



# Mapping the Structure of Semantic Memory

Ana Sofia Morais, Henrik Olsson, Lael J. Schooler

*Max Planck Institute for Human Development, Berlin*

Received 12 January 2010; received in revised form 7 May 2012; accepted 8 June 2012

---

## Abstract

Aggregating snippets from the semantic memories of many individuals may not yield a good map of an individual's semantic memory. The authors analyze the structure of semantic networks that they sampled from individuals through a new snowball sampling paradigm during approximately 6 weeks of 1-hr daily sessions. The semantic networks of individuals have a small-world structure with short distances between words and high clustering. The distribution of links follows a power law truncated by an exponential cutoff, meaning that most words are poorly connected and a minority of words has a high, although bounded, number of connections. Existing aggregate networks mirror the individual link distributions, and so they are not scale-free, as has been previously assumed; still, there are properties of individual structure that the aggregate networks do not reflect. A simulation of the new sampling process suggests that it can uncover the true structure of an individual's semantic memory.

*Keywords:* Individual semantic networks; Small-worlds; Power laws; Scale-free; Snowball sampling

---

## 1. Introduction

The question of how semantic memory is structured has been tackled in recent years through graph-theoretic analyses of semantic networks constructed from word association norms (De Deyne & Storms, 2008b; Steyvers & Tenenbaum, 2005). However, semantic networks based on word association norms do not unambiguously reflect the structural form of the networks of individual participants. The problem is that these networks are based on word association data sets that aggregate across the responses of many people. The structural properties of these aggregate networks need not resemble the properties of an individual's network because various forms of biases can occur when combining data over

---

Correspondence should be sent to Ana Sofia Morais, Max Planck Institute for Human Development, Center for Adaptive Behavior and Cognition and Center for Lifespan Psychology, Lentzeallee 94, 14195 Berlin, Germany. E-mail: morais@mpib-berlin.mpg.de

individuals. In this article, we investigate how the individual's associative semantic network is structured. To this end, we present a new experimental procedure for sampling the associative semantic networks of individual participants and analyze the statistical structure of the resultant individual networks. Each of these individual maps of semantic structure contains thousands of unique words that were sampled during approximately 6 weeks of 1-hr daily sessions. Moreover, we reanalyze existing semantic networks based on group data to examine how their structure compares with the structure of the networks of individuals.

The question of how an individual's enormous reservoir of semantic knowledge is structured is not only interesting in its own right but it also has significant implications for the processes that operate on the structure of semantic memory. In the words of Herbert Simon (1986, p. 299):

A central concern in describing any symbol-processing system is to characterize the structure (...) of its memories. When we know these facts about a brain or a digital computer, we know a great deal about its capabilities and methods of operation.

For instance, the structure of the individual's semantic network can elucidate the processes through which memory grows or develops over the life span. It has been argued by a number of researchers that the structural properties of networks are important consequences of the characteristic ways that network systems grow over time (e.g., Amaral, Scala, Bart-hélémy, & Stanley, 2000; Barabási & Albert, 1999; Watts & Strogatz, 1998). As not all mechanisms of semantic growth may give rise to one type of memory structure, the statistical structure of the individual's semantic network can help restrict the search space of possible growth mechanisms.

### *1.1. Structural properties of aggregate semantic networks*

We start by defining some basic terminology from network analysis and by introducing the structural properties that have been shown to characterize semantic networks based on group data. In an aggregate network of word associations, a set of words is represented as nodes joined by links or connections that represent a nonzero probability of a word being named as an associate by many people in response to a cue word. Two nodes are said to be neighbors if they are connected. A network can be treated as having directed or undirected links, where a directed link represents the direction of the association between two words and an undirected link leaves the direction unspecified. A network with only directed links is said to be directed and a network with only undirected links is said to be undirected.

Networks constructed from word association norms have been shown to have a small-world structure, characterized by high local clustering and short global distances between words (De Deyne & Storms, 2008b; Steyvers & Tenenbaum, 2005). On the one hand, these networks are highly structured locally, having clusters of words that are densely connected to each other by associative relations. On the other hand, there are words that connect semantically distant clusters to one another, making it possible to connect any pair of words by traversing only a few connections. A semantic network with these two properties is called a "small world"

because almost every word in the network is somehow “close” to almost every other word, even those that could be thought of as being very distant in semantic relatedness.

More formally, small-world networks are considered in terms of how they compare with random networks with the same type of links (i.e., directed or undirected), number of nodes, and average number of links across nodes. In a random network, a node is arbitrarily connected to nodes that can lie anywhere. The comparison is made regarding two statistical properties: the average shortest path length  $L$  and the clustering coefficient  $C$ . The average shortest path length  $L$  refers to the average of the shortest path lengths (i.e., the minimum number of links) that separate all pairs of words in the network.<sup>1</sup> The clustering coefficient  $C$  represents the probability that two neighbors of a randomly chosen word will themselves be neighbors (Watts & Strogatz, 1998). If the  $k_i$  neighbors of a given word  $i$  were part of a fully connected neighborhood, there would be  $k_i(k_i - 1)/2$  possible connections between them. The clustering coefficient of word  $i$  is the ratio between the number  $T_i$  of connections that actually exist between the  $k_i$  neighbors of word  $i$  and the number of possible connections between them,

$$C_i = 2T_i/k_i(k_i - 1). \quad (1)$$

The clustering coefficient  $C$  of the whole network is calculated by taking the average of the  $C_i$ 's across all words  $i$ . Because the definitions of  $T_i$  and  $k_i$  are independent of whether the links are directed, the clustering coefficient for a directed network and the corresponding undirected network are equal. Let  $L_g$  be the average shortest path length of the real network  $G$  and  $C_g$  its clustering coefficient, and let  $L_{random}$  and  $C_{random}$  be the equivalent quantities for the corresponding random networks.  $G$  is said to be a small-world network if  $L_g \geq L_{random}$  and  $C_g \gg C_{random}$  (Watts & Strogatz, 1998). Whereas the short distances allow for connecting two words chosen at random via a chain of only a few intermediaries, the high clustering implies that, on average, a word's neighbors are more likely to be connected than two words chosen at random.

