Do Humans Navigate via Random Walks? Modeling Navigation in a Semantic Word Game

Mohammad Isyroqi Fathan[1] **(m576f167@ku.edu), Eli J. Renfro**[1] **(eli.renfro@ku.edu)**
Joseph L. Austerweil[2] **(austerweil@wisc.edu) and Nicole M. Beckage**[1,2] **(beckage@ku.edu)**
[1]Department of Electrical Engineering and Computer Science
University of Kansas, 1520 W 15th St., Lawrence, KS 66045 USA
[2]Department of Psychology, University of Wisconsin-Madison,
1202 W Johnson Street Madison, WI 53706 USA

Abstract

We investigate a method for formulating context- and task-specific computational models of human performance in a constrained semantic memory task. In particular, we assume that memory retrieval can only use a simple process – a random walk – and examine whether the effect of context and task specifications can be captured via a straightforward network estimation method that is sensitive to context and task. We find that a random walk model on the context-specific networks mimics aggregate human performance.

Keywords: Network analysis; Semantic search; Spreading activation; Semantic memory; Random walks

Introduction

What concepts and facts are relevant in a given context? The type of knowledge that is likely to be important in a given situation is dependent on a large number of contextual factors. For example, consider an astronomy professor discussing the motion of planets in the solar system. The relevant facts depend heavily on contextual factors: Is she talking to a graduate student in her laboratory, adults from her community, or her four year old daughter? Is she trying to discuss why Mercury's orbit is more elliptical than would be expected assuming Newtonian physics, or instead trying to explain why the sun is so bright? Her knowledge of the cosmos is an extremely rich and interconnected set of concepts and facts. The concepts and facts relevant to her given situation are likely dependent on a host of contextual factors, ranging from communicative goals to the content of task at hand. How does memory facilitate the retrieval of the appropriate set of concepts and facts adapted to a given context?

Although people recall different facts given different tasks and contexts, most work in retrieval from semantic memory (people's memory for facts and concepts) attempts to mitigate any effects of context or task, rather than account for them. For example, Abbott, Austerweil, and Griffiths (2015) used human performance in a previous study (Nelson, McEvoy, & Schreiber, 2004) of the *free association* task (e.g., "say the first word that comes to mind when you hear 'dog"') to estimate a network representation for semantic memory. This model was then used to capture human performance in a different task: human retrieval of members of a specified category (e.g., "name all of the animals you can in 3 minutes"), which is a *semantic fluency* task. Previous work had used the same network representation as the above to capture human performance in a different constrained memory search task (Griffiths, Steyvers, & Firl, 2007): Say the first word that comes to mind starting with a specific letter. Different processes were assumed to capture human performance in the two different tasks, but they used the same semantic network.

Researchers using spatial representations for semantic memory model different tasks by changing the process used on a single representation. For example, Landauer, Foltz, and Laham (1998) estimated a spatial representation for semantic memory by applying latent semantic analysis to a representative sample of texts read by students. They used that representation with a simple retrieval process (return word with smallest cosine distance to the target) to capture human performance on the synonym portion of the TOEFL exam, a test of English proficiency. Using the same representation, but changing the retrieval process, other researchers captured human performance on the remote association task, which is a complex, constrained semantic search task (Smith, Huber, & Vul, 2013). Different representations are sometimes examined, but typically it is to argue either that other reasonable representations have comparable results (Smith et al., 2013), or that one representation captures how people encode knowledge in semantic memory better than another representation (Jones & Mewhort, 2007). Across all of these studies, the semantic representations do not depend on the context or task.

In this paper, we investigate a method for deriving context and task specific computational models of human performance in a constrained semantic memory task. To do so, we assume that memory retrieval can only use a simple process, a random walk. We describe a method for constructing context- and task-specific networks from human performance. We find that our network estimation technique with a random walk process is able to capture human success rates in a constrained semantic search task.

