

Random Graph Model Simulations of Semantic Networks for Associative Concept Dictionaries

Hiroyuki Akama

Tokyo Institute of Technology
2-12-1 O-okayama Meguro-ku
Tokyo 152-8550, Japan
akama@dp.hum.titech.ac.jp

Terry Joyce

Tama University
802 Engyo Fujisawa-shi
Kanagawa-ken 252-0805, Japan
terry@tama.ac.jp

Jaeyoung Jung

Tokyo Institute of Technology
2-12-1 O-okayama Meguro-ku
Tokyo 152-8550, Japan
catherina@dp.hum.titech.ac.jp

Maki Miyake

Osaka University
1-8 Machikaneyama-cho Toyonaka-shi
Osaka 560-0043, Japan
mmiyake@lang.osaka-u.ac.jp

Abstract

Word association data in dictionary form can be simulated through the combination of three components: a bipartite graph with an imbalance in set sizes; a scale-free graph of the Barabási-Albert model; and a normal distribution connecting the two graphs. Such a model makes it possible to simulate the complex features in degree distributions and the interesting graph clustering results that are typically observed for real data.

1 Modeling background

Associative Concept Dictionaries (ACDs) consist of word pair data based on psychological experiments where the participants are typically asked to provide the semantically-related response word that comes to mind on presentation of a stimulus word. Two well-known ACDs for English are the University of South Florida word association, rhyme and word fragment norms (Nelson et al., 1998) and the Edinburgh Word Association Thesaurus of English (EAT; Kiss et al., 1973). Two ACDs for Japanese are Ishizaki's Associative Concept Dictionary (IACD) (Okamoto and Ishizaki, 2001) and the Japanese Word Association Database (JWAD) (Joyce, 2005, 2006, 2007).

While there are a number of practical applications for ACDs, three are singled out for mention

here. The first is in the area of artificial intelligence, where ACDs can contribute to the development of intelligent information retrieval systems for societies requiring increasingly sophisticated navigation methods. A second application is in the field of medicine, where ACDs could be used in developing systems that seek to prevent dementia by checking higher brain functions with a brain dock. Finally, within educational settings, ACDs can greatly facilitate language learning through the manifestation of inherent cultural modes of thinking.

The typical format of an ACD is to list the stimulus words (cue words) and their response words together with some statistics relating to the word pairing. The stimulus words are generally basic words determined in advance by the experimenter, while the response words are semantically associated words provided by respondents on presentation of the stimulus word. The statistics for the word pairing include, for example, measured or calculated indices of distance or perhaps some classification of the semantic relationship between the pair of words.

In order to mathematically analyze the structure of ACDs, the raw association data is often transformed into some form of graph or complex network representation, where the vertices stand for words and the edges indicate an associative relationship (Joyce and Miyake, 2007). However, to our knowledge, there have been no attempts at mathematically simulating an ACD as a way of determining in advance the architectural design of a dictionary. One reason is that it is a major challenge to compute maximum likelihood estimations (MLEs) or Monte-Carlo simulations for graph data (Snijder, 2005). Thus, it is extremely difficult to predict dependences for unknown

© 2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported* license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>). Some rights reserved.

factors such as the lexical distribution across a predetermined and controllable dictionary framework starting simply from a list of basic words. Accordingly, we propose an easier and more basic approach to constructing an ACD model by combining random graph models to simulate graph features in terms of degree distributions and clustering results.

2 Degree distributions for ACDs

2.1 Typical local skew

It is widely known that Barabási and Albert (1999) have suggested that the degree distributions of scale-free network structures correspond to a power law, expressed as $P(x = d) = d^{-r}$ (where d stands for degree and r is a small number, such as 2 or 3). This type of distribution is also known as Zipf's law describing the typical frequency distribution of words in a document and plots on a log scale as a falling diagonal stroke. However, in the degree distribution of ACDs, there is always a local skew, as a local peak or bump with a low hemline. Figure 1 presents two degree distributions; for the IACD (upper) ($r = 1.8$) and the JWAD (lower) ($r = 2.3$).

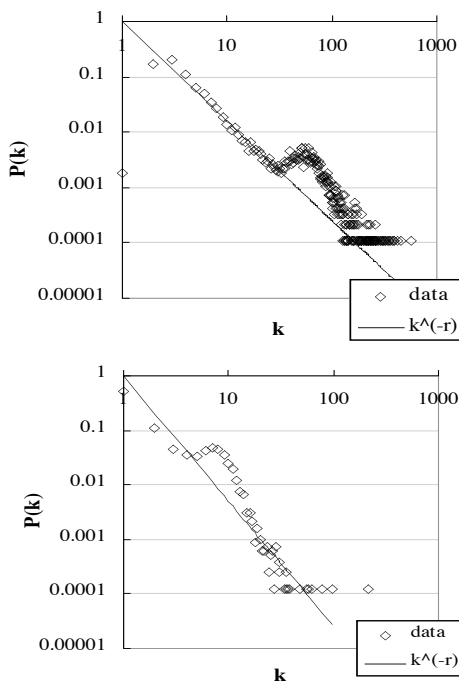


Figure 1. Degree distributions for actual data

The plots indicate a combination of heterogeneous distributions, consisting of a single degree

distribution represented as a bell form with a steep slope on the right side. However, what is most interesting here is that throughout the distribution range the curves remain regular and continuous, with an absence of any ruptures or fractures both before and after the local peaks.

