

Influence analysis of real options collaboration network

Análisis de influencia de la red de colaboración de opciones reales

Hernandes Coutinho Fagundes*
Rodrigo Tavares Nogueira**

* Master in Production Engineering from State University of Northern Rio de Janeiro, Brazil. Technical consultant in oil business, Petrobras (Brasil). [hernandesfagundes@gmail.com], [https://orcid.org/0000-0002-6293-0540] (Corresponding Author).

** PhD in Engineering Sciences/Production Engineering from Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brazil. Associated professor at State University of Northern Rio de Janeiro and Director of the Technology and Science Center (CCT) (Brasil). [nogueirart@gmail.com], [https://orcid.org/0000-0001-7143-5798].

Fecha de recepción: 17 de marzo de 2019.

Fecha de aceptación: 16 de mayo de 2019.

Para citar este artículo:

Fagundes, H. y Nogueira, R. (2019). Influence analysis of real options collaboration network. ODEON, 16, 37-65.

DOI: <https://doi.org/10.18601/17941113.n16.03>

Abstract

Real Options Theory arose as an alternative to value flexibilities entrenched in projects and has acquired popularity since the end of the twentieth century. Through bibliometric methods and graph theory, this paper develops an analysis of the collaboration network comprised of Real Options' researchers, including scientific papers from over the last eighteen years. In this effort, we meticulously identify authors and their co-authorship alliances, finding a distinct topology without a giant component. Developing unweighted and weighted models, the network is unraveled, providing measurement from internationalization propensity and computing different impact metrics, which recognize the most relevant researchers on the subject.

Key words: Real options; bibliometrics; collaboration networks; social network analysis.

JEL classification: D81, Z1, C44, G31.

Resumen

La teoría de opciones reales surgió como una alternativa para valorar las flexibilidades arraigadas en proyectos y ha adquirido popularidad desde finales del siglo XX. A través de métodos bibliométricos y teoría de grafos, este documento crea un análisis de la red de colaboración compuesta por los investigadores de opciones reales, que incluye trabajos científicos de dieciocho años. En este esfuerzo identificamos meticulosamente a los autores y sus alianzas de coautoría, encontrando una topología distinta sin un componente gigante. Al desarrollar modelos no ponderados y ponderados, la red se desenreda y proporciona mediciones a partir de la propensión a la internacionalización y el cálculo de diferentes métricas de impacto, que reconocen a los investigadores más relevantes sobre el tema.

Palabras clave: opciones reales; bibliometría; redes de colaboración; análisis de redes sociales.

Clasificación JEL: D81, Z1, C44, G31.

Introduction

Full comprehension of a specialized field involves the understanding of knowledge structure and the most influential researchers, which is useful in many situations, including the majority of scientific paper bibliographic reviews. Nevertheless, this mapping is not trivial whether it be addressed subjectively or with several independent metrics.

In this paper, we address the previous issue considering Real Options Theory, which influences a wide range of fields, particularly Economics, Finance, Engineering and Management. Real Options refers to optimization problems under uncertainty and was created based on a financial derivative instrument, allowing for the quantitative valuation of flexibilities and managerial decisions embedded in projects. The interest in Real Options quickly increased over the 90's and remains today as a promising tool to deal with both unstable and unpredictable economic scenarios (Dixit & Pindyck, 1995).

Bibliometrics is the measure of the impact of scientific publications (Cervantes & Mena-Chalco, 2010) through a set of mathematical and statistical principles, applied to knowledge production. For the purpose of knowledge development, bibliometrics is a powerful method to obtain systematic data. As described by Groos and Pritchard (1969), bibliometrics was known in the first instance as Statistical Bibliography and its application can be tracked until 1922, in the work of E. Wyndham Hulme. After that, it was latent for a period, thereafter being applied in 1944, twenty-two years later, by Gosnell and in 1962 by L. M. Raisig. The authors discussed the term 'bibliography', commonly employed, arguing that it was not suitable and suggesting the adoption of 'bibliometrics', which became the official designation.

The ability of bibliometrics to objectively evaluate knowledge generation has inspired many researchers to consider it a valuable tool. It helps to achieve a reliable bibliography, determine the "state of the art" of a particular field and study relationships in networks related to scientific communities.

Graph Theory is effective for the analysis of complex connections. It is included in the field of Discrete Mathematics and makes use of vertices (dots) and edges (lines) to represent relations and make problem solving easier to understand. From the depiction of a network, a researcher is capable of extracting important metrics that reveal relevant information and have immediate consequences in robustness and performance (Colizza, Flammini, Serrano, & Vespignani, 2006).

For a complete and formal description of graph construction or any of the following metrics, refer to van Steen (2010).

Two different vertices are connected lines that intersect if there is a path between them. Globally, a graph is connected if every possible pair of vertices is connected. Collaboration network graphs are generally disconnected, with a lot of small fragments or components. A component's size distribution graphically shows the size of each component, according to the number of vertices involved.

According to Newman (2009), a giant component is a single extensive group and 'one of the definitive features of any network' and almost every network in society exhibits a giant component formation. Newman, Watts, & Strogatz (2002) study the giant component creation concerning the network cut-off value and establish that its existence has important implications in social networks.

Concerning vertex importance, Freeman (1978) defines high centrality vertices as focal points, identifying three conceptual foundations of centrality: degree, closeness and betweenness. A vertex's unweighted degree is defined as the number of edges linked to it or, alternatively, the sum of weights in a weighted network (Barrat, Barthelemy, & Vespignani, 2007). It is a simple basic measure of vertex involvement frequently used as an initial step for studies (Opsahl, Agneessens, & Skvoretz, 2010). Nevertheless, it is local and provides no information about the network's topology.

When vertex degrees are organized in decreasing order, in the shape of a plot or list, it represents the Graph's Degree Distribution. Some networks, like the internet and social networks, have a distinct degree distribution including few vertices with particularly high degrees and a lot of vertices with lower degrees. It is a consequence of preferential attachment or "Matthew Effect", discussed by Merton (1968), describing the higher chance that prominent community members have to establish additional connections, in comparison to less distinguished ones (Barabási & Albert, 1999; Tomassini & Luthi, 2007). But even though the internet and social networks share a skewed distribution, they are frequently not the same. Unlike the internet, regularly described by scale-free models, social networks are often fitted to exponential distributions with a cutoff value (Newman, 2004).

Betweenness centrality is a global influence metrics defined as the number of shortest paths crossing through a specific vertex. If there is more than one possible shortest path, the betweenness contribution is proportionally reduced. This indicator represents the most influential researchers in the network that have

the strongest control over the flow of information (Opsahl *et al.*, 2010) and the ones that once removed, will generate the most sensitive impact on network's structure. Newman (2001) suggests that betweenness centrality identifies very clear winners, which means that second betweenness centrality is commonly well behind the highest value and also the difference between second and third place is equally notable. It is a positive aspect to identify most relevant authors, but Opsahl *et al.* (2010) also notice that a lot of vertices will have 0 betweenness, resulting in a poor contrast among the less connected individuals.

A vertex's Closeness Centrality is the inverse sum of the shortest distances to all other nodes (Opsahl *et al.*, 2010) and like betweenness centrality, it is indicative of a vertex's relative importance. The main difference is, while the first is focused on the researcher's control over information, closeness centrality is related to information access. Vertices with higher closeness centrality are more likely to receive information quicker and information originating with them will reach other researchers faster (Newman, 2001). However, standard closeness centrality is not suitable for a disconnected graph, often being applied for components (Opsahl *et al.*, 2010). In this case, comparison consistence requires a concept reformulation, which may vary among authors. We consider Opsahl's solution (Opsahl, 2010) involving the sum of inversed distances, to be that which avoids result annulation and favors vertices inside the largest (and more influential) components.