Prior work

One method for formalizing human retrieval from semantic memory is to model retrieval as a search problem over an associative semantic network (Collins & Loftus, 1975; Lehmann, 1992; Richens, 1956). In these semantic networks, words represent nodes, and edges represent the corresponding words are associated in some manner. Edges can be defined based on a variety of linguistic or cognitive features including synonyms, hierarchical relations, co-occurrence, free-association or shared features. In addition, the edges can further have weights to encode the association strength of the words. Network representations are increasingly used to study and model the interaction of cognition and language. See Baronchelli, Ferrer-i Cancho, Pastor-Satorras, Chater, and Christiansen (2013); Beckage and Colunga (2016) for reviews on the contribution of network science to the study of language and human cognition.

Recent analysis of semantic retrieval tasks, such as verbal fluency where individuals list words from a particular category, suggest that some types of search through semantic memory can be well explained by random walks on a network Abbott et al. (2015); Griffiths et al. (2007). They did so by deriving a network from a large data set of free association results (Nelson et al., 2004). Every cue and associate was a node in the network and an edge was created from node A to node B if the word corresponding to B was said as a response when a participant was given the word corresponding to A as a cue. Edges can be given weights by assigning them to be the number of times that cue-associate pair was produced by a participant. However, it is unclear whether a random walk process can capture how people solve other related tasks. For example, can a random walk on the same semantic network capture how people connect loosely related words through pairs of associated words? Here we test whether this type of directed search, with a specific goal in mind (a target word), can also be captured by a random walk process. To explore this question, we consider a semantic word game where individuals must make a sequence of decisions for navigation.

Previous work has illustrated that even rich human performance in some memory tasks can be modeled using a random walk process. For example, when recalling animals from memory, animals tend to be listed in clusters (e.g., four farm animals followed by two pets). Further, the retrieval of animals is consistent with optimal foraging theory (Hills, Jones, & Todd, 2012). People switch clusters when the time to retrieve a subsequent animal is larger than the overall average time between animal retrievals. To capture this behavior, previous work modeled the production of the list of animals as a random walk on a standard semantic network that emits animals whenever the walk visits a previously unvisited animal node. The animals in the list produced by this process are clustered and the random walk retrieval behavior is also consistent with optimal foraging theory (Abbott et al., 2015). In our work we use a random walk model over a semantic network to capture human performance in semantic navigation task where people connect a source word to a target word. The random walk model is not intended as a full model of human retrieval in this task, but serves as a baseline to compare different possible representations and inspire future work.

Previous work explored a semantic navigation problem where people were given a source and target word that were loosely related (Beckage, Steyvers, & Butts, 2012). Their task was for participants to get from the source to the target word by selecting a word to move to from a set of words strongly associated to the source (or current) word. Whichever word was chosen would become the new "source" word, and this repeated until participants reached the target word. They found that human performance cannot be explained via random guessing or choosing the strongest free associate to the current source word. This suggests that people utilize information present in the network to get closer to the target word. One aspect that was not considered in their analysis is the role of task and context on participant performance — for example, often the current word would be one that the participant observed in a previous round.

We examine whether creating a context- and task-specific network based on a participant's previous choices and experience can be used to explain their performance in future trials. In other words, *can we change the network representation in a systematic way such that a random walk process performs better than in a standard network representation?* To examine this question, we analyze how people navigate from a start word to a specific target word via forced choice between pairs of very associated words. We model this as navigation on a directed network by assuming that edges exist between words that are part of the choice set for a specific word. Our hope is that adapting a network representation to the context and task will capture human performance on the task. If so, this provides a novel avenue for investigating how the mind searches semantic memory in different contexts and for different tasks: Assume a single process, but adapt the representation in a manner sensible to the current context and task.

