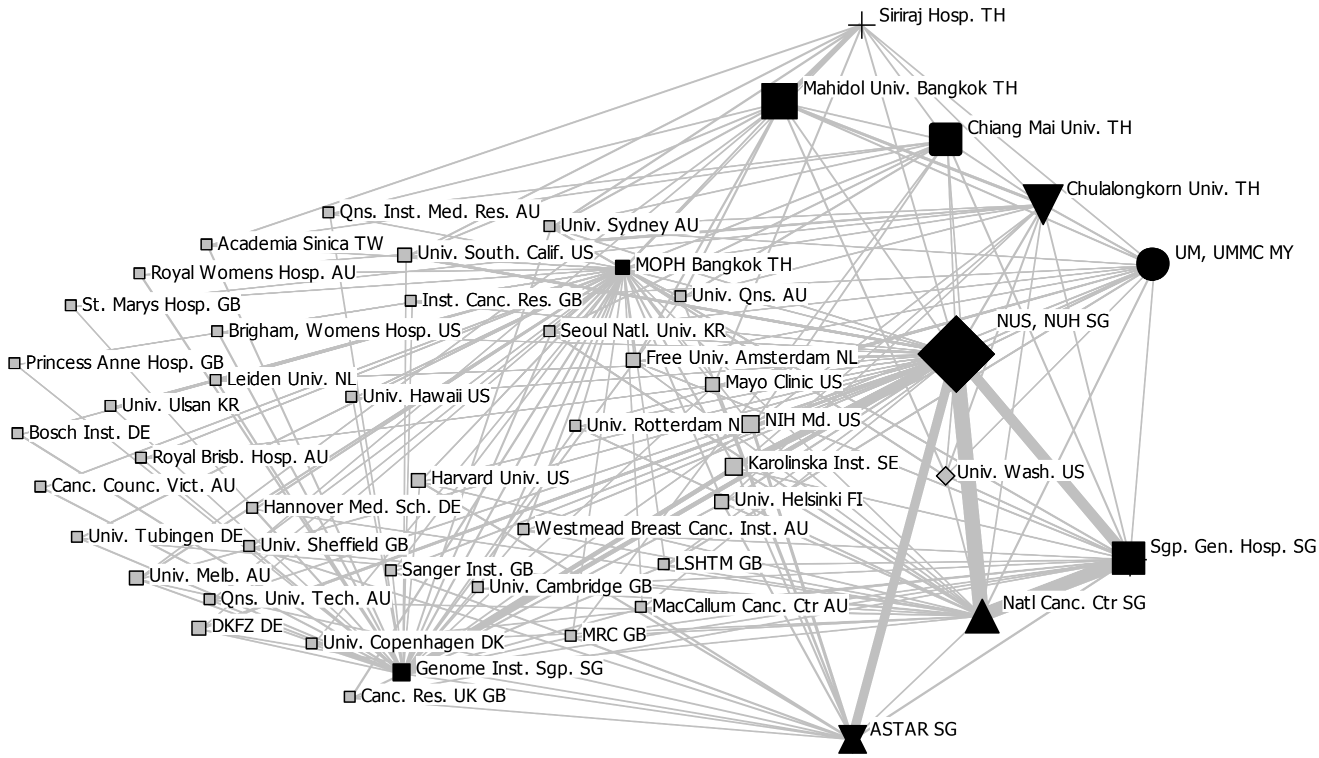

Figure 2: Network map of top collaborative institutions on cancer and neoplasm research (2005-2009). Only edges to or from Southeast Asia are shown. Sizes of nodes are scaled to number of publications for this dataset. Density = 0.0385, Ave Clustering = 2.360, and Girvan-Newman Q on 10th Partition = 0.022.

Only 36.80% of institutions in this network are located within Southeast Asia (Figure 2), and a large cluster is formed by institutions outside the region (Girvan-Newman Q=0.022)(Girvan & Newman, 2002; Borgatti, 2002; Newman, 2006). Local institutions exist in the periphery of the research network. The Karolinska Institute in Sweden, British Council Cancer Agency in Canada, and German Cancer Research Center in Germany are the most frequent foreign partners of Southeast Asia in neoplasm research.