When actual ACD data is examined, one finds that as response words are not linked together, almost all the words located in the skewed part are stimulus words (which we refer to as peak words in this study), while the items before the local peak are less frequent response words that have a strong tendency to conform to a decaying distribution. It is therefore relatively natural to divide all word pairs into two types of graph: either a bipartite graph for new response words that are not already part of the stimulus list and a graph that conforms to Zipf's law for the frequencies of response words that are already present in the stimulus list. For the first type, new response words are represented as nodes only with incoming links, generating a bipartite graph with two sets of different sizes. This bipartite graph would exhibit the decaying distribution due to low-frequency response words prior to the local peak. In the second type of graph, response words are represented as nodes with both incoming and outgoing links. This second type is similar to a scale-free graph, such as that incorporated within the Barabási-Albert (BA) model.

2.2 Bipartite Graph and BA Model

A bipartite graph is a graph consisting of vertices that are divided into two independent sets, S and R, such that every edge connects to one S vertex and one R vertex. The graph can be represented by an adjacency matrix with diagonal zero submatrices, where the values of the lower right submatrices would all be zero were it not for the appearances of some stimulus words as response words. The lower right section is exactly where the extremely high degrees of hubs are produced, which far exceed the average numbers of response words.

Thus, we adopt an approach to generating a scale-free graph that reflects Zipf's law for frequency distributions. According to the BA model, the probability that a node receives an additional link is proportional to its degree. Here, we implement the principle of *preferential attachment* formulated by Bollobás (2003):

$$P(x \in N_{x+1}) = md_i(x) / \sum_{T=1}^t d_i(T) \quad (1),$$

with the addition of one condition that is specific to ACDs, which we explain below. The BA model starts with a small number, m_0 of vertices, and at each time step, T , a new vertex with m edges is added and linked to m different vertices that already exist in the graph. N_{t+1} represents a random set of m early vertices, $d_t(i)$ the degree of vertex i in the process at time t . The probability that a new vertex will be connected to a vertex i depends on the connectivity of that vertex, as expressed by Equation (1). However, we specifically assume that m is a random natural number that is smaller than m_0 , because in actual data the ratio of stimulus words among all responses words for each stimulus word is obviously far from constant.

Moreover, the graph for the BA model here should be regarded as being a directed graph, because the very reason that hubs emerge within semantic network representations of ACDs is that the number of *incoming* edges is much larger than the expected number of nodes for each possible *in-degree*. In contrast, *out-degree* is limited by the number of responses for each stimulus word i , which is represented as $c(i)$. Let $c(i)$ follow a normal distribution with a mean c_m and a small variance value σ^2 (which is not constant but nearly so) to smoothly combine the distribution of the bipartite graph and the power distribution. If a *directed* adjacency matrix for the network exclusively between stimulus words is expressed as $D(B_{ij})$, then the sum of the non-zero values for each row in a random bipartite graph introducing new response words will be $c(i) - \sum_{l^*} D(B_{ij})$ (The vertices of stimulus

words with the subscript j are linked with the vertex of the stimulus word i). Thus, new response words—words that are not stimulus words—will be randomly allocated within a bipartite graph according to Equation (2):

$$P((i,l) = 1) = r^{-1}(c(i) - \sum_{l^*} D(B_{ij})) \quad (2),$$

where r is the approximate number of such words. Equation (2) will yield the lower left and the upper right sections of the complete adjacency matrix A for the ACD model. The subsequent sub-matrix P^t refers to the transposition of the prior sub-matrix P . The adjacency matrix in Equation (3) represents a pseudo bipartite graph structure where the upper left section is a zero sub-matrix (because there are no intercon-

nections among new response words), but the lower right section is not. Here, B_{ij} (not $D(B_{ij})$, but the undirected counterpart to it), which corresponds to the BA model, is taken as a subsection of the adjacency matrix that must be *non-directed* for the whole composition.

$$A = \begin{pmatrix} O & P^t \\ P & B_{ij} \end{pmatrix} \quad (3)$$

The key to understanding Equation (3) is to realize that P is conditionally dependent on B_{ij} , because we assume a normal distribution for the number of non-zero values at each row in the lower section of A .

2.3 Simulation Results

Taking into account the approximate numbers of possible new response words, in other words, the balance in sizes between the two sets in the bipartite graph, we built a composition of partial random graphs that could represent an adjacency matrix of the ACD model. Figure 2 presents one of the results obtained for the following conditions:

$$t = 90, m_0 = 10, m = 3, c_m = 5, \sigma = 1, r = 3000.$$

As the Figure shows, the local peak and the accompanying hemline in the degree distribution are clearly simulated by the complex combination of random graphs.

