Sensitivity to Temporal Community Structure in the Language Domain

Kendra V. Lange (kxl786@psu.edu)
Department of Psychology, The Pennsylvania State University,
Moore Building, University Park, PA 16802 USA

Carol A. Miller (cam47@psu.edu)
Department of Communication Sciences and Disorders, The Pennsylvania State University,
Ford Building, University Park, PA 16802 USA

Daniel J. Weiss (djw21@psu.edu)
Department of Psychology, The Pennsylvania State University,
Moore Building, University Park, PA 16802 USA

Elisabeth A. Karuza (exk521@psu.edu)
Department of Psychology, The Pennsylvania State University,
Moore Building, University Park, PA 16802 USA

Abstract

The interrelatedness of lexical items, typically defined in terms of semantic or phonological overlap, has been shown to influence language learning. Given that language also contains sequential structure, we investigate here whether temporal overlap among words, formalized in graph theoretical terms as displaying the property of *community structure*, might also have consequences for learning. We create a graph organized into clusters of densely interconnected nodes with relatively sparse external connections. After assigning a novel pseudoword to each node in the graph, we generate a continuous sequence of visually-presented items by walking along its edges. Word-by-word reading times suggest that learners are indeed sensitive to temporal overlap. Compellingly, we also demonstrate that prior exposure to sequences organized into temporal communities influences performance on a subsequent word recognition task.

Keywords: network science; statistical learning; language acquisition

Introduction

A foundational question in cognitive science asks how the human brain converts a vast amount of sensory input into usable knowledge. Fortunately for our brains, sensory input, though noisy, tends to be richly patterned. A means of characterizing broad-scale patterns, network science enables the mathematical description of systems as varied as social relationships (Scott, 2017) and neural connectivity (Bassett & Sporns, 2017). Of particular relevance to the present series of experiments, applications of network science to the domain of natural language have dramatically increased our understanding of the organization of phonological (Vitevitch, 2008; Arbesman, Strogatz, & Vitevitch, 2018), syntactic (Ferrer i Cancho, Solé, & Köhler, 2004; Liu, 2008), and semantic systems (Collins & Loftus, 1975; Borge-Holthoefer & Arenas, 2010).

A growing body of evidence suggests that humans use network-level properties when acquiring and accessing linguistic knowledge (for a review, see Karuza, Thompson-Schill, & Bassett, 2016). For example, an index of the extent to which phonological neighbors of a word are themselves neighbors, clustering coefficient has been shown to predict acquisition of novel object labels designed to vary with respect to this property (Goldstein & Vitevitch, 2014). Learners also show sensitivity to lexical islands, or small groups of phonologically related words isolated from a network's "giant component," or the largest group of interrelated words. Siew & Vitevitch (2016) observed that words drawn from lexical islands are recognized and recalled more easily than those from a giant component. For semantic networks, in which nodes representing concepts are linked according to some similarity metric, evidence suggests that densely connected words are most likely to be acquired early in development (Steyvers & Tenenbaum, 2005). In sum, the structural properties of complex language networks may carry important implications for learning.

Outside the language domain, a number of studies have also begun to probe human sensitivity to network topology, generally focusing on community structure in temporally-defined graphs. In these studies, nodes correspond to fractals, glyphs, or button press combinations, and edges mark the transition between two images in a continuous sequence (Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013; Karuza, Kahn, Thompson-Schill, & Bassett, 2017; Tompson, Kahn, Falk, Vettel, & Bassett, 2018; Kahn, Karuza, Vettel, & Bassett, 2018). Response times are typically recorded as participants view an uninterrupted stimulus stream created by "walking" along the edges of a graph comprised of sparsely connected clusters of densely interconnected nodes (i.e., that display the property of community structure; Figure 1). Results point to a signature response pattern associated with the transition between communities: a pronounced increase in learners' processing times when measured against within-community transitions (Karuza et al., 2017).

