Deriving Word Association Networks from Text Corpora

David Galea (dp.galea@student.qut.edu.au) and **Peter Bruza (p.bruza@qut.edu.au)**
Information Systems School
Queensland University of Technology
2 George Street
Brisbane, QLD 4000 AUS

Abstract

This article presents and evaluates a model to automatically derive word association networks from text corpora. Two aspects were evaluated: To what degree can corpus-based word association networks (CANs) approximate human word association networks with respect to (1) their ability to quantitatively predict word associations and (2) their structural network characteristics. Word association networks are the basis of the human mental lexicon. However, extracting such networks from human subjects is laborious, time consuming and thus necessarily limited in relation to the breadth of human vocabulary. Automatic derivation of word associations from text corpora would address these limitations. In both evaluations corpus-based processing provided vector representations for words. These representations were then employed to derive CANs using two measures: (1) the well known cosine metric, which is a symmetric measure, and (2) a new asymmetric measure computed from orthogonal vector projections. For both evaluations, the full set of 4068 free association networks (FANs) from the University of South Florida word association norms were used as baseline human data. Two corpus based models were benchmarked for comparison: a latent topic model and latent semantic analysis (LSA). We observed that CANs constructed using the asymmetric measure were slightly less effective than the topic model in quantitatively predicting free associates, and slightly better than LSA. The structural networks analysis revealed that CANs do approximate the FANs to an encouraging degree.
Keywords: semantic networks; free association networks; corpus-based semantic representation

Introduction

The mental lexicon is a mental dictionary of words, but its structure is founded on the associative links that bind these words together. Such links are acquired through experience and the vast and semi-random nature of this experience ensures that words within the lexicon are highly interconnected, both directly and indirectly through other words. For example, during childhood development and the associated acquisition of English, the word *planet* becomes associated with *earth*, *space*, *moon*, and so on. Even within this set, *moon* can itself become linked to *earth* and *star* etc. Words are so associatively interconnected with each other that they meet the qualifications of a 'small world' network wherein it takes only a few steps to move from any one word to any other in the lexicon (Steyvers & Tennenbaum, 2005). Because of such connectivity individual words are not represented in long-term memory as isolated entities but as part of a network of related words. One approach to extract such network is to employ a target as a cue and collect free associations from human subjects (Nelson, McEvoy, & Schreiber, 2004; Simon, Navarro, & Storms, 2013). For example, Figure 1 depicts such a network where t is the target word and the a_i's denote associates. An arrow, e.g., $t \rightarrow a_1$ represents that associate a_1 was produced in a free association experiment in respect to target t. Table 1 shows the corresponding adjacency matrix for this example network. When collected over a subject pool, the edges can be weighted, e.g., by the probability that a given associate is produced in relation to a cue. Such networks are referred to as free association networks (FANs). FANs have formed the basis of human memory models such as Spreading Activation (Collins & Loftus, 1975) and Processing Implicit and Explicit Representations (PIER) (Nelson, Schreiber, & McEvoy, 1992; Nelson, Kitto, Galea, McEvoy, & Bruza, 2013).

FANs have the following structural characteristics:

R1 The edges are directed, hence allowing for asymmetric associations between words.

R2 The target word has an edge with each associate in the network.

R3 The edges are weighted.

FANs are derived manually which is time consuming and labor intensive. They are therefore restricted in relation to the breadth of vocabulary in human language and challenging to keep up-to-date as language and associations evolve. The aim of this paper is investigate to what degree corpus-based semantic methods can be used to approximate FANs in relation to both their structural network characteristics and their ability to quantitatively predict human word associations. We shall refer to such networks as Corpus Based Association Networks (CANS).

