



This is a repository copy of *Graph analysis of semantic word association among children, adults, and the elderly.*

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/153552/>

Version: Accepted Version

---

**Article:**

Zortea, M., Menegola, B., Villavicencio, A. [orcid.org/0000-0002-3731-9168](http://orcid.org/0000-0002-3731-9168) et al. (1 more author) (2014) Graph analysis of semantic word association among children, adults, and the elderly. *Psicologia: Reflexão e Crítica / Psychology: Research and Review*, 27 (1). pp. 90-99. ISSN 0102-7972

<https://doi.org/10.1590/S0102-79722014000100011>

---

© 2014 The Authors. This is an author-produced version of a paper subsequently published in *Psicologia: Reflexão e Crítica*. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

Análise de Grafos de Associação Semântica de Palavras entre Crianças, Adultos e Idosos

Graph Analysis of Semantic Word Association between Children, Adults, and the Elderly

Área: Processos Básicos em Psicologia

### Abstract

This work uses graph analysis to investigate how age differences modify the structure of semantic word association networks from children and adults and if these networks present a small-world structure and a scale-free distribution typical of natural languages. Three age groups of Brazilian Portuguese speakers (children, young and elderly adults) participated in the experiment. Quantitative and qualitative measures suggested that adults and elderly speakers had similar network structures. The children's network had fewer nodes, connections and clusters, and longer inter-node distances. All networks presented a small-world structure, but they did not show entirely scale-free distributions. These results suggest that from childhood to adulthood, there is an increase not only in number of words semantically linked to a target but also an increase in the connectivity of the network.

Keywords: graph theory; word association; developmental age groups; semantic memory

### Resumo

Este estudo utilizou análise de grafos para investigar como a idade modifica a estrutura das redes de associações semânticas de palavras de crianças e adultos e se estas redes apresentam estrutura small-world e distribuição scale-free típicas da linguagens naturais. Participaram dos experimentos três grupos etários (crianças, adultos e idosos) falantes do Português Brasileiro. Medidas quantitativas e qualitativas sugeriram que adultos e idosos possuem redes semelhantes quanto a número de nós, conexões e agrupamentos. A rede das crianças expressou menor número de nós, conexões e agrupamentos e maiores distancias inter-nós. As redes apresentaram uma estrutura small-world, mas não uma distribuição scale-free completa. Os resultados sugerem que além do número de palavras semanticamente associadas aos alvos aumentar das crianças para os adultos, as redes se tornaram mais conectadas.

Palavras-chave: teoria dos grafos; associação de palavras; grupos etários; memória semântica

## Introduction

The structure and organization of human knowledge in terms of semantic relations have been the topic of much debate in areas such as cognitive sciences, psychology, and computer sciences. Nevertheless, the effects of age in semantic processing are not yet fully understood, mainly due to diverging research results. For example, Rönnlund, Nyberg, Bäckman, and Nilsson (2005), found a decline in semantic fluency and vocabulary tasks, in participants from 60 to 85 years old. By contrast Park et al. (2002) showed an increase in the performance scores of vocabulary tests and relations of synonyms, even when participants were over 60 years old.

In terms of word associations, a method to assess semantic knowledge through free association, the findings are also disparate. While Burke and Peters (1986) did not find differences in terms of the strength of association between words for young and elderly adults, Hirsh and Tree (2001) found a greater strength for the latter than for the former. Although the diversity of results is clear, the variability of tasks and methodologies employed in these studies (Little, Prentice, & Wingfield, 2004; Sauzéon, Lestage, Raboutet, N'Kaoua, & Claverie, 2004) have to be considered.

Hierarchical networks like those presented by Collins and Quillian (1969) have been adopted for understanding organization of categorical knowledge in semantic networks. In recent years and in a different fashion, graph analysis techniques (Albert & Barabási, 2002) were employed for investigating characteristics and structures of language in these networks (Ferrer-i-Cancho & Solé, 2001; Steyvers & Tenenbaum, 2005). The underlying assumption is that the graph structure represents the organization of the associations between elements, for instance, the organization of words in terms of its semantic relations in semantic knowledge investigations. These techniques have been used for analyzing a variety of psycholinguistic tasks in both healthy and clinical conditions (Cabana, Valle-Lisboa, Elvevåg, & Mizraji.

2011; Mota et al., 2012). In language studies, this organization can influence *inter alia* the speed of the cognitive activity, including memory processes (Coronges, Stacy, & Valente, 2007). In semantic memory and word association research, some advantages of this method are that it considers the associative relation between every word produced, focusing on the structure of this associative network. So, graph analysis provides complementary measures to traditional quantitative methods, such as strength found in Hirsh and Tree (2001). For the word associations, nevertheless, little is known about the impact of extra-linguistic aspects of the speakers, such as age (Coronges et al., 2007; Hills, Maouene, Maouene, Sheya, & Smith, 2009), and how these are to be considered in the structure and analysis of the graphs.

The aim of this work is to investigate semantic word associations, looking at the influence of age on language through comparative graph analysis of different age groups. In doing so, we intend to expand the understanding of the development of these associations. We look at psycholinguistic data from native speakers of Brazilian Portuguese, belonging to three age groups: children, young and elderly adults.