Expanding on prior work, which has focused exclusively on non-linguistic visual stimuli, we investigate here whether learners display a comparable sensitivity to community structure when it dictates the order of visually-presented pseudoword sequences. One defining characteristic of linguistic signal is that it unfolds in time. In light of this, we examine whether the temporal overlap between words, not only their phonological and semantic interrelatedness, might steer the learning process. In adapting this paradigm to the language domain, our work makes two additional contributions: first we expand on the size of tested network structures, creating graphs of 40 nodes instead of the 10-15 used in related prior work (Schapiro et al., 2013; Karuza et al., 2017; Tompson et al., 2018; Kahn et al., 2018). Second, we refine an offline measure that allows us to investigate the influence of community structure not only in moment-to-moment processing of novel stimulus streams, but also in accessing previously acquired knowledge in future contexts.

Study 1: Community Structure and Substring Familiarity

We first examine whether learners exhibit cross-community reaction time (RT) increases as they process continuous sequences of unfamiliar linguistic stimuli. We also ask whether sensitivity to community structure will manifest in the expression of knowledge in offline familiarity judgements involving short sequences (substrings) extracted from the original exposure stream. Analyses test the hypothesis that learners prefer substrings drawn from within communities relative to those that span communities.

Materials and Methods

Participants 33 neurologically normal participants (5 male, 28 female; 18-21 years old) participated in this study. They were recruited from the undergraduate psychology research pool at Pennsylvania State University and were granted course credit for their participation. All participants provided informed consent. Three participants were excluded for performance below a pre-determined threshold on an orthogonal cover task (<70% correct; Karuza et al., 2017).

Stimuli

Network properties. Exposure streams were generated via a random walk on a graph featuring five communities of eight nodes each (Figure 1). Each community was connected to two other communities through boundary nodes sharing a single edge with an adjacent community. With the exception of boundary nodes within the same community, which were unlinked, each other node was connected to every other node in their community. Thus, all nodes had equivalent degree, or number of incident edges. Because edges were undirected and unweighted, (1) they could be traversed in any direction and (2) transitions between any two nodes were equally probable. Nodes within the graph corresponded to a unique, pronounceable pseudoword, and edges represented the direct succession of two pseudowords within the stimulus stream.

Pseudoword properties. Pseudowords were selected from the ARC non-word database (Rastle, Harrington, & Coltheart, 2002). Forty orthotactically plausible, single-syllable words were chosen, 20 four-letter words and 20 five-letter words. All words had 5-30 orthographic neighbors and 5-30 phonological neighbors. While metrics such as Coltheart's N