For many real-world networks, the connection probability between any two nodes is increased by the number of common neighbors these two nodes share (Watts & Strogatz, 1998). Newman, *et al.* (2002) observe that a mutual friend may increase probabilities of acquaintance in several orders of magnitude. This idiosyncrasy induces a clustering tendency often observed and measured in network analysis as clustering coefficient.

The clustering coefficient is defined only for vertices with a degree higher than two, as the chances that two of that vertex's neighbors are also neighbors between each other (Latapy, 2008). Globally, a graph's clustering coefficient is the arithmetic mean of each vertex's coefficient.

Pioneer studies in network analysis are assigned to Rapoport and Harary, in the 1940's and 1950's. From this context, Social Networks are organizations composed a group of people with some patterns of interactions between them (Newman *et al.*, 2002). Collaboration Networks are a special case included in Social Network Analysis (SNA), considering authors as vertices and co-authorship

as edges. They represent a ‘copious and meticulously documented record of the social and professional network of scientists (Newman, 2004) and a strategy for investigating social structures (Otte & Rousseau, 2002).

Over the last decades, there has been an increasing interest in network analysis. A crucial event for this enthusiasm was the information revolution that made it possible to analyze large-scale networks (Colizza *et al.*, 2006; Cervantes, Mena-Chalco, De Oliveira, & Cesar, 2013) and also easily applies the visual advantages of graphs to depict complex networks. Recent papers have focused a lot of attention on networks with skewed degree distribution, due to their practical relevance (Newman *et al.*, 2002).

Acknowledging the benefits of collaboration networks, several papers have developed structured analyses of different knowledge fields such as Newman (2004) for Biology, Mathematics and Physics, Luthi *et al.* (2007) for Genetic Programming and Cervantes *et al.* (2012) for Biological Sciences, Exact and Earth Sciences, Engineering, Health Sciences and Agricultural Sciences. Other relevant studies did not focus on field description but examined centrality measures, such as Newman (2001), Borgatti (2005), Borgatti *et al.* (2006) and Opsahl *et al.* (2010).

The practical conclusions regarding a collaboration network are based on the logical inference that authors who share a scientific work are personally acquainted and able to influence each other (Newman 2001, 2004; Cervantes & Mena-Chalco, 2010; Tomassini & Luthi, 2007). The simplest and most common representations of such networks consider co-authorship performing no magnitude distinction and constructing unweighted networks (Cervantes & Mena-Chalco, 2010; Opsahl *et al.*, 2010), but more intricate approaches address differences among co-authorship links such as considering recurrence and proportion (Newman 2001), the sort of scientific production (Cervantes & Mena-Chalco, 2010) or how recent the collaboration occurred (Cervantes *et al.*, 2013).

Despite link strength considerations, Opsahl *et al.* (2010) criticize regular weighted centrality, which focuses on tie weights only and does not take into account the number of ties, the basis of the original unweighted measures. From this point of view, a single tie may be preferred over multiple ties, even if the sum of the weights is the same. To make this consideration possible, they suggest a tuning parameter (α) that rewards fewer intermediates when assuming values between 0 and 1. Limits of the interval represent unweighted networks ($\alpha=0$) or a conventional weighted network, that attributes no cost to the number of intermediates ($\alpha=1$).

According to Cervantes *et al.* (2013), vertex significance and influence is an important topic of Social Semantic Web and Social Network Analysis. In general, the identification of different lines of thought associated with social groups or the most influential scientists on a specific topic is often needed. Some applications are the evaluation of bibliographical research, subject analysis, congress representativeness, academic recruitment or award granting.

The first author importance verification, simple and popular, concerns the absolute number of published works, according some specific criteria or areas of interest. However, it is a local measure, ignoring the relationship with the scientific community and not taking into account any quality aspect.

Recently, h-index was proposed as an alternative to measuring an author's impact by Hirsch (2005). With a simple rule, it considers the number of published papers and also the number of existent citations by others. H-index is gaining popularity, being adopted by major search engines and databases such as Google Scholar, Scopus and Web of Science (Bar-Ilan, 2008), but not without several critiques concerning its adoption, as detracting from young researchers and divergence among different databases (Flaatten, Rasmussen, & Haney, 2016) or index manipulation by strategic self-citation (Bartneck & Kokkelmans, 2011). In addition, h-index incorporates a quality aspect, but it is still local, ignoring the network structure.

From a network perspective, vertex influence quantification commonly relies on centrality measures: degree centrality, closeness centrality, betweenness centrality or similar variations (Cervantes *et al.*, 2013). These measures can be obtained for unweighted and weighted networks as discussed in the previous chapter.

1. Methodology

1.1. Search criteria and data gathering

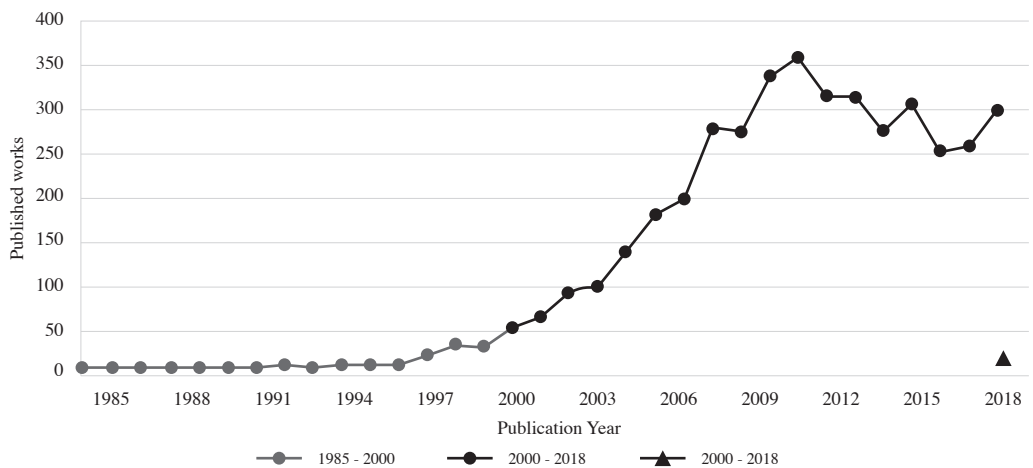
Since the 90's, bibliographic examination intensively exploited online databases to accurately reconstruct large-scale scientific networks. This procedure is an ongoing evolution from previous interview methodology, being capable of an automatic and reliable construction of relatively complete networks (Newman, 2001; Newman *et al.*, 2002; Newman, 2004).

This research was conducted using Elsevier's SCOPUS database (Elsevier, 2018b) on January 31, 2018. SCOPUS is one of the most complete sources (Falagas,

Rasmussen, & Haney, 2008) and self-proclaimed as the largest abstract and citation database of peer-reviewed literature (Elsevier, 2014). Furthermore, comparative studies show a high correlation between results from SCOPUS and Clarivate Analytics' Web of Science (WoS) (Bar-Ilan, Levene, & Lin, 2007; Archambault, Campbell, Gingras, & Larivière, 2009), which contributes additional confidence for query comprehensiveness.

Search input considered the terms “real options”, “investment under uncertainty” or “irreversible investment” in paper title or keywords. An initial examination under these criteria demonstrates that publish ratio started to grow in the 90's and quickly increased over the first decade of the 21st century (Figure 1). This development was boosted by influential studies such as Dixit & Pindyck (1994) and Trigeorgis (1996).

Figure 1: Real Options published scientific papers obtained from SCOPUS using the mentioned search criteria



Source: Created by authors.

To ensure both actuality and representativeness, we establish an additional criterion to include only scientific work published after the year 2000. Thus, 4,103 publications were found, fulfilling search criteria and published between 01/01/2000 and 01/31/2018. Every available publication was included, with an initial distribution containing 67.0% scientific articles, 25.2% conference papers, 3.4% reviews and 4.4% other document types.

