








Who does what to whom? graph representations of action-predication in speech relate to psychopathological dimensions of psychosis

Amir H. Nikzad¹ , Yan Cong¹ , Sarah Berretta¹, Katrin Hänsel² , Sunghye Cho³, Sameer Pradhan³, Leily Behbehani¹, Danielle D. DeSouza^{4,5} , Mark Y. Liberman³ and Sunny X. Tang¹ 

Graphical representations of speech generate powerful computational measures related to psychosis. Previous studies have mostly relied on structural relations between words as the basis of graph formation, i.e., connecting each word to the next in a sequence of words. Here, we introduced a method of graph formation grounded in semantic relationships by identifying elements that act upon each other (action relation) and the contents of those actions (predication relation). Speech from picture descriptions and open-ended narrative tasks were collected from a cross-diagnostic group of healthy volunteers and people with psychotic or non-psychotic disorders. Recordings were transcribed and underwent automated language processing, including semantic role labeling to identify action and predication relations. Structural and semantic graph features were computed using static and dynamic (moving-window) techniques. Compared to structural graphs, semantic graphs were more strongly correlated with dimensional psychosis symptoms. Dynamic features also outperformed static features, and samples from picture descriptions yielded larger effect sizes than narrative responses for psychosis diagnoses and symptom dimensions. Overall, semantic graphs captured unique and clinically meaningful information about psychosis and related symptom dimensions. These features, particularly when derived from semi-structured tasks using dynamic measurement, are meaningful additions to the repertoire of computational linguistic methods in psychiatry.

Schizophrenia (2022)8:58; <https://doi.org/10.1038/s41537-022-00263-7>


INTRODUCTION

Disturbances in speech have been recognized as a key component of both positive and negative symptoms in psychosis¹. Here we define speech as the sum of all acoustic and lexical aspects of spoken communication. Increasingly, speech phenotypes in psychosis can be objectively and reliably measured through automated language analysis for the detection and prediction of psychotic disorders^{2–7}. Speech graphs, derived from transcribed texts, have shown the ability to accurately quantify language disorganization and impoverishment. There were significant relationships between graph features and key psychosis phenotypes, including thought disorder, cognition, global functioning, and brain connectivity changes⁸. Speech graphs are network representations of discourse that treat linguistic elements (words, lexemes, etc.) as nodes and relationships among those elements as the bridging links (edges)⁹. Generally, relationships among linguistic elements may be structural (based on the relative locations of the words, e.g., occurring in sequence or co-occurrence in the same utterance) or semantic (based on the meaning of the utterance, e.g., entity A is acting upon entity B). Quantitative measures of the size, connectedness, and organizational structure of the speech graphs can then be calculated¹⁰. For example, the size of the graph can be quantified by the number of nodes and edges as well as measures of internal distances such as network diameter and average shortest path length. The connectedness of the graph is reflected in average degrees, graph density, and size of the largest connected component. The

degree of organization in the graph can be measured by comparing the statistical similarity of graph features to randomly generated graphs of the same size.

Sequential speech graphs of individuals with psychosis spectrum disorders (PS+) have been characterized as being smaller, less connected, and more disorganized than individuals without psychosis spectrum disorders (PS–). In their pioneering study, Mota and colleagues showed that the PS+ speech graphs exhibited fewer nodes and edges, lower average degrees, and smaller connected components compared to those of healthy controls and patients with mania¹¹. PS+ speech graphs also had a lower average shortest path length and network diameter per fixed word lengths, reflecting shorter internal distances¹². A subsequent study comparing the size of connected components in speech graphs with that of randomly generated graphs of the same size revealed that PS+ graphs have more random-like organization compared to PS–¹³. Palaniyappan et al.⁸ further showed that the graph measures of connectedness (size of connected component) and organization (size of the connected component divided by that of random graphs) were associated with disorganized and impoverished thought disorders as well as clinical measures of schizophrenic symptoms and biological measures of neural connectivity.