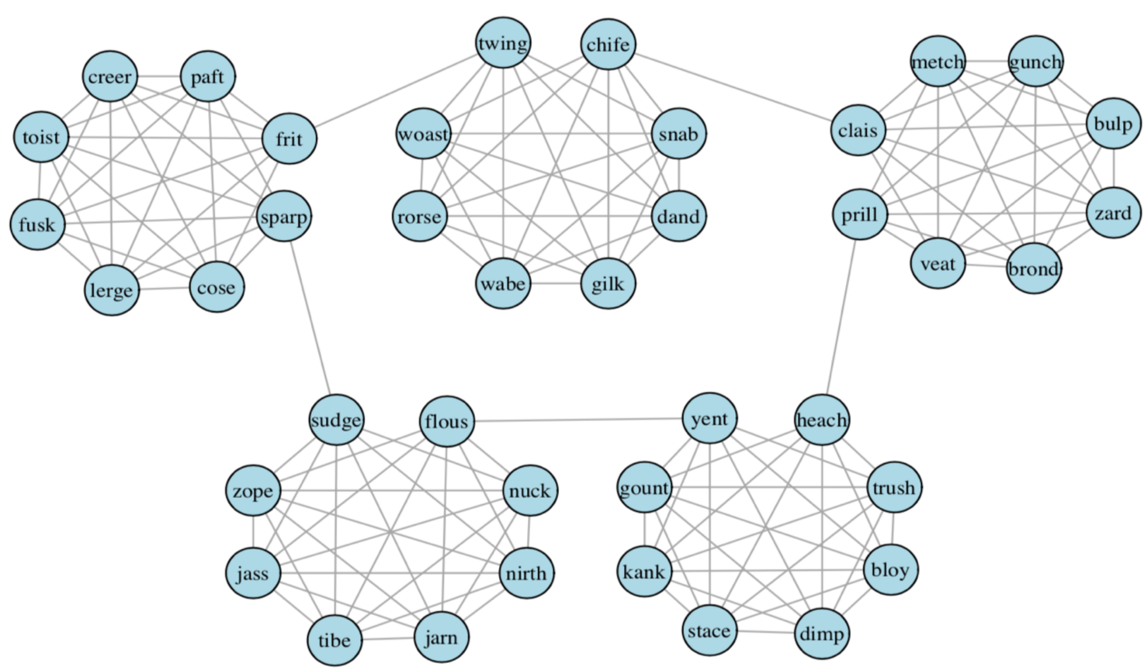

Figure 1. The network architecture used to generate stimulus streams in Studies 1 and 2. Each node represents a pseudoword, and edges represent the co-occurrence of two pseudowords in a continuous sequence

have been shown to have certain limitations (Yarkoni, Balota, & Yap, 2008), and indeed some pseudowords shared surface-level similarities, we stress that any *systematic* phonological or orthographic overlap was minimized by our word-to-node randomization procedure. For the purposes of the cover task (described below), we created a scrambled, unpronounceable version of each pseudoword (e.g., *clais* and *gilk* became *aislc* and *igkl*).

Test items. Eighteen short test sequences (length 5-7 pseudowords) were spliced out of the continuous exposure stream. Half of these substrings consisted exclusively of nodes from within one community, while the other half included traversal of a community boundary. Matched pairs of intra- and inter- community substrings were created by equating length (number of items in the string), number of node repetitions (if any), chunk strength (within one standard deviation of the mean; Meulemans & Van Der Linden, 1997), and general position in the exposure stream (first third, second third, etc.).

Procedure The experiment was composed of four phases: familiarization, exposure, test and debriefing. Participants were randomly assigned to one of four conditions consisting of a unique random walk (i.e., ordering of nodes) in the exposure phase and a unique series of test items. Independent of condition, node-to-pseudoword correspondence was randomized (i.e., "node 1" might correspond to *clais* in one participant and *gilk* in another).

Familiarization phase. Participants were told that they would see a list of made up and scrambled words presented in alphabetical order. They then viewed the list of pseudowords and the scrambled words in a series of 1.5-second trials. They were instructed to press [1] if the word on the screen followed the rules of English (these were the pseudowords) and [2] if the word did not follow the rules of English (these were the scrambled versions). To facilitate their understanding of the task, participants first saw two examples: the pseudoword was *corb*, and the scrambled word was *brco*.

Exposure phase. Following the familiarization phase, participants viewed a 1000-trial continuous sequence of individually presented pseudowords. To obtain RT measures across the entirety of the exposure phase, we instructed participants to complete an orthogonal cover task. At each trial, they were asked to press [1] if the pseudoword appeared in its "regular form" and [2] if the pseudoword appeared scrambled (12% of trials). Each pseudoword was presented for 1.5s with no interstimulus interval. Total duration of the exposure phase was 25 minutes.

Test phase. At the conclusion of the exposure phase, participants were presented with 18 pairs of substrings presented simultaneously on the screen, one above the other (position was randomly determined). They selected which of the two short sequences looked more familiar to them based on what they saw during the previous phase of the experiment. We adopted a familiarity-based approach to judging pairs to promote relatively implicit access of knowledge during the test phase. Unlike the exposure phase, the test phase was self-paced (i.e., both sequences stayed on screen until participants made their selection), with an interstimulus interval of 1.5 seconds.