1.2. Screening

The first screening action addressed 10 documents without a valid identification. In this group, 4 documents owned valid ID's that were not provided due to SCOPUS output misalignment, which was manually fixed. The other 6 did not present any kind of valid identification. In these cases, artificial unique IDs were created and all 10 published papers were reintegrated into the main data.

Also, three publications did not present any authorship information. One of them referred to a Conference (instead of a specific article) and another item pointed to a note without traceable records (Google Scholar, PubMed e WoS were consulted in an effort to find additional data). These two records were excluded from the main data. The third case author was identified using Google's Scholar search and the data was updated and reintegrated. Hence, the final data set was comprised of 4,101 scientific publications.

1.3. Data Identification and Disambiguation

Studies about academic networks are performed focusing on individuals, affiliations and countries, which are sometimes referred to as micro, meso and macro-levels (Hou, Kretschmer, & Liu, 2007). But while it is fairly straightforward to assign an unambiguous identification to scientific production, it is not always clear how to avoid misinterpretation about remaining data. Tomassini & Luthi (2007), Haak, Fenner, Paglione, Pentz, & Ratner (2012) and Fagundes & Nogueira (2017) report difficulties with different spellings or resemblances in authors' names and Cervantes *et al.* (2012) mention problems due to multiple forms of country names.

To consider different authors as one or one influential author as many may produce sensitive changes in network topology, thusly compromising the study goal. Therefore, we gave special attention to this requirement extracting the author's unique identification number (ScopusID) from the records, via an application programming interface (API), entitled Abstract Retrieval, which also informs regarding the author's affiliation ID (Elsevier, 2018a). Data management was performed with RStudio software using Rscopus package (GitHub, 2017). Same API is capable of returning an alternative author's unique ID provided by Orcid (Haak *et al.*, 2012), which may be useful when merging different databases.

At this point, we alert that the simplest choice to directly use an author's name fields (last name and first name) could be unreliable. For the 6,074 authors

in our data set, there were 6,867 different spellings, with 11.2% of authors being designated by more than one way and some with several (as many as nine) different names.

Another discouraging option is to use the author's names and affiliation, extracted through regular output files as info for Author Retrieval API, in order to obtain the author's ID number. This procedure has the opposite bias, agglomerating different authors into a single ID. As an example, for every author named Liu, H., Liu, J., Liu, X. and Liu, Y.-H. the mentioned API associated the same ID, suggesting a significant amount of published works (30), when in fact these names are related to 23 different researchers, none of them with more than 6 published works.

Establishing an affiliation ID database from publication records, we extracted affiliation and country names from the institution profile, through Affiliation Retrieval API, which will be used to evaluate the network in meso and macro-levels.

At this point we find that some authorships did not have an affiliation bond (affecting 3.5%) and some affiliations were not correlated to specific countries (affecting an additional 0.2% of total authorships). These cases were ignored in meso or macro statistics or when identifying author-country connection.

Gathering micro and macro-level data, we were able to identify and compare patterns of international collaboration. To obtain objective metrics on this topic, we use the proposition from Leite *et al.* (2011), to calculate an International Publication Ratio (IPR). In this methodology, each paper is designated international if it involves two or more countries or nationalities and the IPR is calculated as the proportion of international publications. Single author work is considered national and only authors with 3 or more categorized papers are included, to avoid bias. For the purpose of IPR calculation, the entire paper was disregarded when lacking country information.

H-index calculation, considered only the 4,101 Real Options papers, instead of the authors' complete records, and the citation count obtained on 01/31/2018, when the search was performed.

1.4. Graph construction

Initially, the authors' IDs were enumerated as vertices and every single co-authorship relationship listed, in order to provide edge input for the graph creation.

In the simplest model, link weight is always 1, for every existing connection, but for weighted analysis, each co-authorship was correlated with a link strength

(ls) according to Equation 1 (Newman, 2001), where n_k stands for the number of coauthors in paper k and δ_i^k is 1 if author i coauthors paper k and 0 otherwise.

$$ls_{i,j} = \sum_k \frac{\delta_i^k * \delta_j^k}{n_k - 1} \quad (1)$$

Graph visual representation and basic topology measures were obtained with the open-source network analysis software Gephi (Bastian, Heymann, & Jacomy, 2009). Clustering coefficient was also calculated with a Gephi's implementation of Latapy's routine (Latapy, 2008), which applies triangle counting and has improved performance on graphs resembling power-law distribution.

Most centrality measures obtained with Gephi's current version apply to unweighted graphs only. In this manner, betweenness and closeness were obtained using Opsahl *et al.* (2010) solution, which is able to consider link strengths. The method can also handle multi-component closeness, using the sum of inversed distances to all other nodes instead of the inverse of the sum of distances. All weighted distances included in this paper do not distinguish the number of ties in a path, equivalent to set alpha parameter as 1 in Opsahl's model.

In graph representation, vertices were distributed using Fruchterman-Reingold's algorithm. The routine simulates attractive and repulsive forces between nodes to provide 'aesthetically-pleasing' network organization (Fruchterman & Reingold, 1991).