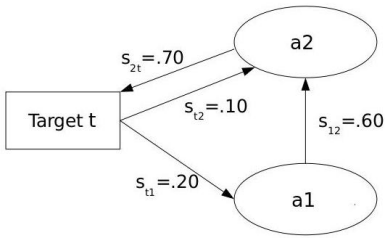

Figure 1: Example of a Free Association Network

Corpus Based Association Networks

A CAN comprises nodes, which correspond to words, and directed weighted edges, which model the associations between words. We begin by describing how the nodes of a CAN are constructed.

	t	a_1	a_2
t	0	0.2	0.1
a_1	0	0	0.6
a_2	0.7	0	0

Table 1: Example adjacency matrix of the FAN depicted in Figure 1

Vector Representations of Words

Each word u, (i.e., a node) has a vector based representation $\underline{u}$, where the vector has been computed from an underlying corpus. There are a variety of strategies to produce such vectors (Bullinaria & Levy, 2007), which are sometimes referred to as "semantic vectors" due to their ability to replicate human semantic association norm data (Dumais, 2004; Lund & Burgess, 1996; Turney & Pantel, 2011).

We used a Positive Pointwise Mutual Information (PPMI) vector representation because of its robust performance across a variety of linguistic and semantic tasks (Bullinaria & Levy, 2007). PPMI vectors are derived from discrete probability distributions built from word co-occurrence statistics. In our case, these discrete probability distributions are built from a modified version of a standard word co-occurrence matrix where the rows correspond to a set of pre-defined target words. The co-occurrence frequencies of a given target word with other words are computed using a sliding window of fixed size (denoted w) across the corpus where sentence and paragraph boundaries are ignored. Context words are those words surrounding the target word when it is centered in the window. The frequency of each context word is accumulated as the window slides across the corpus. In this process, stop words are ignored. The frequencies are subsequently normalized to produce a probability distribution for the given target word. As a consequence all vector elements are positive real values, and thus exist in the first orthant of Euclidean Space. This property has important consequences for the bounds for the word association measures to be discussed in the next section.For this analysis, both target and context words were treated as single tokens. Furthermore the window size was not explored as part of this analysis.

Measures of Association $S(\underline{u},\underline{v})$

The preceding section described how the nodes of a CAN are represented via corpus-based vectors. These vectors are used to compute weighted associations between words thus providing the means to derive edges for CANs. For this paper, we have utilized one well known metric: the cosine metric as well as introducing a new measure of association called the GP measure.

The cosine metric was chosen as a baseline as it is often used to compute vector based associations, e.g., in the Latent Semantic Analysis model where it has shown consistently good performance in computing associations between words across a number of studies and text corpora (Landauer, Foltz, & Laham, 1998).

$$\cos(\underline{u},\underline{v}) = \frac{\langle \underline{u},\underline{v} \rangle}{||\underline{u}||||\underline{v}||} \tag{1}$$

As pointed out previously, PPMI vector representations exist in the first orthant. Consequently the standard boundaries for the cosine metric being $[-1,1]$ are transformed to $[0,1]$ and can be interpreted as a normalized measure of strength, where 0 represents no relationship between words u and v and 1 represents a perfect synonymous relationship. In having a normalized measure, requirement R3 is satisfied. Unfortunately, as cosine is a metric, its associations are necessarily symmetric meaning $\cos(\underline{u},\underline{v}) = \cos(\underline{v},\underline{u})$. This violates characteristic R1 specified above. In order to satisfy R1, a measure is required that permits asymmetric associations between words. The topic model (Griffiths, Steyvers, & Tenenbaum, 2007) used conditional probabilities to achieve this. For example the strength of association from word u to v is computed as $\Pr(u|v)$ and the strength of reverse relation is computed $\Pr(v|u)$. Note that these probabilities need not be the same which thus allows for asymmetry in the associations between these two words. In this paper, however, we will build on a word association measure based on projection (Pothos, Busemeyer, & Trueblood, 2013). Initially, a simple orthogonal vector projection was considered:

$$P(\underline{u},\underline{v}) = \frac{\langle \underline{u},\underline{v} \rangle}{||\underline{v}||} \tag{2}$$