Diabetes Research

In a study by Lewin (2008), publication patterns on diabetes mellitus showed an overall positive growth from 1984 to 2005, which paralleled the increase of diagnosed cases of diabetes patients. In Southeast Asia, diabetes research comprises 822 bibliographic records (2.41% of total output) for 2005-2009. In this period, Thailand, Malaysia, and Singapore are the most productive. Publication output of a country is correlatable to number of collaborations (r=0.899, p=0.000; ρ=0.936, p=0.000) and group influence (r=0.876, p=0.001; ρ=0.927, p=0.000), but not to closeness (r=0.667, p=0.035; ρ=0.973, p=0.000). Patterns of collaboration in diabetes research reveal that Singapore and Malaysia have the most visible partnership within the region. Australia is closest to Singapore, while the United States is closest to Thailand.

At the institutional level, the subset on diabetes research involves a total of 449 institutions, 211 of which are in Southeast Asia. The National University of Singapore, Chulalongkorn University, and University of Malaya are among the most productive. Published output is strongly correlated with degree centrality (r=0.886, p=0.000; ρ=0.779, p=0.000), moderately correlated with eigenvector (r=0.668, p=0.000; ρ=0.615, p=0.000), and weakly correlated with closeness (r=0.422, p=0.000; ρ=0.690, p=0.000).

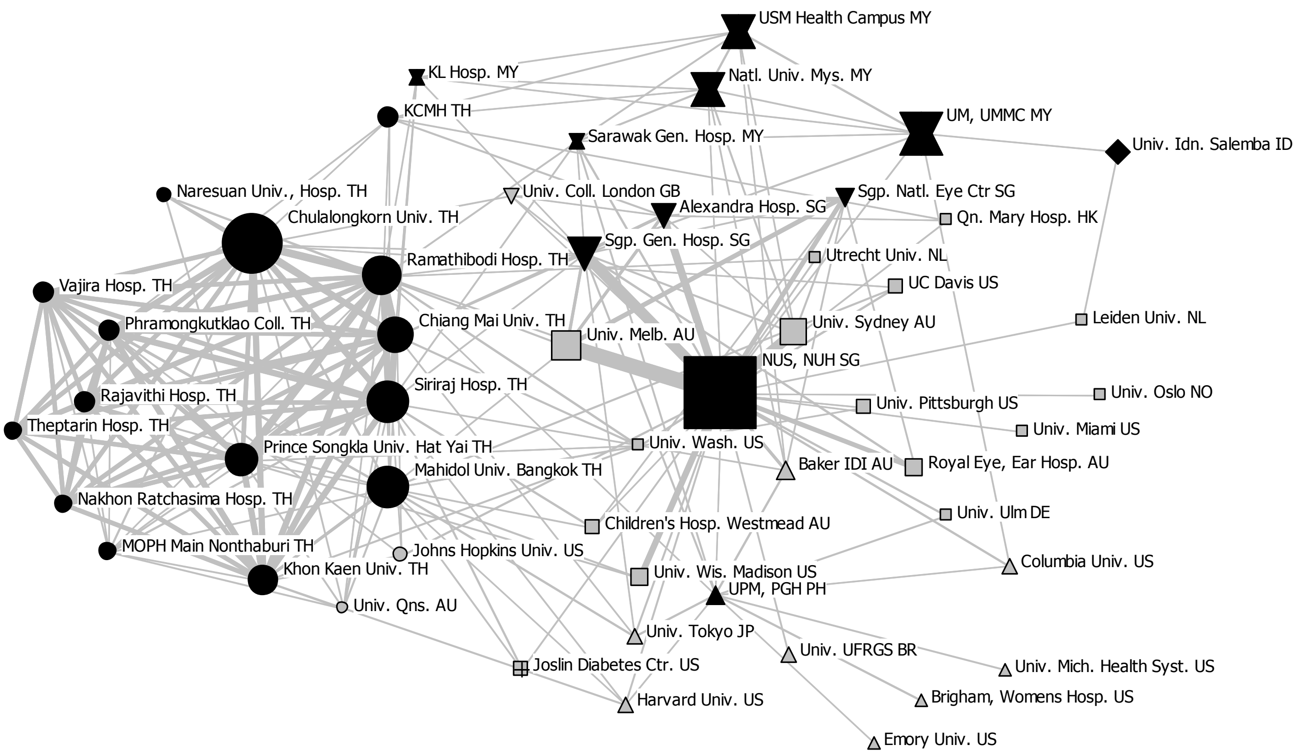

Figure 3: Network map of top collaborative institutions on diabetes research (2005-2009). Only edges to or from Southeast Asia are shown. Sizes of nodes are scaled to number of publications for this dataset. Density = 0.0283, Ave Clustering = 1.925, and Girvan-Newman Q on 7th Partition = 0.378.