2. Results and discussion

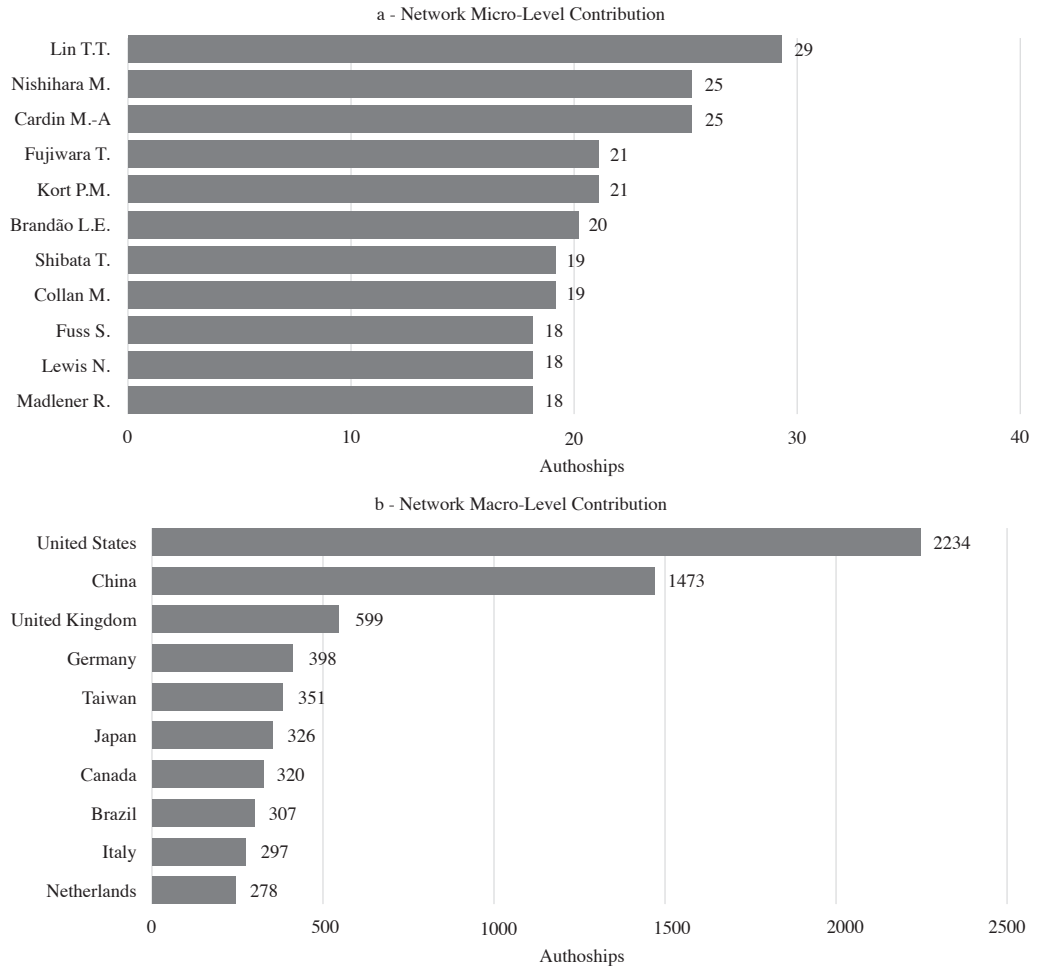
2.1. Network general metrics

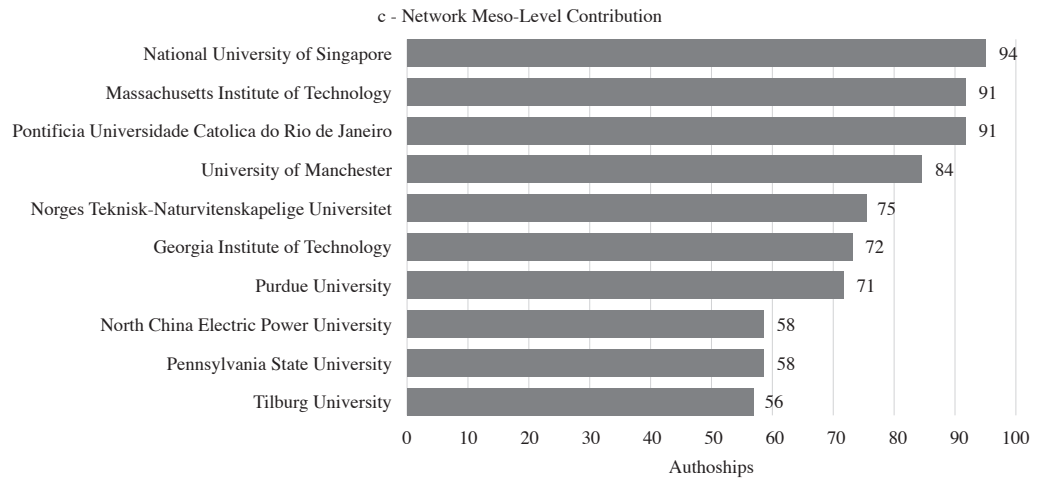
From the gathered data, we found a network composed of 4,101 scientific works and 9,828 authorships. These papers involved 6,074 different authors and 2,041 institutions, spread across 81 countries. The great majority of institutions were universities, but there were also some governmental organizations and private companies such as General Motors, Fuji Xerox and Hewlett-Packard.

From the records, we could identify the most beneficial participations at the individual, affiliation and national levels, as shown for 10 top contributors in Figure 2. At the micro-level, Lin T.T. was the most prolific author, with 29 published works. The National University of Singapore was the most productive

institution, being responsible for 94 authorships, followed closely by MIT (91) and PUC-RIO (91). The macro-level demonstrates a fair lead by American institutions, with 2,234 authorships, followed by Chinese (1,473) and British (599). Productivity, as previously calculated, is sometimes applied as an influence indicator. However, we emphasize that it only categorizes input data, disregarding relevance or global impact, and thus is very limited.

Figure 2: Top 10 Network Contribution in Micro (a), Macro (b) and Meso-Levels (c)





Source: Created by authors.

Main data statistics are summarized in Table 1 with metrics from other studied collaboration networks. We are able to observe that Real Options Networks is smaller than more generic ones analyzed by Newman but have similar publication amounts when compared to Genetic Programming Network, which has more specific subjects and also comparable time coverage.

Table 1: Real Options Data Statistics and Comparison with Other Studied Collaboration Networks

Parameter	Real Options	Biology ^a	Physics ^a	Mathematics ^a	Genetic Programming ^b
Scientific Works	4,101	2,163,923	98,502	NA	4,564
Authors	6,074	1,520,251	52,909	253,339	2,809
Works per author	1.62	6.4	5.1	6.9	3.16
Authors per work	2.40	3.75	2.53	1.45	2.25

^a From Newman (2004); ^b From Tomassini & Luthi (2007); NA – Non-Available, despite author estimate around 1.6 million papers.

Source: Created by authors.

Newman (2004) states that the number of authors per paper are very different between fields and consider the difference a consequence of the performed research mode: Biology occupies one extreme, with 3.75 authors per paper, where endeavors involve a large group of professionals, with a significant share

of experimental contribution. At the other extreme is Mathematics (1.45) with individual, high theoretical research. Real Options relationship (2.40) is similar to Genetic Programming (2.25) and Physics (2.53) Networks, all of which supports Newman's argument that intermediate ratios are identified in fields where theoretical and experimental practices are present.

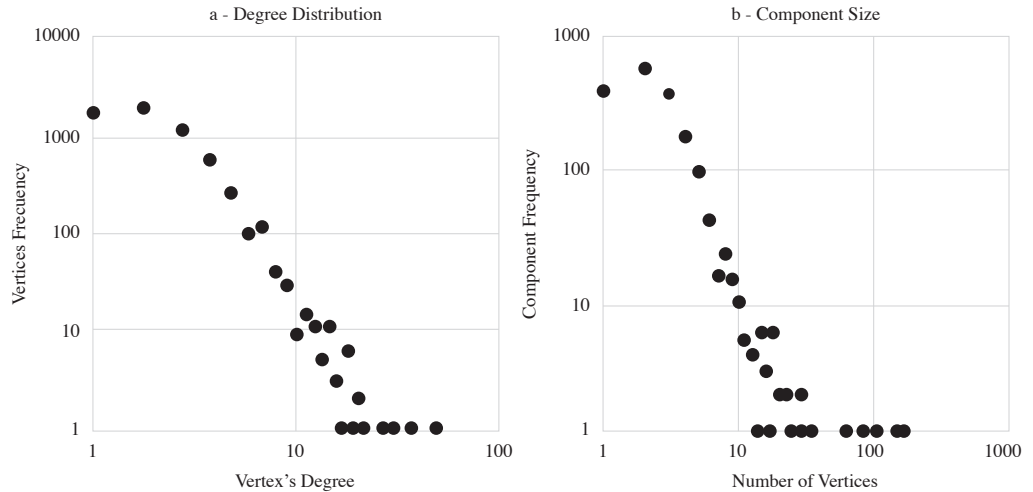
One notable difference is that Real Options Network has a relatively high number of authors (it is the only dataset with more authors than papers) and, as a consequence, the lowest papers per author ratio (1.62). The topology evaluation will explain that it is influenced by a large number of peripheral researchers, with sparse results and connections. Considering that the data provided a representative description of the network, this aspect may be related to practical professionals who are not entirely focused on knowledge production (as Real Options is a very practical tool and numerous papers present application cases) or suggest that the field is often made up of multivalent authors, who also produce scientific work in other knowledge fields.

2.2. Network relational metrics

Initial structure evaluation of networks often relies on degree distribution (Colizza *et al.*, 2006) as shown on the left side of Figure 3, with double logarithmic axes. This figure suggests the existence of a skewed distribution, usually exponential with a cutoff value, after a lower bound. The proper verification is not covered in this paper but may be performed with systematic steps, as discussed in papers such as Clauset, Shalizi, & Newman (2009). Few high collaborative authors are visible in the degree distribution bottom right and are frequently described as high impact scientists. The log scale also omits 388 vertices representing authors that did not collaborate (degree 0) and have minor influence on network topology.

Right plot in Figure 3 represents the component size distribution. It exhibits a highly disconnected network, as already expected in a collaborative setting. But where a giant component was supposed to exist, joining the majority of researchers, we find particular large components that individually involve no more than 2.82% of the community. For detailed investigations, we highlight the five largest components that together represent 9.53% of Real Options authors, as emphasized in Figure 3, bottom right.

Figure 3: Degree distribution (a) and component size distribution (b) for Real Options collaboration network



Source: Created by authors.