Exploration of this measure shows that it is bound between $[0,||\underline{u}||]$, where 0 represents no relationship and $||\underline{u}||$ represents a perfect synonymous relationship. Although not normalized this does preserve rank when comparing multiple $\underline{v}$'s to $\underline{u}$. Unfortunately, when comparing multiple $\underline{v}$'s to different $\underline{u}$'s, say $\underline{u_1}$, $\underline{u_2}$ we arrive at two sets of bounds, $\left[0,||\underline{u_1}||\right]$ & $\left[0,||\underline{u_2}||\right]$, which destroys rank equivalence (unless $||\underline{u_1}|| = ||\underline{u_2}||$). To overcome this undesirable property, the GP measure was developed in which the relative difference between $\underline{v}$ and the length of the projection of $\underline{u}$ onto $\underline{v}$ is taken into account:

$$GP(\underline{u},\underline{v}) = \begin{cases} \frac{P(\underline{u},\underline{v})}{||\underline{v}||} & : P(\underline{u},\underline{v}) < ||\underline{v}|| \\ 1 + \frac{||\underline{v}||}{||\underline{u}||} - \cos(\underline{u},\underline{v}) & : P(\underline{u},\underline{v}) \geq ||\underline{v}|| \end{cases}$$

From a technical point of view, GP is not a metric, but a pre-metric. As was the case with cosine, GP is also bound from $[0,1]$ and can be interpreted as a normalised measure of strength (thus satisfying R3). Furthermore, it permits asymmetric associations between words meaning $GP(\underline{u},\underline{v})$ is not necessarily equal to $GP(\underline{v},\underline{u})$, thus satisfying R1.

Constructing Corpus Based Association Networks

This section describes an abstract algorithm to compute a CAN using the notation shown in Table 3. A CAN is based around a target word t.

The first step is to compute the list of associates t_A based on t. In order to compute this list, the vector representation $\underline{t}$ is compared to the vector representation of all other words, v $(v \in V)$ using a measure of association $S(\underline{u},\underline{v})$, which can

be either cosine, or GP. For an associate to be added to the list, the strength of association must be greater or equal to a threshold value: $S(\underline{u},\underline{v}) \geq S_\tau$. This ensures the target has an association with all associates in t_A thus satisfying requirement R2. The threshold is a parameter which is empirically set per measure (cosine or GP).

A word t's t_N is constructed by taking t's t_A and computing the strengths between each directed pair (u,v) $u \neq v$ and including those strengths in which $S(\underline{u},\underline{v}) \geq S_\tau$. The results are stored in t_M so that $t_M(u,v) = S(\underline{u},\underline{v})$. This process is formalized by Algorithm 0.1

Algorithm 0.1: CAN(t,t_A)

$t_A = t_A \cup t$
for each $u \in t_A$
do { **for each** $v \in t_A$, $v \neq u$
do { **if** $S(\underline{u},\underline{v}) \geq S_\tau$
then $t_M(u,v) = S(\underline{u},\underline{v})$

Consider the following example, where a target word t and the associate list $t_A = \{a_1, a_2\}$ and assume the following two associations are above the threshold: $S(\underline{a_2},\underline{a_1}) = S_{1,2} \geq S_\tau$ and $S(\underline{a_2},\underline{t}) = S_{2,t} \geq S_\tau$ and that all other associations $S(\underline{a},\underline{b}) = 0$. Applying Algorithm 0.1, the first step is to add the target t as a default element to its associate list, i.e., $t_A = \{t, a_1, a_2\}$. The next step is to consider the associations that each member of t_A has with one another and keep those for which $S(\underline{a},\underline{b}) \geq S_\tau$

$$u = t\ ,\ v = a_1 : S(\underline{t},\underline{a_1}) = S_{t,1} \geq S_\tau \rightarrow t_M(t,a_1) = S_{t,1}$$
$$v = a_2 : S(\underline{w},\underline{a_2}) = S_{t,2} \geq S_\tau \rightarrow w_M(t,a_2) = S_{t,2}$$