Behavioral Experiment: The Mindpaths Game

Mindpaths is a semantic word game based on the experimental setup of Beckage, Steyvers and Butts 2012. We derived a network from the USF free association norms (Nelson et al., 2004) using the method described above, but trimming the network to include only words with a minimum in-degree of 3 and a maximum out-degree of 12. In the online version of the game, available at `socientize.eu/pybossa/app/Semantics/`, semi-random start and target words are presented. Individuals are then asked to navigate from the start word to the target word via forced choice. The options presented are words that were directly connected to the current word in the network. An example of the game is shown in Fig 1.

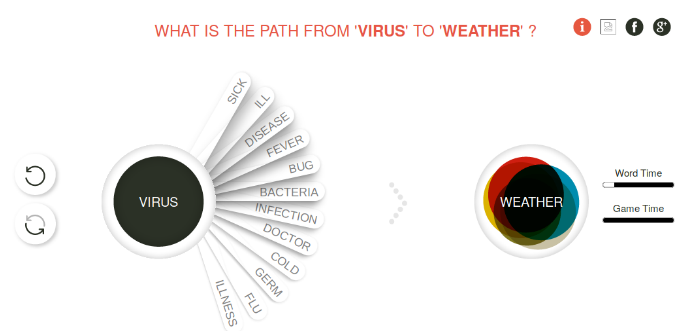

Figure 1: A screenshot of the Mindpaths game.

Fig 1 provides an example where the start word is *Virus* and the target word is *Weather*. Players then choose from the given options, including sick, ill, and disease. After a choice is made, the chosen word replaces the start word and the chosen word's associates are listed. The current dataset analyzed in this paper is publicly available via the game website. To date, the dataset contains 141 different start and target word pairs. We call each unique start and target pair a *game*. Each play of a game we call a *trial*.

In total 141 games were played 11,698 times, with an average of 83 trials per game, with a range of between 2 and 469 trials or plays for each game. The underlying network contained 1,972 different words, and the number of words visited across all tasks is 1,852. The game automatically terminated if the player had not found the target word after 25 choices (these games were not recorded). Players could click the reset button to bring them back to the start word (the number of steps they have taken so far does not reset). Finally, there was a time limit that refreshes on each move and the response time is also recorded. If a player takes over a minute to respond, the trial is not recorded and the page refreshes with a new game.

The players can be identified by their accounts or, if they do not have accounts, by their IP address. To date, there are 365 registered users with 9,237 trials, and 2,461 trials from unregistered users. Based on the data collection mechanism, only successful games are recorded. So, we do not know how often individuals did not solve different games.

Computational Methods

We test whether random walks can account for human performance. Previous work used random walks to capture human performance when listing all the animals they know (Abbott et al., 2015). It is unclear if this will explain participant behavior in this task, as it is a directed search in which individuals have a specific target they must reach. Via comparison to search by random walks, we can quantify the extent to which individuals are making non-random choices (i.e., using additional information or strategies to help them succeed in this game). To summarize random walk performance, we explore different networks derived from the underlying game network and player performance with the aim of capturing human performance.

Nelson Network The free association network, collected by Nelson and colleagues 2004, provides the underlying network used in the game. In this network, words are nodes in the graph. Edges are directed and weighted. The weights encode the frequency at which a given response was said to a specific cue in the free association study. For example, if the cue word *dog* is said, *cat* is the free association response with probability 0.78. Note that MindPaths has trimmed the network based on in-degree and out-degree to make the game more playable. In our simulations, we consider random walkers that are walking on both a weighted and unweighted version of the trimmed *Nelson network* representation. In the weighted version, we normalize the original free association network such that the total probability transitioning out of any node sums to 1.

Traffic Network We define here a novel method for adapting networks to a specific task and context. We call the resulting network a *traffic network*. Traffic networks have the same nodes as the free association network above, but we define the edges based on people's choices in the game. So, if any player chose a move from *cold* to *weather*, there will be an edge from cold to weather in the traffic network. We consider both weighted and unweighted versions of this traffic network. In the case of a weighted traffic network, the frequency at which a game is played will influence how likely it is that a particular cue word is seen and thus the probability of transitioning between a particular cue and response. To estimate a weighted traffic network, we normalize by the number of times a game was played.