However, sequential relationships among words are vulnerable to non-psychopathologic factors such as stylistic preferences of the speaker, passive or active voice, and different grammatical structures in different languages. On the other hand, semantic relationships are tied to the underlying concepts being expressed,

¹Feinstein Institutes for Medical Research, Zucker Hillside Hospital, Department of Psychiatry, Glen Oaks, USA. ²Yale University, Department of Laboratory Medicine, New Haven, USA. ³University of Pennsylvania, Linguistic Data Consortium, Philadelphia, USA. ⁴Winterlight Labs, Inc., Toronto, Canada. ⁵Stanford University, Department of Neurology & Neurological Sciences, Stanford, USA. email: anikzad@northwell.edu; stang3@northwell.edu

and they are consistent across languages and speaking style. For instance, in “The dog chased the cat” and “The cat was chased by the dog,” the same semantic content is expressed with different word sequences. Furthermore, there is evidence that thought disorder is related to disruptions in semantic networks¹⁴. Thus, graphs that utilize the core semantic content may be a more direct method for representing disrupted brain circuits in psychosis. Here, we attempt to build speech graphs upon two semantic relations which have been proposed as universal linguistic relationships¹⁵: (1) Action, which links the actor of the underlying event to its undergoers (dog → cat), and (2) predications, which link the predicate of the utterance to its arguments (chase → dog and cat). These relationships capture the core semantic meaning of the utterance: who does what to whom? Figure 1a illustrates the structures of sequential and action-predication graphs.

In this paper, we aim at (1) introducing a new way to produce semantic speech graphs based on the two universal linguistic relations of action and predication, (2) verifying the validity of size, connectedness, and organization as three non-redundant domains of speech graph features, (3) comparing semantic and structural graph measures of size, connectedness and organization in their associations with the presence of psychosis, and (4) examining the relationship between graph features and clinically rated dimensions of psychosis symptoms. The overall goal is to operationalize semantic speech graph methodology with respect to studying language disturbance in psychosis and guide subsequent studies.

RESULTS

Participant characteristics

In total, speech samples of 205 and 201 participants were collected and transcribed for picture description and open-ended narrative tasks respectively, corresponding to 81 PS+ and 124 PS- participants (Table 1). On average, picture descriptions included 110 ± 66 words and narrative responses included 162 ± 121 words; word counts were not significantly different between PS+ and PS- (p -values were 0.310 and 0.051 for open-ended narrative and picture description tasks respectively). As expected, PS+ participants scored significantly higher in overall psychosis symptoms, negative symptoms, and demonstrated significantly more abnormal speech per clinical ratings.

Speech graphs formation and measurements

The structural graphs were formed by the sequential connection of elements of structural graph entry irrespective of utterance boundaries. The semantic graph representations were created by first tagging verbs, actor-arguments and undergoer-arguments using semantic role labeling, and then combining action relations (actor → undergoer) and predication relations (verb-predicate → actor and undergoer arguments) within each utterance (Fig. 1a). Iterations of the same relationships were captured as the edge weight; the first occurrence was weighted 1 and repetitions added 1 to the edge weight incrementally. The performance of the semantic role labeler was inspected and did not appear to demonstrate any biases (Supplementary Table 2). The code for the formation of semantic graphs is shared in a public repository (<https://github.com/STANG-lab/Semantic-Graphs>). A more comprehensive illustration of the applied method is presented in Supplementary Table 2.

- *Utterance Example*: The kid is grabbing the cookie jar.
- *Structural Graph Connections*: kid → grab + grab → cookie + cookie → jar
- *Semantic Graph Connections*: grab → kid + grab → cookie jar + kid → cookie jar

Size, connectedness, and organization of sequential and action-predication networks were measured by computing relevant graph features. Graph size was quantified with straightforward counts (number of nodes (NN), number of edges (NE)) and internal-distance measures (diameter, average shortest path length between any two nodes (ASPL)). Connectedness was quantified using average weighted degree (i.e., the number of weighted edges for each node; AWD), graph density (realized edges divided by possible edges) and number of nodes in the largest strongly connected component (LSCC). The level of organization in graphs was estimated by computing the z-score of ASPL and LSCC relative to 1000 randomly generated graphs of the same NN and NE. All calculations were performed in Python using the igraph library¹⁶. Random graphs were generated using the built-in Erdős-Rényi algorithm¹⁷.

