A Network Theoretical Approach to Assess Knowledge Integration in Information Studies

José María Díaz-Nafría[1,2], Teresa Guarda[3,4,5], Iván Coronel[3]

[1] BITrum-Research Group, C/ San Lorenzo 2, 24007 León
[2] Munich University of Applied Sciences, Dachauerstr 100a, 80636 Munich, Germany
[3] Universidad Estatal Península de Santa Elena – UPSE, La Libertad, Ecuador´
[4] Algoritmi Centre, Minho University, Guimarães, Portugal
[5] Universidad de las Fuerzas Armadas-ESPE, Sangolqui, Quito, Ecuador
jdian@unileon.es, tguarda@gmail.com, iacoronel@hotmail.com

Abstract. The paper presents a general approach to assess knowledge integration as a basis to evaluate the performance of transdisciplinary and interdisciplinary approaches with respect to their knowledge integration capacity. The method is based on the development of Interdisciplinary-glossaries as tools for the elucidation of the conceptual networks involved in interdisciplinary studies. Such ID-glossaries are used as proxies of the corresponding knowledge integration, which is measured through the structural analysis of the co-occurrence network of terms. This approach is applied to an ID-glossary devoted to the general study of information, called glossariumBITri. The results show the capacity of the approach to detect integration achievements, challenges and barriers. Its qualitative nature is complemented by an enhanced methodology in which both the diversity of disciplines and the knowledge integration can be measured in a bi-dimensional index. To that purpose each contribution to the target ID-glossary is identified by the knowledge domains involved (using a set of knowledge domains adapted from the higher categories of the Universal Decimal Classification), while the integration is measured in terms of the small-world coefficient of the co-occurrence of terms.

Keywords: Interdisciplinarity, Network Analysis, Knowledge integration, Transdisciplinarity, Information Studies.

1 Introduction

A scientific discipline can be characterized by its conceptual network [1] [2]. The systematic relation among the nodes of the network enables the mapping of the objects and problems that such discipline is focused on. The concepts have thus not an isolated absolute value; this is rather gained in virtue of the capacity of the whole. At the same time, each concept enables that a knowledge domain can better approach a specific part of the reality it strives to gather (or provides an operational capacity to the other concepts in such endeavor). If a node is really worth, by taking it away, the whole network would lose its ability to address its field of interest: the network separates (partially or

Draft version of the article: Díaz-Nafría J.M., Guarda T., Coronel I. (2018) A Network Theoretical Approach to Assess Knowledge Integration in Information Studies. In: Rocha Á., Guarda T. (eds) Developments and Advances in Defense and Security. MICRADS 2018. Smart Innovation, Systems and Technologies, vol 94. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-78605-6_31

globally) from the reality that it is attempting to map. These are problems of a scientific domain operating by its own.

Different problems arise when various scientific disciplines need to be gather to join their knowledge with the purpose of addressing a complex issue which none of the isolated disciplines is capable to cope with by its own capacity (for instance, the understanding of the information phenomena across the different levels of reality, from the physical to the social aspects). In the endeavor of merging a set of disciplines, we can achieve different levels of integration. UNESCO distinguishes the following levels, organized from lesser to higher integration degree: multi-, pluri-, cross-, inter- and trans-disciplinarily [3]. At the lowest level, the multidisciplinary represents a simple juxtaposition of disciplines (i.e. they solve by their own the issues they are entrusted with). Therefore, the conceptual network of the domains involved do not interact significantly. Nothing needs to be changed in their respective conceptual network to address the problems tackled. However, transdisciplinarity, at the highest integration level, "assumes conceptual unification between disciplines" [ibid, p. 9]. In other words, the conceptual network of the disciplines involved blends into a unified operative framework. In between, interdisciplinarity embraces coordination and cross-communication among participant disciplines, but "the total impact of the quantitative and qualitative elements is not strong enough to establish a [unified framework,] a new discipline". From the perspective of the conceptual network, the common concepts (for instance, 'communication', 'message' or 'data' in the general study of information) often establish different relations with the rest of the combined conceptual network because the different value of the node (term/concept) at each domain.

