Author(s): Lahti, Lauri

Title: Generation of learning paths in educational texts based on vocabulary co-occurrence networks in Wikipedia and randomness

Year: 2015

Version: Post print

Lahti, L. (2015d). Generation of learning paths in educational texts based on vocabulary co-occurrence networks in Wikipedia and randomness. Proc. Global Learn 2015: Global Conference on Learning and Technology, 16-17 April 2015, Berlin, Germany. Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, USA.

Generation of learning paths in educational texts based on vocabulary co-occurrence networks in Wikipedia and randomness

Lauri Lahti
Department of Computer Science
Aalto University School of Science, Finland

Abstract: We propose a new computational method for generating learning paths in educational texts. The method relies on forming vocabulary co-occurrence networks among articles of Wikipedia online encyclopedia and exploiting random explorations to generate route weighting parameters to form pedagogic co-occurrence networks. Motivated by previous research about scale-free small-world networks we suggest our networks to offer efficient and intuitive properties for knowledge representation and learning. In the context of cell biology we provide experimental results about the properties of the linkage emerging in the vocabulary co-occurrence networks in a set of 175 Wikipedia articles and contrast it with the linkage emerging in the corresponding hyperlink network in Wikipedia. Furthermore we describe formation of the pedagogic co-occurrence network that can be exploited to recommend learning paths for the student.

Introduction

We suggest that various kinds of learning contents (in diverse forms, including for example collections of texts, images, videos and concept maps) can be modeled with a compact efficient hierarchical modular structure relying on scale-free small-word networks. Furthermore we suggest that in almost all everyday knowledge available in school books, newspapers, online resources etc. it is possible to identify structures and dynamics of scale-free small-word networks, or at least this everyday knowledge can be easily transformed to be represented in a form of scale-free small-word networks. Therefore we claim that it is important to carry out analysis about properties of various knowledge structures to identify how the principles of scale-free small-word networks actually emerge, manifest and behave in educational contexts so that these properties could be exploited to recommend efficient and intuitive learning paths for the student.

Previous research

There are well over 54 000 word families in English (Nation & Waring 1997). Knowing 95 percent of words in text can be sufficient for reasonable comprehension and can be reached with a vocabulary of 3000–5000 word families or just 2000–3000 word families (Nation & Waring 1997; Laufer 1989). People respond more quickly to high-frequency words of a language than low-frequency words of a language in respect to for example lexical decision, reading aloud, semantic categorization and picture naming (Duyck et al. 2008). Both in native language and secondary language acquisition words learned earlier in a person's life can be recognized and produced more quickly than words learned later in life and it has been suggested that mappings between orthographic, phonological and semantic representations of words form a network that supports later reconfigurations for new associations but still favours connections learned early in language acquisition (Izura & Ellis 2002; Ellis & Lambon 2000).

Simple English edition of Wikipedia online encyclopedia (http://simple.wikipedia.org) containing about 109 000 articles as of November 2014 is specifically tailored to represent knowledge content with a simple vocabulary and grammar, and thus seems to offer a useful educational resource supplied with a hyperlink network connecting related articles. High dominance of nouns as the titles of Wikipedia articles is fruitful for our aim to parallel knowledge structures of humans and Wikipedia since there is a noun dominance in children's early word learning in both language production and comprehension (Gentner & Boroditsky 2009). Based on a random sample of 1000 Wikipedia articles we estimated that about 13.8 percent of Wikipedia articles describe a topic titled with a common noun (Lahti 2015a).

In structural and functional human brain networks a small-world topology has been indentified (Wang et al. 2010) and in functional brain networks scale-free properties possibly hold (Bullmore and Sporns (2009). Wikipedia holds scale-free small-world properties (Masucci et al. 2011) and represents a hierarchical structure following so called power law, and the distribution of category sizes s has been estimated to be proportional to s-lambda with the parameter lambda having a value of about 2.2 and a similar kind of power law decay emerged in a link-based cluster size distribution (Capocci et al. 2008). The hyperlink network of Wikipedia has been found to be scale-free concerning ingoing links, outgoing links and broken links, and article sizes were lognormal distributed having a linear growing median (Voß 2005). Our previous experiments showed that to connect all 3710 unique nouns we identified in the highest language ability level C2 of English Vocabulary Profile based on hyperlink network of Wikipedia results in 25153 unique links containing 2878 unique nouns (Lahti 2014).