### ***Word association tasks***

Several studies about memory and language have used word association tasks to obtain measures of the development of lexical-semantic knowledge (Hirsh & Tree, 2001; Macizo, Gómez-Ariza, & Bajo, 2000). Studies about word associations use, to a great extent, two types of tasks. The tasks of free association, in which the participant must respond to a stimulus word, called *target*, evoking the first word, called *associate*, that may come to his/her mind (Nelson, McEvoy, & Schreiber, 1999). There are also semantic association tasks, in which the associated word must be related to the meaning of the target (Salles, Holderbaum, & Machado, 2009; Zortea & Salles, 2012). Because of the lack of semantic restriction, free association tasks are more likely to generate pairs that are not semantically

associated, e.g. “cat - hat”. In Brazilian Portuguese there are association norms for children (Salles et al., 2009) and adults (Salles et al. 2008; Stein, Feix, & Rohenkohl, 2006).

### *Graph analyses*

A graph or a network can be operationally defined as a set of nodes (*vertices*) connected through links (*edges* or *arcs*), which results in a network structure. In an *undirected* network links are referred to as edges, and in *directed* networks when there is directionality in the links, they are known as arcs. These characteristics could be related to the direction of the associations in a word association task and, by inference, to the memory retrieval processes and lexical-semantic structure (Steyvers & Tenenbaum, 2005). A graph can have dynamic properties that relate to the evolution or growth process, or static properties that do not consider intermediary steps of the evolution, but only the measures of the resulting network (Callaway, Hopcroft, Kleinberg, Newman, & Strogatz, 2001).

In terms of structure, the topology of the connections of a complex system can vary from regular to random, but the graphs in the context of language tend to have an intermediary structure known as *small-world* (Watts & Strogatz, 1998). This architecture concerns the existence of groups of different words linked to each other, as in regular networks, but with short paths, on average, connecting one word to another in the network, as in random networks (Albert & Barabási, 2002). This structure can be observed in the natural language due to, for instance, the high speed of sentence generation (Ferrer-i-Cancho & Solé, 2001).

Moreover, it has been argued that the organization of language networks can follow a *scale-free distribution*, referring to the fact that there are many nodes that share few links with their neighbors, and a small number of nodes, called hubs, that have many links with other nodes (Albert & Barabási, 2002). A scale-free organization will be revealed if the distribution of links of each node follows a *power-law curve*. In an applied perspective, a word that has a

great number of links will probably be more rapidly processed or accessed in tasks such as naming or lexical decision (Steyvers & Tenenbaum, 2005).

In terms of word association graphs, there are two main methods for establishing the relations between target and associate. The two-mode network consists of targets that are included with the respective associated words connected to them. In other words, each node of the network can represent a target or an associate. In the *one-mode network*, each node represents only associates and targets are not explicitly included (Coronges et al., 2007). If two associates share the same target, they will be linked in the network. Hence, one-mode networks mainly highlight conceptual relations in detriment of target-associate relations. On the comparison between two or more graphs with repeated targets, this mode may be more suitable for showing relations among associates. As a reflection of the increase in the number of connections using the one-mode network, one can expect that the diameter and the average shortest path length would decrease and the clustering coefficient would increase.

#### *Graph analysis on word associations*

Steyvers and Tenenbaum (2005) investigated the development of a network of associated words, using lists of discrete free association (Nelson et al., 1999) with 5,019 targets produced by more than 6,000 adults. They found that the resulting networks had a small-world structure and scale-free organization, and that these same characteristics were present in large corpora, as the WordNet and Roget's thesaurus. These results led them to pose the hypothesis of a representational structure of language similar in corpora and in speakers (Steyvers & Tenenbaum, 2005). According to the small-world and scale-free aspects, the organization of this structure would be based upon the preferential attachment hypothesis. In this sense, the probability of a new word being linked to a word already existing in the network is directly proportional to the number of links that this word already has (Albert & Barabási, 2002).

Other researchers, such as De Deyne and Storms (2008b), Hills et al. (2009), and Coronges et al. (2007), performed similar analyses in other samples. De Deyne and Storms (2008b) used data of a free association task with 1,424 words applied in an adult sample of 10,292 Dutch participants (De Deyne & Storms, 2008a) to create computational graphs in which the targets were used as nodes. The links between two targets were established if the same associate had been said for both the targets by two or more people. Results compatible with those of Steyvers and Tenenbaum (2005) were found, with a small-world structure and the scale-free organization, except for a higher network density (0.02 vs 0.004 in the study by Steyvers and Tenenbaum) due to a larger number of links between its elements. The latter can be explained by De Deyne and Storms' use of a continuous association task considering three associates for each target and to the lower variability of targets. Thus, the nodes have greater probability of connection with each other, thereby making the network denser.

In an age comparison study, Hills et al. (2009) examined which type of relation between the words would present a scale-free distribution pattern in the topological analysis. Examining the acquisition of words by children aged 16 to 30 months, two networks were formed: (1) one from the common semantic attributes between the words (for example, having eyes and being hairy) and (2) another from the target-associate relation in the word association task. The words (all of them nouns) were selected from a development inventory, which handled terms that would supposedly be in the lexicon of children of the same age. Of the networks constructed, the one that used the target-associate relation best fitted the results.