Because of the lack of integration provided by multidisciplinarity, the international panel of experts, convened by the UNESCO in 1985, excluded it as a level of effective knowledge integration, and agreed to consider just three interdisciplinary levels: (i) pluridisciplinarity (the disciplines are just brought together with often few contact), (ii) interdisciplinarity (there is a good knowledge of each other's concepts) and (iii) transdisciplinarity (the conceptual unification is achieved) (ibidem). Hence, while (i) and (iii) represent the extremes, (ii) occupies a broad space in between. It is our purpose to provide means to assess the interdisciplinary level applied to the general study of information, i.e., the effective distance to (i) or (iii), which can also put in terms of the effective attainment in the integration of knowledge. To this end, we rely on: a) the development since 2009 of an interdisciplinary glossary of concepts devoted to foster the integration of knowledge in the general study of information, (named glossariumBITri) supported by an international interdisciplinary network of scientists: b) a network approach to evaluate the effective achievement of the conceptual network deployed through the elucidation process as crystallized in the glossariumBITri edition of 2016 [4]. Building upon these results, an enhanced methodology to assess knowledge integration is presented at the end.

2 Methodology

2.1 Interdisciplinary glossaries as tools for the integration of knowledge and the evaluation of the integration achieved

The concept of interdisciplinary glossary (ID-G), in which the glossariumBITri (gB) is based, differs significantly from the usual glossaries since they aim at clarifying what is meant by the terminology from the disciplinary perspective [4] [5]. The purpose of ID-G is, on the contrary, bringing together the different understandings of common terms from the summoning of various disciplinary perspectives. In the endeavour of transdisciplinarity, the meeting of view-points targets the conceptual unification, but through the elucidation process is possible to find that there are some irreducible understandings that are worth to keep in order to preserve the consistency and integrity of the respective theories.

According to this approach, the gB has been conceived as a tool for the conceptual and theoretical clarification in the study of information. It aims at embracing the most relevant viewpoints concerning information, relying on a board of experts coming from a wide variety of knowledge fields. From a theoretical viewpoint, gB aims to shorten the distances among the different viewpoints and increasing the linkages; while from a meta-theoretical view aims to assess the accomplishment of such integration. In other terms, gB serves as a proxy of the knowledge integration achieved by the interdisciplinary study of information; thus assessing the interdisciplinary degree in a manner that can be generalized for other knowledge integration undertakings, as proposed in the PRIMER initiative to foster interdisciplinary capacities, which is supported by the scholar network that backs up the gB [6].

2.2 Network approach to assess glossariumBITri's knowledge integration

According to the abovementioned characterization of the interdisciplinary dialogue, the evaluation of the interdisciplinarity degree or knowledge integration is based on the scrutiny of the structural properties of glossariumBITri's semantic network. To this purpose, the semantic network structure is derived from the meaning relations established by the authors in their own writings devoted to the elucidation of the conceptual network [7] [8]. In so far as the sentence formed by the speaker implies a unit of sense, the mere syntactic co-occurrence of words (grouped in sets of derivative words) in the space of a sentence establishes a semantic linkage that can be explored in terms of the frequency of such links [9]. For instance, if we observe a high repetition in the co-occurrence of "complexity" and "algorithmic", on the one hand, o "message" and "meaning", on the other, is due to the semantic proximity of the co-occurring terms; in one case because of equivalence relation, in the other, because of consequence relation. In short, the greater or lesser occurrence of terms and links between terms have facilitated the examination of the relevance of different categories and the links between them from the perspective of the interdisciplinary research network. At the same time, the formation of semantic networks in the texts analysed with "small-world" or "scale-free" characteristic structures, whose pertinence has been studied, enables the visualization of both the categories effectively used in the generic articulation of utterances and the grouping of verbal categories circumscribed by the dealing with specific issues, for instance, "complexity" [10] [11].

According to this characterization, the semantic network analysis has been structured in the following phases:

(i) Text refinement, getting rid of those elements not corresponding to (textually) expressed utterances for which a meaningful syntactic-semantic treatment could not be performed.
(ii) Quantitative analysis of the texts by means of the application of computational linguistics "KH Coder" which enables the analysis of the semantic network in terms of the semantic links observed in the texts through the adjacency distance in sentences [11] [12].
(iii) Iterative process of relevant terms refinement according to its significance for the analysed issues which enables reviewing the aprioristic categorisation.
(iv) Co-occurrence mapping extraction of the semantic networks derived from the conceptual elucidation of the glossariumBITri.