In addition to being characterized as small-world networks, semantic networks based on group data have been claimed to possess degree distributions that follow a power law (De Deyne & Storms, 2008b; Steyvers & Tenenbaum, 2005). A network's degree distribution represents the probability that a randomly chosen word will have  $k$  neighbors (i.e., will have degree  $k$ ). The distribution can be estimated based on the frequencies of word degrees found throughout the network. When the associative network is directed, researchers have focused on the number of incoming links to a word (i.e., the word's in-degree). When the network is undirected, a word has a certain degree  $k$ , which is simply the number of links that it has. Network degree data can best be represented by plotting a cumulative degree distribution showing the probability that a randomly chosen word has an (in-) degree equal to or greater than  $k$  (Newman, 2005). Researchers have claimed that aggregate semantic networks, either directed or undirected, are characterized by link distributions across nodes that follow a power law, with most words having relatively few connections joined together through a smaller number of words with many connections (De Deyne & Storms, 2008b; Steyvers & Tenenbaum, 2005). Networks with power-law degree distributions are sometimes referred

to as scale-free networks because power laws have the property of having the same functional form at all scales (Barabási & Albert, 1999). Nevertheless, a methodological remark is warranted here. Studies that claimed that aggregate semantic networks are scale-free used the common method of fitting power laws to binned histograms by performing a least-squares fitting (De Deyne & Storms, 2008b; Steyvers & Tenenbaum, 2005). This method has been shown to generate poor estimates of parameters for power-law distributions and, in addition, gives no indication of whether another distribution might give a fit as good as or better than the power law (Clauset, Shalizi, & Newman, 2009). Therefore, the result that aggregate semantic networks are scale-free needs to be validated by principled statistical methods for detecting power-law behavior in empirical data.

### *1.2. Aggregate and individual semantic networks*

Current associative semantic networks are based on word association data sets that aggregate across the associates of many people. These aggregate networks are a way of representing semantic knowledge that is shared among different speakers of a language. However, aggregate networks need not preserve the statistical properties of the networks of individuals. The degree distribution, in particular, can take a different shape when data are combined over individuals. Simulation work has shown that the power-law model provides a good fit to degree distributions that result when averaging across multiple individual degree distributions, none of which follows a power law (Grünenfelder & Müller, unpublished data). Moreover, combining data over individuals can influence other connectivity properties (e.g., the average degree  $k$ , the average shortest path length  $L$ ) that depend strongly on how the word association data were collected, in particular on how many individuals generated associates for each cue word. In short, the structure of aggregate semantic networks does not unambiguously reflect the structural form of the individual's semantic network.

In this article, we map and study the structure of individuals' semantic memories. Specifically, we present a new experimental procedure for sampling words from individuals' semantic memories and analyze the statistical structure of the resultant networks. Our structural analyses focus on whether the networks of individuals display small-world properties and whether their degree distributions are better described by a power law than by alternative functions. We characterize the degree distributions by applying a statistical framework developed by Clauset et al. (2009) that involves comparing the power law with alternative statistical models. The model comparison can shed light on the formation of human semantic networks, as different degree distributions typically arise from different processes of network development (e.g., Amaral et al., 2000). Moreover, we reanalyze the degree distributions of existing aggregate semantic networks using the same method. This will allow for examining whether aggregate networks are indeed scale-free, as has been previously claimed, and how the connectivity structure of these networks compares with the structure of the individual networks. Finally, we present a computer implementation of the new method for sampling the semantic networks of individuals. Our aim in modeling the new sampling method is to examine whether it yields semantic networks

that are representative of the individuals' true semantic networks, thereby revealing their structural characteristics.

## 2. Statistical analyses of the semantic networks of individuals

In this section, we present the new method for sampling the associative semantic networks of individual participants and characterize the statistical structure of the sample-based networks. The associative semantic networks of individuals were sampled via a new experimental procedure, derived from the snowball sampling technique (Goodman, 1961). This technique is often used in social science research for developing a research sample by recruiting future participants from the acquaintances of existing participants. Similarly, our procedure for sampling the semantic networks of individuals uses the associates generated by an individual as cues for "recruiting" their semantic neighbors such that, over time, the word sample grows much like a rolling snowball. We constructed six semantic networks, based on the word associations of six individuals.

### 2.1. Method

#### 2.1.1. Participants

Six undergraduate students, with ages ranging from 22 to 26 years and an average age of 23, participated in the study. Participants 1, 4, and 6 were males, and Participants 2, 3, and 5 were females. All participants were German native speakers and were paid 10 euros for each session of the experiment in which they took part.

#### 2.1.2. Materials

Ten words were selected to start the sampling process. The procedure for selecting these seeds was threefold. First, we compiled the words of three linguistic databases in which norms of rated age of acquisition are provided: the MRC psycholinguistic database (Coltheart, 1981), the norms of Bird, Franklin, and Howard (2001), and the norms of Stadthagen-Gonzalez and Davis (2006). Subsequently, from the merged set of 6,117 words, we extracted those words rated as acquired before age 9, on average. This procedure led to a set of 4,276 words from which we selected nine words with simple random sampling without replacement to be used as "seeds": *pick*, *forgive*, *teach*, *fast*, *brush*, *angry*, *diet*, *voice*, and *adult*. In addition, we purposely added the word *risk* (also rated as acquired before age 9) to the set of seeds for a separate research project that concerns laypeople's understanding of risk. Finally, we translated the 10 words into German.

We decreased the number of seeds after piloting the experiment with Participant 1 because the experiment took too long to complete (i.e., 61 1-hr daily sessions). Whereas Participant 1 was presented with all 10 seeds on the first session, Participants 2–6 were presented with only five of them (i.e., *voice*, *pick*, *diet*, *forgive*, and *fast*). A smaller number of seeds allowed for a higher number of sampling iterations to be completed within 7 weeks.

In general, the fewer the seeds, the fewer the responses generated and, consequently, the fewer the cues to be presented in the subsequent sampling iteration.

### 2.1.3. Procedure

Six participants were tested over multiple measurement sessions. The sessions took place once a day over consecutive days and lasted approximately 1 hr each. At the beginning of each session, participants were presented with the same task instructions on the computer screen. Participants were told that they would be presented with a series of cue words and were asked to type in the words that each cue word brought to mind. The cue words were presented one at a time, in random order. Participants had 1 min to type in the words that each cue brought to mind. They were instructed not to use this time to generate as many words as possible, but solely to type in the words that came to mind spontaneously and effortlessly. Hence, if no response was given after 10 s, the next cue word was displayed before a full minute had elapsed. There was a pause of 15 s between cue words. Participants who took part in pilot sessions of the experiment reported that these time frames allowed them to comfortably perform the task.