Do Traffic networks differ from the Nelson network? One way to investigate this question is to examine whether the degree of words in each network are comparable. The upper triangle of Fig 2 presents the correlations of the word degrees of Traffic and Nelson networks (via a nonparametric measure, Kendall's tau (τ)) in the upper triangle. The strongest τ (0.72) is between the weighted and unweighted traffic networks. While the minimal correlation, $\tau = 0.4$, is not unsubstantial, it is quite lower, suggesting the same words have substantially different degrees in the different networks. The lower triangle of Fig 2 displays the Person correlation of the network edge weights. The unweighted Traffic and Nelson networks have a high correlation. This suggests that the Traffic network captures much of the connectivity pattern of the Nelson network. However, we find that the correlation between the weighted Traffic network and the weighted Nelson network is small (0.16) suggesting that free association responses are different than choices in the Mindpath game. This provides strong support that context- and task- specific representations are important to completing the game.

Random Walkers

We simulated random walks on four network representations (weighted/unweighted $\times$ Nelson/Traffic) to see which network representations can accurately capture human performance on the Mindpath game (specifically, their rate of success). The difference in performance of the random walks over the networks enables us to evaluate the role of context on human navigation behavior because each network representation includes a different amount of context and task information. Our random walkers are likely not a perfect model of participants' semantic search. They provide an interpretable baseline for evaluating the extent to which human performance can be explained as restricted random guessing over different network representations.

For robustness and to quantify differences between games, we use two approaches to analyze the performance of our random walkers. In all cases we perform k-fold cross validation.

Table 1: Percent success of solving various word games in the minimum number (geodesic) steps. Standard deviation across folds in terms of percentage is also reported.

	overall	geodesic 1	geo. 2	geo. 3	geo. 4	geo. 5
all games	11810	2.93	29.30	45.17	19.59	3.01
human perf	32.6 (0.8)	87.0 (2.8)	25.4 (1.4)	36.5 (1.1)	28.6 (0.6)	19.5 (3.5)
unw. Nelson	10.7 (0.5)	42.0 (5.5)	13.8 (1.4)	10.0 (0.4)	4.24 (0.8)	2.21 (2.1)
w. Nelson	7.87 (0.7)	33.3 (2.8)	9.46 (1.6)	7.09 (1.4)	4.45 (1.3)	2.11 (1.8)
unw. Traffic	10.1 (0.3)	35.0 (5.4)	15.8 (2.8)	12.2 (0.9)	8.08 (2.1)	6.16 (2.6)
w. Traffic	24.1 (0.9)	65.3 (5.8)	25.9 (2.7)	33.1 (1.3)	25.8 (1.7)	15.3 (5.6)

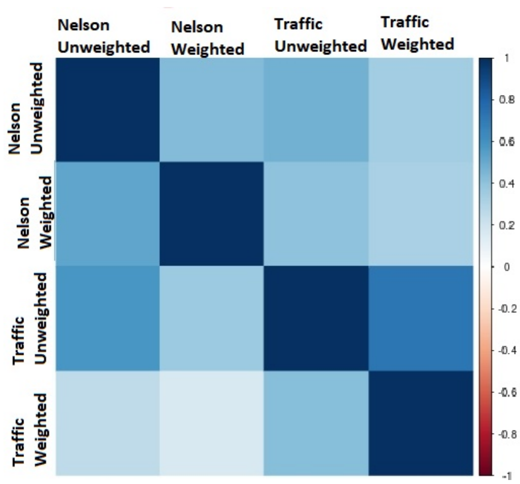

Figure 2: Correlation analysis of the various network representations. The upper diagonal captures Kendall's tau rank correlation of node *degrees* across networks. The lower diagonal captures the Pearson correlation between *edge weights* across the network representations.