It has been identified that inter-retrieval intervals for both paths of animals searching for food with relatively optimal strategies and for human memory category recall resemble random walks called Lévy walks with power law having exponent alpha of about 2 (Thompson et al. 2013). By contrasting recall of concepts belonging to a category of animals and a semantic network model based on 5701 Wikipedia pages about animals it was shown that semantic memory processes can be usefully modelled as searches over scale-free networks (Thompson et al. 2013). During six weeks in 32-54 sessions with a snowball sampling paradigm students could grow individual associative networks reaching 1358-9429 nodes and 3729-27124 directed links showing a small-world structure with average shortest paths between any two nodes being in the range of 5.65-7.05 links (Morais et al. 2013). A thesaurus holds small-world properties and when performing a walk in a corresponding conceptual network always leads to a cycle whose period depends on a desired memory window (i.e. how many preceding visited nodes remain to be avoided at each step) (Kinouchi et al. 2002).

On a synaptic level long-term memory formation can be activated by 4-5 spaced puffs of serotonin (Kandel 2001) or by at least three action potentials separated with at least 10 minutes to make activation to remain for about 30 minutes (Fields 2005), and four brief trains for four days could generate memories that lasted weeks (Kandel 2001). In learning gradually expanding spacing of repetitions does not require identifying the optimal constant spacing time first (Balota et al. 2007) and has been considered to outperform constant spacing if the learner does not get feedback on his retrievals (Thalheimer 2006).

Method

We propose a new computational method for generating exploration paths in educational texts. Our method relies on making analysis about statistical properties of a collection of Wikipedia articles concerning a desired learning topic. Based on a suitable learning topic defined by a student or an educator the method retrieves all Wikipedia articles belonging to a Wikipedia article category corresponding to the given learning topic. Each of these articles defines a concept depicted by the title entry of the article.

Partly following the notation used in our previous work (Lahti 2015a) we now consider that all of the titles of articles belonging to the retrieved set of articles form together a set of concepts called as a *learning topic vocabulary*. Then the method identifies all occurrences of the concepts of the learning topic vocabulary in each article of the retrieved set of articles. Now for each article the method forms a set of conceptual relationships based on links leading from the concept of the title of the article to all occurrences of concepts of the learning topic vocabulary in this article. Then conceptual relationships of all articles together form a conceptual network called as a *vocabulary co-occurrence network*.

For each conceptual relationship the method defines a compact *relation statement* that is extracted from the sentence surrounding the occurrence of the concept of the learning topic vocabulary in the article text to depict the semantic relationship between the concept of the title of the article and the concept of the learning topic vocabulary in the article text.

To generate recommendations for the student about pedagogically rewarding exploration paths, called as *learning paths*, in the vocabulary co-occurrence network the method firstly aims to identify a ranking about the most essential conceptual relationships in the network and then secondly aims to suggest some fertile routes for exploration by chaining these relationships, as we explain next.

The method traverses a random exploration path in the vocabulary co-occurrence network and computes the highest-ranking traversed concepts and the highest-ranking traversed links in this path. Now the method examines the list of the highest-ranking traversed concepts in a descending order of the number of traversals and one by one selects a specific set of links concerning this current concept from the list of the highest-ranking traversed links. This selection is carried out so that for the current concept the method examines the list of the highest-ranking traversed links in a descending order of the number of traversals and selects only links with a gradually increasing spacing in respect to the ranking position of the link in the list of links (for example with a spacing that constantly doubles) and a gradually decreasing number of links per current concept. To keep the amount of concepts and links manageable we suggest that for example 3-5 highest-ranking concepts and 9-20 highest-ranking links become selected.