Static graph features were calculated based on the networks of the whole response for each task, and averaged across each task category for each participant (picture description vs. narrative). We complemented usual static graph features with moving-window measures introduced by Mota and colleagues¹². The length and moving-step of the window were set to 30-tokens length and 1-token step for the sequential graphs based on the best performing models in previous studies which explored this technique^{8,12,13}. Action-predication networks were calculated for 3-utterance segments and 1-utterance steps because these lengths were the closest equivalents to that of structural graphs, considering the mean sentence length of ~10 words. Dynamic graph features were then calculated as the average of graph features over all windows. In total, 36 graph features ($n = 18$ for structural graphs, $n = 18$ for semantic graphs) were analyzed. The full list of graph features is included in Supplementary Table 3. Graph formations and measurements were performed separately for each response (i.e., each picture or open-ended prompt) and then averaged across the task for each participant.

The redundancy of graph features

To evaluate the mutual redundancy of different sets of graph features, we computed variance inflation factors (VIF) at successive layers of analysis¹⁸: first, across domains (size, connectedness, and organization), then across modes (static and dynamic), then for each type of graph (structural and semantic), and finally for each task category—picture description (Table 2) and open-ended narratives (Supplementary Table 4).

Intra-domain VIF analyses revealed that the information in the internal-distance measures was not redundant with respect to general measures of size. All connectedness measures in both graph types survived in the intra-domain analysis. Dynamic organization measures of sequential graphs were not mutually redundant, but only one of the dynamic organization measures survived in the VIF comparison of the action-predication networks. Increasing the level of integration did not lead to exclusion of one entire domain. Notably, dynamic semantic graph features of all three domains of size, connectedness, and organization remained in the final sets in both tasks; this was not observed in other types of graph features.

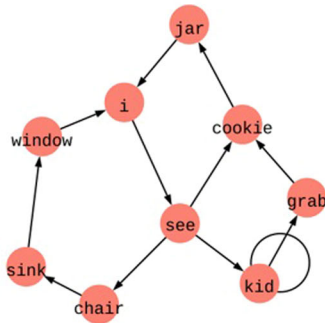
Speech graph features and psychosis

Associations between graph features and the presence of psychosis are presented in Table 3. In general, semantic graphs showed a more task-specific behavior compared to structural graphs, with more significant associations in picture descriptions. Structural graphs performed similarly in picture description compared to open-ended narrative tasks. The moving window method enhanced the significance and effect size of the relationships for both graph types and for all three domains of size, connectedness, and organization. In order to account for the

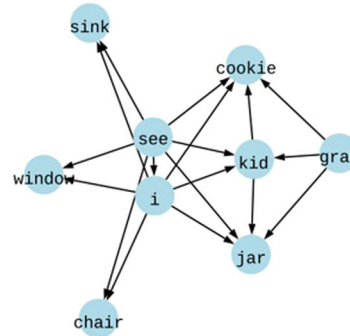
a. Structural and Semantic Graphs

Example:

“I see a cookie jar. I see a chair a sink a window. I see a kid. The kid is grabbing the cookie jar.”