Coronges et al. (2007) studied word associations for 16 targets of 1,097 seventh-grade students (aged between 12 and 13 years old), compared with university adults of the study of Nelson et al. (1999). Topological analyses referring to the total number of associates in the network showed that the children had a greater number of associates. However, when considering only the associates expressed by two or more people, the adults produced a



greater number of associates. Consequently, the group of children generated a larger number of idiosyncratic words (produced by only one participant), which led to a greater variability of answers in this group. Coronges et al. (2007) believe that this is due to the fact that children are less inhibited and restricted in associating words and/or to the fact that the group, being at an earlier language development stage, has more diverging responses between one individual and another.

Concerning the graphs generated, Coronges et al. (2007) noted similarities between the groups related to, for example, the density, the average number of links between the words and the average number of highly clustered words. Differences were reported in terms of the centralization of the words in the network. The children's graph contained 17.9% of words distributed around one or a small number of central nodes, whereas in the adults this percentage was 27.3%. The authors attribute this to an increase in the degree of sophistication of the organization of the memory for words in adults. However, Coronges et al. (2007) also point out some methodological limitations that may have contributed to the differences found between the two groups, e.g. the level of schooling, the sociocultural context and the high rate of bilingualism in the children group.

Besides the analysis of the quantitative parameters of the graphs, it is possible to perform qualitative analyses of the network. Tonietto (2009) employed this type of analysis in a network of words produced by 57 children, aged between two and four and a half years old, in a task of naming actions from 17 films. Using human judges to rate the degree of generality and specificity of the words named by the participants in a longitudinal test and retest method, the authors observed that boys had an increase in the use of more specific words two years after the first evaluation.

Given this context, on the one hand, the works of Steyvers and Tenenbaum (2005), De Deyne and Storms (2008b) and Hills et al. (2009) agree on some of the characteristics of word

association networks, such as the small-world structure and a scale-free distribution found for these networks. On the other hand, for other aspects like age-related differences, there is still a lack of clear theoretical hypotheses, although some differences had been found on the structure of graph of children and adults (Coronges et al., 2007). Therefore, in this article we report on an investigation that may contribute not only to corroborating some of these findings, but also to explore issues such as the influence of age on word association networks based on semantic relations.

We aim to analyze the structure and organization of semantic association graphs of words produced by different age groups: children, adults, and elderly participants. The specific objectives are: a) comparing the measures of the topological analysis of graphs among the three groups; b) checking if these graphs have a small-world structure and scale-free distribution; and c) performing a qualitative analysis comparing the global organization (nodes groupings or concentrations, density, and general spatial structure) of the networks of each age group with the quantitative measures of their topological analysis.

## Method

### *Participants*

The participants (total  $n = 171$ ) are all native speakers of Brazilian Portuguese, living in the same region (the city of Porto Alegre), and are divided into three groups: children, adults, and elderly people. Demographic data are showed in Table 1.

---

Insert Table 1

---

The criteria for being included in the sample were either having Portuguese as the mother tongue or being declared fluent in this language. In the case of the elderly participants, an absence of symptoms of depression, as assessed by the *Geriatric Depression Scale* – GDS-

15 (Almeida & Almeida, 1999; Yesavage et al., 1983) and of dementia, as measured by the *Minimal State Examination* – MMSE (Chaves & Izquierdo, 1992; Folstein, Folstein, & McHugh, 1975), was also required. A transversal contrasting groups design was used. The research was approved by the Ethics Committee in Research of Psychology of the Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

### ***Instruments and procedures***

In order to obtain the word associations, a list of 87 target words, selected from the norms of Salles et al. (2008), with different psycholinguistic characteristics was used. The frequency of the words varied from 2 to 45,625 occurrences ( $M = 3838.63$ ;  $SD = 6520.00$ ). This measure was obtained from didactic, technical, and scientific books, encyclopedias, journals, magazines, juridical and academic materials, and the like (Kuhn et al. 2000). The degree of concreteness was found for 43 targets, varying from 2.35 to 6.85 ( $M = 5.17$ ;  $SD = 1.45$ ), maximum value of 7, according to the norms of Janczura, Castilho, Rocha, Van Erven, and Huang (2007). The length of the words ranged from 3 to 10 letters ( $M = 5.79$ ;  $SD = 6.02$ ). In terms of grammatical class, the list includes 63 nouns, 12 adjectives, 5 adverbs and 7 ambiguous words that can be classified as a noun or an adjective, according to the context.

For the semantic word association task, the following instruction was given: “you have to respond as quickly as possible with the first word that comes to your mind (associate) that is semantically related to the word that the examiner will express orally (target)”. Three examples were given to make sure the participants understood the activity. The children were assessed in a quiet room at the school, individually and gave their response aloud in order to avoid large response times. The adult and elderly participants were assessed in different classrooms at the university, in the majority, collectively and the responses were written. Data from 13 children were originated from Salles et al. (2009)’s study and data from adults were obtained entirely from Salles et al. (2008)’s study. The answers from all samples were

preprocessed to remove those that were blank, illegible, unrelated (e.g., “I don’t know”), identical to the target or when more than one word was produced as a response to the target. Moreover, owing to semantic similarity, the canonic form of associated words was used when they allowed morphological inflections, such as gender [e.g., caneco (cup) in caneco vs. caneca], number [e.g., operário (worker) in operário vs. operários], and degree [e.g., teatro (theater) in teatro vs. teatrinho]. However, these words were grouped only if their meanings were equivalent. Some examples of target-associate pairs are: aberto-fechado (opened-closed) given by children; dente-boca (tooth-mouth) by adults; and raiva-ódio (anger-hate) by elderly people.