If the already gathered collection of links does not yet form a fully connected entity of links the method still begins to examine the list of the highest-ranking traversed links in a descending order of the number of traversals and selects such a minimal amount of additional links that enables all links to form a fully connected entity. After that the method may still select additional links if many concepts of the links seem to be connected by only one or two links. We call the fully connected entity of the links as a *pedagogic co-occurrence network.* To identify how to emphasize links belonging to a pedagogic co-occurrence network in exploration by the students the method traverses a random exploration path in the pedagogic co-occurrence network and computes the highest-ranking traversed concepts and the highest-ranking traversed links in this path. Before staring to explore the network with a random path it may be necessary to make all links bidirectional (or to allow rolling back in exploration when arriving to an dead-end). The proportional number of traversals per each concept and the proportional number of traversals per each link in this random exploration path form together a set of values that we call as *route weighting parameters*.

The method then traverses a set of random exploration paths in the pedagogic co-occurrence network and tries to generate such a random exploration path that most closely matches the *route weighting parameters* and in addition follows the principles of spaced learning so that traversing the same concept or the same link again later along the same path has an optimal spacing in respect to recommendable values identified in previous research about the principles of spaced learning. These two concurrent requirements are measured with a sum of two utility functions $UF_{weighting}$ and $UF_{spacing}$.

$UF_{weighting}$ measures an inverse value of the sum of proportional differences between the number of traversals of concepts (or links) along the current random exploration path and the corresponding route weighting parameters. $UF_{spacing}$ measures an inverse value of the sum of proportional differences between the spacing distance of traversing the same concept (or the same link) again later along the same path in the current random exploration path and the spacing distance based on the proportion of the number of traversals of concepts (or links) along the path and the number of different concepts (or links) in the network. The most matching random exploration path (i.e. the path reaching the highest value of the sum $UF_{weighting} + UF_{spacing}$) is then selected by the method and provided to the student as a recommendable exploration path that we call as a *randomness-based learning path*.

Experiment

We retrieved from Simple English edition of Wikipedia all articles belonging to the category of "Cell biology" (http://simple.wikipedia.org/wiki/Category:Cell_biology) in January-February 2015, altogether 175 articles. Due to space constraints details about the results we report here and all 175 articles are shown in publications (Lahti 2015a, 2015b). Each of 175 articles defines a concept depicted by the title of the article. Thus we consider that these 175 articles together form a vocabulary co-occurrence network of corresponding 175 concepts and furthermore enables to generate a pedagogic co-occurrence network linking these concepts. We first report some measures about the textual contents of the articles we estimated with semi-automated and thus somewhat coarse statistical methods. All articles together contained about 49000-54000 words of which 5527 words were unique or after eliminating some redundant -s/-es suffixes 4829 words were unique. All articles together contained about 3017 sentences. There were 1-84 sentences (average 17.24 and median 13) per article, and 26-1646 words (average 308.41 and median 229) per article. If assuming a typical reading speed of 200 words per minute then reading an average article takes 1.54 minutes and reading all the articles takes 269.86 minutes (about 4.50 hours).

For the collection of 175 concepts we contrasted the linkage emerging in the vocabulary co-occurrence network based on Wikipedia article texts with the linkage emerging in the hyperlink network based on Wikipedia article texts.

Vocabulary co-occurrence network: There are 1338 unique links of co-occurring words among 175 concepts, this means 1132 unique conceptual pairs that are linked (either unidirectionally or bidirectionally). In addition there are 164 links from an article back to itself (i.e. the title of article becomes mentioned in the article text for 164 concepts of 175 concepts). These 1338 unique links contain 173 of 175 concepts, and thus two concepts that did not appear in links are catabolism and photobleaching. In links of co-occurring words there are 147 unique end concepts of 175 concepts and 173 unique start concepts of 175 concepts. An article has in its text 1-29 unique different concepts of 175 concepts mentioned (average 7.73 and median 6). Thus in links each start concept has on average 7.73 (median 6) different unique end concepts. Each of 175 concepts is mentioned in texts of 1-141 unique different other articles of 175 articles (average 9.10 and median 4). Thus in links each end concept has on average 9.10 (median 4) different unique start concepts. When making all unique 1338 links bidirectional and removing duplicates there is a network of 2264 unique links of co-occurring words.