Sequential graph



Action-prediction graph

b. Dynamic Graph Features

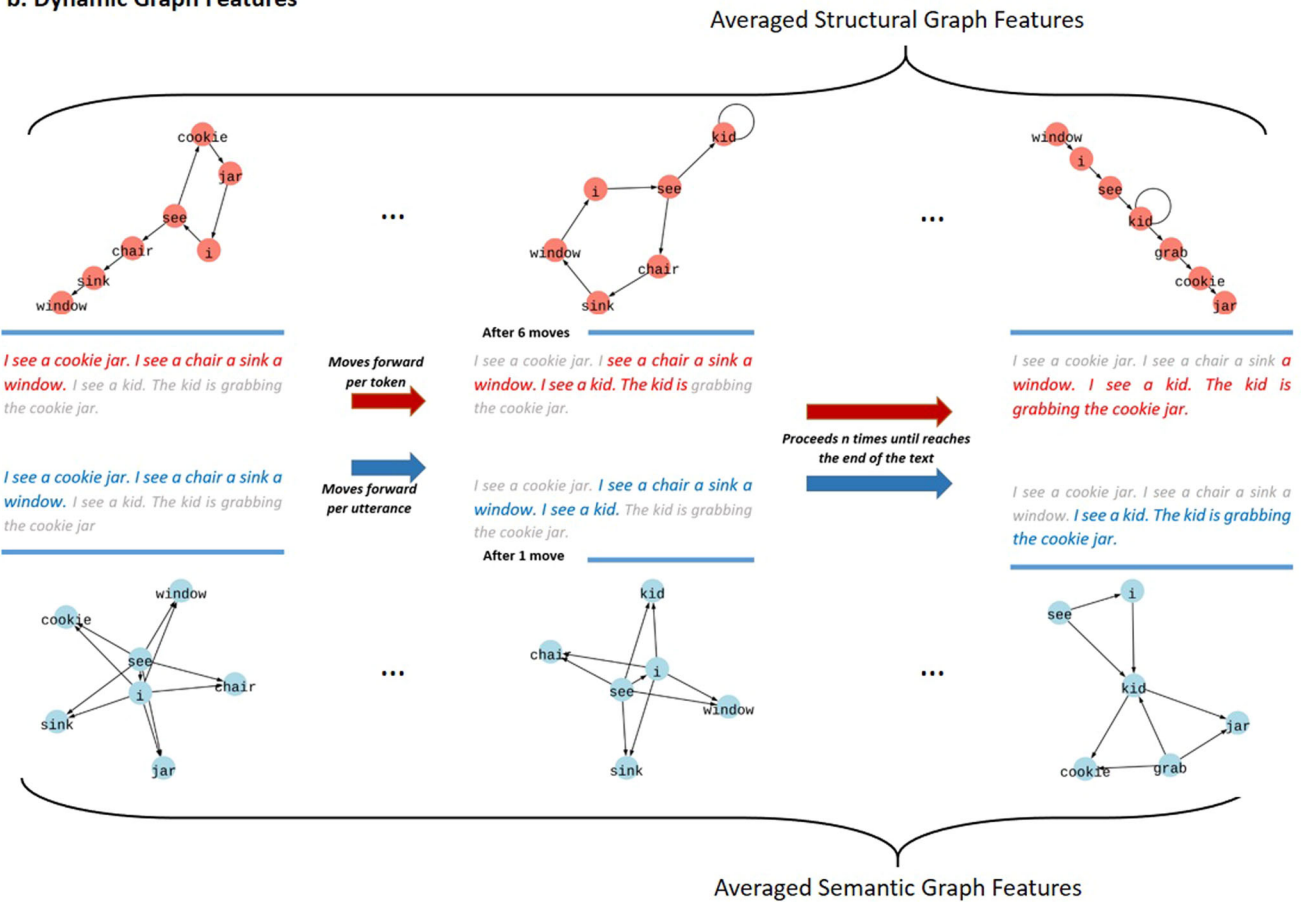


Fig. 1 Speech graph methodology. **a** Structural and semantic graph representations of a given text are illustrated. The structural representation is produced based on sequential relations between lemmatized content words (e.g., I→see→cookie→jar, etc.). The semantic representation is produced by connecting elements that act upon each other (e.g., I→chair; kid→cookie jar), and linking verb predicates to their arguments (e.g., see→I; see→chair; grab→kid; grab→cookie jar). **b** Dynamic graph features are computed by sliding a window of fixed length throughout each sample to produce *n* instances of graph representations. Subsequently, each graph feature is calculated as the mean value of *n* features each belonging to a particular instance. Successive sequential graphs progress one word at a time in windows of 30 words, and semantic graphs slide one utterance at a time in windows of three utterances.

Table 1. Demographic characteristics of the participants.

	Overall	PS+	PS–	<i>p</i> value
<i>n</i>	205	81 (40%)	124 (60%)	
Age		25.9 ± 6.3	25.2 ± 5.9	0.20
Sex				<0.01**
Female	118 (58%)	35	83	
Male	87 (42%)	46	41	
Race				<0.01**
African American	53 (26%)	31	22	
Asian	24 (12%)	11	13	
Caucasian	89 (43%)	20	69	
Other	39 (19%)	19	20	
Education		14.1 ± 2	15.5 ± 1.9	<0.001***
Caregiver education		14.6 ± 3	15.3 ± 2.3	0.07
BPRS total score		42 ± 14	22 ± 4	<0.001***
SANS total score		30 ± 15	4 ± 4.5	<0.001***
TLC total score		16 ± 14	4 ± 5	<0.001***

The significant associations are tagged by asterisks (*<0.05, **<0.01, and ***<0.001). *P*-values were calculated using Pearson's chi-squared test (for sex and race), and the Mann–Whitney *U* test (for age, education, caregiver education, and BPRS, SANS, and TLC total scores).