In the first approach, we define a fold as 20% of the total trial data. Thus the model must extend to unseen trials but not necessarily unseen games. We call this the *trial variant*. In the second approach, we assign 20% of the total unique games to each fold (e.g. 20% of 141 games). We call this the *game variant*. We predict that these two variants will have minimal effect in the performance and evaluation of the Nelson network. However, there is potentially a large performance difference in the case of the Traffic networks if these networks capture how an individual's search is influenced by the given target word. If the target word plays a role in the choices individuals make, the game variant using the traffic network should capture human performance with significantly less accuracy as compared to the trial variant.

For cross-validation, we build unweighted and weighted Traffic networks based on the training set (e.g. all folds except the testing fold). The maximum path length for our random walk simulation is 25 (as it was for human participants). We run simulations until we have as many successes as we see in our human game play data. In the game variant traffic network, we found 3 games with no path between start and end words. We take this as evidence that there are some game specific contextual information that individuals are using to navigate.

Results & Discussions

Our goal is to build a random walk model that captures human performance, defined as the distribution of choices made by participants for each game. Optimal performance solves the task using the fewest number of transitions in the game network. For example, virus to weather can be solved optimally, or geodesically, by transitioning from virus to cold to weather. An example of a suboptimal path is moving from virus to ill to cold. Because the players are navigating a network, we can compute the optimal, or geodesic, path distance. Note that this geodesic distance can only be computed if the whole underlying graph is known, an unlikely situation for our human players. Because participants probably cannot directly access the full graph structure, performance that is close to optimal suggests participants are using information other than their original semantic network (e.g., adapting the network based on previously played games).

To analyze human performance, we consider games that can be optimally completed in between 3 and 8 steps. Out of 11,698 trials, 35% of trials were solved with the minimum number of steps, and 71% of the trials were solved within two steps of optimal. This shows that human are quite good at navigating the semantic space, despite not having global knowledge of the semantic space. Table 1 summarizes estimated geodesic performance based on 5-fold cross validation.

Table 1 shows that humans outperform random walks at finding the geodesic path, regardless of the underlying network representation. Note that the data set is subject to selection bias: We cannot tell how many individuals started the game and did not finish. We attempted to match this bias by conditioning on a successful path being found within 25 steps for both human and random walks. Though humans are still more accurate, when we exclude games where the target word is in the first option set (i.e., games whose optimal solution distance is one), we see comparable performance between the weighted Traffic network and human performance. This suggests that the weighted Traffic network may have captured aspects of the task and context in a manner such that a random walk process can succeed on the task. In future analyses, we plan to examine whether the actual solutions found by the random walks are comparable to those made by participants (and not just being of comparable length). Regardless, this result suggests that the context- and task-specific semantic representation use by people may be similar to our weighted Traffic networks. The failure of the Nelson network suggests that semantic networks need to be adapted to task and context, or a different search process needs to be used.

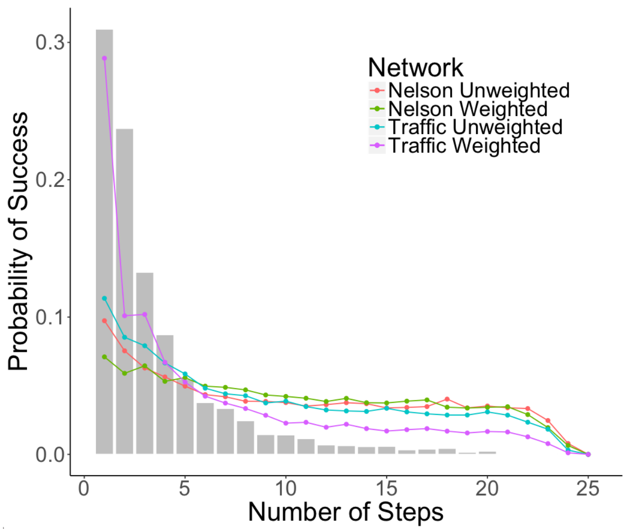

Figure 3: Performance of human players (grey bars) and random walk models in the *trial variant*, conditioned on successful paths and excluding games with geodesic one. The x-axis is the number of steps taken beyond the geodesic distance.