Graph statistics are presented in Table 2 along with data from comparable collaboration networks. Real Options is a more recent subject than fields evaluated by Newman and it is possible to verify that its largest component is notably less comprehensive than any other. Genetic programming network owns a relatively larger giant component than Real Options, although exhibiting a significantly reduced proportion in comparison with more classic fields. Evidence that smaller giant components are identified in both recent networks may suggest we are describing an emerging network that will still reach a steady topology. This inference differs from a statement by A. L. Barabási *et al.* (2002), which states that authors group into a single giant component in very early stages of the field but aligns with findings from Tomassini & Luthi (2007) for the Genetic Programming Network.

Tomassini & Luthi (2007) also noted for Genetic Programming Networks that the average degree jumps from about 2 to 4, in the time of giant component formation and has a tendency to increase over time. This verification suggests that the largest component proportion displays a direct correlation with the average degree. Real Options Network has, in fact, a lower average degree (2.45) that increases inside the largest components (mean of 3.96) which also incites investigations about network dynamic conditions.

Table 2: Real Options Graph Statistics and Comparison with Other Studied Collaboration Networks

Parameter	Real Options	Biology ^a	Physics ^a	Mathematics ^a	Genetic Programming ^b
Average degree (unweighted)	2.45	18.1	9.7	3.9	4.17
Largest component	2.82%	92%	85%	82%	36.5%
Average distance (unweighted)	4.06	4.6	5.9	7.6	4.74
Largest distance (diameter)	15	24	20	27	NA
Clustering coefficient	0.875	0.066	0.43	0.15	0.665

^a From Newman (2004); ^b From Tomassini & Luthi (2007); NA - Non-Available

Source: Created by authors.

The largest distance inside Real Options Network (15) is significantly lower than diameters found on those studied by Newman (2004), between 20 and 30. This distance is also measured in the second largest component instead of the largest one (maximum eccentricity in the largest component is 11). Along with the low average distance (4.06), results suggesting it is easier to reach other scientists in the Real Options Network than in social groups from other fields.

Clustering coefficient is markedly higher and is biased by numerous small components. However, the clustering coefficient is still high even inside the largest components, revealing a tendency of acquaintance between mutual collaborators, as shown in Table 3. According to Newman (2004) differences in the clustering coefficient may indicate distinct collaboration patterns, but the analysis is not obvious and shall be investigated.

Table 3: Five largest component statistics from Real Options Collaboration Network

	Component Number				
	1	2	3	4	5
Authors	171	154	105	85	64
Works per Author	2.84	2.74	2.27	2.27	2.95
Authors per Work	2.6	2.2	2.9	3.0	2.8
Institutions	96	116	37	50	32
Countries	18	23	13	7	12
Clustering Coefficient	0.781	0.686	0.797	0.820	0.741
Average Degree (unweighted)	4.19	2.96	4.02	4.21	4.41

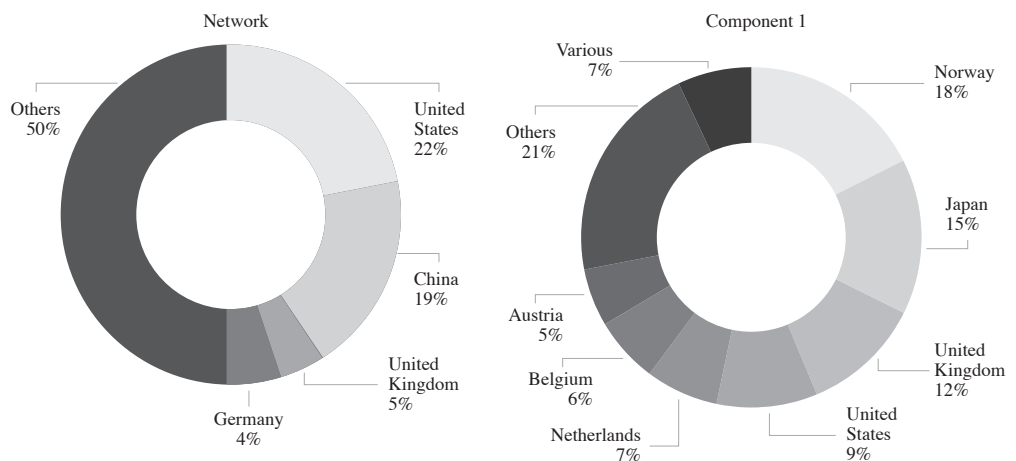
Source: Created by authors.

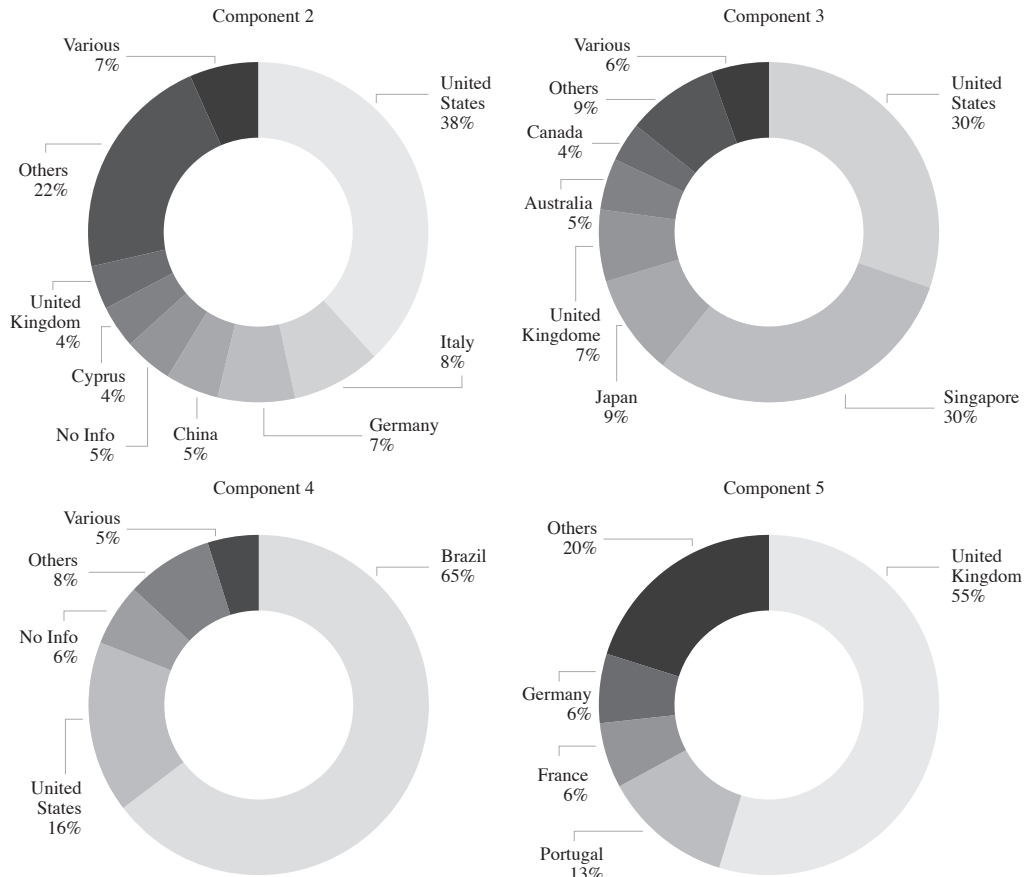
2.3. Component data comparison

Tomassini & Luthi (2007), state that giant component formation in social networks does not have a definitive explanation, but fragmentation may be associated to discipline and geographical boundaries or even human behavior. The consequences of a giant component absence are also obscure. It is possible to hypothesize that different large groups may characterize a decentralized generation of knowledge, beneficial to fast-paced and disruptive innovations, but Newman (2004) also argues that intellectual isolation from the mainstream area ‘cannot often be a good thing’ due to support limitations.

