

Structural Differences in the Semantic Networks of Simulated Word Learners

Aida Nematzadeh, Afsaneh Fazly, and Suzanne Stevenson

Department of Computer Science
University of Toronto
{aida,afsaneh,suzanne}@cs.toronto.edu

Abstract

A learner’s semantic network represents the learner’s knowledge of words/concepts and the relations among them. The structure of this network is significant as it might reveal aspects of the developmental process that leads to the network. In this work, we use computational modeling to examine the structure of semantic networks of different simulated word learners. We find that the learned semantic knowledge of a learner that simulates a normally-developing child reflects the structural properties found in adult semantic networks of words. In contrast, the network of a late-talking learner — one that simulates a child with a marked delay in vocabulary acquisition — does not exhibit these properties. We discuss the implications of this result for understanding the process of vocabulary acquisition and delay.

Introduction

Semantic knowledge includes the knowledge of word-to-concept mappings, as well as relations among the words and/or concepts. Much research shows the importance of different aspects of semantic knowledge in vocabulary acquisition (*e.g.*, Jones, Smith, & Landau, 1991; Colunga & Smith, 2005; Sheng & McGregor, 2010; Colunga & Sims, 2011). A long-standing question is how the overall structural properties of semantic knowledge impact how words are learned and processed in semantic memory (Collins & Quillian, 1969; Collins & Loftus, 1975; Steyvers & Tenenbaum, 2005).

Semantic knowledge is often represented as a *semantic network* in which the nodes correspond to words or concepts, and the edges specify semantic relationships among them (*e.g.*, Collins & Loftus, 1975; Steyvers & Tenenbaum, 2005). Steyvers and Tenenbaum (2005) argue that semantic networks created from adult-level knowledge of words exhibit a *small-world* and *scale-free* structure: an overall sparse network with highly-connected local sub-networks, where these sub-networks are connected through high-degree *hubs* (nodes with many neighbours). Through mathematical modeling, they argue that these properties arise from the developmental process of semantic network creation, in which word meanings are differentiated over time.

The work of Steyvers and Tenenbaum (2005) raises very interesting follow-on questions: To what degree does children’s developing semantic knowledge of words exhibit a small-world and scale-free structure? How do these properties arise from the process of vocabulary acquisition in children? The work of Beckage, Smith, and Hills (2011) is suggestive regarding these issues. They compare semantic networks formed from the productive vocabulary of *normally-developing* children and from that of *late talkers* — children who show a marked delay in their vocabulary acquisition (Ellis-Weismer & Evans, 2002). Beckage et al. (2011)

show that the network of vocabulary of late talkers exhibited a small-world structure to a lesser degree than that of the normally-developing children. However, while this work suggests some preliminary answers to the first question above, it cannot shed light on the relation between the process of word learning and the small-world and scale-free properties. Specifically, the networks considered by Beckage et al. (2011) only include productive vocabulary, not the many words a child will have partial knowledge of, and the connections among the words are determined by using co-occurrence statistics from a corpus, not the children’s own knowledge or use of the words. In order to shed light on how the small-world and scale-free properties arise from the developmental process of word learning, we need to consider the structure of semantic networks formed from the (partially) learned meanings of the words in the child’s environment.

In this work, we take advantage of a computational model to simulate normally-developing (ND) and late-talking (LT) learners, enabling us to examine the properties of semantic networks that include all the vocabulary a learner has been exposed to (*i.e.*, even those partially learned), and that has connections based on the actual learned knowledge of those words. The model is a probabilistic cross-situational learner which incrementally acquires word-to-meaning mappings through exposure to naturalistic input. When parameterized to reflect normal and deficit scenarios, this computational model has been shown to replicate several patterns of results observed in normally-developing and late-talking children (Nematzadeh, Fazly, & Stevenson, 2011, 2012), making its learned knowledge suitable as a simulation of the knowledge of such children. We create semantic networks based on the learned knowledge of this model in the ND and LT modeling scenarios, and investigate their structural differences with respect to having a small-world and scale-free structure.