### ***Graph analysis***

The targets and their associates were used to form three graphs, one for each age group, in order to allow comparisons. The graphs were constructed as one-mode networks, which includes in the network only the associated words, linking them if they have been said by more than one participant for the same target. The resulting networks were indirect (with no directionality in the links). Following Steyvers and Tenenbaum (2005), the static analyses performed used as dependent variables for comparison among the groups:

*Number of nodes ( $n$ ):* total number of nodes (associates) found in the network;

*Number of edges ( $m$ ):* total number of links between these nodes in the network;

*Degrees ( $k$ ):* number of links of each node in the network or number of neighboring nodes connected to each node of the network (this measure will be important in analyzing the distribution of degrees);

*Average degree ( $\langle k \rangle$ ):* average number of connections of all the nodes in the network, calculated from the distribution of degrees or dividing the number of edges by the number of nodes;

*Diameter ( $L$ ):* maximum distance, measured by number of edges, between any two nodes in the network;

*Average shortest path length ( $l$ ):* average of the shortest distances between any two nodes in the network;

*Clustering coefficient ( $C$ ):* refers to the probability of finding a node linked to another two nodes that are also linked to each other, forming a triangular structure. This measure indicates the presence of groups of highly interconnected nodes in the network.

The comparisons were made according to the nature of the variables, and the data available. For frequency measures, such as  $n$ ,  $m$  and  $L$ , and for the average shortest path length variable ( $l$ ), descriptive analyses were made. For the average degree ( $\langle k \rangle$ ), a one-way ANOVA test was performed, having as factor the 3 different groups.

Following Steyvers and Tenenbaum (2005), comparisons were made for  $l$  and for  $C$  between the networks of each group (associative-semantic networks) and a random network, to determine if the networks have a small-world structure. The latter was created using the same number of nodes and edges as the original networks, but the links between the words were defined by a random criterion. Thus, if the  $l_{random}$  is similar to that found in the associative-semantic networks, but the  $C_{random}$  is clearly smaller, there are arguments to infer that these networks have a small-world structure (Steyvers and Tenenbaum 2005).

In order to check the distribution of degrees, we used curve estimation analysis, where for the network to have a scale-free organization, it must fit a power-law curve. Following Albert and Barabási (2002) and Hills et al. (2009), we adopted cumulative distributions, which represent the probability of a node having a number of degrees greater than or equal to the other nodes of the network. We also performed a qualitative analysis of each of the networks. These enabled us to compare the groups in terms of the global organization and

structure of the associates, as well as to contrast the quantitative data of the topological analysis with the qualitative data.

### Results

Table 1 shows the values of the variables for the network of each group and for the random network for their topological analysis.

---

Insert Table 2

---

According to the results of graph analysis, the semantic word association networks of young adults and the elderly presented similar values in several measures, such as  $n$ ,  $m$ ,  $C$ , and  $L$ . In a descriptive manner, the network of the children's semantic word associations had, fewer nodes and links, one extra step for  $L$ , and lower  $C$  value compared to the graphs of the two other groups. Moreover, the average shortest path length ( $l$ ) was distinct for each group. A one-way ANOVA showed that the differences in the  $\langle k \rangle$  were significant [ $F(2, 1700) = 51.92$ ;  $p < 0.001$ ] and the post-hoc Fisher's Least Significant Difference (LSD) and the Tukey tests indicated that all groups differ from each other ( $p < 0,001$ ).

In order to investigate the small-world property of the graphs, we compared the values of *average shortest path length* ( $l$ ) and clustering coefficient ( $C$ ) in the associative-semantic networks and in the random networks. The  $l$  of the associative-semantic networks is similar to that of the random networks ( $l_{random}$ ), which indicates short distances between the nodes in both kinds of networks. Nevertheless, there are few local clusterings in the random networks, as indicated by the low values of  $C_{random}$ . This contrast was observed for all the three groups.

Figure 1 illustrates the analysis of degree distribution. The horizontal axis represents the degree of the nodes in the network and the vertical axis the probability of finding by chance a node whose degree is equal to or greater than the other nodes of the network. The

results show that for all the graphs the majority of nodes have a low degree. Nodes with high degrees are infrequent, according to the tail of the distribution. In addition, the sharp drop of the tail also suggests that the probability falls exponentially with the degree of the node.

---

Insert Figure 1

---

The power-law function in a graph with logarithmic scales has the format of a straight line. The curve estimation analysis yielded an  $R^2 = 0.76$  for the children's network (Figure 1A)<sup>1</sup>, which means that 76% of this distribution can be explained by a power-law function. According to an ANOVA test for goodness-of-fit, this function fitted significantly better [ $F(1, 56) = 174,6; p = 0,005$ ] the data as compared to the mean. However, an exponential function, represented by a curved line, explained 99% of the variance in this group, and this function significantly fitted the data [ $F(1, 56) = 8.448,4; p < 0,001$ ] as well. The panorama was similar for the other two groups, where although the power-law curve had an  $R^2 = 0.86$  and  $R^2 = 0.87$  in the cumulative distribution for the adult (Figure 1B) and elderly (Figure 1C) groups respectively, the exponential function had  $R^2 = 0.98$  for both groups. All of them were significantly better predictions than using the mean ( $p < 0,005$ ).