As geographical boundaries represent a logical influence, we make an effort to identify location heterogeneities among components. For this purpose, we linked authors with their institution’s country in order to categorize contributions inside each one of the five largest components. As an example, authors who published in American institutions were assigned to a class named United States and researchers who published in several institutions, located in more than one country, were assigned to the category ‘Various’. Authors assigned as ‘No info’ did not hold location information in any of their published work. Reiterating that class refers to the country where the author’s institution is situated and may not correspond to the author’s nationality. Results are shown in Figure 4, exhibiting countries with a contribution greater or equal to 4%.

Figure 4: Countries’ contribution for Real Options Collaboration Networks and its five largest components





Source: Created by authors.

Despite not representing the exact definition from macro-level evaluation, network distribution indicates some major contribution from American and Chinese institutions, followed by British and German, although components examination enlightens different inner patterns. The largest component shows a more balanced global relationship involving 18 countries. Authors with exclusive Norwegian affiliations are responsible for 18% of the network, Japanese for 15%, British for 12% and American for 9%. The second largest component encompasses more countries (23 in total) but is less homogeneous, with American authors representing 38%, Italian authors 8% and German 7%.

The other three analyzed components depict even more concentrated networks. Component number 3 is mainly represented by American and Singapore-

ean institutions, with 30% representativeness for each, and also 9% of Japanese contribution.

The fourth component is primarily Brazilian (65%) with a distinguishing American contribution (16%). The fifth component also shows a major contribution from a single country, where British affiliations are responsible for 55%, followed by Portuguese (13%).

It is possible to verify that China's contribution is significant in the entire network (19%) which is not found in the major components analysis. A detailed examination clarifies that only 13, from the 1,131 Chinese authors, belong to the set containing the five largest components, instead, Chinese affiliations are spread over 389 smaller components. This fact may suggest very secluded groups or a network that is still in formation, with the possibility of establishing future connections thusly becoming a more significant component.

In summary, it is possibly to conclude that components present significant heterogeneity due to country contributions, implying that the institution's nationality shall be a significant variable to predict a vertex component. In this sense, the increase of international collaboration may benefit a larger component establishment.

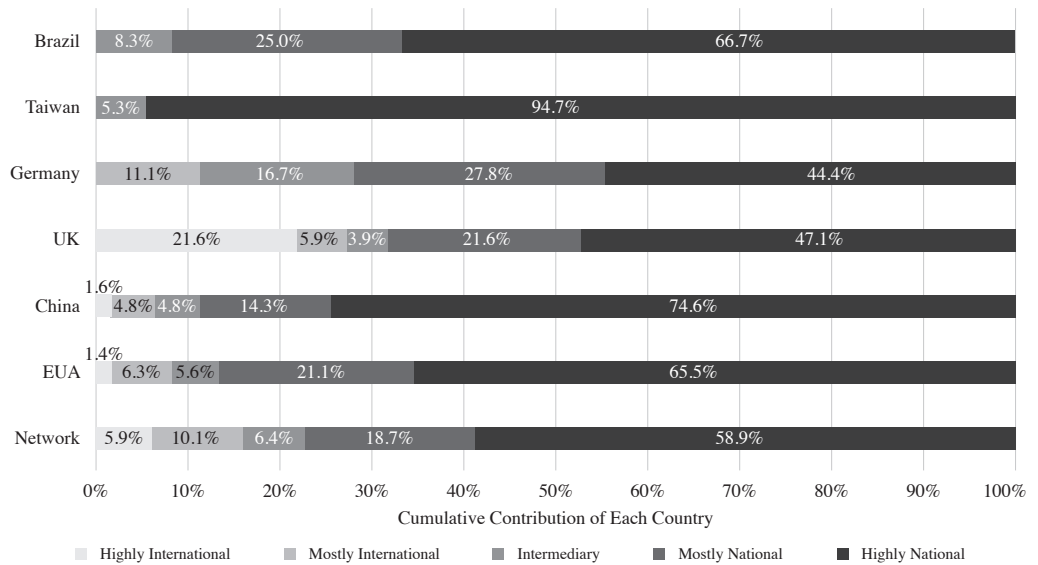
Even, if there is no standard metrics for an international collaboration metric, Leite *et al.* (2011) developed an international production ratio (IPR) to categorize author publications according to their internationalization. Their data included only Brazilian researchers and several knowledge areas. Real Options is a multidisciplinary subject present especially in Social Sciences (Economics and Management)¹, where authors report a significant national IPR (highly national or mostly national around 90%) when compared to exact (25%) and biological (40%) sciences, fields analyzed by Newman and Tomassini.

Through Leite's approach, we selected authors with 3 or more defined papers (national or international), setting up a group of 674 researchers (11.1%) and calculated an IPR for the entire Network and the six most productive countries (still related to the author's affiliation location), shown in Figure 5. The categories represent the participation of international work in the author's records, being highly national if this share belongs to the interval [0%,20%], mostly

1 Real Options subject has a lesser but still significant contribution from Engineering, which exhibits an IPR similar to the biology field, with highly and mostly national authors representing about 40%.

national if in (20%,40%], intermediary if in (40%,60%], mostly international if in (60%,80%] and highly international if in (80%,100%].

Figure 5: International Production Ratio for Real Options Network



Source: Created by authors.

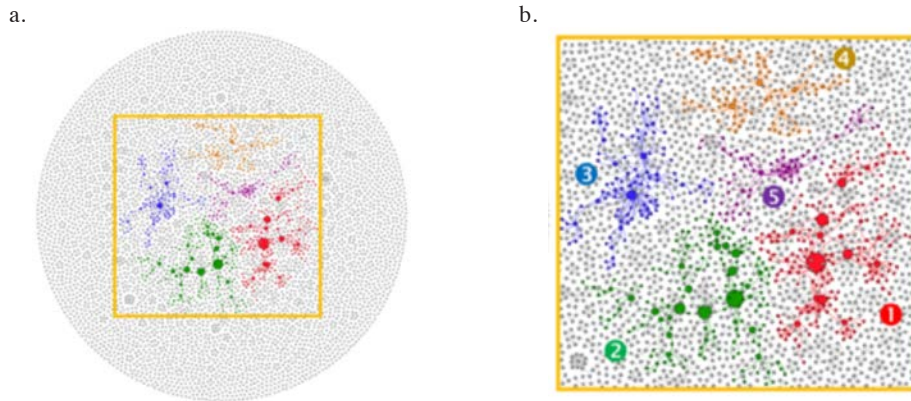
We are able to verify that a large proportion of Real Options authors are highly or mostly national (77.6%), with a range from 68.6% to 94.7% inside components. The Brazilian set has a proportion of 91.7% highly or mostly national authors, very close to Leite's findings for social sciences.

Evidence indicates that Real Options present a low internationalization rate, which may affect collaboration network dynamics. In this sense, some additional effort to extend the IPR results to a global scale may be useful to understand why there is still a large difference in a country's contribution inside Real Options components and a delay in the giant component formation.

2.4. Influence network indexes

In Figure 6, we show Real Options network representation, highlighting the five largest components. The visual construction displays vertex diameter directly proportional to its unweighted betweenness centrality, assisting identification from individuals who have a crucial role in the control of information flow.

Figure 6: Real Options' Collaboration Network representation (a) and a 40% magnified area, highlighting the five largest components (b)



Source: Created by authors.

As discussed in the introduction chapter, there are several influence metrics. Simpler ones are data based, while most intricate evaluations consider weighted global impact across the boards. The following discussion will consider results from different discussed metrics inside Real Options collaboration network, with results shown in Table 4.