3 Findings

3.1 Network analysis of the *glossariumBITri - 2nd Edition*

The result shown in figure 1 illustrates a relevant characteristic of the glossariumBITri: the statistical degree distribution of the semantic network exhibits the properties of the free-scale networks. This means that the subsidiarity properties discussed by Díaz-Nafría [8] can be applied to glossariumBITri's semantic network. The recursive character of the corresponding structure entails a disciplinary clustering of issues that, at the same time, are well connected to the rest of the network from a semantic perspective.

The statistics of the semantic distances observed in the network and the study of the clustering offers an innovative methodological road to strengthen the interdisciplinary study of information linked to the development of the gB.

Figures 2, 3 and 4 shows the results of the gloossariumBITri's semantic network analysis. Each term/concept is represented by a node whose size is proportional to its occurrence frequency, while the thickness of links among terms is proportional to the frequency co-occurrence of the corresponding terms in the sentences of the whole text. Only the terms and links whose frequency surpasses the thresholds indicated in the figure caption are visible. At the same time, the result of the analysis of term clusters determined by intermediation distances is represented using different colours (terms with the same colour are at distances below a threshold).

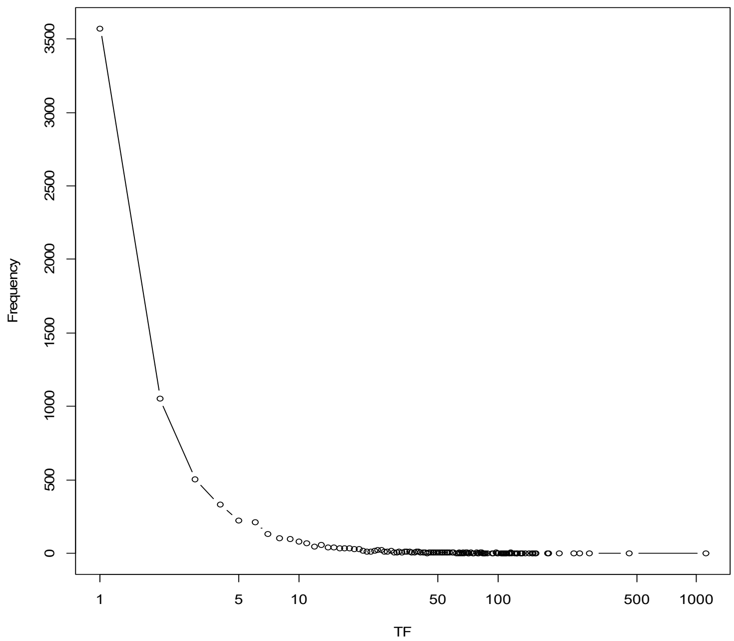

Fig. 1. Distribution of the frequency of word occurrence in the glossariumBITri 2016 edition. The statistic parameters denote that the semantic network is of the type small-world and free-scale.

As we can observe in Fig. 2 (in which the 130 most frequent terms are represented), “information” is the most dominant terms, as it could be expected. Under this nuclear term we can find other outstanding terms: ‘theory’, ‘communication’, ‘knowledge’, ‘use’, and ‘concept’. They reflect, on the one hand, the general objective of the glossariumBITri (concept, theory), on the other, a significant weight of theoretical terms as communication, knowledge, use. We can also observe 4 important clusters, corresponding to domains with capacity to concentrate some specific aspects that have experience a deeper development. In addition, we only find two dominant authors, Shannon and Kolmogorov. However, while Shannon appears at a relatively central position and with a high degree of interconnectedness with the rest of the network, Kolmogorov is located at the central position of a cluster which is less connected and is more peripheral, linked to important theoretical terms as algorithm, complexity, object and other more mathematically oriented as fuzzy, set, function, etc. This cluster corresponds to one of the theoretical domains which has been incorporated in the 2016 glossariumBITri edition. Its relative disconnection with other relevant terms points to the need to devote efforts in developing missing links in order to achieve a more integrated elucidation.