We now consider whether random walks on our various network representations can reproduce the rate of successful games completed in a specific number of steps. This includes not only geodesic path lengths but the exact number of steps it took individuals to find the target word. Because our results above suggest that games with a geodesic of one (e.g. the target is in the first option set) are unique from the other games, we consider only games with geodesic distances greater than one. Fig 3 depicts human performance (bar chart) as compared to the Nelson networks and Traffic networks (colored lines). Along the x-axis are the number of steps taken beyond the geodesic distance.

Fig 3 shows that the distribution of successful path lengths on the weighted Traffic network is similar to the frequency of successful path lengths made by human players. This suggests that the frequency by which a previous player traversed a particular edge contains enough information so that the success rates of search by random transitions is qualitatively similar to human success rates.It is important to note that this is not simply overfitting – the evaluation is on different trials in a test set. The ability of the weighted Traffic network to model participant behavior is significant especially given the fact that the traffic network aggregates across all games, thus the resulting representation is specific to the context of MindPaths, but not a specific game (target and source word). This suggests that individual differences may be minimal in this task and an aggregate contextual network contains much of the needed information for a random walk process to complete this task successfully.

In the *game variant* simulations, we examine how the performance of past players affects the models' success rates on unseen games. We first find that some of the games were not solvable using the Traffic networks. When selecting random games, we find that the resulting traffic network is a disconnected graph with some start and target words in different components making it impossible for the Traffic random walker to find a path between nodes. On average (from the 5 folds), 3 of 28 unseen games in each test set were unsolvable.

Fig 4 compares participant success rate to our random walks across unseen games, conditioning on those games that are solvable by a random walk over a Traffic network. Now the weighted traffic network has successful paths that are similar in frequency to the random walks on the various network representations. Human players are much more successful at solving the task efficiently than any of our random walker models. Thus, further work is needed to capture human performance in this case. Random walks over the Nelson and Traffic semantic networks are insufficient.

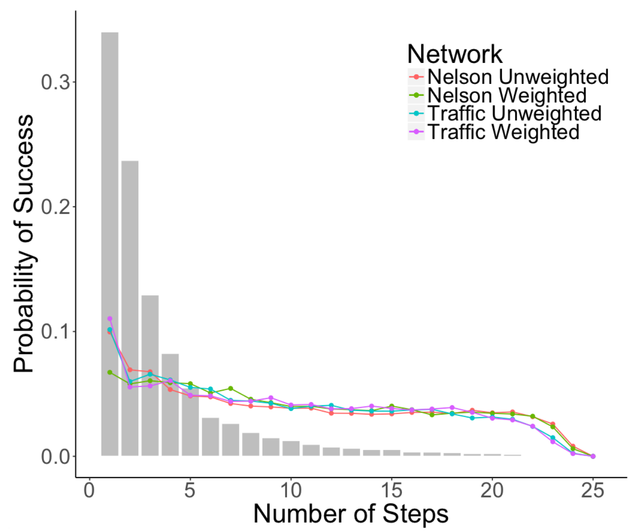

Figure 4: Performance of human players compared to *game variant* random walk models, conditioned on successful paths and excluding games with geodesic one.

We next consider if the structure of a word in the network influences the probability of success for our random walkers. We correlate the betweenness centrality of a word with the success probabilities. We examined betweenness centrality in particular because it captures the amount of information flowing through a node. If this type of information flow is useful for succeeding in this task, we expect high correlations of betweenness centrality and success probability by our ran-

Table 2: Correlation between success probability and betweenness centrality; geodesic probability and betweenness centrality for each network and each simulation variant.