$$u = a_1\ ,\ v = t : S(\underline{a_1},\underline{t}) = 0 \rightarrow t_M(a_1,t) = 0$$
$$v = a_2 : S(\underline{a_1},\underline{a_2}) = 0 \rightarrow t_M(a_1,a_2) = 0$$

$$u = a_2\ ,\ v = u : S(\underline{a_2},\underline{t}) = S_{2,w} \geq S_\tau \rightarrow t_M(a_1,t) = S_{2,t}$$
$$v = a_1 : S(\underline{a_2},\underline{a_1}) = S_{2,1} \geq S_\tau \rightarrow t_M(a_1,a_2) = S_{2,1}$$

The matrix returned by the algorithm t_M is,

Table 2: Adjacency Matrix (t_M) for t

	t	a_1	a_2
t	0	$S_{t,1}$	$S_{t,2}$
a_1	0	0	0
a_2	$S_{2,t}$	$S_{2,1}$	0

Empirical Evaluation

The evaluation aims to address two questions: To what degree CANs approximate FANs with respect to (1) their ability to quantitatively predict human word associations and (2) their structural network characteristics.

Table 3: Notation

u	A word u
$\underline{u}$	The vector representation for u
u_A	The set of associates for u.
mna	The maximum number of associates permitted in u_A
$S(\underline{u},\underline{v})$	Method to measure the strength between $\underline{u},\underline{v}$
S_τ	Minimum threshold value for for $S(\underline{u},\underline{v})$.
u_N	Word Association Network for u
u_M	Adjacency Matrix used to represent u_N
t	A target word
T	Set of Target Words, $T \subset V$
V	Vocabulary of Words

Quantitative Prediction of Word Associations

In order to evaluate the quality of associations in CANs we analyzed the degree to which free associates from the USF norms were appearing in the associate list t_A for a all targets t. To this end we adopt the approach and corpus used to evaluate the Topic Model (Griffiths et al., 2007).

Materials In generating the vector representations, the Touchstone Applied Science Associates (TASA) corpus. was used with a standard stop word list. This corpus comprises 916060 documents. The set of target words T comprised the full 4068 target words present in the University of South Florida (USF) word association norms (Nelson et al., 2004). The baseline models for comparison are the Topic Model (Griffiths et al., 2007) and Latent Semantic Analysis (LSA) (Dumais, 2004). The Topic Model is a corpus based approach to semantic representation which ascribes probabilities to words with respect to latent contexts called "topics". The model allows asymmetric words associations to be computed and has been evaluated on the USF word association norms. The LSA Model was chosen as it a common corpus based benchmark that uses the cosine metric.

Procedure The procedure involves taking each of the 4068 target words and computing the PPMI vector representation using the method described in section "Vector Representations of Words". The size of the resulting vocabulary V was 47059 words, which is the dimensionality of the vector representations. The vocabulary was constructed by taking all words in the TASA corpus (not including stop words) and only considering those with a term frequency greater than 10 (as used with the Topic Model). Thereafter, the associate strength between the target and all words of the vocabulary is computed. This list is then sorted (in descending order) by associate strength and then the rank/position of the target word's first associate is found. The first associate is the associate of the target word (from the USF data) that has the strongest forward relationship. For example, in Fig. 1, a_1 has the strongest forward relationship to t being $S(\underline{t},\underline{a}_1) = 0.2$ and thus would be the first associate for t. The probability of

finding the first associate within the top m associates is computed using: $\Pr(m) = \frac{n_m}{n_T}$, where n_m is the number of first associates produced whose rank $\leq m$ and n_T is the number of words in the corpus.

The cosine and the GP pre-metric were evaluated in this way for 6 different values of m ($m \in M = \{1,5,10,25,50,100\}$) and the results compared with published results of LSA and the Topic Model documented in (Griffiths et al., 2007). In order to determine the best performance a simple method was introduced which sums the probabilties across the different values of m: $P = \sum_{m \in M} Pr(m)$. Best performing results for CAN (cosine) are reported with window size $w = 3$. For CAN (GP) the best performing results were achieved with $w = 6$.

