Cooperation in health: A cluster analysis of 190 research institutions

Pamela Barreto Lang*, Fabio C. Gouveia¹, Jacqueline Leta²

Coordination of Social Communication, Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil, ¹Museum of Life, Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil, ²Medical Biochemistry Institute, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

ABSTRACT

Introduction: Webometrics has been recognized as a science with great potential for application in various fields. However, studies in this area are still restricted to the large field of informetrics. Moreover, the debate surrounding the relation between the web and the health field has been dedicated to addressing issues of socioeconomic nature related to universal access and a new division between north and south. Objectives: With the aim of contributing to the broadening of webometrics to other areas of knowledge, this study intends to map on the web, in the light of international cooperation models, relations between major health research institutions. Results: Results showed predominantly thematic groupings, the reflection of a north-south cooperation model and a low expression of relations between institutions of south countries.

Keywords: Cooperation in health, webometrics, World Health Organization

INTRODUCTION

The web has been widely recognized as a valuable source of information. Studies that investigate the structure and the different uses of the web are increasingly frequent and being applied to different areas. Though considered a recent research field, webometrics has focused on exploratory studies, mainly on academic websites with the purpose of testing and improving methods for collecting, processing, and analyzing data.

A prominent research and an enthusiast of the idea of applying webometrics to fields outside informetrics, Thelwall has argued in recent years the potential webometric studies to the field of Social Sciences, Communication and Health, where some researches have already shown clear applications. Particularly in the field of health, Thelwall found a significant gap, considering it a promising area for the development of future webometric studies.

Aiming to contribute to the expansion of webometrics as an applied science, this study intends to map on the web the relationship between the main institutions of health research in the world. The sample was selected based on the World Health Organization (WHO) Collaborating Centers. Currently, there are about 900 collaborating centers, distributed across more than 90 countries and six regions where WHO maintains offices: Western Pacific: 21%; Americas: 21%; Southeast Asia: 10%; Eastern Mediterranean: 6%; Africa: 4%; Europe: 37%. In the Americas, the largest concentration is in the U.S., with 99 centers, followed by Canada with 25 and Brazil with 21 centers.

METHODOLOGY

Dataset

In October 26, 2009, information regarding the 768 active collaborating centers at the time was retrieved from WHO’s database. A total of 89 countries were represented.
in the list. For each collaborating center, WHO’s database made the following information available: Name (of the collaborating center), theme (of the collaboration), contact, institution, address, city, country, designation date, last designation, and website.

Due to some inconsistencies and given the focus of the study, the list of collaborating centers was revised and many centers were excluded. This occurred, for example, when the website provided did not match the name of the institution appointed by WHO or when the website had changed or ceased to exist. Collaborating centers that are not exclusively dedicated to the health field, such as universities, for example, were also excluded from the sample. Previous researches have shown evidence that the motivations for the generation of links between university websites are much broader than scientific collaboration (Seeber et al., 2012).

As this study has an institutional focus, departments or organizational structures designated as collaborating centers, but lacking their web domains, were excluded. In this case, the option was to select the websites of the institutions to which they belonged. The final list for data collection consisted of 354 institutions (357 websites, considering that some institutions had more than one website) from 52 countries.

**Data Collection**

Data collection occurred between November 7th and 9th 2009. Interlinks between all 357 websites were collected using the software LexiURL Searcher (now known as Webometric Analyst), which relied mostly upon hyperlink data from Yahoo! Search API (http://lexiurl.wlv.ac.uk). For interlinking analysis, an asymmetric matrix was generated.

At this stage, the studied centers list was, once again, reduced by using the adhesion to the sample criterion. The used filter excluded the centers whose sum of the line number was lower than the total number of centers of the sample (n) divided by two. This process was done successively until 190 centers remained, representing 42 countries.