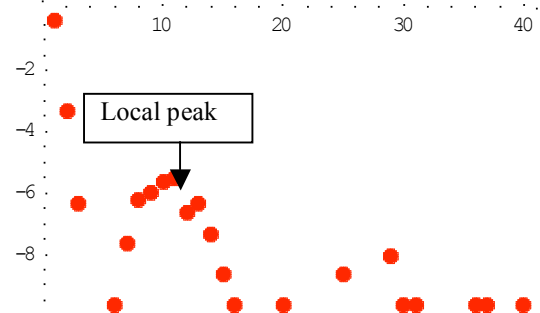


Figure 2. Degree distribution of an ACD model

The degree distribution for the artificial network is consistent with the features observed for actual ACD data, where more than 96% of the stimulus words in each data set are distributed across the peak section of the degree distribution, which is why we have referred to them as peak words. Moreover, it is easy to verify that without the assumption of a normal distribution for $c(i)$, distinct fractures emerge in the artificial curve where new response words in the bipartite struc-

ture would be distinguished from stimulus words located at initial points of the local peak.

3 Markov Clustering of ACDs

3.1 MCL

This section introduces the graph clustering method that is applied to both the real and artificial ACD data in order to compare them. Markov Clustering (MCL) proposed by Van Dongen (2001) is well known as a scalable unsupervised cluster algorithm for graphs that decomposes a whole graph into small coherent groups by simulating the probability movements of a random walker across the graph. It is believed that when MCL is applied to semantic networks, it yields clusters of words that share certain similarities in meaning or appear to be related to common concepts.

3.2 MCL Results

The clustering results for the ACD model created by combining random graphs reveal that each of the resultant clusters contains only one stimulus word surrounded by several response words. This result is somewhat strange because there are dense connections between stimulus words, which would lead us to assume that clusters would have multiple stimulus word. However, the results of applying MCL clustering to the graph for the ACD model are in reality highly influenced by the sub-structure of the bipartite graph and less dependent on the scale-free structure.

Nevertheless, the result is quite similar to results observed with real data. On examining MCL clustering results for different ACD semantic networks, we have observed that MCL clusters tend to consist of one word node with a relatively high degree and some other words with relatively low degrees. On closer inspection of the graph, it is possible to see several supporter nodes that gather around one leader node, forming a kind of small conceptual community. This suggests that the highest degree word for each cluster becomes a representative for that particular cluster consisting of some other low degree words. In short, MCL clustering is executed based on such high degree words that tend to have relatively low curvature values (Dorow, 2005) compared to their high average degree values.

4 Conclusion

In this paper, we have proposed a basic approach to simulating word association dictionary data through the application of graph methodologies. This modeling is expected not only to provide insights into the structures of real ACD data, but also to predict, by manipulating the model parameters, possible forms for future ACDs. Future research will focus on constructing an exponential random graph model for ACDs based on Markov Chain Monte Carlo (MCMC) methods.

References

- Barabási, Albert-László and Réka Albert. 1999. *Emergence of scaling in random networks*, Science. 286:509-512.
- Bollobás, Béla. 2003. *Mathematical Results on Scale-free Random Graphs*, <http://www.stat.berkeley.edu/~aldous/Networks/bol11.pdf>
- Dorow, Beate et al. 2005. *Using Curvature and Markov Clustering in Graphs for Lexical Acquisition and Word Sense Discrimination*, MEANING-2005,2nd Workshop organized by the MEANING Project, February,3rd-4th.
- Joyce, Terry and Maki Miyake. 2007. *Capturing the Structures in Association Knowledge: Application of Network Analyses to Large-Scale Databases of Japanese Word Associations*, Large-Scale Knowledge Resources. Construction and Application, Springer Verlag:116-131.
- Kiss, G.R., Armstrong, C., Milroy, R., and Piper, J. 1973. *An associative thesaurus of English and its computer analysis*, In Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (Eds.), *The Computer and Literary Studies*, Edinburgh University Press.
- Nelson, Douglas L., Cathy L. McEvoy, & Thomas A. Schreiber. 1998. *The University of South Florida word association, rhyme, and word fragment norms*, Retrieved August 31, 2005, from <http://www.usf.edu/FreeAssociation>
- Okamoto, Jun and Shun Ishizaki. 2001. *Associative Concept Dictionary and its Comparison Electronic Concept Dictionaries*. PACLING2001-4th Conference of the Pacific Association for Computational Linguistics:214-220.
- Snijders, Tom A.B., Philippa E. Pattison, Garry L. Robins, Mark S. Handcock, 2005. *New Specifications for Exponential Random Graph Models*, <http://stat.gamma.rug.nl/SnijdersPattisonRobinsHandcock2006.pdf>
- Steyvers, Mark and Josh Tenenbaum. 2005. *The Large-Scale Structure of Semantic Networks, Statistical Analyses and a Model of Semantic Growth*, Cognitive Science. 29 (1):41-78.