Figure 2 shows the networks of each group of the study. The black nodes represent association contexts, in which several associates are linked to a target. Words that connect two or more target words are indicated by white nodes in the network. In the children's network (Figure 2A), there are several isolated nodes and only a few nodes that provide the main connections between the words, in a sparsely populated network. The networks of the adult (Figure 2B) and elderly (Figure 2C) groups have a greater general density, with more nodes and more vertices, and a lower number of vertices that are isolated or with a low degree. Their nodes are more connected compared to those in the children's network. One of

the salient differences between the adult and elderly groups is that the former has two large concentrations of relatively separated vertices (one in the upper right part and the other in the lower part), while the latter does not have such clear separate concentrations.

---

Insert Figure 2

---

### **Discussion**

Regarding the topological analysis of graphs, the children's network showed quite noticeable features that differentiate it from those of the other two adult groups. They include a lower number of nodes (associates) with a lower number of links among them, a smaller average internal connectivity and a low probability of finding clusters of semantically associated words. In the qualitative analysis there were also differences, where the networks of the adult and elderly groups were denser and more connected and with fewer isolated nodes.

In some aspects the results found in this article differ from those in other studies, such as Coronges et al. (2007), who reported equivalent networks of children and adults (university students) in terms of density, number of connections and clustering coefficient. One possible explanation for these differences is the age range of the children involved, as for Coronges et al. (2007) they were aged between 12 and 13 years old (seventh-grade children) and in our study they were aged between 8 and 12 years old. According to Sauz on et al. (2004), from 11 to 12 years old the semantic knowledge measured by the number of taxonomic categories produced in a semantic fluency task becomes more developed in terms of organization, and no salient differences are expected from 12 to 16 years old. On the other hand, as the 57 children of the present study are, on average, approximately nine years old, their semantic-



associative lexical knowledge is still undergoing changes (Macizo et al., 2001; Sauz on et al., 2004) and may be different from the adult standard.

Notwithstanding, Coronges et al. (2007) found that the networks of the seventh-grade students had fewer clusterings and more disperse words than the network of adults, which had words more centralized in clusters. Some results from the present study indicate the same trend. In particular, the clustering coefficient, related to the probability of groups of three interlinked nodes occurring, was lower for the children, who had a sparser network structure than the two adult groups. Motter, Moura, Lai, and Dasgupta (2002) discussed that this aspect is related to the associative nature of human memory, through which a person easily retrieves information by connecting it to similar concepts.

Another variable that might be related to the differences found here between children and adults is educational level. Although we did not find studies focusing on the impact of education on semantic word associations or graph analysis on these associations, it is well-accepted that higher educational levels contribute to higher scores in semantic memory tests (Brucki & Rocha, 2004; Ruff, Light, Parker, & Levin, 1996). Then, the lower number of nodes and links in children's graph could be related to the reduced individual production and group variability rates. Research with low education adult samples could shed light on this issue, in order to reduce differences of educational level between groups. Although some children had grade repetitions, the authors do not consider this characteristic had an effect on the results. First, the number of children with grade repetition was low ( $n = 6$ ), and second, their responses were oral, avoiding the interference of reading or writing difficulties.

Comparing the adult and elderly networks, they had similar number of nodes and links, diameter and clustering coefficients. The nodes and association contexts, which represent a group of associates linked to a target, are, generally speaking, analogously distributed. These results are compatible with the relevant literature (Burke & Peters, 1986;

Little et al., 2004; Rönnlund et al., 2005), which suggests that the organization of the semantic knowledge of young and elderly adults does not seem to have significant changes. Authors like Light (1991) and Little et al. (2004) explain that although the contents of semantic knowledge can undergo changes after 60 years old, its structure remains relatively constant. The same could be inferred for the semantic associations between words.

For these two groups educational level does not play a potential role as, in average, it was similar. Although structural differences are not expected, the slightly higher number of nodes in the elderly sample might be related to a predominance of women in the group (they constituted 98.2% of the group) with a possibly higher word production rate. This gender effect was verified in high educational adults (Ruff et al., 1996) and for some semantic categories (Capitani, Laiacona, & Barbarotto, 1999), though these studies used fluency tasks. Although it is expected, in national research with elderly people, women volunteer more than men, future research could increase the percentage of elderly men in order to examine if any gender effect is present.

Besides the similarities, significant differences between the networks in terms of the average number of connections were verified, in that the associations (nodes) of the adults are more connected than those of the elderly network. There is also, from a descriptive perspective, a lower average minimum distance between any two nodes in the former than in the latter. In fact, the adult network had at least two large concentrations of nodes, possibly representing different wide semantic contexts. The greater number of links and shorter distances for the adults could be explained by the presence of more hubs linking the different clusters. Considering that a hub represents a node with many connections, the young adults had 10 nodes with a degree higher than 100, whereas the elderly group had only 7 such nodes. Moreover, the node with highest degree for the elderly counted 158 links, and for the young adults 168 links.

Additionally, we should stress that the task and sample size adopted in the present work was different than that of other studies (Coronges et al., 2007; De Deyne & Storms, 2009b; Steyvers & Tenenbaum, 2005), which makes direct comparisons difficult. In relation to the word association task, we used a semantic word association task, whereas other works used free association. The former includes the instruction that the associate must have a semantic or meaningful relation with the target, which may have resulted in less variability of responses. This last methodological option, as well as the manner in which the network was modeled, may partially explain the results found in the degree distribution analysis.