**Clustering Analysis**

For better understanding, the web relation through the interlinking process, clustering analysis was performed at the institutional level. Researchers use different values to fill the asymmetric matrix diagonal. For the present study, the sum of each line was used, an approach proposed by Gouveia and Kurtenbach. Statistica 7.0 (a statistics and analytics software package developed by StatSoft) was used for data treatment and Ward’s method was applied for amalgamation, with 1-Pearson r distance measures. In order to obtain more information about the institutions and possible clustering motivations, all 190 websites as well as additional co-related websites were accessed.

**RESULTS AND DISCUSSION**

The 1.55 distance measure sets the interlink analysis results apart in 30 clusters [Figure 1]. Half of them are formed entirely by institutions from the same country. In this case, local collaborations in specific research themes seem to be fairly stronger than international ones, most likely due to their common research themes, as institutions are engaged in solving local health problems or play an important role in the country as governmental agencies. This phenomenon is observed in the following 15 clusters.

Cluster 1 is formed only by French institutions: National Research Institute for Transports and Security (INRETS-FR), National Institute for Health and Medical Research (INSERM-FR), National Institute for Agricultural Research (INRA-FR), Research Institute for Development (IRD-FR) and the Tropical Medicine Institute of the French Armed Forces Medical Service (IMTSSA-FR). Four of them are grouped into two sub-clusters, formed by INSERM and INRA, which appear strongly related, and IRD and IMTSSA with a slightly larger linkage distance.

In clusters 3 and 5, the institutions share very similar research themes. In cluster 3, two Indian institutions conduct several studies together and form a well-known partnership in Tuberculosis research: The National Tuberculosis Institute (NTI-IN) and the National Institute for Tuberculosis Research, former Tuberculosis Research Center (TRC-IN). In cluster 5, Switzerland is represented by two institutions with research collaborations in the same theme area: Addiction Suisse (AIS-CH), a Swiss institute for the prevention of alcohol and other drug problems, and Research Institute for Public Health and Addiction (ISGF-CH).

Four Netherlands institutions compose cluster 9: The Netherlands Institute for Health Promotion and Disease Prevention (NIGZ-NL), the National Institute for Public Health and the Environment (RIVM-NL), the Netherlands Institute of Primary Health Care (NIVEL-NL) and the...
Netherlands Vaccine Institute (NVI-NL). Although, it keeps a larger linkage distance from the rest, NVI has a strong relation with RIVM in real life. NVI is an agency under the responsibility of the Ministry of Health, Welfare and Sport, but its public tasks, such as research, development, procurement, storage and distribution of vaccines, have been recently transferred to RIVM. The latter, however, form a subcluster with another long term partner: NIVEL. The institutes have developed various studies such as on people with chronic conditions as well as on virology (surveillance of respiratory virus infection), on monitoring methodology on the effects of emergencies on public health, and on changes in morbidity and mortality in general practices. In 2010, NIVEL, in collaboration with RIVM, published a health system review for the European Observatory on Health Systems and Policies.

Regarding clusters 11 and 12, they are composed by two Japanese institutions each: Research Institute of Tuberculosis (RIT-JP) and the National Institute of Infectious Diseases (NIH-JP), former NIH Japan, and the National Cancer Center (NCC-JP) and the National Institute of Public Health (NIPH-JP), respectively. The fact that these websites do not have versions in other languages made a deeper analysis on clustering motivations more difficult.

In clusters composed by Chinese institutions, language also represented a barrier for further analysis. It is the case of clusters 14 and 15. The first one shows a strong web relation between the Institute of Information on Traditional Chinese Medicine and the Institute of Medical Information and Library, both integrants of the Chinese Academy of Medical Sciences. And in cluster 15 the Center for Disease Control and Prevention of Guangdong Province (CDCP-CN) and the Chinese Center for Diseases Control and Prevention (CDC-CN) appear together.