We find that the semantic network of the ND learner — created from all the words in the input to the model, and the learned meanings of those words — exhibits a small-world and to some extent a scale-free structure, whereas the corresponding LT network does not. Moreover, we find that a “gold-standard” network — which uses ground-truth meanings rather than the learned meanings from the model — less clearly exhibits these two important properties. This suggests that properties of the learned knowledge may actually aid the learner in making appropriate connections among words. We conclude that considering the learned knowledge of vocabulary is important in understanding how the structure of children’s semantic networks is related to, and might arise from, word learning processes.

$apple: \{ \text{FOOD}:1, \text{SOLID}:.72, \dots, \text{PLANT-PART}:.22, \text{PHYSICAL-ENTITY}:.17, \text{WHOLE}:.06, \dots \}$

Figure 1: true meaning of the word *apple*.

The Simulated Semantic Networks

In this section, we first explain how our computational model learns and represents word meanings, and how our normally-developing (ND) and late-talking (LT) learners differ in their word learning mechanism. Next, we describe our approach to the construction of the semantic networks from the learned knowledge of the model in these two scenarios.

Simulating Different Learners

The model of Nematzadeh et al. (2011, 2012) learns from a sequence of input utterance–scene pairs, corresponding to what a child is exposed to in her natural learning environment. Each input pair consists of an utterance (what a child hears), represented as a set of words, and its corresponding scene (what a child perceives upon hearing that utterance), represented as a set of semantic features; e.g.:

Utterance: $\{she, drinks, milk\}$
Scene: $\{ANIMATE, PERSON, FEMALE, CONSUME, DRINK, SUBSTANCE, FOOD, DAIRY-PRODUCT, \dots\}$

The utterances are taken from the child-directed speech portion of the Manchester corpus (Theakston et al., 2001, from CHILDES MacWhinney, 2000). The corresponding scene representation for each utterance is generated using a *gold-standard* lexicon. In this lexicon the gold-standard meaning (*true meaning*) of each word is represented as a set of features, taken from a pool of F semantic “primitives” that comprise the F -dimensional space of semantic meanings. Each feature for a word is associated with a score that reflects the *strength* of association of the word and feature, and the *specificity* of each feature to that word (see Figure 1 for an example).¹ Given this gold-standard lexicon, a scene is probabilistically generated by sampling from the full set of features in the true meanings of the words in the utterance, according to the score of each feature.²

As a probabilistic cross-situational learner, the model uses observations of co-occurrences of words and features in the utterance and scene inputs to update its hypotheses of the meaning of each word over time. Specifically, for each word w in the input, the model maintains a probability distribution over all F possible semantic features; this probability distribution represents the model’s *learned meaning* of w at any given point in the input sequence. Initially, the probability distribution for w will start out uniformly distributed over

¹For details on the semantic features and data generation see Nematzadeh et al. (2012). We also note that a naturalistic word learning scene contains rich information beyond word meanings such as social cues and context. We cannot represent this information automatically, and given the scale of our data, it is not feasible to manually annotate such information. Our scene representation is thus limited in that it only includes the semantics of words in each utterance.

²Note that the gold-standard lexicon is only used for generating the input data, and is not used in the learning algorithm of the model.

all possible meaning features, but gradually the features that consistently occur in the presence of w in the input will rise in probability, while less relevant features will decline. Note that one consequence of the cross-situational approach is that features of *other words* that frequently occur in the context of w may also have some non-negligible probability mass in the meaning representation of w .

The model incorporates an attentional mechanism that gradually improves over time, enabling it to focus (more or less) on the relevant features to a word. We simulate normally-developing (ND) and late-talking (LT) learners by parameterizing the rate of development of this mechanism, such that ND has a faster rate. Because the attentional mechanism impacts the learning algorithm of the model, the ND and LT learners differ in the quality of their learned meanings; specifically, the LT meanings tend to have a more uniform distribution over semantic features at a given point in development.