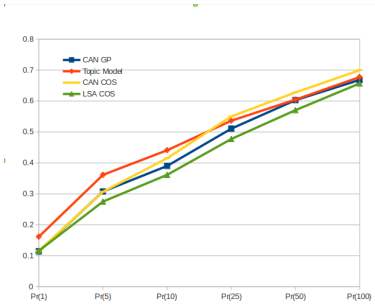

Figure 2: Probabilities for producing the first USF associate modulo the size of the associate list m

Results The results are presented in the Fig 2, the P values for each of the four methods are: $P_{CAN-COS} = 2.7155$, $P_{LSA-COS} = 2.4568$, $P_{Topic-Model} = 2.7818$, $P_{CAN-GP} = 2.5932$.

Of the four, the Topic Model produces the best results followed closely by the CAN (cosine). In comparing both of the baseline methods, CAN (cosine) outperforms LSA. In comparing the asymmetric measures, the Topic Model is slightly more effective than CAN (GP). Given that we are primarily interested in the asymmetric measures of association, we observe that the performance of the Topic Model for first associates for lower m values is considerably better than CAN (GP), however this behavior is not continued for larger m values in which the CAN (GP) approaches and then slightly supersedes the effectiveness of the Topic Model.

Comparison of CANs vs FANs using structural network characteristics

Materials The corpus used for testing was Wikipedia 2008 which comprises 61998051 documents. Wikipedia was chosen and it allows the CAN algorithm to be tested on a very large corpus of text. The set of target words T used was the 4068 target words present in the University of South Florida (USF) word association norms (Nelson et al., 2004). Each word has a corresponding PPMI vector representation using the method described in section . The baseline for comparison are the 4068 FANs in the USF norms.

Procedure A PPMI vector representation for each target word was computed using the method described in section "Vector Representations of Words". The size of the resulting vocabulary V was 255460 words, which is the dimensionality of the vector representations. The procedure involved generating a CAN for each target word using Algorithm 0.1 with GP as the measure used to compute the associations. (CANs were not constructed with cosine as this measure is symmetric) The CANs were generated with $mna \leq 50$, where mna refers to the maximum number of associates a target can have in it's CAN. This value was chosen because it is the maximum number associates encountered across all target words in the USF word association norms.

The structural network characteristics (see Table 4) used for evaluation are derived from the CAN's adjacency matrix (t_M). These characteristics are well known in network analysis and have been used to analyze the USF word association norms (Steyvers & Tennenbaum, 2005) The mean, median and standard deviation (sample size=4068) are calculated for each of these network characteristics. The standard deviation is used to assess the stability of the mean and median.

Table 4: Structural Characteristics

n	The number of nodes in the network.
d	The network density.
L	The average minimum distance between nodes.
$< k >$	The average number of connections for each node.
C	The clustering coefficient for the network.

Table 5: Network Dimension (n)

	USF	GP
Mean	14	16.23
Median	14	14
St Dev	4.7	10.89

Results Table 5 shows that the GP measure has strengths and weakness in replicating the Network Dimension n of the FANs. Whilst CANs over-fit the mean, they produce a perfect median value. There is a quite large standard deviation, which may be due to the fact that it is much easier to establish associations in corpus based processing than humans are able to in free association experiments. We can conclude that whilst the CANs ability to replicate FANs is quite good, there is a larger spread in the numbers of nodes.

Table 6 shows that the mean and median Network Density d of FANs is closely matched by the CANs. Not only it is a great predictor of the mean and median, it's standard deviation is relatively small indicating stability.