Hyperlink network: There are 816 unique hyperlinks among 175 concepts, this means 705 unique conceptual pairs that are linked (either unidirectionally or bidirectionally). These 816 unique hyperlinks contain 167 of 175 concepts, and thus eight concepts that did not appear in links are catabolism; glial cell; growth factor; guard cell; karl wilhelm von nägeli; lipid bilayer; photobleaching and robert remak. Thus connectivity of hyperlinks is about 61 percent (816/1338) of connectivity of links of co-occurring words and conceptual coverage of hyperlinks is about 97 percent (167/173) of conceptual coverage of links of co-occurring words. Thus the number of links of co-occurring words is 1.6 times the number of hyperlinks. In hyperlinks there are 166 unique end concepts of 175 concepts and 105 unique start concepts of 175 concepts. In hyperlinks each start concept has on average 4.92 (median 4) different unique end concepts. In hyperlinks each end concept has on average 7.78 (median 5) different unique start concepts. When making all unique 816 hyperlinks bidirectional and removing duplicates there is a network of 1410 unique hyperlinks.

Discussion and future work

The vocabulary co-occurrence network and the hyperlink network share 778 links and when making all the links bidirectional they share 1380 links thus meaning that both networks have 690 shared conceptual pairs that are linked (either unidirectionally or bidirectionally). These 690 shared conceptual pairs represent about 61 percent of the conceptual pairs of the vocabulary co-occurrence network and about 62 percent of the conceptual pairs of the hyperlink network.

In the links of the vocabulary co-occurrence network some start concepts having the highest number of end concepts are mitochondria (29), organelle (29), meiosis (28), eukaryote (24) and plant cell (22), and some end concepts having the highest number of start concepts are cell (141), gene (79), chromosome (59), dna (51) and membrane (45). In the links of the hyperlink network some start concepts having the highest number of end concepts are meiosis (22), organelle (18), chromosome (17), mitochondria (17) and eukaryote (16), and some end concepts having the highest number of start concepts are chromosome (49), cell (47), dna (46), cytoplasm (30) and organelle (30).

We generated a random exploration path of 300 000 traversals bidirectionally for both the vocabulary co-occurrence network and the hyperlink network thus identifying the following overlap (see Table 1). Among 10 most traversed links both networks shared 1 conceptual pair (between lysosome and vacuole), among 50 most traversed links they shared 3 conceptual pairs (lysosome and vacuole, cell differentiation and stem cell, mutation and toll-like receptor) and among 100 most traversed links they shared 9 conceptual pairs (lactive transport and cell membrane, lysosome and vacuole, cell and toll-like receptor, cell differentiation and stem cell, mutation and toll-like receptor, prokaryote and vacuole, cell membrane and macrophage, b cell and lymphocyte, meiosis and somatic cell). Among 10 most traversed concepts both networks shared 7 concepts, among 50 most traversed concepts they shared 42 concepts and among 100 most traversed concepts they shared 88 concepts.

Table 2 shows how based on a random exploration path in the vocabulary co-occurrence network (300 000 traversals bidirectionally) we generated a pedagogic co-occurrence network containing altogether 18 links and 12 concepts. First, based on the highest-ranking traversed concepts in a descending order of the number of traversals we selected for the concept "cell" from the list of the highest-ranking traversed links the links having ranks 1, 3, 6 and 11, for the concept "gene" the links having ranks 1, 3, 6, and for the concept "chromosome" the links having ranks 1 and 3. After that we added four links to enable all links to form a fully connected entity and then – since many concepts of the links were connected by only one or two links – still five additional links to further increase connectivity.

To define the route weighting parameters for the pedagogic co-occurrence network we carried out a random exploration path (100 000 traversals bidirectionally), and Table 2 shows the number of traversals and their rankings for the links of the pedagogic co-occurrence network. The number of traversals and their rankings for the concepts of the pedagogic co-occurrence network in a descending order of ranking are: gene (21454), cell (17735), chromosome (10877), y chromosome (7279), cell theory (7251), plastid (7083), flagellum (7080), cytokine (7045), haploid (3644), amoeboid movement (3558), cyanobacteria (3523) and thomas hunt morgan (3472).

We suggest that that the spacing of traversals of the same concepts and the same links during exploration along the learning path could rely on for example 3-5 repetitions about 10 minutes apart during several

Table 1. Ten most traversed links and ten most traversed concepts in a random exploration path of 300 000 traversals bidirectionally for both the vocabulary co-occurrence network and the hyperlink network among a set of 175 Wikipedia articles about cell biology.