Finally, it is possible to observe in the blue cluster (Fig. 2) a particularly cohesive group, composed by the terms: ‘society’, ‘media’, ‘technology’, ‘communication’, and ‘critic’. Here we observe the weight of a field developed in depth, the critical theory of information focused on human, social and political aspects of information technologies.

It is worth mentioning that the article "critical theory of information" is the most read article in the interactive-glossariumBITri as determined by Internet traffic analytics.

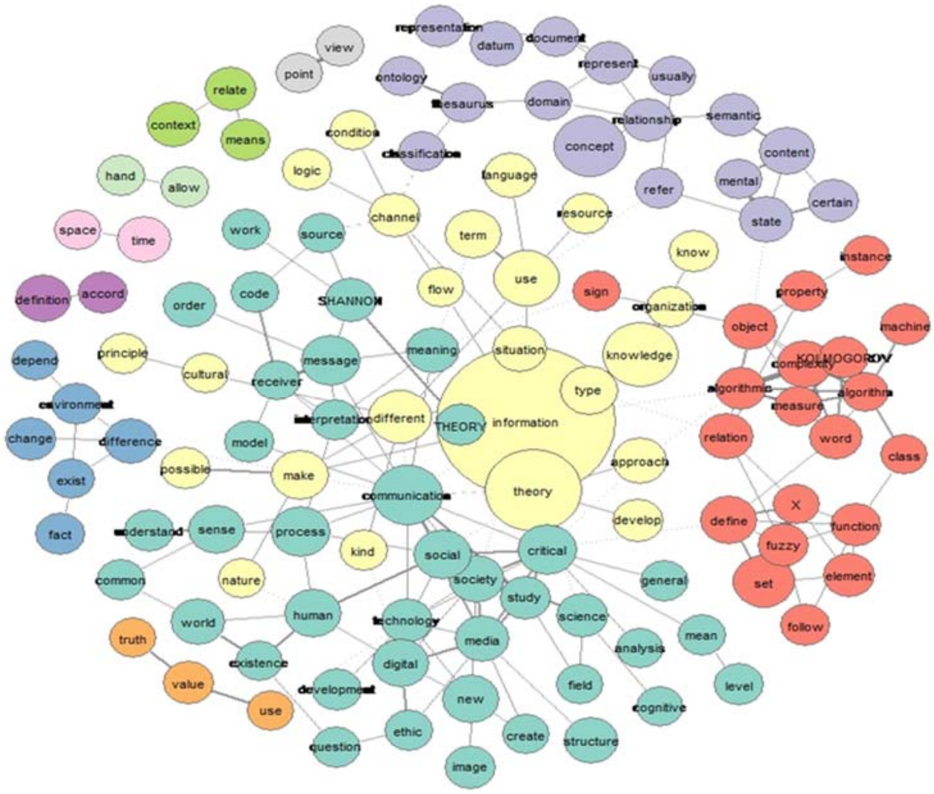

Fig. 2. glossariumBITri's Co-occurrence network. Term frequency occurrence > 50; Number of nodes (words-concepts): 200 most frequent ones; Colours: semantic clusters determined by intermediation measurements. Adverbial and prepositional categories are excluded.

Figure 3 corresponds to the same co-occurrence network in which only the 58 most frequent terms (nodes) are visualized (with a frequency over 75) and the 100 most frequent co-occurrences (links). According to the clustering analysis, the largest cluster is again the one aligned to the critical theory of information. At the same time the well cohesive cluster of terms related to algorithmic information theory and the General Theory of Information. At this level it can be noted that the red cluster (Fig. 2) originates from the convergence of the cluster composed by set, function, and other terms that was extensively developed in the previous edition (e.g. fuzzy logic). In both this figure and the previous one, it is interestingly possible to observe the presence of Bateson's conceptual approach to information, stated in the famous formula: "*information* is a *difference* that *makes* a *difference*" [13], which over time has gained most general support among the varied community of information studies. As we can see, this conception establishes relevant links to "environment" which reflect the concern spread along the community of information studies to go beyond the de-contextualization which is inherent to Shannon's perspective [13] [14].