Four German institutions are grouped in cluster 18: German Institute for Medical Documentation and Information (DIMDI-DE), Friedrich-Loeffler-Institut (FLI-DE), Paul-Ehrlich-Institute (PEI-DE), Biocenter, Johann Wolfgang Goethe-University (BIO.UNI-DE). FLI and PEI, however, show a stronger web relation between
them as they form subcluster with a short linkage distance. Together with the Paul Ehrlich Institute, an Agency of the
German Federal under the Ministry of Health, FLI, a federal research institute for animal health under the Federal Ministry for Food, Agriculture and Consumer Protection, serves as the licensing and approval authority for serums, vaccines, antigens, and detection methods used in veterinary medicine.

Cluster 24 is formed by five Australian institutions: The Murdoch Children’s Research Institute (MCRI-AU), The Royal Children’s Hospital (RCH-AU), the University of Melbourne’s Key Center for Women’s Health in Society (KCWH.UNIMELB-AU), St. Vincent’s Hospital (SVHM-AU) and the Telethon Institute for Child Health Research (ICHR.UWA-AU). Except from SVHM and the Key Center for Women’s Health in Society, these Australian institutions are theme related with their research focus on child health and that may be the main reason for their linkage distance. The RCH and KCWH present the strongest relation in the cluster. In fact, RCH’s Center for International Child Health currently collaborates with the latter on training and education.

Cluster 22 is formed by three Italian theme related institutes: The Person Centered Approach Institute (IACP-IT), the National Institute for Occupational Safety and Prevention (ISPESL-IT) and the Salvatore Maugeri Foundation (FSM-IT). In this case, clustering motivations seem to be once more theme related as they all have been collaborating in the area of occupational health, being IACP Italy’s focal point for the International Labour Organization (ILO) on Safety and Health at Work and the Environment.

Another reason, however, for clusters being composed entirely by institutions from the same country may be linked to structural issues, whether it is web-related or only institutional-related, as observed in the four clusters below:

The University of Wisconsin School of Medicine and Public Health (MEDSCH.WISC-USA) and the Pain and Policy Studies Group, University of Wisconsin Carbone Cancer Center (PPSG.WISC-USA) compose cluster 16. In this case, as they share the same web domain, it is difficult to separate research motivations from web structure itself. The same phenomenon can be observed in cluster 17 is formed by the Education Development Center (EDC-USA) and the Health and Human Development Programs (HHD-USA), being the latter a division of EDC and probably motivating the web aggregation.

Cluster 21 is formed by the University of Texas Medical Branch (UTMB-USA) and two other centers of the same university: The Sealy Center on Aging (SCA.UTMB-USA) and the Center for International Health (CTIH.UTMB-USA). The reason they were not consolidated into UTMB-USA was that both centers had their own web domains and have a very diverse research focus. They ended up all together in the same cluster, probably because the structural links between them were stronger than an external collaboration.

Finally, cluster 29 is composed only by John Hopkins’ institutions: Bloomberg School of Public Health (HSPH-USA), Johns Hopkins Medical Institutions (JHMI-USA), The Johns Hopkins Hospital and Health System (HOPKHOSP-USA), Johns Hopkins Medicine (HOPKMED-USA). It is important to observe that many of the countries cited above are also represented in other clusters characterized by international collaborations, meaning that some research themes, due to their local importance, tend to motivate collaborations within the same country. In this context, the 15 other clusters presented represent international collaborations. When it comes to international collaboration in health, socioeconomic division between the economically developed, industrialized countries, collectively known as the north, and the low- and middle-income countries, known as the south, characterizes an important debate in the area and represents a major challenge in global health.

Though the north-south relation has been criticized over the years for creating an unidirectional dependence, in which the process of high-end technology transfer does not generate the infrastructures needed for the development of the local health system and health policies, this relation dynamics are still common in many cooperation programs. On the web, north-south collaboration was predominant, being present in 12 clusters.

Belgium and Thailand are present in cluster 2, represented by the Walloon Agriculture Research Center (CRA-BE) and the Asian Institute of Technology (AIT-TH), respectively.