In a network growth study, Steyvers and Tenenbaum (2005) found a correlation between the degree of a word and the time the word was introduced in the network, which leads to a power-law degree distribution. This finding is compatible with the preferential attachment rule. But that was not found on the present study. For all the groups the function that best fitted the cumulative degree distribution, based on the amount of variance explained, was the exponential one. Amaral, Barthélemy, Scala, and Stanley (2000) and Callaway et al. (2001) consider that a degree distribution with a power-law regime but with an exponential tail characterizes a distinct subtype of small-world graph, namely a *broad-scale* distribution. According to Amaral et al. (2000), the sharp drop in the tail to the right of a cumulative degree distribution, as found here, represents a restriction to links made to the nodes over time. This restriction can arise from an “aging” or inactivation of the nodes or from a cost to the new links, so that the new nodes incorporated in the network will not necessarily be connected to already existing nodes, as suggested by the preferential attachment rule.

In terms of the cognitive aspects, this distribution suggests a situation in which new words do not necessarily create associations with the words with a high degree of connectivity (a great number of semantic relations with other words, usually corresponding to very frequent and polysemous cases), as would be expected if a preferential attachment

regime was in place. In this case other factors also seem to be playing a role such as recency effects, by which new words would be linked to other recent ones, in certain cases, or conventionality, where certain words are preferred regardless of their frequency or specificity. Nevertheless, the fact that the targets used in the task did not have a semantic relation with each other *a priori* could partly explain why the resulting networks had some isolated words on the one hand, and groups of words on the other hand. Another relevant point is that the word association data used had few targets and associates compared to databases of other studies (e.g. Steyvers & Tenenbaum, 2005; De Deyne & Storms, 2009b)

The exponential distribution found in the present study is similar to the results obtained by Vitevitch (2008), which investigated the structure and organization of the mental lexicon of adults through relations of phonological similarity between words, obtained from an English language dictionary. Although the degree distribution was attributed to the organization of the knowledge of the form of the words (phonological, orthographical or, by and large, lexical organization), we found compatible results in this investigation about the knowledge of the meaning of the words (semantic organization). A distinction between lexical and semantic aspects of a word is shared by the relevant literature (Hillis, 2001; Miller, 1999). So, further investigation with larger sample sizes and more analyses is required to determine whether the degree distribution is the same in the organization of both phonological and semantic knowledge.

Finally, the small-world structure reported for languages in general (Ferrer-i-Cancho & Solé, 2001) and for word associations in particular (Steyvers & Tenenbaum, 2005), were found in the networks used in this study. Thus, this analysis indicated that some nodes are organized in clusters, that is, they are connected to each other, and there is a high probability of finding a node whose neighbors are also interconnected. Even higher probabilities were found in other studies using the same technique (Coronges et al., 2007), where there was a

larger number of answers linked together for each target, possibly due to the fact that more participants gave valid responses for each target.

### **Conclusions**

This study aimed to investigate the structure and organization of semantic association graphs of words produced by children, young and elderly adults. The method used in this study is an attempt to expand the view on word associations compared to the study of strength of association and set size by Burke and Peters (1986), Hirsh and Tree (2001), and Zortea and Salles (2012). As the results and expected age differences observed suggest, graph analysis is an important tool to be used to investigate lexico-semantic knowledge. Its use could be extended to longitudinal studies to examine intragroup changes over time. Also, it may contribute to the study of the semantic impairment related to some diseases, such as Alzheimer's Disease and semantic dementia. Nevertheless, further investigations are needed to determine other factors that contribute to the organization of the networks of these semantic associations.

### **Acknowledgements**

We would like to thank the support of projects CNPq 482520/2012-4, 478222/2011-4, 312184/2012- 3, 551964/2011-1 and 312077/2012-2.

## References

- Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*, 47-97.
- Almeida, O. P. & Almeida, S. A. (1999). Confiabilidade da versão brasileira da Escala de Depressão em Geriatria (GDS) versão reduzida. *Arquivos de Neuropsiquiatria*, *57*(2-B), 421-426.
- Amaral, L., Barthélemy, M., Scala, A., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*(21), 11149–11152.
- Brucki, S. M. D. & Roch, M. S. G. (2004). Category fluency test: effects of age, gender and education on total scores, clustering and switching in Brazilian Portuguese-speaking subjects. *Brazilian Journal of Medical and Biological Research* *37*, 1771-1777.
- Burke, D. M. & Peters, L. (1986). Word associations in old age: evidence for consistency in semantic encoding during adulthood. *Psychology and Aging*, *1*(4), 283-292.
- Cabana, A., Valle-Lisboa J. C., Elvevåg, B., & Mizraji, E. (2011) Detecting order–disorder transitions in discourse: Implications for schizophrenia. *Schizophrenia Research*, *131*, 157–164.
- Callaway, D. S., Hopcroft, J. E., Kleinberg, J. M., Newman, M. E. J., & Strogatz, S. H. (2001). Are randomly grown graphs really random? *Physical Review*, *64*, 1-7.
- Capitani, E., Laiacona, M., & Barbarotto, R. (1999). Gender affects word retrieval of certain categories in semantic fluency tasks. *Cortex*, *35*, 273-278.
- Chaves, M. L. & Izquierdo, I. (1992). Differential diagnosis between dementia and depression: a study of efficiency increment. *Acta Neurologica Scandinavica*, *85*(6), 378-382.