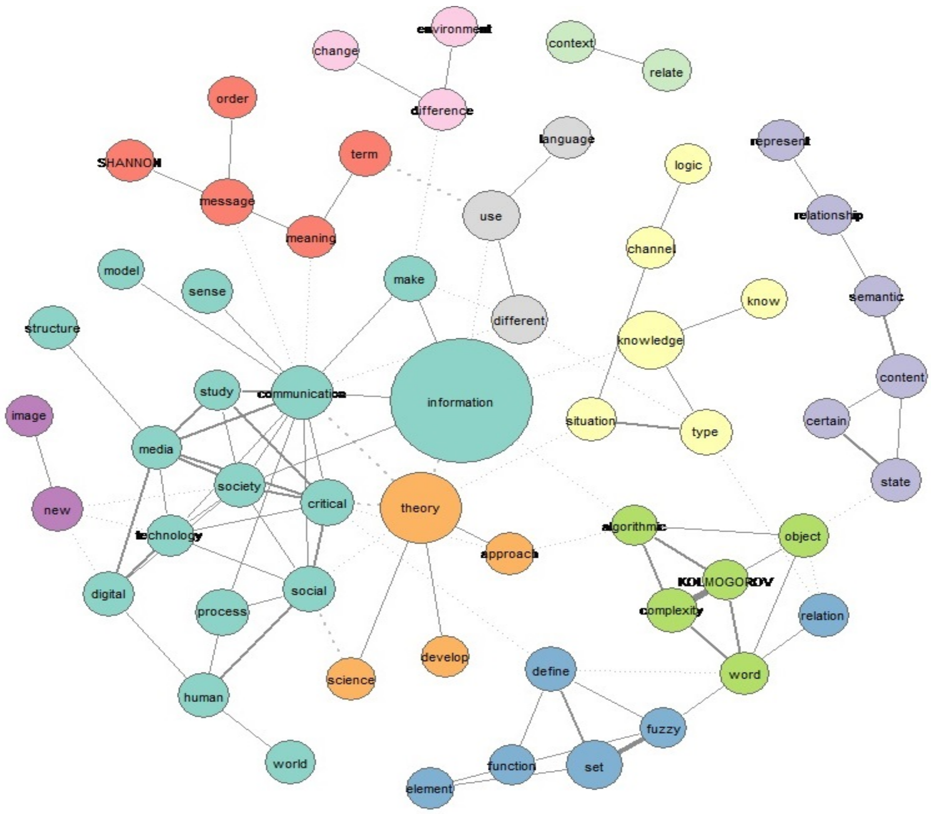

Fig. 3. glossariumBITri's Co-occurrence network. Term frequency occurrence > 50; Number of nodes (words-concepts): 200 most frequent ones; Colors: semantic clusters determined by intermediation measurements. Adverbial and prepositional categories are excluded.

Finally, Fig. 4 corresponds to a further refinement of the previous co-occurrence network in which only the 6 most frequent terms (nodes) and the 100 most frequent co-occurrence (links), which in the figure are reduced to the 12 existing among the 6 visualised terms. We observe here the 4 heaviest conceptual terms upon which the rest of the conceptual elucidation is articulated, as well as two meta-theoretical terms (theory and concept) which manifest the very goal of the gB itself. It is also worth mentioning at this level between communication and use, what shows that the gB, effectively accomplish the objective of giving account of the pragmatic aspects that was missing in the Mathematical Theory of Communication from which Shannon forged the scientific concept of information. From the inspection of the three co-occurrence networks an absence pointing to a direction of further development and improvement of the gB: a more specific and broader consideration of metaphors. In a network structural perspective, the benefit of metaphors relies on their capacity to reduce average distances in the whole conceptual network as discussed by Díaz-Nafría [8].