Analysis and Results

Scrambled Word Detection Participants generally succeeded in distinguishing between pseudowords and their scrambled versions (95.8% accurate, SD = 2.4, excluding the three participants who scored below threshold).

Data Exclusions In the exposure phase, data were prepared for analysis by first eliminating any implausible RTs (i.e., less than 100 ms), then by removing RTs that were greater than three standard deviations away from the mean (4.5% of total data). We also removed all scrambled word trials (12% of total data) and any incorrect responses (4.2% of total data). As we were particularly interested in the RT cost associated with crossing between communities, data were then subset to include only nodes corresponding to entry into a new community (transition nodes), as well as boundary nodes immediately prior to that transition (pre-transition nodes).

Exposure Phase In a linear mixed effects model (library *lme4* 1.1-19 in R 3.5.1), RTs were regressed onto the main effects and interaction of Node Type (pre- vs. transition) and Trial (1-1000). All transitions were included in analysis. The model included the fullest random effects structure that allowed the model to converge: a random intercept for participant and a by-participant random slope for Node Type, Trial, and their interaction.

We observed a significant main effect of Node Type (β = 10.080, t = 2.310, p = 0.022), indicating a processing cost for transition nodes compared to pre-transition nodes. The main effect of Trial (β = -22.798, t = -3.191, p = 0.003) was also significant, an expected finding given that participants were likely to become faster overall at executing button presses. No interaction between Node Type and Trial was observed (β = -6.911, t = -1.477, p = 0.150).

Test Phase Accuracy scores from the posttest did not differ significantly from chance (t(29) = 1.161, p = 0.255). When a post-hoc analysis (mixed logit model) was run to determine whether accuracy was affected by the length of sequences (5 vs 6 word sequences: β = 0.103, z = 0.959, p = 0.341; 5,6 vs 7 word sequences: β = 0.054, z = 0.860, p = 0.390), position on the screen (top or bottom) (β = -0.107, z = -1.208, p = 0.227), or trial number (β = 0.027, z = 0.305, p = 0.761), we continued to observe no significant effects.

Study 2: Community Structure and Word-Level Recognition

Study 1 offers evidence of a cross-community RT increase as learners viewed sequences of written pseudowords. Online measures, collected during the exposure phase, serve to demonstrate learners' expectation that words within a community should co-occur in time. When that expectation was violated by entry into a new community, RTs reflected a processing penalty. Despite these promising results, we found no evidence that participants applied this knowledge offline as they made substring-level familiarity judgements. Successful language acquisition requires not only the accumulation of statistical regularities, but also accessing that accumulated knowledge in varied contexts. Therefore, the focus of Study 2 was on a post-exposure measure that would speak to the role of community structure in the latter process.

Materials and Methods

Participants 37 neurologically normal participants (9 male, 28 female; ages 18-21) participated in this study. They were recruited from the undergraduate psychology research pool at Pennsylvania State University and were granted course credit for their participation. All participants provided informed consent. Four participants were excluded for cover task performance below the pre-determined threshold used in Study 1.

Stimuli Pseudowords and the graph used to generate the exposure streams were identical to those used in Study 1. However, we increased the length of random walk by 40% in order to ensure participants were receiving sufficient exposure before completing a post-test. For the test phase, we developed a new approach to evaluating the influence of network architecture on retrieval of knowledge following initial learning.

Test items. Our method represents an extension of a classic paradigm developed by Meyer & Schvaneveldt (1971). In that pioneering study, participants completed a lexical decision task on various pairs of words and pseudowords. Compellingly, RTs for pairs of semantically related words were significantly faster than RTs for pairs of semantically unrelated words. Instead of asking whether *semantic* similarity influences retrieval processes, we ask instead whether community structure, or *temporal* similarity, influences retrieval. Here, we test the hypothesis that participants will be faster to make old/new judgements on pairs of words drawn from the same community relative to those drawn from distant communities.