The experiment entailed a succession of sampling iterations. In the first iteration, each participant was presented with  $n$  seeds as cue words. Next, in the second iteration, the associative responses to the seeds were presented as cue words to the participant who generated them. Likewise, in the third iteration, the responses given by the participant in iteration two were presented to him or her as cue words. This procedure was repeated over additional sampling iterations, with the associative network of each participant growing like a rolling snowball. In the process of assembling the responses to be presented as cues in the following iteration, words that had already been presented as cues in an earlier iteration were not presented again, as that would have inflated the connectivity of those words in the participants' networks. We terminated the word sampling process whenever the subsequent iteration would exceed the maximum duration established for the experiment—approximately 7 weeks of 1-hr daily sessions.

Due to interindividual differences in the number of responses generated, the overall duration of the experiment differed between participants. Participants 1, 3, and 5 completed five sampling iterations in 61, 36, and 54 sessions, respectively; Participants 2 and 6 completed six iterations in 35 and 49 sessions, respectively; and Participant 4 completed seven iterations in 32 sessions.

## 2.2. Results

We constructed a directed and an undirected associative network based on the word sample of each participant. In the directed network, each cue word is linked by an outgoing link to all of the associative responses it evoked. In the undirected network, the words are joined by an undirected link if they are associatively related regardless of the associative direction. Table 1 shows the statistics for each participant's directed and undirected networks. These statistics are restricted to the words sampled in the course of five sampling iterations—the number of iterations completed by all participants.



Table 1  
 Statistics for the semantic network of each participant

	Number of Links	$k$	$n_{cc}$	$L$	$L_{random}$	$C$	$C_{random}$
P 1 ( $n = 9,429$ )							
Undirected	20,224	4.29	9,429	5.76	6.40	0.10	1.56E-03
Directed	21,631	2.28	3,877	7.03	6.66	0.10	1.56E-03
P 2 ( $n = 2,303$ )							
Undirected	4,805	4.17	2,303	5.30	5.50	0.20	6.88E-03
Directed	5,308	2.30	973	6.24	5.57	0.20	6.88E-03
P 3 ( $n = 5,100$ )							
Undirected	8,904	3.87	5,100	5.84	6.40	0.18	3.48E-03
Directed	10,847	2.12	2,014	7.05	6.51	0.18	3.48E-03
P 4 ( $n = 1,358$ )							
Undirected	3,271	4.88	1,358	4.89	4.69	0.32	1.16E-02
Directed	3,729	2.73	680	5.66	4.75	0.32	1.16E-02
P 5 ( $n = 9,129$ )							
Undirected	22,800	5.47	9,129	5.19	5.55	0.18	2.69E-03
Directed	27,124	2.96	3,643	5.65	5.29	0.18	2.69E-03
P 6 ( $n = 3,239$ )							
Undirected	5,738	4.18	3,239	5.65	5.73	0.24	4.04E-03
Directed	7,828	2.40	1,707	6.78	6.20	0.24	4.04E-03

*Note.* For the directed networks, all of the remaining measures were calculated on the largest connected component:  $L$ , the average shortest path length;  $L_{random}$ , the average shortest path length for the random graphs;  $C$ , clustering coefficient;  $C_{random}$ , the clustering coefficient for the random graphs. P, participant;  $n$ , the number of nodes in the network;  $k$ , the average (in-) degree of all nodes;  $n_{cc}$ , the number of nodes in the largest connected component.

### 2.2.1. Sparseness and connectedness

The size of the networks varies substantially between participants, ranging from 1,358 to 9,429 unique nodes. The number of links also varied considerably between participants. The number of links is always higher in the directed networks than in the undirected networks because, in directed networks, two words can be connected by two links that go in opposite directions. Given the large size of the networks and the average number of connections, it can be observed that all networks are sparse: On average, a node is connected to a very small percentage of other nodes. For example, in the undirected and directed networks of Participant 4, a word is connected on average to only 0.3% or 0.2% of the total number of words, respectively.

Despite their sparseness, the directed networks have a large strongly connected component in which any word can be reached from any other word while taking the direction of association into account. These components include a substantial percentage of all nodes, ranging from 39% to 53%. The directed networks also have additional connected components that contain a very small number of words. We restricted all further analyses of the directed networks to their largest connected components. In the undirected networks, the whole structure is connected.

### 2.2.2. Average shortest path length $L$

All networks display relatively short average path lengths relative to the number of nodes in the networks. For instance, although the network of Participant 1 has more than 9,000 words, only six undirected connections are required on average to move from any node to any other node. Even when the associative directions are taken into account, the average shortest path length in the network's connected component is only seven. Moreover, the short path lengths observed in the network of each participant are similar to those observed in random graphs with size and mean connectivity  $k$  equal to that observed in the network of each participant. Such short global distances allow for connecting two word nodes chosen at random via a chain of only a few intermediaries.

### 2.2.3. Clustering coefficient $C$

The definition of the clustering coefficient  $C$  proposed by Watts and Strogatz (1998) is independent of whether the links are directed, and so  $C$  is the same for each participant's undirected and directed networks. The results show that  $C$  is above zero in all networks, implying that, on average, a node's neighbors are more likely to be directly connected than two nodes chosen at random. Moreover, in the network of every participant,  $C$  is several orders of magnitude larger than would be expected in a random graph of equivalent size and density. The high clustering, in combination with the short average path lengths reported above, indicates that these networks have a small-world structure.

### 2.2.4. Degree distribution $P(k)$

With the purpose of achieving the best possible parameter estimates, we based our estimates of  $P(k)$  on the maximum number of sampling iterations completed by each participant. Our analyses of the degree distributions in the networks of individuals focus on the in-degree distributions in each participant's directed network. The in-degree of a word, or the number of words for which that word is produced as an associate, is a natural predictor of the prominence of words in memory and has been used as such in a number of studies (McEvoy, Nelson, & Komatsu, 1999; Nelson, Dyrda, & Goodman, 2005).

We applied the principled statistical framework developed by Clauset et al. (2009) that involves comparing the power law with alternative models.<sup>2</sup> The degree distribution follows a power law if it is drawn from a probability distribution

$$P(k) \propto k^{-\alpha}, \quad (2)$$

where  $\alpha$  is a parameter of the distribution known as the exponent or scaling parameter. We considered two alternative models for the data, both of which have been shown to be characteristic of two alternative classes of small-world networks (Amaral et al., 2000). One model is the exponential distribution

$$P(k) \propto e^{-\lambda k}, \quad (3)$$

where  $\lambda$  is the exponential rate parameter. Networks whose degree distribution decays exponentially exhibit only one scale of node connectivity, with highly connected nodes being



essentially nonexistent. A second alternative model is the power-law distribution with exponential cutoff

$$P(k) \propto k^{-\alpha} e^{-\gamma k}. \quad (4)$$

This model is a power law multiplied by an exponential function, where  $\gamma$  is the exponential rate parameter. The cutoff form and the pure power law are similar with respect to the property that most words are relatively poorly connected, while a minority of words is very highly connected. However, the cutoff indicates that the most connected words have a smaller degree than would be expected in a purely power-law distributed network. Essentially, for a power-law distribution with exponential cutoff, the tails are fatter than would be expected according to an exponential distribution, but not as fat as would be expected by a power-law distribution.