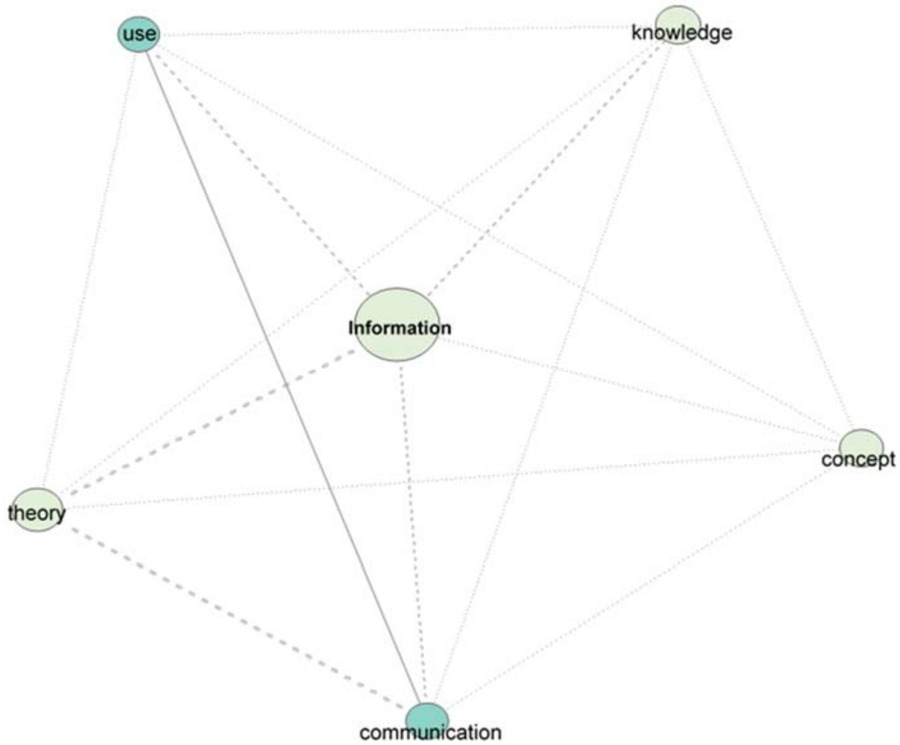

Fig. 4. glossariumBITri's co-occurrence network. Term frequency occurrence > 200; Number of nodes (words-concepts): 6 most frequent ones; Number of links: 100 most frequent; Colors: semantic clusters determined by intermediation measurements.

3.2 Enhanced Methodology to Assess Knowledge Integration

The previous results exhibit the capacity of the network approach to qualify interdisciplinarity within the broad margin left by the UNESCO's classification (referred to in section 1), i.e. how far apart is from transdisciplinarity. However, this approach has not addressed how diverse the integration of knowledge is with respect to scientific knowledge in general. In addition, it provides a rather qualitative assessment that hinders the possibility of an objective evaluation. To fill the gap, building upon the referred approach, we propose–for future development–the assessment of the quality of the knowledge integration, based on two general aspects: the diversity of the disciplines involved (the more disciplines the larger the integrated knowledge); and the effective integration achieved through the meeting of different perspectives (if each discipline treats separately different aspects, the integration will be weak; if the theoretical construct gets to be merged into a general understanding of the involved phenomena, the integration will be strong)

Discipline Diversity Index. In the first place, the granularity level in the distinction of disciplines have to be determined. This can be done, in a first approximation, by fixing the number of relevant digits of the Universal Decimal Classification (UDC) used to distinguish the knowledge areas involved in a particular research [16] [17] [18]. Though the UDC offers a good and well-accepted coverage of knowledge in general, an adaptive implementation need to be introduced in the categorization of Knowledge Domains (KD): (i) some UDC categories have to be disregarded (for instance those which are not related to knowledge but to document types), (ii) other categories should be ascended from a lower granularity level in virtue of its relevance for the problems under study, and (iii) some category groups should be merged because they represent different aspects of the same knowledge, for instance, theoretical and applied.

Assuming the number of relevant KD is N, the diversity of participating disciplines can be determined through Shannon Diversity Index weighted by the maximal diversity achieved through a similar participation of the N KD, i.e. log2 N. By that means, if the N KD are homogenously distributed (i.e. they contribute equally –situation of maximal diversity) the index will be 1; and 0 in case that only one KD is contributing. Generally, the more KD are contributing in a more distributed way, the index will be closer to 1.

Calling pi the frequency of occurrence of a contribution from the ith KD (or probability that a contribution taken at random belongs to such a discipline), the diversity index will be:

$$ID = \frac{1}{\log_2 N} \sum_{i=1}^{N} p_i \log_2 \left({}^{1}\!/_{p_i} \right) \qquad (1)$$

Integration of Disciplines. Nevertheless, it can be the case that despite having achieved the meeting of very diverse knowledge, its theoretical constructs do not merge at all in the explanation of the phenomena concerned, and instead each discipline devote itself to refer a different aspect of the object or problem under study. In this case the integration would be null. In the extremely opposite case, all the theoretical constructs from each involved discipline are interrelated in the explanation of the phenomena concerned.