We created 75 new pseudowords which were not seen in the exposure phase ("new words"). We then selected 15 non-boundary pseudowords ("old words") from the exposure phase (3 from each community). These old words were combined exhaustively to form 95 pairs in which items varied by community distance. Next, each old word was paired once with three new words (45 pairs). Finally, the 30 remaining old words were then paired with each other (15 pairs). In total, 165 pairs were created.

For the purposes of analyses, distance between items in a pair was construed as follows: a community distance of 0 meant the pair came from the same community (e.g. *creer* and *toist* in Figure 1). A community distance of 1 meant that the nodes were drawn from adjacent communities (e.g. *creer* and *twing*). A community distance of 2 meant that the nodes were two communities apart (e.g. *creer* and *metch*). There could be no measurement of community distance between old and new words, as the new words were not present in the exposure stream.

Procedure With the exception of the test phase, described below, procedures for Study 2 mirrored that of Study 1. Due to the increased number of trials presented during the exposure phase, its duration was 35 minutes.

Test phase. Participants were simultaneously presented with both items in a pair, one word above the other. Participants pressed [f] for "familiar" if both items had been seen in the exposure phase and [n] for "not familiar" if one or both of the items were new. All new words were only presented once to minimize confusion during the test phase. Trials were self-paced and separated by a 1.5 second blank screen. The order of the pairs and the position (top or bottom) of all pseudowords was randomized across participants.

Analysis and Results

Scrambled Word Detection Participants generally succeeded in distinguishing between pseudowords and their scrambled versions (93.5% accurate on average, SD = 5.6, excluding the four participants below threshold).

Data Exclusions For the exposure phase, data trimming techniques were identical to those described in Study 1 (17.0% of total data removed). Similarly, we subset trials to include only transition and pre-transition nodes.

For the test phase, we removed data corresponding to incorrect trials and any RTs greater than 3 standard deviations from the mean (total data loss = 11.3%).

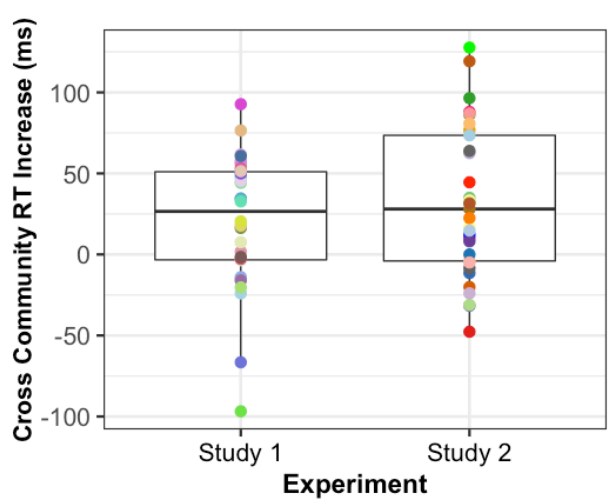

Figure 2. Cross community RT increase for Studies 1 and 2. Values included in the boxplot were calculated by subtracting, for each participant, mean RTs for pre-transition nodes from mean RTs for transition nodes.

Exposure Phase Similar to the previous study, RTs were regressed onto the main effects and interaction of Node Type (pre- vs. transition) and Trial (1-1400). The model included the fullest random effects structure that allowed the model to converge: a random intercept for participant and a by-participant random slope for Node Type, Trial, and their interaction.

Again, we observed a significant main effect of Node Type (β = 15.582, t = 3.782, p = 0.0002), indicating a processing cost for transition nodes compared to pre-transition nodes. The main effect of Trial (β = -18.251, t = -2.609, p = 0.014) was also significant. As in Study 1, no significant interaction between Node Type and Trial was observed (β = -0.924, t = -0.226, p = 0.821). Cross-community RT increases from both Study 1 and Study 2 are presented in Figure 2.