Constructing a Learner’s Semantic Network

We train each learner (ND and LT) on an identical sequence of utterance–scene pairs, and then use their learned lexicons to build a semantic network for each. Unlike Beckage et al. (2011), we do not want to restrict the network to productive vocabulary, which eliminates much semantic knowledge of the learner (e.g., Benedict, 1979; Woodward & Markman, 1998). We thus assume all the words that the model has been exposed to during training are part of the learner’s semantic network. This reflects our assumption that an important aspect of a learner’s semantic knowledge is that it (perhaps imperfectly) captures connections among even words that cannot yet be fully comprehended or produced.

To establish the connections among nodes in the network, we examine the semantic similarity of the meanings of the corresponding words. Specifically, we measure semantic similarity of two words by turning their meanings into vectors, and calculating the cosine of the angle between the two vectors. We connect two nodes if the similarity of their corresponding words is higher than a pre-defined threshold.

This process yields two networks, Net-ND and Net-LT, each of which contains nodes for all the words in the input, with the edges determined by the semantic similarity of the word meanings represented within the ND and LT learners, respectively.

For comparison, we also build a gold-standard semantic network, Net-GS, that contains the same words as Net-ND and Net-LT (i.e., all the words in the input), but relies on the true meanings of words (from the gold-standard lexicon) to establish the connections. Note that the structure of this network does not depend on the learners’ knowledge of word meanings, but only on the similarity of the true meanings.

In order to further explore the importance of the knowledge of (partially) learned meanings to the structure of the resulting networks, we also consider a variation on Net-ND and Net-LT. Like Beckage et al. (2011), we can consider only a subset of the best-learned words of the learners, and see whether the

vocabulary itself – as opposed to what the learner has learned about that vocabulary – exhibits the *small-world* and *scale-free* properties. Recall that Beckage et al. (2011) create semantic networks connected on the basis of corpus-based co-occurrence statistics that are the same for both groups of children – i.e., it is the make-up of the vocabulary, rather than the learner’s knowledge of that vocabulary, that differs across the two types of networks. In our approach, this corresponds to using the true meanings from the gold-standard lexicon to connect the words in the network.

Hence, we form additional networks, Net-ND_{acq} and Net-LT_{acq} as follows. We take “productive” vocabulary in our model to be a subset of words which are learned better than a predefined threshold (by comparing the learned meaning to the true meaning in the gold-standard lexicon). We then build semantic networks that contain these *acquired* words of our ND and LT learners, connected by drawing on the similarity of the true meanings (that are the same for both learners). We can then use these networks to further explore the importance of the partially learned knowledge of words in our original networks in contributing to *small-world* and *scale-free* networks.

To summarize, we consider the following networks:

1. **Net-GS:** The nodes of the network are all the words in the input, and the edges are based on the similarity of the true meanings of the words.
2. **Net-ND** and **Net-LT:** The nodes are all the words in the input, and the edges are based on the similarity of the learned meanings of the words in each of the modeling scenarios.
3. **Net-ND_{acq}** and **Net-LT_{acq}:** The nodes are the acquired words (those best learned) in each scenario, and the edges are based on the similarity of the true meanings of those words.

Evaluating the Networks’ Structural Properties

A network that exhibits a small-world structure has certain connectivity properties – short paths and highly-connected neighborhoods – that are captured by various graph metrics (Watts & Strogatz, 1998). Below we explain these properties, and how they are measured for a graph with N nodes and E edges. Then we explain the requirement for a network to yield a scale-free structure.

Short paths between nodes. Most of the nodes of a small-world network are reachable from other nodes via relatively short paths. For a connected network (*i.e.*, all the node pairs are reachable from each other), this can be measured as the average distance between all node pairs (Watts & Strogatz, 1998). Since our networks are not connected, we instead measure this property using the median of the distances (d_{median}) between all node pairs (*e.g.*, Robins et al., 2005), which is well-defined even when some node pairs have a distance of ∞ .