This density of semantic relation can be measured in terms of the semantic network conformed by the theoretical terms used by the different disciplines convened in the process of interdisciplinary elucidation of concepts, metaphors, theories and problems concerning a shared problem. The development of an interdisciplinary glossary (ID-G), according to the model discussed in section 2 and devoted to a specific topic, will serve to the interdisciplinary theoretical confrontation of different points of view. In such constructive process the agents establish relations between two terms whenever they occur in the same sentence, in lesser degree if they belong to subordinate ones. Thus the statistical analysis of the distance between terms in function of its grammatical function enables the analysis of the semantic distances among terms.

Quantitative assessment of knowledge integration, following the methodology depited above. The study of the minimal average distance between any two words provides a measure of the integration achieved. In the case of natural language, taken an extended vocabulary of 66000 words, Sigman and Cecci [19] determined that the average minimal distance between any two words was around 7. However, when the knowledge is not well integrated, the distances increase at the same time that disconnected clusters can be identified. Thus, clustering coefficient and average minimal distance offer a characteristic of the integration achieved. Indeed, its ratio compared with the equivalent ratio for random networks, provides the small-world coefficient:

$$\sigma = \frac{C/C_{rand}}{L/L_{rand}} \tag{2}$$

from which we can evaluate, with a single parameter, whether the network satisfy or not the condition of a small-world network $\sigma > 1$ and how well integrated it is [20].

Qualitative assessment of knowledge integration. The network analysis, as the one described in section 3.1, facilitates a qualitative evaluation, distinguishing specific theoretical clusters that are not well integrated, fields or concepts that are misrepresented, etc. This evaluation provides guidance for the further development of the research concerned (e.g., what disciplines need to be strengthened, what dialogue should be open up, etc.).

4 Discussion and Conclusions

In spite of the international efforts to boost interdisciplinary research, one of the barriers has been the lack of qualification criteria of interdisciplinarity itself [20] [21] [23]. This has often caused the funding assessment inefficient, disregarding promising ID research projects to which disciplinary criteria were applied. For this reason, the development of assessment criteria has been one of the objectives marked by national and international research funding agencies [21] [23]. The results discussed above shows the interest of the ID-glossaries in combination with the network analysis as a promising approach to qualify interdisciplinarity. But the benefit is not only meta-theoretic, at the level of the knowledge integration itself, it constitutes a useful tool in the advancement of knowledge integration, as it could be shown in the discussion of the co-occurrence networks with respect to the evolution of the glossariumBITri between consecutive editions. If over time the network analysis is performed to facilitate a comparative assessment of the evolution of the knowledge integration achieved, it is expected that this approach will serve to guide the theoretical work, as illustrated with a few examples in the results argued above.

The possibility to use this approach to the assessment of educational processes has been discussed by one of the authors, showing its capacity to detect the development of soft skills for which formal education is practically blind [23]. Its application to the development of knowledge integration skills, as intended in the abovementioned PRIMER initiative [6], is straightforward from the methodology and results discussed herewith.

Nevertheless, the approach, on which the results presented is based, do not provide a quantitative evaluation, which prevents an objective assessment. To circumvent this limitation, the enhanced methodology discussed in section 3.2 fills the gap with a bi-dimensional measure in which both the diversity of disciplines and the effective integration is measured at the same time. An ongoing international project devoted to the enhancement of the glossariumBITri and the creation of an Encyclopaedia of Systems and Cybernetics Online (ESSCO) is currently applying the described categorization of knowledge domains to deploy the described approach.