The Institute of Psychiatry (IOP.KCL-UK), from King’s College in the UK, has a few ongoing or recently concluded projects with the Institute for Population and Social Research (IPSR.MAHIDOL-TH), at Mahidol University
in Thailand. During the period of the year between 1999 and 2004, researchers from the IPRS collected information about population change in an annual census of a large sample of households in Kanchanaburi province for a study conducted by IOP on depression, disability, and socio-economic position among older adults “left behind” by out-migration. The Thai Institute has also advised on the design of a study of adolescent health conducted in Zhejiang Province, China. The 3 years project started in 2004 and concluded in 2009. Both institutions are grouped together in cluster 7.

Eleven institutions from eight different countries are grouped in cluster 10: National Administration of Laboratories and Health Institutes (ANLIS-AR); National Institute for Health and Welfare (THL-FI); International Center for Pesticide Safety and Health Risk Prevention (ICPS-IT); Canadian Center for Occupational Health and Safety (CCOHS-CA); Finnish Institute of Occupational Health (TTL-FI); Federal Institute of Occupational Safety and Health (BAUA-DE); NRW Institute of Health and Work (LIGA-DE); National Institute of Occupational Safety and Health (JNIOSH-JP); Foundation on Occupational Safety and Health Researches and Studies (FUNDACENTRO-BR); National Research and Safety Institute (INRS-FR); Institute for Work and Health (IST-CH).

However, this cluster comprises five subclusters with relatively weak web relations, considering their linkage distance. The strongest subcluster is formed by BAUA and LIGA, both from Germany. Since January 1, 2012, LIGA no longer exists and its tasks were transferred to newly established institutions: Landesinstitut für Arbeitsgestaltung des Landes Nordrhein-Westfalen - LIA.NRW (http://www.lia.nrw.de) and Landeszentrum Gesundheit Nordrhein-Westfalen - LZG.NRW (http://www.lzg.ge.nrw.de).

It is also worth noticing that the National Institute for Health and Welfare (THL) is a research and development institute under the Finnish Ministry of Social Affairs and Health. THL opened in January 2009 after the merger of the National Public Health Institute (KTL) and the National Research and Development Center for Welfare and Health (Stakes). Web information on three sites was collected and interlinks from Stakes and KTL were consolidated into THL.

Cluster 23 is formed by six institutions, from five different countries (Thailand, Sweden, Uganda, Colombia and China): Armed Forces Research Institute of Medical Sciences (AFRIMS-TH), Institute of Biomedicine (BIOMED-SE), Injury Control Centre Center (ICC-UG), Karolinska Institute (KI-SE), Institute for Peace Promotion and Injury/Violence Prevention (CISALVA.UNIVALLE-CO) and Institute of Materia Medica (IMM-CN). The strongest web relation in this cluster is between KI and CISALVA. The KI is the largest research institution in Sweden and plays a major role in international collaboration. However, it is very clear that the web linkage distance in this case is related to the WHO, as they are both collaborating centers on Community Safety Promotion, being CISALVA directly affiliated to KI for their newly created Safe Community Support Center. Both websites exchange a considered number of links recognizing each other's work in the area.

It is also worth observing that the Injury Control Centre (ICC-UG) web links show an important connection with Karolinska Institut. This web relation may be a reflection of the existing research collaborations between KI's Department of Public Health Sciences and ICC, who plays a distinguished role in the region by being designated as the Secretariat for Injury Prevention Initiative for Africa (IPIFA), a non-governmental agency founded in 1997 that unites 13 African countries.

Cluster 27 aggregates the Centers for Disease Control and Prevention (CDC-USA), the National Sanitation Foundation (NSF-BE), the Center for Public Health (CPH-UK), the National Institute of Epidemiology (NIE-IN), the Clinical and Laboratory Standards Institute (CLSI-USA) and the National Institute of Communicable Diseases (NICD-IN). Two subclusters can be clearly identified in this group: CDC and NSF; CPH and NIE, which shows that, in this case, research collaborations may represent a stronger motivation than geography for uniting these institutions. NIE is the sole institution designated by WHO as a collaborating center on Leprosy Research and Epidemiology and the Institute.