PS+ individuals with psychosis spectrum disorders, PS– individuals without psychosis spectrum disorders, BPRS Brief Psychiatric Rating Scale, SANS scale for assessment of negative symptoms, TLC scale for the assessment of thought, language, and communication.

uneven distributions of sex and race in PS+ and PS–, we selected two subsamples of participants matched for sex and race, respectively. The correlations between graph measures and psychosis status were consistent with the overall findings in each matched subsample (Supplementary Table 5).

Size. Speech from PS+ yielded significantly smaller structural and semantic graphs as reflected in multiple correlations between measures of network size and psychosis (Table 3-A). The moving-window technique improved the correlations in both graph types with ~0.1 increase in mean RBC for sequential and action-predication graphs respectively. General measures of the network size (NE and NN) manifested the strongest and most significant correlations in psychosis for both graph types and in both tasks.

Connectedness. Psychosis was associated with smaller connected components and higher graph densities (Table 3-B). This pattern was observed in both structural and semantic graphs. However, average weighted degrees were higher in the structural graphs of psychotic speech and lower in semantic networks compared to non-psychotic counterparts. The most informative semantic graph features in this domain were dynamic graph density and LSCC for open-ended and picture description tasks respectively. Dynamic density in sequential graphs showed the highest effect size in both tasks (picture description: $p < 0.001$, RBC = -0.33 ; narrative: $p < 0.001$, RBC = -0.31).

Organization. Structural and semantic speech graphs showed different patterns of organization with respect to random graphs with negative mean z-scores in semantic graphs and positive z-scores for structural graphs. This pattern was consistent for both measures (LCCZ and ASPLZ) and in both tasks. These findings suggest that action-predication graphs of speech are organized in smaller connected components with shorter inter-nodal pathways compared to random graphs of the same size. Conversely, sequential graphs produce larger connected components with

nodes that were further apart. However, in both cases, PS+ graphs incline toward the more random-like patterns of organizations compared to those of PS–, i.e., exhibiting higher z-scores in semantic graphs and lower z-scores in structural graphs (Table 3-C).

Speech graphs and dimensional clinical characteristics

Figure 2a represents task-wise correlation plots for structural and semantic graph features vs. clinical measures of language disturbance (TLC speech disorganization and speech poverty factors), disease severity (BPRS total score) and psychopathological dimensions (BPRS anxiety/depression, hostility/suspiciousness, thought disturbance and withdrawal/psychomotor retardation factors; SANS affective flattening, alogia, avolition, and asociality/anhedonia global scores). In general, graph features generated from the picture description tasks were more closely related to clinical measures than those from narrative tasks. Within the picture description task, the dynamic action-predication graph features outperformed the static features as well as both static and dynamic sequential graph features (Fig. 2a). Among graph measures of the narrative task, a dynamic measure of structural graph organization (D_SEQ LSCCZ) showed the strongest connections with clinical characteristics.

Clinical measures of speech disturbance. Speech disorganization was associated with increased connectedness and decreased organization in both tasks. This relationship is more prominent in structural features with more Bonferroni survived correlations and higher correlation coefficients (absolute $\rho = 0.36$ – 0.39). Speech poverty was correlated with the density and size of graphs of both type, with impoverished speech having denser and smaller graphs; this correlation was more clearly observed in the semantic networks than in the structural graphs (absolute $\rho = 0.40$ – 0.41).

Overall disease severity. The strongest correlates of total BPRS score were dynamic measures of size ($\rho = -0.54$ to -0.58) and organization ($\rho = 0.51$ – 0.54) of semantic graphs in the picture description task. For narratives, the same relationships were reproduced, but there was a stronger correlation between overall disease severity (BPRS total score) and the dynamic organization measure (LSCCZ) of structural graphs ($\rho = -0.45$).