References

1. C. G. Hempel, Fundamentals of Concept Formation in Empirical Science, vol. 2, University of Chicago: University of Chicago Press, 1952.
2. J. Losee, A Historical Introduction to the Philosophy of Science, New York: OUP Oxford, 2001.
3. L. d'Hainault, Interdisciplinarity in General Education, Paris: UNESCO, 1986.
4. J. M. Díaz-Nafría, F. Salto and M. Pérez-Montoro, Glossarium BITri 2016 : Interdisciplinary Elucidation of Concepts, Metaphors, Theories and Problems Concerning Information, La Libertad (Ecuador)-León (Spain): Santa Elena : Universidad Estatal Península de Santa Elena, 2016.
5. L. R. Lattuca, Creating Interdisciplinarity: Interdisciplinary Research and Teaching Among College and University Faculty, Chicago: Vanderbilt University Press, 2001, pp. 1-20.
6. PRIMER consortium, "PRomoting Interdisciplinary Research and Education," BITrum-Research Group, 2018. [Online]. Available: http://primer.unileon.es. [Accessed 20 1 2018].
7. P. Drieger, Semantic network analysis as a method for visual text analytics. *Procedia - Social and Behavioral Sciences*, vol. 79, pp. 4-17, 2013. https://doi.org/10.1016/j.sbspro.2013.05.053
8. J. M. Díaz-Nafría, Cyber-Subsidiarity: Toward a Global Sustainable Information Society. In E.G. Carayannis et al Handbook of Cyber-Development, Cyber-Democracy, and Cyber-Defense, Berlin: Springer Verlag, 2017, pp. 1-30.
9. K. M. Jackson and W. M. Trochim, "Concept Mapping as an Alternative Approach for the Analysis of Open-Ended Survey Responses," Organizational Research Methods, vol. 5, no. 4, pp. 307-336, 2002.
10. A.-L. Barabási, The New Science Of Networks, vol. 30, Cambridge: Perseus Publishing, 2002, p. 288.
11. K. Higuchi, KH Coder 3 Reference Manual, Ritsumeikan University, Kioto (Japan), 2016. [Online]. Available: https://goo.gl/CeWSeQ [Accessed 20 1 2018].
12. S. Anzai, & Ch. Matsuzawa, "Missions of the Japanese National University Corporations in the 21st Century: Content analysis of mission statements", *Academic Journal of Interdisciplinary Studies*, vol. 2, no. 3, pp. 197-207, 2013. https://doi.org/10.5901/ajis.2013.v2n3p197
13. G. Bateson, Mind and Nature: A Necessary Unity, New Jersey Hampton: Hampton Press, 2002.
14. J. M. Díaz-Nafría, What is information? A multidimensional concern, TripleC , vol. 8, no. 1, pp. 77-108. [online] Available: http://www.triple-c.at/index.php/tripleC/article/view/76, [Accessed 20 01 2018]

15. J. M. Díaz-Nafria, Information: A multidimensional reality, in E. Curras and N.L. Romero (eds.), Systems Science and Collaborative Information Systems: Theories, Practices and New Research: Theories, Practices and New Research, IGI Global, 2011, pp. 37-70.
16. I. C. McIlwaine, The Universal Decimal Classification: A Guide to Its Use, UDC Consortium, 2007.
17. A. Slavic, Innovations in Information Retrieval: Perspectives for Theory and Practice., London: Facet Publishing, 2011, pp. 23-48.
18. C. UDC, "UDC Fact Sheet. In UDCC website, The Hague," 2017. [Online]. Available: http://www.udcc.org/index.php/site/page?view=factsheet.
19. M. Sigman and G. Cecchi, "Global organization of the Wordnet lexicon," vol. 99, no. 3, pp. 1742-1747, 2002. https://doi.org/10.4018/978-1-61350-201-3.ch003
20. Q. K. Telesford, K. E. Joyce and S. Hayasa, "The Ubiquity of Small-World Networks. Brain Connectivity," Brain Connectivity, vol. 1, no. 5, pp. 367-375, 2011.
21. R. Frodeman, J. Thompson Klein and R. Car, The Oxford Handbook of Interdisciplinarity, Oxford, UK: Oxford University Press, 2017.
22. DEA-FBE, Thinking Across Disciplines – Interdisciplinarity in Research and Education, Australia: Danish Business Research Academy (DEA) & Danish Forum for Business Education (FBE)., 2008.
23. European Research Advisory Board (EURAB), Interdisciplinarity in Research, European Commission - EURAB, 2004.
24. E. Díez-Gutiérrez and J. Díaz-Nafría, "Ubiquitous learning ecologies for a critical cybercitizenship," Comunicar, vol. XXVL, no. 54, pp. 54, 49-58, 2018.