The fitting procedure was as follows. We followed the practice of assessing only the tail of the distribution for values greater than a minimum  $k_{min}$ , as most naturally occurring distributions only follow a power-law distribution above some lower bound (Clauset et al., 2009). All models were fitted to the data using the method of maximum likelihood. The value of  $k_{min}$  was chosen to make the probability distributions of the data and the best-fitting power law as similar as possible above  $k_{min}$ . The difference between these two distributions is determined by the Kolmogorov–Smirnov (KS) statistic. The KS statistic was also used when testing the goodness-of-fit of the power law. The  $p$ -values were obtained by generating power-law distributed synthetic data sets with a scaling parameter  $\alpha$  and a lower bound  $k_{min}$  equal to those that best fit the data. Each synthetic data set was fit to its own power-law model, and a KS statistic was calculated for each fit. The  $p$ -value was the fraction of the time that the resulting statistic was larger for the synthetic than for the empirical data.

The approach developed by Clauset et al. (2009) frames the model selection problem in terms of how the power law compares with each of the alternative models. In our analyses, we conducted two model comparisons: one between the power law and the exponential distribution and another between the power law and the power-law distribution with exponential cutoff. The focus on the power-law distribution when addressing the model selection problem can be justified by the large number of data sets, such as the aggregated word association data sets, which have been conjectured to follow a power law. The statistical framework of Clauset et al. (2009) employs a method proposed by Vuong (1989) to select between the models. Vuong's method uses Kullback–Leibler Information Criterion to produce likelihood ratio-based statistics for the null hypothesis that the two models are equally close to the truth against the alternative hypothesis that one of the models is closer. The method can be applied to both nonnested and nested model comparisons. Following Clauset et al. (2009), we calculated the normalized log-likelihood ratio for the nonnested comparison between the power law and the exponential. The log-likelihood ratio was normalized with the estimated standard deviation of the ratio. For the nested comparison between the power law and the power law with cutoff, we calculated the actual log-likelihood ratio.

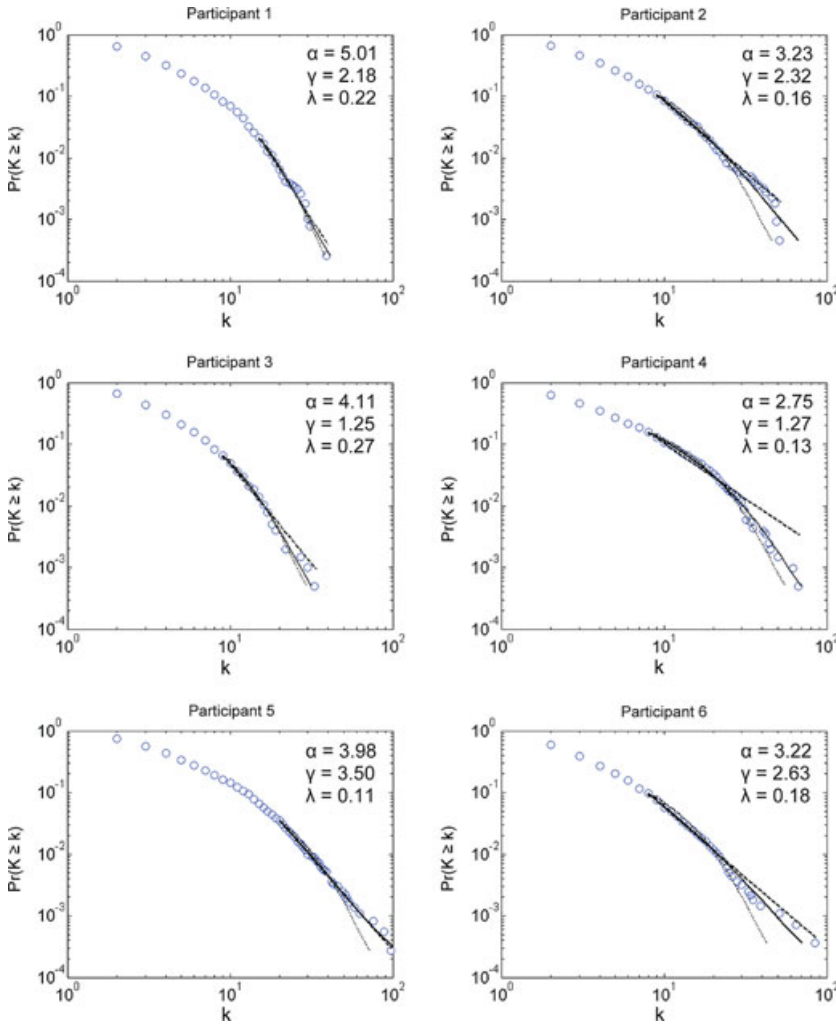


Fig. 1. The cumulative in-degree distribution of each participant's network. The distributions are shown in log–log coordinates, with the lines showing the best-fitting power-law distribution (dashed), power-law distribution with cutoff (solid), and exponential distribution (dotted).  $\alpha$  = exponent for the best-fitting power law;  $\gamma$  = exponent for the best-fitting power law with exponential cutoff;  $\lambda$  = rate parameter for the best-fitting exponential.

To avoid the potential bias associated with how the data are binned in log–log plots (Newman, 2005), we plotted the degree of each directed network relative to its cumulative distribution—showing the probability that a randomly chosen node is of in-degree equal to or higher than  $k$ . Fig. 1 plots the cumulative in-degree distribution in the directed network of each participant, with the best-fitting power-law distribution, power-law distribution with cutoff, the exponential distribution, and the exponents or scaling parameters for each distribution ( $\alpha$ ,  $\gamma$ ,  $\lambda$ , respectively).