Cluster 19 groups three Spanish-speaking countries: The Center for the Development of Evaluation and Technology in Public Health (CEDETES-CO), the National Institute of Public Health (INSN-MX) and the Carlos III Health Institute (ISCIII-ES). In this case, it was not clear whether linguistics or north-south relation played a more important role. One could speculate that, besides the possibility of thematic motivations, there may be a linguistic motivation for generating links between them. Gouveia and Kurtenbach[10] investigated the relationship
between museums and science centers in Latin America and observed a clear isolation of Brazilian museums due to a language barrier. It is possible, therefore, that Spanish bringing these institutions together.

Singapore and Cyprus are represented by single institutions: The National Cancer Center Singapore (NCSS-SG) and the Thalassaemia International Federal (TIF-CY). Though TIF has been in official relations with the non-communicable Diseases/Human Genetics Department of the WHO since 1996 and involves 108 national thalassaemia associations and other members from over 55 countries across the world, its web connection seems very limited and it is not reflecting the Scopus of its international collaborations. Both institutions provide links to NIH-USA and NCSS-SG also links to SPH. EMORY-USA. On cluster analysis, TIF is connected to other American institutions in cluster 28. NCSS appears with Rollins School of Public Health and the Department of Emergency Medicine (School of Medicine), both from Emory University in the United States, which may be the strongest reason they appear together in cluster 6.

Except from one Swiss institution, cluster 25 comprises five south institutions, each from a different country: Rosario Center of Perinatal Studies (CREP-AR), Geneva Foundation for Medical Education and Research (GFMER-CH), National Institute for Research in Reproductive Health (NIRRH-IN), National Family Planning Coordination Board (BKKBN-ID) and Shanghai Medical College of Fundan University (SHMC. FUNDAN-CN). Aside from the latter, they are all WHO Collaborating Centers in human reproduction. Therefore, theme-related. as for Shanghai Medical College, a number of reasons could explain its presence in this cluster. One possible speculation is the collaboration with the Shanghai Institute of Planned Parenthood Research (SIPPR), both of Fundan University. The SIPPR established in 1978 was the first comprehensive research institute in family planning and reproductive health research in China. The Institute was first designated as a WHO Collaborating Center for Research in Human Reproduction in 1983 and has become the largest WHO Collaborating Center among all WHO/Human Reproduction collaborating centers in the Asia-Pacific region. Despite its relevance to the field, SIPPR was not included in the sample because it lacks self-web domain.

And cluster 26, aside from Argentina, represented by one institution (National Scientific and Technical Research Council (CONICET-AR), brings together Italy, Poland, and Sweden by gathering four institutions: The Center for the Evaluation of the Effectiveness of Health Care (CEVAS-IT), Nofer Institute of Occupational Medicine (IMP-PL), the National Institute of Health (ISS-IT) and the Uppsala Monitoring Center (UMC-SE). Despite the presence of two Italian institutes, geographical motivations do not seem to play any role in this cluster. The strongest web relation occurs between ISS and IMP.

Despite a slow progress on social justice and poverty reduction over the last decades, the Millennium Development Goals have achieved some important outcomes in public health and many programs have been created, such as the Union of South American Nations (UNASUL), in which Brazil plays a strategic role, the Asia Pacific Academic Consortium for Public Health (APACPH), with members from East and South Asia, and the African-South American (ASA).