Table 4: Top 10 most influential authors in Real Options collaboration network according to different metrics

	Productivity			H-index			Degree (unweighted)			Degree (weighted)		
	Author	CN	MV	Author	CN	MV	Author	CN	MV	Author	CN	MV
1	Lin T.T.	S	29	Madlener R.	S	10	Cardin M.-A.	3	43	Lin T.T.	S	25
2	Nishihara M.	1	25	Folta T.B.	2	10	Fleten S.-E.	1	33	Cardin M.-A.	3	23
3	Cardin M.-A.	3	25	Kort P.M.	1	9	Lin T.T.	6	27	Kort P.M.	1	21
4	Kort P.M.	1	21	Fuss S.	1	9	Brandão L.E.	4	25	Brandão L.E.	4	20
5	Fujiwara T.	S	21	Szolgayova J.	1	9	Kort P.M.	1	20	Madlener R.	S	18
6	Brandão L.E.	4	20	Obersteiner M.	1	9	Fuss S.	1	19	Nishihara M.	1	18
7	Shibata T.	1	19	De Neufville R.	3	9	De Neufville R.	3	19	Fuss S.	1	18
8	Collan M.	S	19	Cardin M.-A.	3	8	Howell S.D.	5	18	Lewis N.	S	18
9	Fuss S.	1	18	Fleten S.-E.	1	8	Madlener R.	S	17	Fleten S.-E.	1	17

10	Madlener R.	S	18	Trigeorgis L.	2	8	Collan M.	S	17	Collan M.	S	17
11	Lewis N.	S	18	Siddiqui A.	1	8	Musshoff O.	S	17
12	Tong T.W.	2	8	Wen F.	S	17
13	Blanco G.	S	17
14	Johnson P.V.	S	17
	BC (unweighted)			BC (weighted)			CC (unweighted)			CC (weighted)		
	Author	CN	MV	Author	CN	MV	Author	CN	MV	Author	CN	MV
1	Fleten S.-E.	1	9700	Fleten S.-E.	1	10489	Fleten S.-E.	1	83.8	Trigeorgis L.	2	87.6
2	Trigeorgis L.	2	8364	Trigeorgis L.	2	8364	Fuss S.	1	71.0	Reuer J.J.	2	82.0
3	Reuer J.J.	2	6284	Siddiqui A.	1	6433	Siddiqui A.	1	68.4	Martzoukos S.H.	2	80.4
4	Siddiqui A.	1	5175	Reuer J.J.	2	6284	Hagspiel V.	1	68.4	Tong T.W.	2	80.4
5	Alessandri T.M.	2	4632	Alessandri T.M.	2	4632	Chronopoulos M.	1	67.2	Kort P.M.	1	80.0
6	Hagspiel V.	1	4057	Hagspiel V.	1	4249	Szolgayova J.	1	67.0	Fleten S.-E.	1	79.6
7	Fuss S.	1	3924	Fuss S.	1	4016	Cardin M.-A.	3	66.3	Smit H.	2	79.4
8	Cardin M.-A.	3	3833	Cardin M.-A.	3	3978	Kort P.M.	1	61.8	Siddiqui A.	1	79.3
9	Koussis N.	2	3651	Koussis N.	2	3651	Andersson A.M.	1	60.6	Koussis N.	2	78.7
10	Takashima R.	1	3623	Takashima R.	1	3650	Elverhøi M.	1	59.9	Huisman K.J.M.	1	78.0

- BC = Betweenness Centrality; CC = Closeness Centrality; CN = Component Number; MV = Metric Value.

- Component number is assigned according to its size, as component 2 is the 2nd largest component. - S represents that author belongs to a component that is smaller than the 5th component.

- List was extended when ties occurred in 10th position.

Source: Created by authors.

From data-based measures, we calculate productivity, which was already referred to as a micro-level measurement and h-index that correlates the number of publications and their impact, through citations. Authors with higher productivity are Lin T. T. (29 published works), Nishihara M. (25) and Cardin M.-A. (25) while authors with higher h-index are Madlener R. and Folta T.B., both with h-index equal to 10. It is interesting to notice that the most distinguished authors, according to these criteria Lin T. T. and Madlener R., belong to smaller components with 35 and 18 members respectively.

Vertex degree is the first and simpler (still local) centrality measure, establishing the most collaborative authors. From an unweighted perspective, degree measurement is the number of associated collaborators while weighted degree introduces link strength considerations, as discussed in the methodology chapter. A scientist's weighted degree is similar to his productivity, with the difference being that productivity is influenced by single authored papers while weighted degree is not. In Real Options Network, the most collaborative authors are Cardin M.-A. (43), Fleten S.-E. (33) and Lin T.T. (27) and the authors with the highest weighted degree are Lin T.T. (25), Cardin M.-A. (23) and Kort P.M. (21). Lists with both (unweighted and weighted) degrees hold similarities, but strength consideration was sufficient to raise Lin.T.T. from 3rd to 1st position and Kort from 5th to 3rd. Even when compared with productivity, weighted degree shows significant differences as exemplified by Nishihara M., the 2nd most productive author but 6th position in weighted degree rank.

Global influence measures are more consistently obtained using betweenness and closeness centrality, emphasizing that both can also be calculated with same unweighted and weighted considerations. Authors with the highest betweenness centrality are the individuals with the most control over information, which in an unweighted model are Fleten S.-E. (9700), Trigeorgis L. (8364) and Reuer J.J. (6284). Strength addition provides very similar results, ranking Fleten S.-E. (10489), Trigeorgis L. (8364) and Siddiqui A. (6433) in the top positions.

The last evaluated centrality measure is Closeness Centrality, indicating researchers with greater access to information. Fleten S.-E. (83.8), Fuss S. (71.0) and Siddiqui A. (68.4) are scientists with higher closeness index in an unweighted network, all of them from the largest component. But unlike betweenness, the list is significantly different when link strength is introduced. In a weighted model, Trigeorgis L. (87.6), Reuer J.J. (82.0) and Martzoukos S.H. (80.4) have higher indexes and it is remarkable that for the top 4 positions, weight considerations replaced component 1 members by component 2 scientists.

From the mentioned global measures, we were able to identify the most influential authors (mainly positioned in components 1 or 2). But since local clusters were identified, it is important to provide information about authors relevance inside the other three large components, according to weighted measures. In component 3, authors with the highest control over information (higher weighted BC) are Cardin M.-A. (BC=3978) and De Neufville R (BC=1999) who are also the ones with higher information access (CC is 62.6 and 56.0 respectively).

Inside the fourth component, with a clear Brazilian influence, authors with higher weighted BC are Dias M.A.G. (BC=2434) and Teixeira J.P. (BC=1883) and researchers with higher closeness centrality are Brandão L.E. (CC=50.9) and Dyer J. (CC=45.8). For the fifth component, a British dominant group, Moriarty J. (BC=1202) and Howell S.D. (BC=751) are the authors with the most control over information and Johnson P.V. (CC=35.7) and Howell S.D. (CC=35.5) have more efficient information access.

It is relevant to verify that there are dissimilar conclusions when using different influence metrics, especially between local and global measures. Notice that higher global centrality measures are correlated to the largest components which effectively provide more influence in the community. Also, the weighted model is capable of distinguishing a single, big group co-authorship from a recurrent strong relationship and, in several ways, may provide a more complete description.

3. Conclusions

In this paper, we researched more than 18 years of scientific work about Real Options, extracting reliable information at the micro, meso and macro-levels. Using graph theory, we constructed a Real Options collaboration network, identifying an expected skewed distribution but also verifying that there was not a giant component formation and the largest connected group gathers only 2.82% of Real Options authors.

In component analysis, we demonstrated that even the largest components have significant differences regarding country contribution, suggesting that geographic boundaries may be a relevant factor to explain so many disconnected components. We showed that the United States and other productive countries have a significant contribution in the largest components while China is the second most productive country in the network but occupying a less distinctive presence inside the five most relevant groups.

Using an International Publication Ratio, we verified that 77.6% of Real Options authors are highly national or mostly national. From previous research we identified that social sciences may present a larger proportion of national researchers in comparison to biology and exact sciences and hypothesize that these features may influence dynamic considerations about the component's idiosyncrasies and giant component delay.