Particularly, institutions in Thailand appear to have the greatest group influence in the diabetes research network (Figure 3). A tight cluster of Thai medical centers and universities is observed to form a heavy lattice in one portion of the network map. Smaller clusters exist for institutions in Malaysia, Singapore, and the Philippines. The University of Melbourne and University of Sydney are the most frequent foreign collaborators in diabetes research.

Infectious Disease Research

A few general trends in public health research have been generated in previous studies using bibliometric approaches. For instance, Western Europe and the United States were shown to have had the greatest impact on parasitic research for 1995-2003 (Durando, Sticchi, Sasso, & Gasparini, 2007; Falagas, Papastamataki, & Bliziotis, 2006; Glover & Bowen, 2004). Meanwhile, research production in Asia, excluding Japan, remained low despite the major incidence of parasitic diseases in some areas. Takahashi-Omoe, Omoe, & Okabe (2009) surveyed infectious disease research in Asia specifically, and found that Japan, China, India, and Taiwan were greatly productive for the period 1998-2006. It is important to quantify the frequency of output and co-authorship on infectious disease research by Southeast Asian countries, given that the combined regions of Southeast Asia and Western Pacific have the largest mortality rate in infectious and parasitic diseases next to Africa (WHO, 2008).

It is observed that the subset of infectious disease research during 2005-2009 comprise 4,420 bibliographic records and 120 countries. The largest research producer and most frequent collaborator is Thailand. High output is strongly correlated with high degree ($r=0.988$, $p=0.000$; $\rho=0.961$, $p=0.000$) and eigenvector centrality ($r=0.976$, $p=0.000$; $\rho=0.861$, $\rho=0.001$), and weakly correlated with closeness centrality ($r=0.765$, $p=0.010$; $\rho=0.912$, $p=0.000$). In terms of co-authorship, the United States, United Kingdom, and Japan provide the most instances of collaboration. Within the region, Thailand is a significant partner to six of its neighbors, suggesting a central role for Thailand in infectious disease research in the Southeast Asia.

At the institutional level of aggregation, the network is dominated by universities and research agencies in Thailand, e.g., Mahidol University, Chulalongkorn University, and Chiang Mai University. The Centers for Disease Control and Prevention in the United States and Churchill Hospital in the United Kingdom are the most frequent foreign partners in infectious disease research. Research productivity in the field is strongly correlated with degree centrality ($r=0.950$, $p=0.000$; $\rho=0.796$, $p=0.000$), moderately correlated with eigenvector ($r=0.799$, $p=0.000$; $\rho=0.635$, $p=0.000$), and weakly correlated with closeness ($r=0.337$, $p=0.000$; $\rho=0.653$, $p=0.000$). As mentioned above, distances between nodes in the infectious disease research network are the shortest, and has the highest likelihood of cohesion compared to other institutional subsets.

Malaria Research

In previous bibliometric studies on malaria research, Thailand was identified as one of the leading countries on malaria research during 1996-2000 (Lewison, Lipworth, & De Francisco, 2002) and in 1990 and 2000 (Garg, Dutt, & Kumar, 2006). Mahidol University was also named as one of the most prolific producers globally during the same period. The dataset confirmed these observations, which are important to funders particularly in directing future resource allocations for malaria research.

The subset on malaria research in 2005-2009 comprises 2.78% of the total biomedical output of the region. At the country level of aggregation, malaria research has the shortest average distance and greatest cohesion among all datasets. Thailand and Indonesia lead the region in terms of output and collaboration, while Cambodia and Lao PDR have the third and fourth highest instances of co-authorship. High research output is strongly correlated with degree ($r=0.989$, $p=0.000$; $\rho=0.863$, $p=0.003$) and eigenvector ($r=0.995$, $p=0.000$; $\rho=0.858$, $p=0.003$), and weakly correlated with closeness ($r=0.879$, $p=0.002$; $\rho=0.725$, $p=0.027$). The United Kingdom and United States provide the most instances of cooperation in malaria research. Within the region, Thailand is an important collaborator of Indonesia, Cambodia, Lao PDR, Singapore, Malaysia, and Myanmar, thus assuming a central position in this network.