Highly-connected neighborhoods. The neighborhood of a node n in a graph consists of n and all of the nodes that are connected to it. A neighborhood is maximally connected

if it forms a complete graph —*i.e.*, there is an edge between all node pairs. Thus, the maximum number of edges in the neighborhood of n is $k_n(k_n - 1)/2$, where k_n is the number of neighbors. A standard metric for measuring the connectedness of neighbors of a node n is called the *local clustering coefficient* (C) (Watts & Strogatz, 1998), which calculates the ratio of edges in the neighborhood of n (E_n) to the maximum number of edges possible for that neighborhood:

$$C = \frac{E_n}{k_n(k_n - 1)/2} \quad (1)$$

The *local clustering coefficient* C ranges between 0 and 1. To estimate the connectedness of all neighborhoods in a network, we take the average of C over all nodes, *i.e.*, C_{avg} .

Small-world structure. A graph exhibits a small-world structure if d_{median} is relatively small and C_{avg} is relatively high. To assess this for a graph g , these values are typically compared to those of a random graph with the same number of nodes and edges as g (Watts & Strogatz, 1998; Humphries & Gurney, 2008). The random graph is generated by randomly rearranging the edges of the network under consideration (Erdos & Rényi, 1960). Because any pair of nodes is equally likely to be connected as any other, the median of distances between nodes is expected to be low for a random graph. In a small-world network, this value d_{median} is expected to be as small as that of a random graph: even though the random graph has edges more uniformly distributed, the small-world network has many locally-connected components which are connected via *hubs*. On the other hand, C_{avg} is expected to be much higher in a small-world network compared to its corresponding random graph, because the edges of a random graph typically do not fall into clusters forming highly connected neighborhoods.

Given these two properties, the “small-worldness” of a graph g is measured as follows (Humphries & Gurney, 2008):

$$\sigma_g = \frac{\frac{C_{avg}(g)}{C_{avg}(random)}}{d_{median}(g)} \quad (2)$$

where *random* is the random graph corresponding to g . In a small-world network, it is expected that $C_{avg}(g) \gg C_{avg}(random)$ and $d_{median}(g) \geq d_{median}(random)$, and thus $\sigma_g > 1$.

Note that Steyvers and Tenenbaum (2005) made the empirical observation that small-world networks of semantic knowledge had a single connected component that contained the majority of nodes in the network. Thus, in addition to σ_g , we also measure the relative size of a network’s largest connected component having size N_{lcc} :

$$size_{lcc} = \frac{N_{lcc}}{N} \quad (3)$$

Scale-free structure. A scale-free network has a relatively small number of *high-degree* nodes that have a large number

of connections to other nodes, while most of its nodes have a small degree, as they are only connected to a few nodes. Thus, if a network has a scale-free structure, its degree distribution (*i.e.*, the probability distribution of degrees over the whole network) will follow a power-law distribution (which is said to be “scale-free”). We evaluate this property of a network by plotting its degree distribution in the logarithmic scale, which (if a power-law distribution) should appear as a straight line.

Evaluation

Set-up

We train our learners on 10,000 utterance–scene pairs taken from the input data of Nematzadeh et al. (2012). Recall that our ND and LT learners differ in the rate of attentional development that is a parameter of the model (c). Following Nematzadeh et al. (2011), we use $c = 1$ for ND and $c = 0.5$ for LT. We use only nouns in our semantic networks: since we draw on different sources for the semantic features of different parts of speech (POS), we cannot reliably measure the similarity of two words from different POS’s. To determine the subset of “acquired words” for Net-ND_{acq} and Net-LT_{acq}, we follow Fazly et al. (2010) and use a threshold of 0.7 for similarity between the learned and true meaning of a word. Finally, when building a network, we connect two word nodes with an edge if the similarity of their corresponding meanings is higher than 0.6.