Table 2

Tests of power-law behavior in the in-degree distributions of each participant's network

	Power Law		Power Law vs. Power Law With Cutoff		Power Law vs. Exponential	
	KS	<i>p</i>	Log LR	<i>p</i>	Log LR	<i>p</i>
P 1	0.047	.27	-0.74	.22	-0.46	.74
P 2	0.040	.13	-2.35	.03	5.09	.34
P 3	0.058	.02	-2.43	.02	-2.14	.43
P 4	0.059	.00	-11.51	1.61E-06	-7.00	.30
P 5	0.030	.83	-0.21	.52	4.79	.20
P 6	0.031	.38	-1.54	.08	11.78	.21

*Note.* The first two columns present the results of the Kolmogorov–Smirnov statistic for the fit to the power-law model. The remaining columns present the results of the likelihood ratio tests comparing the power law with each of the alternative distributions (i.e., the power-law distribution with cutoff and the exponential distribution). For the power-law distribution with exponential cutoff, we give the actual log-likelihood ratio, whereas for the exponential distribution, we give the normalized log-likelihood ratio. The ratio was normalized with the estimated standard deviation of the ratio. P, participant; KS, Kolmogorov–Smirnov statistic; LR, likelihood ratio.

The power law fits the majority of the degree distributions reasonably well, with the exponents  $\alpha$  varying between 2.75 and 5.01. Yet the distribution in the network of Participant 4 shows a considerable deviation from the power-law form. Fig. 1 also shows that the exponential distribution is overall as good a fit as the power law. Finally, the power law with exponential cutoff is a reasonable alternative model for the data. Whereas the power law with cutoff fits the data of Participants 1 and 5 as well as the power law, it yields a better fit for Participants 2, 3, 4, and 6.

Table 2 presents the KS statistic and the corresponding *p*-value for the fit to the power-law model. According to Clauset et al. (2009), the power law is a plausible hypothesis for the data if  $p > .10$ . Table 2 also displays the results of the likelihood ratio tests comparing the power law with each of the two alternative models. The likelihood ratio test accounts for differences in model complexity by using the chi-square distribution with  $df_1 - df_2$  degrees of freedom, where  $df_1$  and  $df_2$  are the number of free parameters in models 1 and 2, respectively. The logarithm of the likelihood ratios should be positive or negative depending on whether the power law or the alternative model, respectively, is better. The accompanying *p*-values indicate whether the observed sign is statistically significant. As indicated by Clauset et al. (2009), if  $p < .10$ , the sign is a reliable indicator of which model better fits the data; otherwise, neither model is favored over the other.

In line with Fig. 1, the low *p*-values for the fit to the power-law model indicate that the power law is not a plausible fit to the degree distributions in the networks of Participants 3 and 4. In addition, the negative log-likelihood ratios indicate that the power-law distribution with an exponential cutoff is in general a better fit to the degree distributions than the pure power-law form—a result that was also observed for the undirected networks. However, for the networks of Participants 1 and 5, the large *p*-values indicate that there is no statistical

reason to prefer the cutoff form over the alternative model. The results for the exponential distribution are more ambiguous. For Participants 1, 3, and 4, the negative log-likelihood ratios suggest that the exponential is a superior fit compared to the power law, but the large  $p$ -values indicate that the results of the tests are inconclusive. For Participants 2, 5, and 6, the power law is favored over the exponential, but the large  $p$ -values show that the exponential cannot be ruled out as a plausible fit.

### 2.3. Discussion

We have shown that the associative semantic networks of individuals have a small-world structure, characterized by short average path lengths and high local clustering. Moreover, the analyses of the degree distributions in the semantic networks of individuals conveyed two important messages. First, in none of the participants' networks did the power law appear to be truly convincing. The power law with a cutoff was clearly favored in half of the distributions well fit by the pure power law and was also a plausible hypothesis for the other half. Additionally, the exponential distribution could not be ruled out as a plausible hypothesis for all of the distributions well fit by the power law. The second message is that the power law with a cutoff was overall a more reasonable distributional form for the degree distributions in the semantic networks of individuals. The cutoff form was undoubtedly favored over the pure power law for four out of six degree distributions and was also a plausible fit for the remaining two distributions. In brief, our analyses of the semantic networks of individuals suggest that they have a small-world structure, characterized by degree distributions that have a power-law regime followed by an exponential cutoff. This last result means that, in individual networks, most words have relatively few neighbors and a minority of words has a very high, yet bounded, number of neighbors.

## 3. Reanalysis of aggregate associative semantic networks

The finding that the semantic networks of individuals possess a small-world structure is in line with past analyses of semantic networks based on group data (De Deyne & Storms, 2008b; Steyvers & Tenenbaum, 2005). But while the degree distributions in the networks of individuals are best described by a power law truncated by an exponential cutoff, past studies have claimed that aggregate networks have degree distributions that follow a pure power law. However, these claims of power-law behavior have been made solely by observing the approximately straight-line behavior of a histogram on a log-log plot without comparing the fit of alternative models (De Deyne & Storms, 2008b; Steyvers & Tenenbaum, 2005). Moreover, the figures presented in these studies for the degree distributions of directed aggregate networks show that these distributions deviate slightly from the power-law form. Hence, we have applied the statistical framework of Clauset et al. (2009) to existing aggregate networks to examine whether past claims of power-law behavior in these networks are warranted and how their connectivity structure compares with the structure of the individual networks.

### 3.1. Method

We constructed two directed networks, each based on a word association data set that aggregates across the responses of many people. One network was constructed from English word association norms collected at the University of South Florida (Nelson, McEvoy, & Schreiber, 1999). A second directed network was built from Dutch word association norms collected at the University of Leuven (De Deyne & Storms, 2008a). The two networks—which we refer to as the Florida and Leuven networks, respectively—matched the networks analyzed in past studies in size, average connectivity, average shortest path lengths, and clustering (for details, see De Deyne & Storms, 2008b; Steyvers & Tenenbaum, 2005).

### 3.2. Results

A first result is that the aggregate networks are considerably less sparse than the individual networks. The mean number of incoming connections is 12.7 and 16.7 in the Florida and Leuven networks, respectively. In the directed networks of individuals, however, the mean in-degree ranges from 2.12 to 2.96. The aggregate networks also have higher connectedness as indicated by the size of the largest connected components. While in the Florida network, for instance, the largest connected component consists of 96% of all words, in the networks of individuals the component contains only 39% to 53% of all words. Finally, whereas each node in the connected component of the Florida and Leuven networks can be reached from any other node through an average of three and four connections, respectively, in the networks of individuals the average number ranged from five to seven connections. Comparable differences in sparseness and connectedness are observed when the associative directions in the networks are not taken into account.