- Collins, A. M. & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-248.
- Coronges, K. A., Stacy, A. W., & Valente, T. W. (2007). Structural comparison of cognitive associative networks in two populations. *Journal of Applied Social Psychology*, 37(9), 2097-2129.
- De Deyne, S. & Storms, G. (2008a). Word associations: norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, 40(1), 198-205.
- De Deyne, S. & Storms, G. (2008b). Word associations: network and semantic properties. *Behavioral Research Methods*, 40(1), 213-231.
- Ferrer-i-Cancho, R. & Solé, R. V. (2001). The small-world of human language. *Proceedings of The Royal Society of London. Series B: Biological Sciences*, 268(1482), 2261-2265.
- Folstein, M. F., Folstein, S. & McHugh, P. R. (1975). Mini-mental state. *Journal of Psychiatry Resources*, 12, 189-198.
- Hillis, A. E. (2001). The organization of the lexical system. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology* (pp. 185-210). Philadelphia: Psychology Press.
- Hills, T. T., Maouene, M., Maouene, J. Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks. *Psychological Science*, 20(6), 729-739.
- Hirsh, K. W. & Tree, J. J. (2001). Word association norms for two cohorts of british adults. *Journal of Neurolinguistics*, 14, 1-44.
- Janczura, G. A., Castilho, G. M., Rocha, N. O., Van Erven, T. J. C., & Huang, T. P. (2007). Normas de concreitude para 909 palavras da língua portuguesa. *Psicologia: Teoria e Pesquisa*, 23(2), 195-204.
- Kuhn, D. C. S., Abarca, E. & Nunes, M. G. V. (2000). *Corpus Nilc de português escrito no Brasil*. Retrieved November, 29 2005, from <http://www.nilc.icmc.usp.br/nilc/publications.htm#TechnicalReports>

- Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42*, 333-76.
- Little D. M., Prentice, K. J., & Wingfield, A. (2004). Adult age differences in judgments in semantic fit. *Applied Psycholinguistics, 25*, 135-143.
- Macizo, P., Gómez-Ariza, C. J., & Bajo, M. T. (2000). Associative norms of 58 spanish words for children from 8 to 13 years old. *Psicológica, 21*, 287-300.
- Miller, G. (1999). On Knowing a word. *Annual Review of Psychology, 50*, 1-19.
- Mota, N. B., Vasconcelos, N. A. P., Lemos, N., Pieretti, A. C., Kinouchi, O, Cecchi, G. A., Copelli, M., & Ribeiro, S. (2012) Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis. *PLoS ONE 7(4)*, e34928.  
doi:10.1371/journal.pone.0034928
- Motter, A. E., de Moura, A. P. S., Lai, Y. C., & Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E, 65*, 065102.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). *The University of South Florida word association, rhyme and fragment norms*. Retrieved September 26, 2008, from <http://luna.cas.usf.edu/~nelson/>
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17(2)*, 299–320.
- Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, Growth, and Decline in Adult Life Span Development of Declarative Memory: Cross-Sectional and Longitudinal Data From a Population-Based Study. *Psychology and Aging, 2(1)*, 3-18.
- Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1996). Benton Controlled Oral Word Association Test: Reliability and updated norms. *Archives of Clinical Neuropsychology, 11(4)*, 329-338.



- Salles, J. F., Holderbaum, C. S., & Machado, L. L. (2009). Normas de associação semântica de 50 palavras do português brasileiro para crianças: Tipo, força de associação e *set size*. *Revista Interamericana de Psicologia*, *43*(1), 57-67.
- Salles, J. F., Holderbaum, C. S., Becker, N., Rodrigues, J. C., Liedtke, F. V., Zibetti, M. R., & Piccoli, L. F. (2008). Normas de associação semântica para 88 palavras do português brasileiro. *Revista PSICO*, *39*(3), 362-370.
- Sauzéon, H., Lestage, P., Raboutet, C., N’Kaoua, B., & Claverie, B. (2004). Verbal fluency output in children aged 7–16 as a function of the production criterion: Qualitative analysis of clustering, switching processes, and semantic network exploitation. *Brain and Language*, *89*, 192-202.
- Stein, L. M., Feix, L. F., & Rohenkohl, G. (2006). Avanços metodológicos no estudo das falsas memórias: construção e normatização do procedimento de palavras associadas. *Psicologia: Reflexão e Crítica*, *19*(2), 166-176.
- Steyvers, M. & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41-78.
- Tonietto, L. (2009). *Desenvolvimento da convencionalidade e especificidade na aquisição de verbos: relações com complexidade sintática e categorização* (Doctoral dissertation). Retrieved from <http://hdl.handle.net/10183/17823>
- Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech Language and Hearing Research*, *51*(2), 408-422.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440-42.
- Yesavage, J. A., Brink, T. L., Rose, T. L., & O. Lurn, O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatry*, *17*, 37-49.

Zortea, M. & Salles, J. F. (2012). Semantic word association: comparative data for Brazilian children and adults. *Psychology & Neuroscience*, 5(1), 77-81.

## Footnote (page 14)

<sup>1</sup> In this network, a node was excluded from the curve estimation analysis because it had no link. The power-law equation does not accept zero values for the independent variable, in this case, the number of degrees ( $k$ ).