Table 7 reveals that the mean and median average minumum distance between nodes in FANs is under-fitted by the CANs, but produces a stable result. This is to be expected given the structure of the USF FANs. These FANs are generally quite sparse except in two areas, firstly all associates have

Table 6: Network Density (d)

	USF	GP
Mean	0.23	0.2
Median	0.21	0.15
St Dev	0.11	0.14

Table 7: Average Minimum Distance Between Nodes (L)

	USF	GP
Mean	1.79	1.19
Median	1.76	1.05
St Dev	0.36	0.32

a forward association to the target (as per R2) and secondly it is a common theme that the backward relationships (to the target) also exist (though these can be of very low weight). Consequently, the majority of associates in a USF FAN are connected to the target in both a forward and backward connection and thus allow for an easier traverse between any two nodes in the FANs resulting in a low L value. The pattern of forward connections is replicated by the CANs (R2) and is strongly desired when replicating FANs (small world behavior). The lower L value for the GP generated CANs indicates that traversal between nodes in a CAN is easier than in a FAN. Given that the densities for FANs and CANs are almost identical (as illustrated in Table 6), and that both have forced forward connections to the target, the difference in structure probably lies in the non-target nodes being, on average, more interconnected in the CANs, than in the FANs. This higher degree of interconnectedness provides more opportunities for traversal through the network and thus a lower L value. Table

Table 8: Average Number of Connections ($< k >$)

	USF	GP
Mean	1.12	2.34
Median	1.14	1.81
St Dev	0.15	1.94

8 shows that the mean and median average number of connections of FANs is over-fitted by the CANs and is quite unstable. On average, the number of associate to associate relationships is greater for CANs than for FANs, which is consistent with our preceding conjecture that the non-target nodes of CANs are more interconnected than in FANs. Again, a possible explanation is that in corpus-based techniques it is generally much easier to establish associations between words. Whether this is a result of the PPMI representation, the large size of the corpus and/or a consequence of the GP pre-metric is currently under investigation.

Table 9 shows that the mean and median Clustering Coefficient C of FANs are under-fitted by the CANs. The Clustering Coefficient measures the average density for localized sub-networks for each node in the network. Although we have observed that words appear to be more connected in CANs over FANs (as observed in Table 7 and 8), there is therefore likely to be, on average, more sub-networks in CANs. However, the density of these sub-networks around a node is smaller than in FANs. The direct cause of this is unknown at this stage.

Table 9: Clustering Coefficient (C)

	USF	GP
Mean	0.44	0.31
Median	0.43	0.32
St Dev	0.10	0.16

Discussion

The first component of analysis evaluated the degree to which CANs can quantitatively predict human word associations. Two models were used as baselines for comparison - the Topic Model and LSA. The results revealed the following findings.

CANs extracted using both the cosine metric and the GP pre-metric outperform LSA though the differences are small. The Topic Model outperforms CAN (GP pre-metric) and CAN (cosine) at higher levels of precision. At lower levels of precision CAN (cosine) outperforms the Topic Model. That being said, all models are poor at generating FANs' first associate at maximal precision (i.e., when $m = 1$). The cosine metric in conjunction with corpus-based vectors like PPMI has shown in many studies to have a predisposition to compute semantic associations (e.g., (Lund & Burgess, 1996; Dumais, 2004)). As there are many cases where the first associate is not semantically associated with the target, it is therefore challenging for such associates to be ranked first based on a PPMI representation. Clearly the asymmetry of GP pre-metric could not mitigate the predisposition of the PPMI vector representations to compute associations of a semantic nature. Conversely, the Topic Model is better at predicting first associates perhaps because the conditional probabilities pick up associations which are broader in nature than semantic associations.

Currently the CAN method creates vector representations for words in Euclidean space. In doing so, established metrics of Euclidean Space (i.e., the cosine metric) can be used to compute word associations. These metrics must satisfy four axioms being (1) $d(a,b) = d(b,a)$, (2) $d(a,a) = 0$, (3) $d(a,b) \geq 0$ and (4) $d(a,b) \leq d(a,c) + d(c,b)$, where $d(a,b)$ denotes the distance between points a and b in the space. Tversky challenged this assumption and found empirical evidence that symmetry (1) and the triangle inequality (4) are violated. Tversky argued that these violations implied that words do not act like points in Euclidean space (Tversky & Gati, 1982). Although the vectors for the CANs are in Euclidean space, the GP pre-metric does not base the degree of association on the distance between points in the space, but

rather on the degree of projection between the respective vectors.