Mahidol University is the most important node at this institutional level of aggregation. The National Institute of Health Research and Development and US Naval Medical Center Research Unit No. 2 in Indonesia, Armed Forces Research Institute for Medical Sciences in Thailand, and Institut Pasteur in Cambodia also comprise the most collaborative institutions in malaria research. High research output in this dataset is strongly correlated with degree ($r=0.974$, $p=0.000$; $\rho=0.773$, $p=0.000$) and eigenvector ($r=0.945$, $p=0.000$; $\rho=0.710$, $p=0.000$), and weakly correlated with closeness ($r=0.337$, $p=0.000$; $\rho=0.684$, $p=0.000$). The most frequent foreign collaborators in malaria research are the University of Oxford and Churchill Hospital in the United Kingdom, and Charles Darwin University in Australia.

Tuberculosis Research

An overview of tuberculosis research for 1997-2006 has been previously described, where Thailand was identified as among the most prolific producers of tuberculosis research output (Ramos, Padilla, Masia, & Gutierrez, 2008). More importantly, Ramos and his colleagues noted that countries with a greater burden of tuberculosis cases have published less research on tuberculosis, while industrialized countries such as those in Western Europe and the United States were significantly more productive.

Using the biomedical research data, Thailand is the most productive among Southeast Asian countries. High research output is strongly correlated with degree (r=0.945, p=0.000; ρ=0.815, p=0.004) and eigenvector (r=0.942, p=0.000; ρ=0.767, p=0.010), and moderately correlated with closeness (r=0.809, p=0.005; ρ=0.821, p=0.004). The United States, Netherlands, and Japan are the most frequent foreign collaborators in the tuberculosis research network.

At the institutional level of aggregation, Mahidol University is the most productive in tuberculosis research while the Ministry of Public Health in Thailand is the most collaborative institution. Frequency of output is moderately correlated with degree (r=0.853, p=0.000; ρ=0.745, p=0.000) and eigenvector (r=0.627, p=0.000; ρ=0.549, p=0.000), and weakly correlated with closeness (r=0.434, p=0.000; ρ=0.541, p=0.000).