Vocabulary co-occurrence network				Hyperlink network			
Ten most traversed links		Ten most traversed concepts		Ten most traversed links		Ten most traversed concepts	
Link	*Traversals*	*Concept*	*Traversals*	*Link*	*Traversals*	*Concept*	*Traversals*
membrane→light-dependent reaction	179	cell	18640	mitochondria→symbiosis	260	cell	11317
white blood cell→lymphocyte	173	gene	10503	monocyte→red blood cell	258	chromosome	10985
cell membrane→syncytium	172	chromosome	8072	transcription (genetics)→dna	257	dna	10499
cell biology→molecular biology; cell theory→dna; plant cell→plasmodesma	170	dna	6998	gene expression→eukaryote	253	organelle	7931
pseudopodia→cilia	168	organelle	6442	adenosine triphosphate →mitochondrial dna; centriole→plant cell; molecular biology →transcription (genetics); organelle→molecular biology	251	meiosis	7500
bacterial microcompartment →membrane; down syndrome→patau syndrome; edmund beecher wilson→walter sutton; membrane→receptor (cell biology); t cell→symbiosis	167	t cell	6080	adenosine triphosphate→dna	250	eukaryote	7131
		eukaryote	6060	cell→gene; chromosome→allele; lymphocyte→phagocyte; mutation→clone	249	mitochondria	6852
		membrane	6003			cytoplasm	6631
		rna	5498			cell division	6128
		meiosis	5324			gene	5481

Table 2. The order of selecting links for the pedagogic co-occurrence network based on the most traversed links of the vocabulary co-occurrence network and the number of traversals and their rankings in random exploration paths bidirectionally.

Reason for including a link of the vocabulary co-occurrence network to the pedagogic co-occurrence network	*Link*	*Traversals in a random path in the vocabulary co-occurrence network*		*Traversals in a random path in the pedagogic co-occurrence network*	
		Traversals (n=300000)	*Ranking among the most traversed links*	*Traversals (n=100000)*	*Ranking among the most traversed links*
cell based link having rank 1	cell→cyanobacteria	164	8 (shared)	3523	11 (shared)
cell based link having rank 3	cell→cell theory	157	14 (shared)	3634	5
cell based link having rank 6	amoeboid movement→cell	153	18 (shared)	3558	9 (shared)
cell based link having rank 11	cell→flagellum	151	20 (shared)	3514	13
gene based link having rank 1	cytokine→gene	162	9 (shared)	3505	15
gene based link having rank 3	gene→thomas hunt morgan	158	13 (shared)	3472	16
gene based link having rank 6	plastid→gene	152	19 (shared)	3517	12
chromosome based link having rank 1	chromosome→haploid	165	7 (shared)	3644	4
chromosome based link having rank 3	y chromosome →chromosome	160	11 (shared)	3652	3
to enable a full connectivity	chromosome→gene	160	11 (shared)	3609	6
to enable a full connectivity	plastid→flagellum	152	19 (shared)	3566	8
to further increase the connectivity	gene→cytokine	149	22 (shared)	3539	10
to further increase the connectivity	cell theory→gene	147	24 (shared)	3724	1
to further increase the connectivity	cell→cytokine	146	25 (shared)	3506	14
to further increase the connectivity	cyanobacteria→cell	146	25 (shared)	3523	11 (shared)
to further increase the connectivity	gene→chromosome	144	27 (shared)	3582	7
to further increase the connectivity	gene→y chromosome	144	27 (shared)	3655	2
to further increase the connectivity	cell→amoeboid movement	142	29 (shared)	3558	9 (shared)

consecutive days, as motivated by findings of Kandel (2001) and Fields (2005), and also gradually expanding spacing of repetitions, as motivated by findings of Balota et al. (2007) and Thalheimer (2006).

We do not know any previous work proposing methods and results similar to our work especially in respect to developing intuitive recommendations for learning paths to explore efficiently educational text collections. Please note that further details about our results are available in publications (Lahti 2015a, 2015b). Motivated by our results we suggest that future research should actively develop methods helping to build and exploit educational resources relying on the efficient form of scale-free small-world networks. We hope that the empirical measures we have now provided about vocabulary co-occurrence networks and pedagogic co-occurrence networks relying on texts freely available in Wikipedia can help to identify fertile structures and dynamics of texts to be applied in diverse educational activities by learners and educators.