Our main analyses focused on the distributions of node in-degrees in the directed versions of the Florida and Leuven networks. Fig. 2 plots the cumulative in-degree distribution in

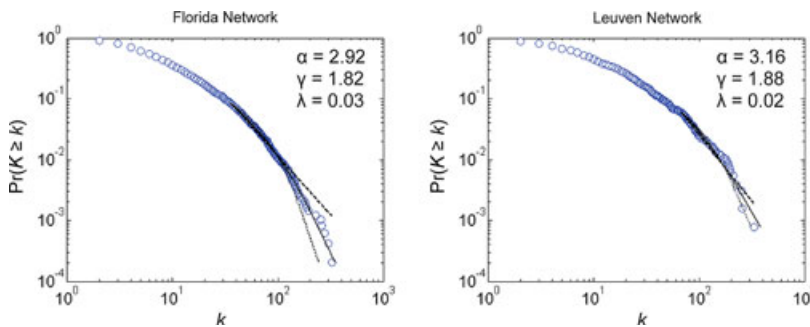


Fig. 2. The cumulative in-degree distributions of aggregate directed semantic networks. The distributions are shown in log–log coordinates with the lines showing the best-fitting power-law distribution (dashed), power-law distribution with cutoff (solid), and exponential distribution (dotted).  $\alpha$  = exponent for the best-fitting power law;  $\gamma$  = exponent for the best-fitting power law with exponential cutoff;  $\lambda$  = rate parameter for the best-fitting exponential.

Table 3

Tests of power-law behavior in the in-degree distributions of aggregate semantic networks

	Power Law		Power Law vs. Power Law With Cutoff		Power Law vs. Exponential	
	KS	<i>p</i>	Log-LR	<i>p</i>	Log-LR	<i>p</i>
Florida network	0.053	.00	-8.04	.00	3.04	.64
Leuven network	0.050	.64	-1.39	.10	0.15	.52

*Note.* The first two columns present the results of the Kolmogorov–Smirnov statistic for the fit to the power-law model. The remaining columns present the results of the likelihood ratio tests comparing the power law with each of the alternative distributions (i.e., the power-law distribution with exponential cutoff and the exponential distribution). We give the actual log-likelihood ratio for the power-law distribution with exponential cutoff, while for the exponential distribution we give the normalized log-likelihood ratio. The ratio was normalized with the estimated standard deviation of the ratio. KS, Kolmogorov–Smirnov statistic; LR, likelihood ratio.

these networks, with the best-fitting power-law distribution, power-law distribution with exponential cutoff, and exponential distribution. As Fig. 2 illustrates, both the power law and the exponential distribution fit the degree distribution in the Florida network rather poorly, but they provide a reasonably good fit to the distribution in the Leuven network. Additionally, Fig. 2 shows that the power law with exponential cutoff is clearly a better fit to the degree distribution in the Florida network and also a reasonable alternative model for the distribution in the Leuven network.

The *p*-values displayed in Table 3 for the fit with the power-law distribution are in line with Fig. 2, indicating that the model is not a plausible fit to the degree distribution in the Florida network. Table 3 also shows the results for the likelihood ratio tests comparing the power law with each of the alternative models. The negative log-likelihood ratios indicate that the power-law distribution with an exponential cutoff is overall a better fit to both degree distributions than the pure power-law form. However, there is no statistical reason to prefer the cutoff form in the case of the Leuven network. The results for the exponential distribution are similar for both networks: The power law is overall a better fit than the exponential, but the latter cannot be ruled out as a possible fit. The likelihood ratio tests held the same pattern of results when the links in the two networks were treated as undirected.

### 3.3. Discussion

Previous studies have claimed that the Florida and Leuven networks exhibit power-law degree distributions. However, our analyses have shown that the power law was not convincing for either network: Even when it was a good fit to the data, the alternatives fit at least as well. For the Florida network, the power law was not a good fit to the data and was ruled out when compared with the power law with exponential cutoff. With regard to the Leuven network, the degree distribution was a plausible power law, but it was also a plausible truncated power law and a plausible exponential. Moreover, our analyses have shown that the Florida and Leuven networks were overall better described by a power-law distribution



truncated by an exponential cutoff. The cutoff form was clearly favored over the pure power law for the Florida network and was also a plausible fit for the Leuven network. This result suggests that aggregate and individual semantic networks display similar degree distributions. Despite this resemblance, aggregate semantic networks are denser and more highly connected than the networks based on individual data.

#### 4. Computer simulation of snowball sampling

We have used the snowball sampling method to uncover the latent structure of the semantic networks of six individuals by sampling a subset of the words or concepts known by each individual. But to what extent do our empirical network samples reveal the structure of the individuals' *true* semantic networks? In other words, can evidence be provided that the new sampling method, as implemented in our experiment, yields network samples that are representative of the individuals' true semantic networks? In this section, we tackle this question by means of computer simulations. We drew snowball network samples from a network representation of memory whose structure was known (hitherto, the memory network), and then examined whether the structure of the sample-based networks resembled the structure of the memory network. We begin by describing how we constructed the memory network. After that, we describe our implementation of the snowball sampling process.

We generated a synthetic network to be used as the memory network, as existing semantic networks were too small for our purposes. The network was generated using a model for the growth of semantic networks as proposed by Steyvers and Tenenbaum (2005). In particular, we used the version of the model that grows a directed synthetic network by the gradual addition of nodes and directed links. In the original article, the model is referred to as Model B. Let  $n$  be the size of the network that we wish to grow. The model starts with a small network of  $M$  nodes ( $M < n$ ), where each node is connected to every other node by a directed link. At each time step, a new node is added to the network by attaching to an existing node  $i$ , chosen with probability proportional to the total number of incoming and outgoing links that it has. Subsequently, the new node connects to  $M$  randomly chosen nodes in

Table 4  
Results of the snowball sampling simulations

Variable	Memory Network	Sampling Iterations				
		2	3	4	5	6
$n$	50,000	26 (3)	393 (193)	3,595 (2,115)	15,027 (7,403)	32,871 (9,879)
$k$	6	18.00 (6.58)	14.30 (2.63)	10.20 (0.58)	8.02 (0.88)	6.75 (0.54)
$L$	6.14	2.41 (0.21)	4.09 (0.60)	5.01 (0.22)	5.50 (0.26)	5.92 (0.19)
$C$	0.32	0.27 (0.08)	0.24 (0.03)	0.25 (0.01)	0.27 (0.01)	0.30 (0.01)

*Note.* Mean statistics across 20 simulations. Standard deviations are given between parentheses.  $n$ , the number of nodes in the network;  $k$ , the average degree of all nodes in the network;  $L$ , the average shortest path length;  $C$ , clustering coefficient.

the neighborhood of node  $i$ ; in the simplest version of the model, the connection probabilities are distributed uniformly over the neighborhood of node  $i$ . Finally, the direction of each new link is chosen randomly and independently of the other links, pointing toward the older node with probability  $\alpha$  and toward the new node with probability  $1 - \alpha$ . In our simulation of the network growth model, we set  $n = 50,000$  as an approximation of the size of an individual's lexicon of known words. Moreover, we set  $M = 3$ , with all nodes in the neighborhood of node  $i$  being equally likely to be chosen. An  $M$  value of 3 was chosen to achieve an average degree that was similar to the average degree of the networks of individuals. Following Steyvers and Tenenbaum (2005), we set  $\alpha = 0.95$ , corresponding to the assumption that 19 of 20 new links point from a new node toward an existing node. When modeling the sampling process, we adopted the simplifying assumption that the memory network was undirected by ignoring the direction of the links.<sup>3</sup> The first column of Table 4 presents the mean statistical properties of the resultant undirected network, our memory network, across 20 simulation runs.