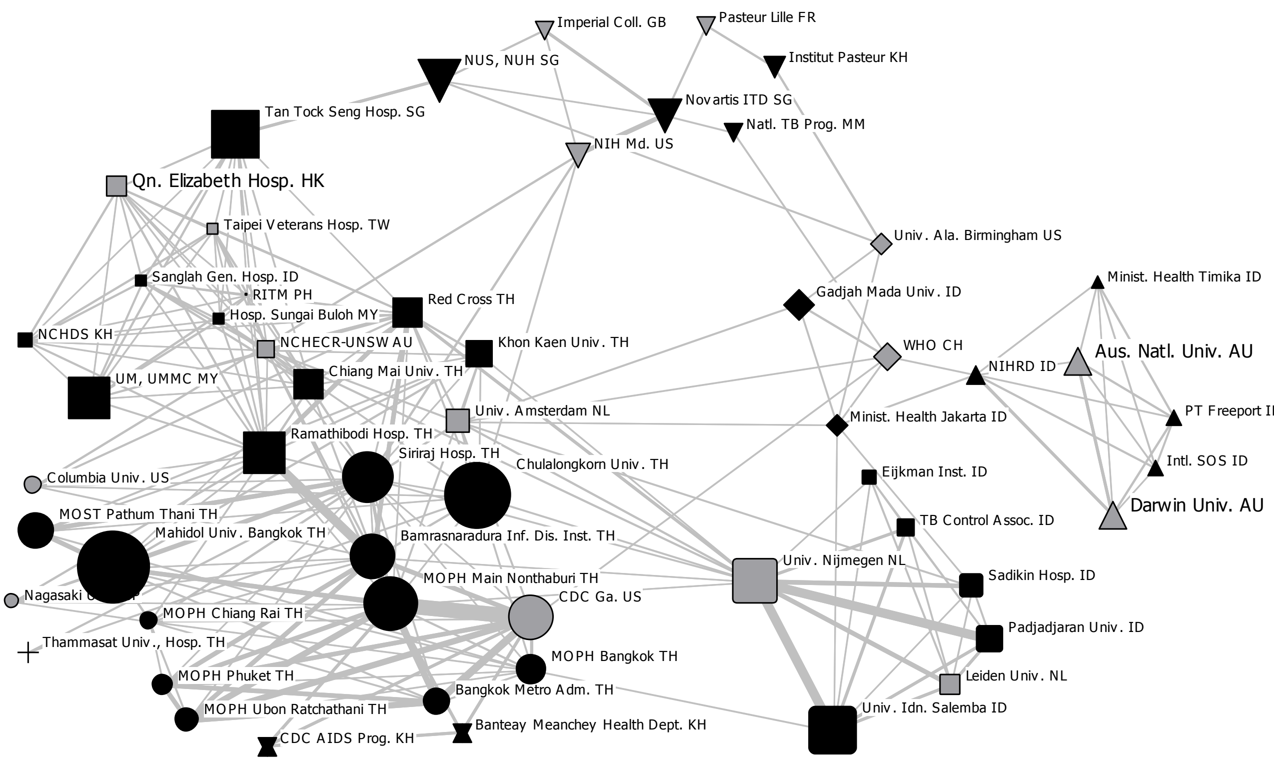

Figure 4: Network map of top collaborative institutions on tuberculosis research (2005-2009). Only edges to and from Southeast Asia are shown. Sizes of nodes are scaled to number of publications for this dataset. Density = 0.0195, Ave Clustering = 1.513, and Girvan-Newman Q on 8th Partition = 0.473.

Radboud University Nijmegen and University Medical Center in the Netherlands and Centers for Disease Control and Prevention in the United States are the most frequent foreign collaborators in the tuberculosis research network. The network map of institutions in tuberculosis research displays a distinctive clustered layout (Figure 4).

One cluster adjoins Dutch and Indonesian institutions together, while another subgroup comprises institutions in Australia and Indonesia. Two larger clusters are mediated primarily by medical centers, universities, and research agencies in Thailand.

The patterns of co-authorship illustrated here will be constantly affected by a number of factors influencing the decision of scientists to collaborate. At the country level of aggregation, co-publication patterns may reveal the political affinity and economic status of a country (Glanzel, 2001), such as neo-colonial ties described above. At the institutional level, collaboration may be influenced by the need for multiple study sites, such as for clinical field trials. At a personal level, a scientist may decide to collaborate with his past affiliations and academic peers. Another reason of a scientist to opt for joint research is the need for multiple disciplines and expertise to interpret separate portions of his research project.

Indeed, scientific collaboration is an elaborately complex phenomenon, where personal motivations of authors overlap with research opportunities and gaps existing in their institution and country. The patterns of cooperation reported in this study will change over time. Nonetheless, the combination of network analysis tools and bibliometric methods which was demonstrated in this paper will be useful as an objective methodological framework for quantifying scientific collaboration in Southeast Asia in future studies.

CONCLUSIONS

Based on the results, several salient observations can be drawn. Firstly, the region as a whole collaborates most actively with American counterparts, while some Southeast Asian nations maintain preferential partnerships to specific European knowledge hubs. Geographical distance, therefore, is not a limiting factor. Neither does regional proximity assure of closer collaborations. Secondly, the central role of Singapore and Thailand is common in biomedical research and its subsets. Thailand is distinctly central in the domains of infectious disease research. Thirdly, distinct institutional clusters or subgroups have formed within networks. A large component of foreign institutions is present in neoplasm research. Tight cooperation among Thai researchers exists in malaria research, while preferential associations of Indonesia to the Netherlands and Australia are described in tuberculosis research. Lastly, the bibliometric output of a country or institution is strongly correlatable with collaboration count and group influence, but weakly correlatable with maximal data flow. There is an emerging need to increase joint research work among Southeast Asian countries and institutions, in order to deepen our understanding of local health problems that are specific to the region. A collaborative atmosphere at a closer perspective may hasten innovative and locally-adapted solutions for reducing ill health and improving the quality of life in our respective communities.

COMPETING INTERESTS

The author declares having no competing interests.

ACKNOWLEDGMENTS

Sciverse Scopus data have been previously downloaded as part of a completed project funded by the Special Programme for Research and Training in Tropical Diseases (TDR) of the World Health Organization (WHO). The author would like to thank Dr. Bernadette Ramirez of WHO TDR, Dr. Jaime Montoya & Ms. Carina Rebulanan of the Philippine Department of Science and Technology, and Dr. Alan Feranil of the University of San Carlos, Cebu, Philippines for their technical advice.

REFERENCES

Barabasi, A., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications, 311*(3-4), 590-614.

Batagelj, V., & Mrvar, A. (1998). Pajek - Program for Large Network Analysis. *Connections, 21*(2), 47-57.

Bonacich, P. (1972). Factoring and Weighting Approaches to status scores and clique identification. *Journal of Mathematical Sociology, 2*(1), 113-120.

Borgatti, S. (2002). *Netdraw Network Visualization*. Harvard, Massachusetts, United States: Analytic Technologies.

Borgatti, S. (2005). Centrality and network flow. *Social Networks, 27*(1), 55-71.

Borgatti, S., Everett, M., & Freeman, L.C. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Harvard, Massachusetts, United States: Analytic Technologies.