However, emerging south-south collaborations involving donor countries such as India, South Africa, Malaysia, Korea, and China are still not reflected on the web. Only one cluster is entirely composed by south countries, being two Brazilians and a Cuban institute: The National Cancer Institute (INCA-BR), the Oswaldo Cruz Foundation (FIOCRUZ-BR) and the National Institute of Hygiene, Epidemiology, and Microbiology (INHEM-CU). Fiocruz is the major link between the two other institutes as it maintains cooperation projects with both of them. With INCA, Fiocruz has a well-known cooperation history in research projects, being both linked to the Ministry of Health. In 2005, the Brazilian government established the National Policy for Oncology Care and the two institutions gathered together through a Cooperation Program to face the challenges related to developing cancer research. The program includes five research networks in Evaluation of Oncology Care, Diagnostic and Therapeutic Implementation, Identification of Markers for the Diagnosis and Prognosis, Development of Therapeutic Technologies and Identification of Genetic and Environmental Risk Factors. As for INHEM, Fiocruz and the Cuban institute have been working together since 2010 to establish an official cooperation between Cuba and Brazil in health, work and environment.

The largest cluster on this analysis is cluster 4, composed by a total of 83 institutions, including north and south countries. Many of the institutions in this cluster have a low expression considering the number of links
distributed, which may be the reason why they are gathered in the same group.

As for north-north collaborations, there are two clusters composed by economically developed countries. One that gathers in cluster eight Belgium and the US. The Center for Research on the Epidemiology of Disasters (CRED-BE) and the International Research Institute for Climate Prediction (IRI.COLUMBIA-USA), from Columbia University, in the US. IRI and CRED are collaborating on a project focused on improving the documentation of drought disasters in the EM-DAT database, the only publicly available global disaster event database on the Epidemiology of Disasters. And cluster 13, composed by two Australian and one New Zealand institution, that share theme-related researches: National Drug Research Institute (NDRI.CURTIN-AU), the National Drug and Alcohol Research Center (NDARC.UNSW-AU) and the Center for Social and Health Outcomes Research and Evaluation (SHORE-NZ).

**CONCLUSION**

With the exception of the four groups that may have been conformed due to the structure of their websites, all other clusters gather institutions sharing the same lines of research or that have been collaborating for some time. Fifteen of the 30 groups found gather local institutions from the same country. Though theme-related researches are considered a strong motivation for theses clusters formation, it is possible that in such cases, local partnerships aiming to solve national health problems have higher impact than the international relations established by these institutions.

As for the models of international cooperation, it was observed in this sample the predominance of north-south relations. Only one group was composed entirely by South institutions, represented by Brazil and Cuba. And emerging cooperation within south-south involving donor countries such as India, South Africa, Malaysia, Korea, and China are not yet reflected on the web. Similar results were observed for analyzing the same institutions using the technique of Social Network Analysis (SNA). [13]

It is important to consider, however, that despite becoming a fascinating source of information, the web has its limitations. In webometrics, the data should be interpreted with caution. When you seek the motivations behind clusters formation, for example, one should not assume a deterministic position. Several studies have tried to analyze the links between universities to seek the motivations behind these connections. [14-16] However, the performance of classification of types of links proved to be a complex exercise with inconsistencies and discrepancies regarding its categorization. [16] The use of the web cannot be determined by technology, but each case will depend on the specific context in which it appears. [17]

Another limitation may be related to the sample itself, since it was necessary to exclude those institutions whose websites did not have their own domain as well as those websites sharing a very low number of links. These exclusions certainly reduced the number of institutions studied and many representatives of south-south cooperation may have been left out of the final sample.

Despite such limitations, the results presented in this study reveal an important scenario of relations between institutions in the field of health on the web and contribute to the advancement of webometric studies applied to fields outside the Information Science.

**REFERENCES**

6. Zuccala A. Author cocitation analysis is to intellectual structure as web co-link analysis is to…? J Am Soc Inf Sci Technol 2006;57:1487-502.
12. Patrick WK. The Asia Pacific Academic Consortium for Global Public Health and medicine: Stabilizing south-south academic


16. Thelwall M. What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation.