We reviewed and calculated the most relevant influence metrics, from local data centered measures to global weighted network evaluations, identifying substantially different results, being crucial to understanding metrics outlines and limitations, in order to reach a reliable conclusion. In global measures, our results showed that link strength considerations have an extensive influence on closeness centrality and a minor effect on betweenness centrality.

A more complete verification, using global centrality measure in a weighted model, found that Stein Erik Fleten is the author with most control over information, holding the highest betweenness centrality score in Real Options Collaboration Network, and Lenos Trigeorgis is the researcher with the widest information access and the highest closeness centrality. We also, identified and measured the most relevant centrality scores inside each one of the five biggest components, highlighting local distinguished scientists.

Finally, it would be inaccurate to state that non-collaborative authors cannot influence each other. However, a co-authorship relationship is effectively stronger than other measurable ones (like citations) and assurance of a broader giant component may benefit field evolution.

References

- Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing Bibliometric Statistics Obtained from the Web of Science and Scopus. *Journal of the Association for Information Science and Technology*, 60(7), 1320-1326.
- Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509-512.
- Barabási, A. L., Jeong, H., Néda, 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), 590-614.
- Bar-Ilan, J. (2008). Which H-index? - A Comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.
- Bar-Ilan, J., Levene, M., & Lin, A. (2007). Some Measures for Comparing Citation Databases. *Journal of Informetrics*, 1(1), 26-34.

- Barrat, A., Barthelemy, M., & Vespignani, A. (2007). The Architecture of Complex Weighted Networks: Measurements and Models. In *Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science* (pp. 67-92): World Scientific.
- Bartneck, C., & Kokkelmans, S. (2011). Detecting h-index Manipulation Through Self-Citation Analysis. *Scientometrics*, 87(1), 85-98.
- Bastian, M., Heymann, S., & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *International AAAI Conference on Weblogs and Social Media. Association for the Advancement of Artificial Intelligence, 2009*.
- Borgatti, S. P. (2005). Centrality and Network Flow. *Social Networks*, 27(1), 55-71.
- Borgatti, S. P., Carley, K. M., & Krackhardt, D. (2006). On the Robustness of Centrality Measures under Conditions of Imperfect Data. *Social Networks*, 28(2), 124-136.
- Cervantes, E. P., & Mena-Chalco, J. P. (2010). A New Approach to Detect Communities in Multi-Weighted Co-Authorship Networks. In *2010 XXIX International Conference of the Chilean Computer Science Society, Antofagasta, Chile, 15-19 Nov. 2010* (pp. 131-138): IEEE.
- Cervantes, E. P., Mena-Chalco, J. P., & Cesar, R. M. (2012). Towards a Quantitative Academic Internationalization Assessment of Brazilian Research Groups. In *2012 IEEE 8th International Conference on E-Science, Chicago, IL, USA, 8-12 Oct. 2012* (pp. 1-8): IEEE.
- Cervantes, E. P., Mena-Chalco, J. P., De Oliveira, M. C. F., & Cesar, R. M. (2013). Using Link Prediction to Estimate the Collaborative Influence of Researchers. In *2013 IEEE 9th International Conference on eScience, Beijing, China, 22-25 Oct. 2013* (pp. 293-300): IEEE.
- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law Distributions in Empirical Data. *SIAM Review*, 51(4), 661-703.
- Colizza, V., Flammini, A., Serrano, M. A., & Vespignani, A. (2006). Detecting Rich-Club Ordering in Complex Networks. *Nature physics*, 2(2), 110-115.
- Dixit, A. K., & Pindyck, R. S. (1994). *Investment Under Uncertainty*. Princeton: Princeton University Press.

- Dixit, A. K., & Pindyck, R. S. (1995). The Options Approach to Capital Investment. *Harvard Business Review*, 73(3), 105-115.
- Elsevier (2014). Scopus Quick Reference Guide. Available at: https://www.elsevier.com/__data/assets/pdf_file/0005/79196/scopus-quick-reference-guide.pdf.
- Elsevier (2018a). Elsevier Developers. Available at <http://api.elsevier.com>.
- Elsevier (2018b). Scopus. Available at: <http://www.scopus.com>.
- Fagundes, H. C., & Nogueira, R. T. (2017). *Analyzing the Collaboration Network of Real Options Authors*. Paper presented at the 21st Annual International Conference on Real Options, Boston.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses. *The FASEB journal*, 22(2), 338-342.
- Flaatten, H., Rasmussen, L. S., & Haney, M. (2016). Publication Footprints and Pitfalls of Bibliometry. *Acta Anaesthesiologica Scandinavica*, 60(1), 3-5.
- Freeman, L. C. (1978). Centrality in Social Networks Conceptual Clarification. *Social Networks*, 1(3), 215-239.
- Fruchterman, T. M., & Reingold, E. M. (1991). Graph Drawing by Force-Directed Placement. *Software: Practice and experience*, 21(11), 1129-1164.
- GitHub (2017). R Package to Interface with Elsevier and Scopus APIs. Available at: <https://github.com/muschellij2/rscopus>.
- Groos, O. V., & Pritchard, A. (1969). Documentation Notes. *Journal of Documentation*, 25(4), 344-349.
- Haak, L. L., Fenner, M., Paglione, L., Pentz, E., & Ratner, H. (2012). ORCID: a System to Uniquely Identify Researchers. *Learned Publishing*, 25(4), 259-264.
- Hirsch, J. E. (2005). An Index to Quantify an Individual's Scientific Research Output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569.

- Hou, H., Kretschmer, H., & Liu, Z. (2007). The Structure of Scientific Collaboration Networks in Scientometrics. *Scientometrics*, 75(2), 189-202.
- Latapy, M. (2008). Main-Memory Triangle Computations for Very Large (Sparse [Power-Law]) Graphs. *Theoretical Computer Science*, 407(1-3), 458-473.
- Leite, P., Mugnaini, R., & Leta, J. (2011). A New Indicator for International Visibility: Exploring Brazilian Scientific Community. *Scientometrics*, 88(1), 311.
- Luthi, L., Tomassini, M., Giacobini, M., & Langdon, W. B. The Genetic Programming Collaboration Network and its Communities. In *Proceedings of the 9th annual conference on Genetic and Evolutionary Computation, 2007* (pp. 1643-1650): ACM.
- Merton, R. K. (1968). The Matthew Effect in Science. *Science*, 159(3810), 56-63.
- Newman, M. E. (2001). Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks and Centrality. *Physical Review E*, 64(1), 016132.
- Newman, M. E. (2004). Coauthorship Networks and Patterns of Scientific Collaboration. *Proceedings of the national academy of sciences*, 101(suppl 1), 5200-5205.
- Newman, M. E. (2009). Random Graphs with Clustering. *Physical review letters*, 103(5), 058701.
- Newman, M. E., Watts, D. J., & Strogatz, S. H. (2002). Random Graph Models of Social Networks. *Proceedings of the national academy of sciences*, 99(suppl. 1), 2566-2572.
- Opsahl, T. (2010). Closeness Centrality in Networks with Disconnected Components. Available at: <https://toreopsahl.com/2010/03/20/closeness-centrality-in-networks-with-disconnected-components/>.
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. *Social Networks*, 32(3), 245-251.
- Otte, E., & Rousseau, R. (2002). Social Network Analysis: A Powerful Strategy, also for the Information Sciences. *Journal of information Science*, 28(6), 441-453.
- Tomassini, M., & Luthi, L. (2007). Empirical Analysis of the Evolution of a Scientific Collaboration Network. *Physica A: Statistical Mechanics and its Applications*, 385(2), 750-764.

Trigeorgis, L. (1996). *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. Cambridge (MA): MIT Press.

van Steen, M. (2010). *Graph Theory and Complex Networks - An Introduction* (vol. 144). United States: van Steen, Maarten.