The second component of analysis was to assess the structural similarities of the FANs with the CANs. A set of well known network characteristics were employed to measure the performance. It was found that the CANs built using the GP pre-metric performed encouragingly well at replicating the structural features of the FANs, however issues of stability and under/over fitting the network characteristic need to be investigated in more detail. Structural analysis of the USF norms has been performed previously (Griffiths et al., 2007), however instead of analyzing the individual networks (as done in this analysis), the networks were aggregated into a single global network which was then subjected to network analysis. The focus of this study was different; we were interested in how well FANs based on *individual* target words can be structurally replicated. For this reason, the small world network characteristic γ (used in $P(k) = k^{-\gamma}$) was not investigated because this characteristic is more meaningfully applied to a global network rather than small individual networks.

The brute force style strategy employed to isolate the optimal parameters for the structural analysis could be improved. Whilst it does converge to the optimal set of solutions, it is computationally inefficient and does not explore the stability of each set of solutions, nor does it assign weightings to individual parameters. Lastly, the USF norms collected over three decades and were primarily sourced from students who attended the University of South Florida. As a consequence, the corpus suffers from temporal and geographical bias. To overcome the temporal and geographical bias, a new collection of FANs built by the University Of Leuven could be used as a more comprehensive and contemporary baseline of human word association data (Simon et al., 2013).

Conclusion

The aim of this paper is to investigate to what degree corpus based semantic methods can be used to derive weighted networks of words which approximate human free association networks (FANs) in relation to both structural network characteristics and the ability to quantitatively predict human word associations. We conclude that corpus-based methods can approximate the structural characteristics of FANs to an encouraging degree when a thresholded asymmetric measure based on vector projection is used to construct the network.

The degree to which the corpus-based procedures can replicate human word associations is still questionable. When benchmarked against two corpus-based models, CANs produced similar effectiveness. At this stage we conclude that when term co-occurrence statistics are used to provide vector representations, the performance of the symmetric cosine metric can't be differentiated from an asymmetric measure based on vector projection. The difference in performance between CANs and the benchmark models is small from which we can conclude that CAN (cosine and GP) do show promise for further development.

References

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*, 510-526.

Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428.

Dumais, S. (2004). Latent Semantic Analysis. *Annual review of information science and technology*, *38*, 189-200.

Griffiths, T., Steyvers, M., & Tenenbaum, J. (2007). Topics in semantic representation. *Psychological Review*, *114(2)*, 211-244.

Landauer, T., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2&3), 259-284.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behaviour Research Methods, Instruments & Computers*, *28*(2), 203–208.

Nelson, D., Kitto, K., Galea, D., McEvoy, C., & Bruza, P. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory & Cognition*, *41*(6), 797-819.

Nelson, D., McEvoy, C., & Schreiber, T. (2004). The university of South Florida, word association, rhyme and word fragment norms. *Behavior Research Methods, Instruments & Computers*, *36*, 408-420.

Nelson, D., Schreiber, T., & McEvoy, C. (1992). Processing implicit and explicit representations. *Psychological Review*, *99*(2), 322-348.

Pothos, E., Busemeyer, J., & Trueblood, J. (2013). A quantum geometric model of similarity. *Psychological Review*, *120*(3).

Simon, D., Navarro, D., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*, 480-498.

Steyvers, M., & Tennenbaum, J. (2005). The large scale structure of semantic networks: statistical analyses and a model of semantic growth. , *21*, 41–78.

Turney, P., & Pantel, P. (2011). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141-188.

Tversky, A., & Gati, I. (1982). Similarity, separability and the triangle inequality. *Psychological Review*, *89*, 123–154.