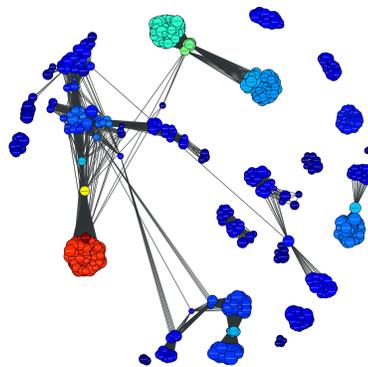
Results and Discussion

Table 1 contains the graph measures for all the semantic networks we consider here. The table displays the number of nodes (N) and edges (E) in each network, as well as the measures that capture characteristics of a small-world structure. We first discuss these measures, and the indicator of scale-free structure, for our primary networks, Net-GS, Net-ND, and Net-LT, and then consider the networks formed without using the learned knowledge of the words, Net-ND_{acq} and Net-LT_{acq}.

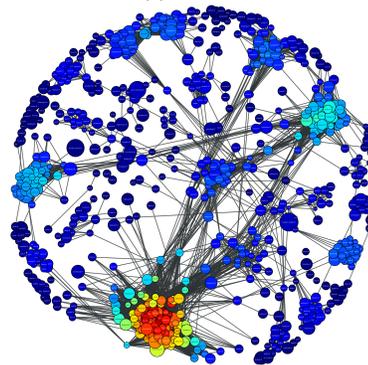
Small-world and scale-free structure in the learners’ networks. We first compare the structure of Net-GS and Net-ND (rows 1 and 2 in the table), and then turn to Net-LT (row 3).

According to the values of σ_g , we can see that both Net-GS and Net-ND yield a small-world structure, although the structure is more clearly observed in Net-ND: $\sigma_g(\text{ND}) = 5.5$ versus $\sigma_g(\text{GS}) = 3.1$. This is especially interesting since both networks have the same nodes (all the words), but Net-ND uses learned meanings to connect the nodes, whereas Net-GS uses the true meanings (from the gold-standard lexicon).

A closer look reveals that Net-ND has a structure in which many more nodes are connected to each other ($\text{size}_{lcc}(\text{ND}) = .90$ vs. $\text{size}_{lcc}(\text{GS}) = .72$) by using substantially fewer edges ($E(\text{ND}) = 12,704$ vs. $E(\text{GS}) = 26,663$). Net-ND achieves this by a better utilization of *hubs*: each hub node connects to many nodes, and in turn to other hubs, ensuring a high-degree of connectivity with a relatively small number of edges. Note



(a) Net-GS

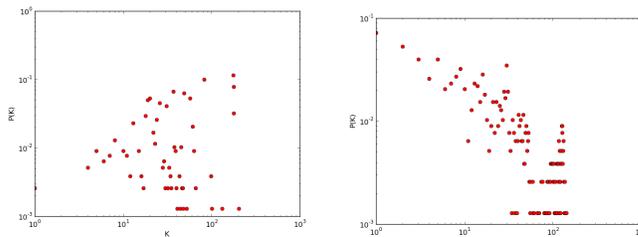


(b) Net-ND

Figure 2: (a) The gold-standard network, and (b) the network of ND with all words connected by learned meanings.

that these hubs are one of the main characteristics of a small-world structure. The different structures of Net-GS and Net-ND are evident from their visualizations in Figure 2. We can see that in Net-GS there are a number of isolated components that are not connected to the rest of the network.

We also examine Net-GS and Net-ND for having a scale-free structure by looking at their degree distributions in the logarithmic scale (see Figure 3). According to these plots, Net-ND to some degree exhibits a scale-free structure (with the plot roughly following a straight line), but Net-GS does not.



(a) Net-GS

(b) Net-ND

Figure 3: The degree distributions of Net-GS and Net-ND in the logarithmic scale.

Now, looking at the characteristics of Net-LT (row 3 of the table), we can see that it does not clearly show a small-world

	Networks	N	E	size_{lcc}	C_{avg}	d_{median}	σ_g
1	Net-GS (gold-standard)	776	26,633	0.72 (1)	0.95 (0.09)	7 (2)	3.1
2	Net-ND	776	12,704	0.90 (1)	0.70 (0.04)	6 (2)	5.5
3	Net-LT	776	239,736	1.00 (1)	0.97 (0.81)	1 (1)	1.2
4	Net-ND_{acq}	512	12,470	0.67 (1)	0.96 (0.10)	∞ (2)	0
5	Net-LT_{acq}	84	423	0.23 (1)	0.81 (0.11)	∞ (2)	0