In what follows, we describe a simple model of the snowball sampling process as implemented in our experiment. We used the model to draw network samples from the memory network and then examined whether the structure of the sample-based networks resembled the structure of the memory network. If the snowball method is a reliable technique for sampling the semantic memories of individuals, then the sample-based networks should display the same structural characteristics as the memory network.

#### *4.1. Description of the sampling process*

Snowball sampling is an iterative process. It starts by selecting a sample of seed-nodes, with simple random sampling without replacement, from the memory network. In the first iteration, the neighbors of each seed are sampled or retrieved from the memory network. The seeds, their neighbors, and the links between them constitute the beginning of the sample-based semantic network. In the second iteration, the neighbors of each node selected in the first iteration are retrieved from the memory network and connected to that node in the sample-based network. In a given iteration, however, only nodes that have not been retrieved before (and hence not already included in the sample-based network) can be retrieved. The process can be iterated as often as desired.

The snowball sampling model has three parameters. In our simulations, we set these parameters according to our empirical work to examine whether the snowball method, as implemented in our experiment, is able to recover the structural characteristics of the memory network. One parameter in the model is the number of seed-nodes. We used five seed-nodes in our simulations because that was the number of words used as seeds when sampling the majority of our participants' networks. A second parameter is how many neighbors of a node can be sampled. We sampled all of a node's neighbors in our simulations based on our participants' reports that they felt they had generated all associates of a word they could possibly think of. Finally, we iterated the snowball process six times—the same number of iterations that were used when sampling the empirical networks. If the snowball sampling process yields close agreement with the statistical properties of the memory network with

these parameter values, then there is good reason to believe that our empirical networks give a reliable indication of the true structure of participants' semantic memories.

#### 4.2. Results and discussion

We compared the sample-based network with the original memory network in terms of three statistical properties: the average degree  $k$ , the average shortest path length  $L$ , and the clustering coefficient  $C$ . These properties were calculated for the sample-based network after each of six iterations of the snowball process. Simulations were repeated 20 times, each time with a different set of five seed-nodes. Table 4 shows the statistics of the memory network, the mean statistics, and standard deviations for the sample-based network from iterations 2 to 6 over 20 simulation runs. We do not give the statistical properties for iteration 1 because the sample-based network had very few nodes at that point in the simulation.

The results show that the sample-based network grew quickly over time. For instance, the percentage of nodes sampled increased from 30% to 66% from iteration 5 to iteration 6. Moreover, the snowball sampling process led to an overestimation of the real average degree  $k$  in early iterations—a result that has been found in past simulation work (Illenberger, Flötteröd, & Nagel, unpublished data; Lee, Kim, & Jeong, 2006). By definition, well-connected nodes have a higher probability to be sampled compared with poorly connected nodes because they have more connections along which they can be discovered. In addition, nodes that are predominantly linked to well-connected nodes are also more likely to be sampled. Yet our results show that, after iteration 6, the estimate of  $k$  for the sample-based network approximated the true value in the memory network. The results also show that the average shortest path length  $L$  and the clustering coefficient  $C$  were underestimated in early iterations, but that reasonably good estimates were obtained by iteration 6. In short, our simulation shows that the snowball sampling process yielded close agreement with the statistical properties of the memory network, especially after the sixth sampling iteration. This result suggests that the statistical properties observed in the empirical networks of individuals can be considered to be approximating the properties of the individuals' true networks, particularly for those participants who completed six or seven sampling iterations.

### 5. General discussion

The present work characterized the structure of semantic networks based on individual data and examined how their structure compares with the structure of existing aggregate networks. We have shown that the semantic networks of individuals have a small-world structure with short distances between words and high local clustering. The degree distributions in these networks follow a power law truncated by an exponential cutoff, meaning that most words are relatively poorly connected and a minority of words has a very high, although bounded, number of connections. Furthermore, our analyses showed that the power law with an exponential cutoff is also a more reasonable distributional form for the aggregate

networks. This result indicates that aggregate semantic networks are not scale-free, as has been previously claimed, but have degree distributions that are similar to those observed in the individual networks. Nevertheless, there are properties of individual structure that the aggregate networks do not reflect. Specifically, the aggregate networks have a comparatively higher average degree, larger connected components, and shorter distances between words. The structural differences between individual and aggregate networks indicate that semantic networks may take a different form when data are combined over individuals. Finally, our simulations of the snowball method give us reason to believe that our individual maps of semantic structure reliably depict the structure of semantic memory. Future simulation efforts may examine whether the snowball method gives accurate estimates of network structure under different parameterization and more realistic assumptions about how memory works. Such simulations could support the development of a more refined snowball method that would use fewer seeds and iterations, being therefore faster and cheaper to use in the laboratory. In what follows, we (a) consider some methodological issues in the estimation of degree distributions, (b) discuss the psychological implications of semantic network structure, and (c) conclude by discussing possible applications of the snowball sampling method in future research.

### *5.1. Estimation of degree distributions*

We have relied on the approach of Clauset et al. (2009) to detect and characterize power-law behavior in network degree distributions. Our analyses indicate that the power law with an exponential cutoff is, *overall*, a better model for the degree distributions of both individual and aggregate semantic networks. However, there were cases where the likelihood ratio test was unable to distinguish power-law from non-power-law behavior. In the networks of Participants 1 and 5, and in the aggregate Leuven network, it was not possible to distinguish between the pure power law and the truncated power-law form. Also, for none of the networks analyzed in this work, could the likelihood ratio test distinguish between the power law and the exponential distributions. Still, we can say with certainty that the power-law hypothesis was never a truly convincing model for the individual or the aggregate networks. Even when the power law was a good fit to the data, there was another distribution that gave a fit as good as or better than the power law. In future work, it may be worthwhile to develop other principled approaches for comparing between alternative distributions, such as Bayesian inference, cross-validation, or minimum description length. These methods could be used not only to compare the power law with each of its alternatives, as in the approach of Clauset et al. (2009), but also to compare between the alternative models themselves. As we discuss later, these developments would be particularly useful to researchers interested in the mechanisms underlying the formation of human semantic networks, where it may matter greatly whether the degree distribution follows a power law or some other form.