Repetition priming. Prior work examining the influence of community structure on RT patterns has addressed the potential for perceptual priming effects (e.g., Karuza et al., 2017; Kahn et al., 2018). It is well known that humans are faster to process a stimulus that they have seen recently. Though we propose that priming can in fact be considered a form of learning (see e.g., Chang, Dell, Bock & Griffin, 2000), we make contact with prior work by adding to both exposure phase models (Studies 1 and 2) the following measures of repetition priming: Lag10 and Recency. Lag10 indexes the number of times a particular node has been seen in the last 10 trials. Recency indexes the number of trials that have elapsed since a given node was last seen in the exposure stream. When adding these new predictors to our models, the main effect of Node Type was no longer significant (Study 1: β = 2.345, t = 0.441, p = 0.660; Study 2: β = 5.906, t = 1.104, p = 0.273).

Test Phase As in Meyer & Schvaneveldt (1971), our dependent measure of interest was RT for the old/new judgements. Given our lengthy exposure phase, and the fact we never repeated any of the "new words" during the test phase, participants attending to the test phase should have been able to easily and accurately make judgements about the novelty of items in the word pairs. Accuracy scores, though high (88.0% correct overall, SD = 9.9), were not our measure of interest. Rather, we were interested in whether RTs would vary as a function of the distance between nodes in a pair, with the fastest RTs for nodes within the same community. Thus, we imposed a cut-off of 75% accuracy on the test phase to exclude participants who were not complying with this relatively simple task, resulting in the exclusion of three additional participants. We note that without the exclusion of these participants, the significant results reported below do not hold.

Response times from the old/new judgments were regressed onto main effects of Community Distance (reverse-Helmert coded to reflect an increase in processing cost as distance increased) and Trial (1-165; intended to capture general task adaptation). Results are summarized in Table 1. Participants were fastest to respond to pseudoword pairs drawn from the same community relative to pseudowords drawn from the two adjacent communities. Unsurprisingly, participants were faster when responding to pseudowords pairs when they had seen both pseudowords before, compared to pairs in which when one or both of those pseudowords was new (Figure 3).

Table 1: Coefficients, t-values, and p-values for each predictor in a model examining the effect of Community Distance and Trial on participants' RTs for old/ new judgments (Study 2).

Predictor	Results
Community Distance (1 vs. 0)	β = 0.018, t = 2.062, p = 0.041 *
Community Distance (2 vs. 0,1)	β = 0.005, t = 1.040, p = 0.302
Community Distance (new vs. 0,1,2)	β = 0.024, t = 4.480, p = 0.0001 ***
Trial	β = -0.018, t = -1.626, p = 0.115

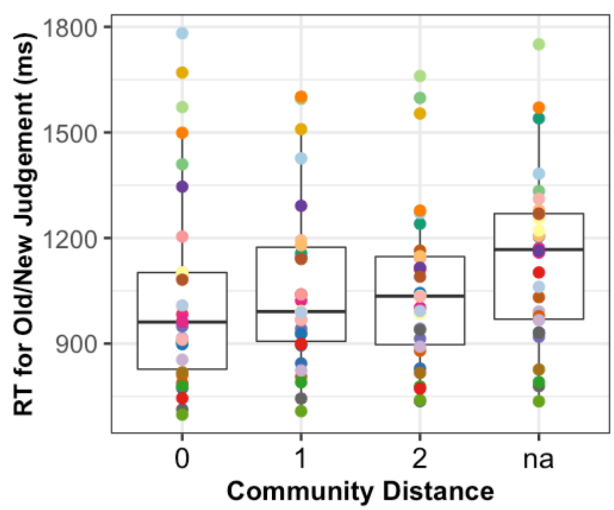

Figure 3. Boxplot of RTs for old/ new judgments on word pairs (Study 2). Values included in the boxplot were calculated by averaging, for each color-coded participant, mean RTs for nodes within a community (community distance = 0), from adjacent communities (= 1) and from non-adjacent communities (= 2). "NA" signifies that at least one word in the pair had not been seen by participants during exposure.