Cambrosio, A., Keating, P., Mercier, S., Lewison, G., & Mogoutov, A. (2006). Mapping the emergence and development of translational cancer research. *European Journal of Cancer, 42*(18), 3140-3148.

Durando, P., Sticchi, L., Sasso, L., & Gasparini, R. (2007). Public health research literature on infectious diseases: coverage and gaps in Europe. *European Journal of Public Health, 17*(suppl 1), 19-23.

Elsevier. (2010). *Sciverse Scopus*. Retrieved from http://scopus.com.

Falagas, M.E., Papastamataki, P.A., & Bliziotis, I.A. (2006). A bibliometric analysis of research productivity in Parasitology by different world regions during a 9-year period (1995–2003). *BMC Infectious Diseases, 6*(1), 56.

Falagas, M., Pitsouni, E., Malietzis, G., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB Journal, 22*(2), 338-34.

Freeman, L.C. (1979). Centrality in networks: Conceptual clarification. *Social Networks*, *1*(3), 215–239.

Garg, K., Dutt, B., & Kumar, S. (2006). A preliminary scientometric investigation of malaria research. *Annals of Library and Information Studies*, *53*(1), 43-53.

Gauffriau, M., Larsen, P., Maye, I., Roulin-Perriard, A., & von Ins, M. (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics*, *73*(2), 175-214.

Girvan, M. & Newman, M.E.J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821–7826.

Glanzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics, 51*(1), 69–115.

Glover, S., & Bowen, S. (2004). Bibliometric analysis of research published in Tropical Medicine and International Health 1996–2003. *Tropical Medicine and International Health, 9*(12), 1327-1330.

Glynn, R.W., Chin, J.Z., Kerin, M.J., & Sweeney, K.J. (2010). Representation of Cancer in the Medical Literature - A Bibliometric Analysis. *PLoS ONE, 5*(11), e13902.

Jacso, P. (2005). As we may search–Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science, 89*(9), 1537-1547.

Jarneving, B. (2010). Regional research and foreign collaboration. *Scientometrics, 83*(1), 295–320.

Kulkarni, A., Aziz, B., Shams, I., & Busse, J. (2009). Comparisons of Citations in Web of Science, Scopus, and Google Scholar for Articles Published in General Medical Journals. *Journal of the American Medical Association, 302*(10), 1092-1096.

Lewin, H.S. (2008). Diabetes mellitus publication patterns, 1984–2005. *Journal of the Medical Library Association*, *96*(2), 155-157.

Lewison, G., Lipworth, S., & De Francisco, A. (2002). Input indicators from output measures: a bibliometric approach to the estimation of malaria research funding. *Research Evaluation*, *11*(3), 155-163.

Leydesdorff, L. (2008). *Scopus.exe*. Retrieved from www.leydesdorff.net/software/scopus/index.htm.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.936v1 | CC-BY 4.0 Open Access | rec: 27 Mar 2015, publ: 27 Mar 2015

Leydesdorff, L , & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, *2*(4), 317-325.

Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 404-409.

Newman, M.E.J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8577-8582.

Nwaka, S., Ilunga, T. B., Da Silva, J. S., Rial Verde, E., Hackley, D., De Vré, R.,…Ridley, R.G. (2010). Developing ANDI: A Novel Approach to Health Product R&D in Africa. *PLoS Med*, *7*(6), e1000293.

Ramos, J.M., Padilla, S., Masia, M., & Gutierrez, F. (2008). A bibliometric analysis of tuberculosis research indexed in PubMed, 1997–2006. *International Journal of Tuberculosis and Lung Disease*, *12*(2), 1461-1468.

Takahashi-Omoe, H., Omoe, K., & Okabe, N. (2009). New journal selection for quantitative survey of infectious disease research: application for Asian trend analysis. *BMC Medical Research Methodology*, 9, 67.

Tas, F. (2008). The contribution of countries and world regions in productivity of oncological publication. *Annals of Oncology*, *19*(11), 1962–1968.

Travers, J. & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, *32*(4), 425-443.

UNESCO. (2010). *UNESCO science report 2010*. Paris, France.

WHO. (2007). I*nternational statistical classification of diseases and related health problems, 10th revision, version for 2007*. Geneva, Switzerland.

WHO. (2008). *The global burden of disease: 2004 update*. Geneva, Switzerland.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.936v1 | CC-BY 4.0 Open Access | rec: 27 Mar 2015, publ: 27 Mar 2015