Table 1

*Age, Education, and Gender Characteristics of the Three Groups*

Groups	n	Age		Education		gender	
		Range (years)	M(DP)	Range (years)	M(DP)	Male n(%)	Female n(%)
Children	57	8 to 12	9.32(0.97)	3*	-	26(46.6)	31(54.4)
Adults	57	17 to 45	21.8(5.54)	11 to 13	-**	17(29.6)	40(70.4)
Elderly people	57	60 to 87	70.89(6.87)	5 to 19	12.91(3.37)	1(1.8)	56(98.2)

*Note.* \* = all were third grade students of elementary school; \*\* = data on the frequencies of participants within each year of education were missing.

Table 2

*Measures of Dependent Variables Obtained by Graph Analysis for Each Age Group*

Variable	Children	Young adults	Elderly adults
Number of nodes ( $n$ )	411	620	672
Number of edges ( $m$ )	7,738	19,306	18,372
Average degree ( $\langle k \rangle$ )	18.83	31.14	27.34
Diameter ( $L$ )	5	4	4
Average shortest path length ( $l$ )	2.58	2.34	2.47
Clustering coefficient ( $C$ )	0.372	0.687	0.741
Average shortest path length of random graphs ( $l_{random}$ )	2.35	2.15	2.27
Clustering coefficient of random graphs ( $C_{random}$ )	0.046	0.051	0.041

*Note:*  $n$ ,  $m$ ,  $L$  = absolute values;  $\langle k \rangle$ ,  $l$ ,  $l_{random}$  = average values;  $C$ ,  $C_{random}$  = probability values.

Figure 1. Log-Log Plot of the Cumulative Degree Distributions of the A) Children, B) Young Adults, and C) Elderly Adults Networks. Note:  $\blacklozenge$  = Cumulative Frequencies Observed; Straight Line = Power-Law Function; Curved Line = Exponential Function.

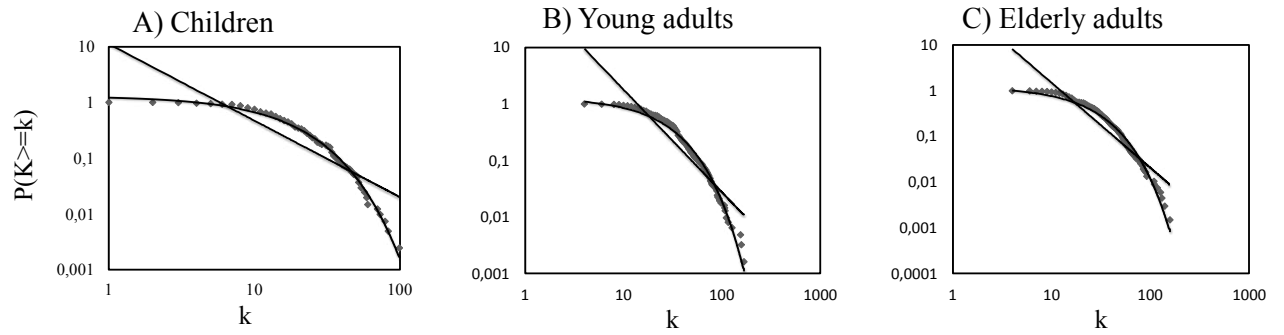
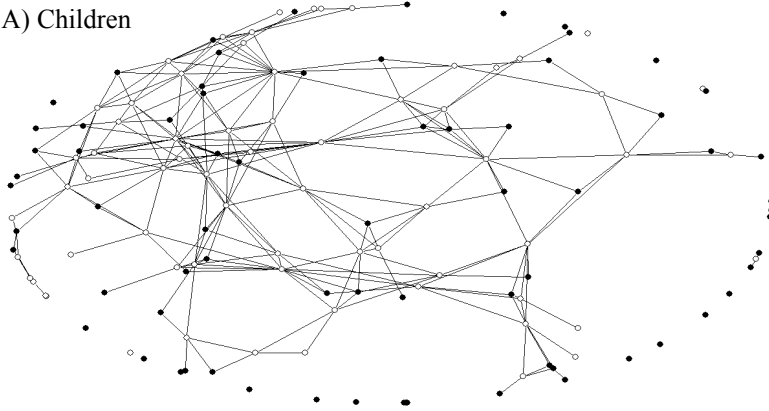
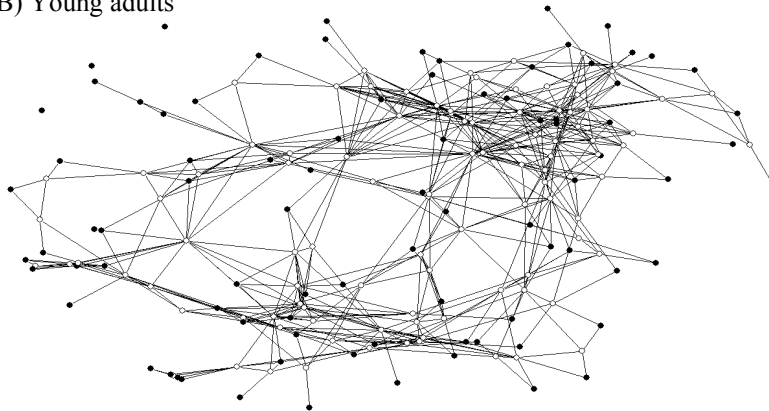


Figure 2. Graphic Representation of the A) Children, B) Young Adults, and C) Elderly Adults Networks. White Vertices Represent Associates and Black Vertices Represent Contexts (Targets).

A) Children



B) Young adults



C) Elderly adults