### *5.2. Psychological implications of semantic network structure*

The question of whether the observed degree distributions follow a power law or some other form has important psychological implications. For instance, the observed

distributions can be considered as an indication of the limited capacity of the cognitive system to store and process information. Such limits are sensible in light of the fact that connections between concepts come with costs, requiring resources to be learned and maintained. Much in the same way, social and biological systems that operate under resource constraints often display a limit on the number of connections that nodes can have (e.g., Amaral et al., 2000; Jeong, Mason, Barabási, & Oltvai, 2001; Newman, 2001). By reflecting statistical associations between words that can be used to retrieve information that is likely to be useful in the current context, limited word connectivity could reduce the cost of processing irrelevant information (Schooler & Anderson, 1997).

Furthermore, the structural properties of our individual maps of semantic structure, such as the shape of their degree distributions, have implications for the processes through which semantic memory grows or develops over time. Since not all mechanisms of semantic growth may give rise to the observed type of memory structure, this structure can help restrict the search space of possible growth mechanisms. One model of network growth that has been widely studied is the preferential attachment model, also known as the rich-get-richer model (Barabási & Albert, 1999; see also Simon, 1955). In the model, the probability of a word acquiring new associates over time is directly proportional to the number of associates that it already has. This equates to well-connected or richer words being more likely to acquire new associates, while poorly connected words are disproportionately likely to remain poor. The preferential attachment model gives rise to power-law degree distributions (e.g., Amaral et al., 2000), but it is unable to generate structures like truncated power-law degree distributions or clustered neighborhoods of densely connected nodes, both of which characterize the semantic networks of individuals. In other words, our structural analyses of individual networks indicate that a pure preferential attachment model cannot account for the growth of a person's semantic knowledge.

Using the current structure of networks to make inferences about their past growth leaves many possible growth models in play. In fact, apart from the pure preferential attachment model, there are a number of other growth mechanisms that could give rise to the truncated power-law distributions that characterize the semantic networks of individuals. Next, we discuss how the snowball method may provide the researcher with empirical data suitable for distinguishing between alternative models of semantic growth.

### *5.3. Further applications of the snowball sampling method*

In this article, we have used the snowball method to take “snapshots” of the semantic networks of individuals at one time point in the life span. Future research on semantic development could use the snowball method to take multiple snapshots over time that can be sequenced into a “movie” of how the individual's network develops over time. These individual movies could cover the time from childhood to adulthood or a more restricted period of life, such as a single year at school. The developing networks could then be used to test alternative predictions as to what words are more likely to be learned by the individual or to enter his or her network of already known words at each period in the life span. For instance, the likelihood that an unknown word is learned at a given developmental period may be

proportional to the number of associates that the word has in the language-learning environment or, alternatively, in the individual's network of already known words (Hills, Maoune, Maoune, Sheya, & Smith, 2009). These longitudinal analyses of individual networks could also elucidate how interindividual differences in word acquisition may lead to adult semantic networks with distinct structural properties. As our statistical analyses have shown, the structural properties of our network snapshots were not entirely homogeneous across individuals. Interindividual differences in network structure may arise from distinct processes of word acquisition or from the same process running at different rates. Finally, aging researchers could use the snowball method to sample the semantic network of the aging individual at fixed time intervals throughout the later stages of life. These movies of aging networks could shed light on how aging impacts the structure of adult semantic networks and why some network topologies may be less resilient than others to the impact of aging.

## 6. Concluding remarks

The time and expense of using the snowball method to map the structure of even a small portion of a person's semantic memory is not cheap. Yet this investment comes with a guaranteed return. In a few weeks, we can learn with unparalleled precision about the structure of an individual's semantic memory, the residue left from years of life experience.

## Notes

1.  $L$  must be computed on networks or network components in which any node can be reached from any other node by traversing a finite number of connections. In directed networks, however, not all nodes can be reached from a given node. For this reason, statistical analyses of directed networks are typically restricted to the largest connected component in the network. A connected component is a set of nodes where every node can be reached from any other node.
2. In our analyses, we used the method implementations made available by Aaron Clauset and Cosma R. Shalizi at <http://www.santafe.edu/~aaronc/powerlaws/>.
3. By ignoring the link directionalities from Model B (the directed model) at the output stage, we made it equivalent to Model A (the undirected model). As shown by Steyvers and Tenenbaum (2005), Model B reproduces the results of Model A when all directed links are converted to undirected.

## Acknowledgments

This work was supported by a fellowship to Ana Sofia Morais from the International Max Planck Research School on the Life Course (LIFE). The authors thank Soo-Youn Lee for editing the data and the fellows and faculty of LIFE for helpful comments and discussions.



## References

- Amaral, L. A. N., Scala, A., Barthélemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*, 11149–11152.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*, 73–79.
- Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*, 661–703.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 447–505.
- De Deyne, S., & Storms, G. (2008a). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*, 198–205.
- De Deyne, S., & Storms, G. (2008b). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*, 213–231.
- Goodman, L. A. (1961). Snowball sampling. *Annals of Mathematical Statistics*, *32*, 148–170.
- Hills, T., Maoune, M., Maoune, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, *20*, 729–739.
- Jeong, H., Mason, S., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*, 41–42.
- Lee, S. H., Kim, P.-J., & Jeong, H. (2006). Statistical properties of sampled networks. *Physical Review E*, *73*, 016102.
- McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What's the connection between true and false memories: The different roles of inter-item associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1177–1194.
- Nelson, D. L., Dyrdal, G., & Goodmon, L. (2005). What is preexisting strength? Predicting free association probabilities, similarity ratings, and cued recall probabilities. *Psychonomic Bulletin & Review*, *12*, 711–719.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). *The University of South Florida word association norms*. Available at: <http://w3.usf.edu/FreeAssociation>. Accessed October 23, 2008.
- Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*, 404–409.
- Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*, 323–351.
- Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology*, *32*, 219–250.
- Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*, 425–440.
- Simon, H. A. (1986). The parameters of human memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities* (pp. 299–309). Amsterdam: North Holland.
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*, 598–605.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*, 307–333.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442.