Table 1: The calculated graph metrics on each of the semantic networks. The number in the brackets is the measure for the corresponding random network. The value of N and E are the same for each network and its random graph.

structure. The value of $\sigma_g(\text{LT})$ is very close to 1 because the value of C_{avg} for Net-LT is very similar to its corresponding random graph (cf. Eqn. 2). This is mostly due to the existence of a very large number edges in this network, which reflects the un informativeness of the learned meanings of LT for identifying meaningful similarities among words. Specifically, the meanings that the LT learns for semantically unrelated words are not sufficiently distinct, and hence almost all words are taken to be similar to one another. Net-LT consequently also does not show a scale-free structure, since the nodes across the network all have a similar number of connections (resulting in a bell-shaped rather than a power-law degree distribution).

What underlies the small-world and scale-free findings?

To summarize, we find that Net-ND shows a small-world and (to some degree) a scale-free structure, while Net-LT does not. This is consistent with the findings of Beckage et al. (2011) who observed that a network of vocabulary of normally-developing children had more of a small-world structure than a network of late-talkers’ vocabulary. However, by using the *simulated knowledge* of ND and LT learners, and comparing it to a representation of the “true meanings” in Net-GS, we can go beyond their work and address the question we raised in the introduction: How do these properties arise from the process of vocabulary acquisition in children?

The fact that Net-ND exhibits a small-world and scale-free structure more clearly than Net-GS suggests that the probabilistically-learned meanings of our model capture important information beyond the true meanings. Recall that our model learns the meaning of each word w by gradually associating w with semantic features that consistently co-occur with it across its usages. We noted above that this probabilistic cross-situational approach can lead to a “contextualization” of meaning representation for w : i.e., if another word w' consistently co-occurs with w (e.g., due to semantic relatedness), then the learned meaning of w can include semantic features of w' . This contextualized meaning representation essentially makes the learned meanings of the two co-occurring words more similar than their true meanings. This “blurring” of meanings entails that, even though Net-ND has fewer edges than Net-GS, those edges form connections across hubs that achieve a greater small-world structure.

On the other hand, the lack of a small-world structure in Net-LT clearly arises from the lack of differentiation of mean-

ings achieved by that learner. The relative deficit in attention in our LT learner entails that the learner cannot focus on the most relevant meaning features, yielding a network that fails to distinguish relevant clusters of meaning around “hubs”.

Clearly, this is data from a computational model, and not the actual semantic memory representation of children. However, it does lead to interesting predictions about the relationship between the small-world and scale-free properties and the process of vocabulary acquisition: specifically, that the contextualization of otherwise (at least moderately) distinguishable meanings is a crucial outcome of successful vocabulary acquisition, and one that leads to the formation of semantic networks with the overall structural properties found in representations of adult semantic knowledge.

A further look at the role of learned meanings. We suggest above that the small-world and scale-free properties of Net-ND arise due to qualitative differences in its learned knowledge of words, compared to both Net-LT or Net-GS. However, Beckage et al. (2011) found differences in the degree of small-world structure in their ND and LT networks that differed only in the vocabulary used as nodes in the network – that is, even though both networks used the same external knowledge to create edges among those nodes. Hence we also examine two additional networks, Net-ND_{acq} and Net-LT_{acq}, formed from the best-acquired words of the learners and the similarity of the true meanings of those words. This can help reveal whether it is the make-up of the vocabulary or the specific learned knowledge of words that plays a role in our results.

The graph measures for Net-ND_{acq} and Net-LT_{acq} are shown in rows 4 and 5 of Table 1. We see that neither of these networks exhibits a small-world structure ($\sigma_g = 0$), mainly because they have many isolated sub-networks, resulting in d_{median} having a value of ∞ (i.e., most node pairs are not connected to each other).