Discussion

We present data from two related studies demonstrating that learners are attuned to the network architecture underpinning continuous streams of linguistic stimuli. Specifically, we show that participants exhibited an increase in processing times when transitioning from one community of words to the next, suggesting that their expectations about upcoming input were influenced by the presence of element clusters in the sequence. As previous investigations into learners' sensitivity to network architectures have taken place exclusively in the visuomotor domain (Schapiro et al., 2013; Karuza et al., 2017; Kahn et al., 2018; Tompson et al., 2018), one notable contribution of the present work is that it speaks to the potential domain-generality of this learning mechanism.

Complex network analysis of natural language has consistently revealed that, among other properties, community structure may be essential to the organization of the mental lexicon. To varying degrees, real-world networks in which edges represent phonological overlap, semantic relatedness, and temporal co-occurrence, display this property (De Deyne, Verheyen, & Storms, 2016). The present experiments break new ground in that they demonstrate that community structure is not only an emergent property of language, but also a form of high-level regularity that can guide sequence-level learning. Perhaps of greatest interest, we show through a post-exposure measure in Study 2 that temporal overlap can be translated into an accessible representation, as evidenced by the influence of community distance as participants completed a subsequent word recognition task.

At first blush, it is potentially surprising that we observe no significant interaction between Node Type and Trial. In other words, the magnitude of the cross-community RT increase did not change significantly over the course of exposure. However, these results align with previous findings suggesting that sensitivity to community structure may emerge very early in exposure (e.g., Karuza et al., 2017; Karuza, Kahn, & Bassett, 2019). To be clear, we do find a key point of divergence between the present findings and the existing literature on community structure in visuomotor sequences. Specifically, the effect of traversing an inter-community edge was substantially weakened by the inclusion of nuisance regressors intended to account for repetition priming. While there are several possible explanations for this pattern of results, we narrow in on two of them. First, we studied stimulus streams generated from a significantly larger graph than those used in related experiments (i.e., a total of 40 nodes relative to 15). Participants therefore observed far fewer unique edge traversals throughout the course of the experiment. Perceptual priming may have an inflated effect when learners are exposed to more varied stimulus streams in which nodes are repeated only a handful of times. Second, the choice to include pronounceable pseudowords with relatively few real-word orthographic and phonological neighbors meant that these features of our stimuli may have also exerted an undue influence on processing times (Vitevitch, Chan, & Roodenrys, 2012). This source of noise, coupled with some phonological overlap between the pseudowords themselves (e.g., *wabe* and *woast*), may have also contributed to null results obtained for the substring comparison post-test of Study 1. We reiterate that the randomization of word-to-node mapping should have minimized these effects. Nevertheless, evaluation of the full impact of phonological and orthographic neighborhood, defined in terms of the extent of overlap with existing English words as well as among stimulus items themselves, will be an important area of future study. It is possible, for example, that cross-community RT effects shift in magnitude in cases where pseudoword stimuli have an extremely high number of real-word neighbors.

Taken together, this set of results opens up a number of intriguing future directions not limited to investigations into learners' sensitivity to multiple layers of structure (e.g., through the construction of multiplex networks that simultaneously take into account phonological and temporal overlap; Stella, Beckage & Brede, 2017). In a broader context, formalization of the relationship between linguistic network structure and learning could add substantially to discussions regarding how language networks change with development (Ke & Yao, 2008) or why they display certain characteristic properties in special populations (Beckage, Smith & Hills, 2011). On a final note, decreased sensitivity to statistical associations has been linked to disorders ranging from Broca's aphasia (Goschke, Friederici, Kotz, & van Kampen, 2001) to dyslexia (Schmalz, Altoe, & Mulatti, 2017) and developmental language disorder (Lammertink, Boersma, Wijnen, & Rispens, 2017). Extending these lines of inquiry to reveal potential impairments in the extraction of network-level patterns could have powerful consequences, not only in terms of informing rehabilitative practices but also in deepening our understanding of language acquisition more generally.