References

Balota, D., Duchek, J., & Logan, J. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. Psychology Press, 83-107.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186-198.

Capocci, A., Rao, F., & Caldarelli, G. (2008). Taxonomy and clustering in collaborative systems: the case of the on-line encyclopedia Wikipedia. Europhysics Letters, 81(2).

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. (2008). The frequency effect in second-language visual word recognition. Psychonomic Bulletin & Review, 15(4), 850-855.

Ellis, A., & Lambon R. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. Journal of Experimental Psychology: Learning, Memory and Cognition, 26, 1103-1123.

Fields, R. (2005). Making memories stick. Scientific American, 292 (February 2005), 74-81.

Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: evidence from Navajo. In Gathercole, V. (ed.), Routes to Language: Studies in honor of Melissa Bowerman, 5-36. Taylor & Francis, New York, NY, USA. http://www-psych.stanford.edu/~lera/papers/navajo.pdf

Izura, C., & Ellis, A. (2002). Age of acquisition effects in word recognition and production in first and second languages. Psicológica, 23, 245-281. http://www.uv.es/revispsi/articulos2.02/4.IZURA%26ELLIS.pdf

Kandel, E. (2001). The molecular biology of memory storage: a dialog between genes and synapses. Nobel Lecture, 8 December 2000. Bioscience Reports, 21(5).

Kinouchi, O., Martinez, A., Lima, G., Lourenco, G., Risau-Gusman, S. (2002). Deterministic walks in random networks: an application to thesaurus graphs. Physica A: Statistical Mechanics and its Applications, 315(3-4), 665-676.

Lahti, L. (2014). Computational method for supporting learning with cumulative vocabularies, conceptual networks and Wikipedia linkage. International Journal for Cross-Disciplinary Subjects in Education (IJCDSE), 5(2), June 2014, 1632–1644. Infonomics Society, UK. ISSN 2042-6364. http://www.infonomics-society.org/IJCDSE/Computational%20Method%20for%20Supporting%20Learning.pdf

Lahti, L. (2015a). Computer-assisted learning based on cumulative vocabularies, conceptual networks and Wikipedia linkage. Doctoral dissertation, Department of Computer Science, Aalto University School of Science, Finland.

Lahti, L. (2015b). Supplement to doctoral dissertation "Computer-assisted learning based on cumulative vocabularies, conceptual networks and Wikipedia linkage". Department of Computer Science, Aalto University School of Science, Finland.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren and M. Nordman (eds.), Special Language: From Humans Thinking to Thinking Machines. Multilingual Matters, Clevedon, UK.

Masucci, A., Kalampokis, A., Eguíluz, V., & Hernández-García, E. (2011). Wikipedia information flow analysis reveals the scale-free architecture of the semantic space. Public Library of Science ONE (PLoS ONE), 6(2), e17333.

Morais, A., Olsson, H., & Schooler, L. (2013). Mapping the structure of semantic memory. Cognitive Science, 37, 125-145.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In Schmitt, N., & McCarthy, M. (eds.), Vocabulary: Description, Acquisition, Pedagogy. Cambridge University Press, New York, USA, 6-19.

Thalheimer, W. (2006). Spacing learning over time: what the research says. Will Thalheimer, Work- Learning Research Inc., Somerville, Massachusetts, USA (published March 2006, reformatted 2010). Online available at http://willthalheimer.typepad.com/files/spacing_learning_over_time_2006.pdf

Thompson, G., Kello, C., & Montez, P. (2013). Searching semantic memory as a scale-free network: evidence from category recall and a Wikipedia model of semantics. Proc. 35th Annual Meeting of the Cognitive Science Society. Cognitive Science Society, Austin, TX, USA, 3533-3538. ISBN 978-0-9768318-9-1.

Voß, J. (2005). Measuring Wikipedia. Proc. 10th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2005). http://eprints.rclis.org/bitstream/10760/6207/1/MeasuringWikipedia2005.pdf

Wang, J., Zuo, X., & He, Y. (2010). Graph-based network analysis of resting-state functional MRI. Frontiers in Systems Neuroscience, 4:16.