We conclude that in our simulations of child knowledge, it is the actual meaning representation that is important to yielding a small-world and scale-free structure, not simply the particular words that are learned. Our finding differs from that of Beckage et al. (2011), who found small-world structure even when using simple corpus statistics to similarly connect the vocabulary of each type of learner. It could be that our “best-learned” words do not correspond to the productive vocabulary of children; we also note that forming network connec-

tions based on similarity of our true meanings is much stricter than compared to the simple co-occurrence statistics used by Beckage et al. (2011).

More importantly, we think our simulated networks can turn attention around these issues to the actual (developing) knowledge that different learners are bringing to the task of word learning and semantic network creation. Specifically, Beckage et al. (2011) conclude that the semantic networks of late talkers might be less connected because they use a word-learning strategy that favors semantically-dissimilar words. It is not clear, however, how such children could follow a strategy of semantic *dissimilarity* when they do not have an adequate representation of semantic *similarity*. To the extent that the semantic knowledge of children is similar to the simulated knowledge in our model – in being partial, probabilistic, and contextualized – our experiments point to a different explanation of late talkers' disconnected vocabulary: Not that it is purposefully disconnected, but that due to the lack of meaningful semantic differentiation, it is accidentally so – i.e., late talkers have simply failed to exploit the contextualized meanings that help normally-developing children formulate helpful connections among words.

Summary and Future Direction

We use a computational model to simulate normally-developing (ND) and late-talking (LT) learners, and examine the structure of semantic networks of these learners. We compare the networks of ND and LT learners with that of a gold-standard (GS) network that has access to ground-truth meanings. Our goal is to investigate whether the simulated learned meanings of words reflected in ND and LT networks yield a small-world and scale-free structure, as observed in adult semantic networks (Steyvers & Tenenbaum, 2005).

Our results show that while Net-GS and Net-ND exhibit a small-world and (to some extent) a scale-free structure, the less differentiated meanings of Net-LT does not. We also observe that Net-ND shows a stronger small-world and scale-free structure compared to Net-GS. We attribute this interesting observation to the way our model learns word meanings: Unlike the true meanings, the learned meanings capture contextual semantic knowledge, which brings in an additional and helpful source of information for identifying semantic relatedness among words.

An interesting future direction is to model the actual development of a semantic network over the course of word learning. This would allow us to examine the underlying mechanisms that might be involved in the growth of a semantic network, and how the developing knowledge of word meanings interacts with the formation of network connections.

References

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS one*, 6(5), e19348.
Benedict, H. (1979). Early lexical development: Compre-

hension and production. *Journal of child language*, 6(2), 183–200.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407.
Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8(2), 240–247.
Colunga, E., & Sims, C. (2011). Early talkers and late talkers know nouns that license different word learning biases. In *Proc. of CogSci'11*.
Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, 112(2), 347–382.
Ellis-Weismer, S., & Evans, J. L. (2002). The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders*, 22(3), 15–29.
Erdos, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5, 17–61.
Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34(6), 1017–1063.
Humphries, M. D., & Gurney, K. (2008). Network small-world-ness: a quantitative method for determining canonical network equivalence. *PLoS One*, 3(4), e0002051.
Jones, S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, 62(3), 499–516.
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed., Vol. 2: The Database). Erlbaum.
Nematzadeh, A., Fazly, A., & Stevenson, S. (2011). A computational study of late talkers in word-meaning acquisition. In *Proc. of CogSci'11*.
Nematzadeh, A., Fazly, A., & Stevenson, S. (2012). Interaction of word learning and semantic category formation in late talking. In *Proc. of CogSci'12*.
Robins, G., Pattison, P., & Woolcock, J. (2005). Small and other worlds: Global network structures from local processes. *American Journal of Sociology*, 110(4), 894–936.
Sheng, L., & McGregor, K. K. (2010). Lexical-semantic organization in children with specific language impairment. *J. of Speech, Lang., & Hearing Research*, 53, 146–159.
Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, 29(1), 41–78.
Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *J. of Child Language*, 28, 127–152.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *nature*, 393(6684), 440–442.
Woodward, A. L., & Markman, E. M. (1998). Early word learning.