Acknowledgements

The authors are grateful to Joezette Gray and Alex Ferraccio, for assistance with data collection and stimulus creation, and to David Wiegand for helpful comments on this work.

References

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, *20*(03), 679-85.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience, 20*(3), 353-64.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS one*, *6*(5), e19348.

Borge-Holthoefer, J., & Arenas, A. (2010). Semantic networks: Structure and dynamics. *Entropy*, *12*(5), 1264-1302.

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research, 29,* 217–30.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-28.

De Deyne, S., Verheyen, S., & Storms, G., (2016). Structure and organization of the mental lexicon: a network approach derived from syntactic dependency relations and word associations. In Mehler, A., Lücking, A., Banisch, S., Blanchard, P., & Job, B., *Towards a Theoretical*

Framework for Analyzing Complex Linguistic Networks. Berlin: Springer Berlin Heidelberg.

Ferrer i Cancho, R. F., Solé, R. V., & Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review E*, *69*(5), 051915.

Goldstein, R., & Vitevitch, M. S. (2014). The influence of clustering coefficient on word-learning: how groups of similar sounding words facilitate acquisition. *Frontiers in Psychology*, *5*, 1307.

Goschke, T., Friederici, A. D., Kotz, S. A., & van Kampen, A. (2001). Procedural learning in Broca's aphasia: Dissociation between the implicit acquisition of spatio-motor and phoneme sequences. *Journal of Cognitive Neuroscience, 13(3)*, 370-388.

Kahn, A. E., Karuza, E. A., Vettel, J. M., & Bassett, D. S. (2018). Network constraints on learnability of probabilistic motor sequences. *Nature Human Behaviour*, *2*(12), 936.

Karuza, E. A., Kahn, A. E., & Bassett, D. S. (2019). Human sensitivity to community structure is robust to topological variation. *Complexity*, *2019*.

Karuza, E. A., Kahn, A. E., Thompson-Schill, S. L., & Bassett, D. S. (2017). Process reveals structure: How a network is traversed mediates expectations about its architecture. *Scientific Reports*, *7*(1), 12733.

Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2016). Local patterns to global architectures: influences of network topology on human learning. *Trends in Cognitive Sciences*, *20*(8), 629-40.

Ke, J., & Yao, Y. A. O. (2008). Analysing language development from a network approach. *Journal of Quantitative Linguistics*, *15*(1), 70-99.

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis*. Journal of Speech, Language, and Hearing Research*, *60*(12), 3474-3486.

Liu, H. (2008). The complexity of Chinese syntactic dependency networks. *Physica A: Statistical Mechanics and its Applications, 387*(12), 3048-58.

Meulemans, T., & Van der Linden, M. (1997). Does the artificial grammar learning paradigm involve the acquisition of complex information? *Psychologica Belgica, 37*(1-2), 69-88.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227-34.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. The Quarterly Journal of Experimental Psychology: A, 55(4), 1339-62.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486-492.

Schmalz, X., Altoe, G., & Mulatti, C. (2017). Statistical learning and dyslexia: A systematic review. *Annals of Dyslexia, 67(2),* 147-162.

Scott, J. (2017). *Social Network Analysis.* London: SAGE Publications.

Siew, C. S. Q., & Vitevitch, M. S. (2016). Spoken word recognition and serial recall of words from components in the phonological network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 394-410.

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, *7*, 46730.

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.

Tompson, S. H., Kahn, A. E., Falk, E. B., Vettel, J. M., & Bassett, D. S. (2018). Individual differences in learning social and nonsocial network structures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(2), 253-71.

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research, 51*(2), 408-22.

Vitevitch, M. S., Chan, K. Y., & Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language*, *67*(1), 30-44.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971-979.