	Success x Start		Success x End		Geodesic x Start		Geodesic x End	
	Trial	Game	Trial	Game	Trial	Game	Trial	Game
unw. Nelson	0.077	0.112	0.818	0.799	-0.018	0.131	-0.025	0.128
w. Nelson	-0.012	0.0982	0.604	0.693	0.027	0.04	0.026	0.211
unw. Traffic	0.002	0.0996	0.784	0.798	0.033	0.215	0.015	0.314
w. Traffic	0.215	0.0627	0.370	0.608	0.262	0.125	0.185	0.217

dom walkers. We correlate betweenness centrality of start and end words with a) the success probability and b) the probability of navigating successfully in geodesic length. Table 2 shows these correlations for *trial variant* and *game variant*. The first and second columns in each block shows the correlation between the success probability and the betweenness centrality of the start word and the end word, respectively. Table 2 suggests that there is small to no correlation between a start words' betweenness centrality and success probability in both the *trial variant* and *game variant* cases. This suggests that betweenness centrality of the start word may not be all that important for solving the task. However, the correlations with end words are particularly high on these networks, except for the weighted Traffic network. This suggests that the more *central* the target words are, the greater the chance for the random walk to succeed on these networks.

The pattern of target word correlations with the weighted Traffic network is interestingly different. The correlation is low for the *trial variant* test, but high for the *game variant* test. This suggests that the *centrality* of the target word is not as important as the random walk gets closer in qualitative fit to human players. This may be because in *trial variant* tests, the weighted Traffic network captures contextual information (arising from the edge weights and the network structure) in MindPaths game that the other random walkers cannot. Thus, the weighted Traffic network random walker may not need to rely on the centrality of the target words to navigate successfully in *trial variant* tests. But, it still relies on centrality for *game variant* tests.

Conclusion and future work

We described a technique for constructing a context-specific semantic network, which we call a weighted Traffic network. We showed that random walkers on a weighted Traffic network are able to capture much of the relevant contextual information in the *trial variant* with performance that qualitatively mimics human success rates. We also found that a set of games produces a unique context that is specific to each game and does not necessarily allow the model to succeed at unseen games with performance matching human players.

Our results provide promise for examining constrained semantic search within a semantic network framework. We plan to adapt random walks to include structural information of the network (e.g., betweenness centrality). Further, the data set is extremely rich, containing both human choices and their decision times. While we evaluated the ability of the random walks to successfully navigate to the target word conditioned on the number of steps, we did not attempt to compare the paths taken by the walkers and by people. Through these analyses, we hope to understand how the current context affects human semantic search across different tasks.

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psych Rev*, *122*(3), 558–569.

Baronchelli, A., Ferrer-i Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in cognitive sciences*, *17*(7).

Beckage, N. M., & Colunga, E. (2016). Language networks as models of cognition. In *Towards a theoretical framework for analyzing complex linguistic networks.*

Beckage, N. M., Steyvers, M., & Butts, C. T. (2012). Route choice in individuals—semantic network navigation. *Proc 34th Cognitive Science Society*, 108 - 113.

Collins, A., & Loftus, E. (1975, 11). A spreading activation theory of semantic processing. , *82*, 407-428.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting with pageRank. *Psych Sci*, *18*(12), 1069–1076.

Hills, T. T., Jones, M. N., & Todd, P. (2012). Optimal foraging in semantic memory. *Psych Rev*, *119*, 431–40.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psych Rev*, *114*(1), 1–37.

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

Lehmann, F. (1992). Semantic networks. *Computers & Mathematics with Applications*, *23*(2), 1 - 50.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *BRM*, *36*(3), 402–407.

Richens, R. H. (1956). Preprogramming for mechanical translation. *Mechanical Translation*, *3*, 20-25.

Smith, K. A., Huber, D. E., & Vul, E. (2013). Multiply-constrained semantic search in the remote associates test. *Cognition*, *128*, 64–75.