Dimensional measures of psychosis. Figure 2b shows the significant relationships among non-redundant graph features and clinical measures. Patterns were task-dependent. Semantic graph features derived from picture description tasks were consistently more informative for dimensional measures of psychosis than structural features. Multiple connections were observed between semantic graph features from picture description and multiple psychopathological dimensions including hostility/suspiciousness, thought disturbance, withdrawal/psychomotor retardation, affective flattening, and avolition. Thought disturbance was strongly correlated with almost all dynamic graph features in the Action-Predication network ($\rho = -0.4$ to -0.62). Negative factors of withdrawal/psychomotor retardation and affect flattening were connected to static and dynamic measures of connectedness in semantic graphs. Dynamic measure of structural graph organization (LCCZ) was connected with multiple clinical dimensions in open-ended narrative task, including avolition, thought disturbance, and hostility/suspiciousness (absolute $\rho = 0.4$ – 0.45). The only domain that remained uncorrelated in both tasks for all graph features was the anxiety/depression factor.

DISCUSSION

Here, we presented an automated speech analysis method that objectively measured speech quantity, connectedness, and

Table 2. Survived graph features in sequential VIF comparisons in picture description task.

Graph Features (n = 36)	Picture Description Task		
	1. Domain	2. Type	3. Task
Static Action-Predication Graph Size (n = 4)	S _{AP} NN, S _{AP} ASPL	S _{AP} Density, S _{AP} LSCC,	S _{AP} LSCC, D _{AP} ASPL,
Static Action-Predication Graph Connectedness (n = 3)	S _{AP} AWD, S _{AP} Density, S _{AP} LSCC	D _{AP} ASPL, D _{AP} Density, D _{AP} LSCC, D _{AP} ASPLZ	D _{AP} Density, D _{AP} LSCC, D _{AP} ASPLZ,
Static Action-Predication Graph Organization (n = 2)	S _{AP} LSCCZ		S _{SEQ} NE, S _{SEQ} Density, S _{SEQ} LSCCZ,
Dynamic Action-Predication Graph Size (n = 4)	D _{AP} NN, D _{AP} ASPL		D _{SEQ} AWD, D _{SEQ} LSCCZ, D _{SEQ} ASPLZ
Dynamic Action-Predication Graph Connectedness (n = 3)	D _{AP} AWD, D _{AP} Density, D _{AP} LSCC		
Dynamic Action-Predication Graph Organization (n = 2)	D _{AP} ASPLZ		
Static Sequential Graph Size (n = 4)	S _{SEQ} NE, S _{SEQ} ASPL	S _{SEQ} NE, S _{SEQ} Density, S _{SEQ} LSCCZ,	
Static Sequential Graph Connectedness (n = 3)	S _{SEQ} AWD, S _{SEQ} Density, S _{SEQ} LSCC	D _{SEQ} AWD, D _{SEQ} LSCCZ, D _{SEQ} ASPLZ	
Static Sequential Graph Organization (n = 2)	S _{SEQ} LSCCZ, S _{SEQ} ASPLZ		
Dynamic Sequential Graph Size (n = 4)	D _{SEQ} NE, D _{SEQ} ASPL		
Dynamic Sequential Graph Connectedness (n = 3)	D _{SEQ} AWD, D _{SEQ} Density, D _{SEQ} LSCC		
Dynamic Sequential Graph Organization (n = 2)	D _{SEQ} LSCCZ, D _{SEQ} ASPLZ		

Columns within each segment accommodate survived graph features. VIF comparison was conducted and features of the highest VIF were excluded successively until a set of features all showing VIF < 5 was attained. Survived features were then passed to the next column on right for another comparison on a more integrated level. 1. Domain column shows results of intra-domain comparisons. 2. Type column presents the features integrated on graph-type level, i.e., semantic vs structural graph features. 3. Task column combines all graph features per each task. For the dynamic semantic graph, features belonging to three domains of size, connectedness and organization remained in the final set. Graph features of different methods are color coded. VIF comparisons for open-ended narrative task is provided in Supplementary Table 4. More details on graph features are available in Supplementary Table 3.

S_{AP} static action-predication graph feature, D_{AP} dynamic action-predication graph feature, S_{SEQ} static sequential graph feature, D_{SEQ} dynamic sequential graph feature, NN number of nodes, NE number of edges, diameter graph diameter, ASPL average shortest path length, AWD average weighted degree, density graph density, LSCC size of largest strongly connected component, LSCCZ z-score of LSCC compared to 1000 random graphs, ASPLZ z-score of ASPL compared to 1000 random graphs.

organization using graph metrics. These features were quantitative and objective, and they captured semantic relationships of action (between actors and undergoers of utterance) and predication (between the predicate and arguments).

We found that the semantic graph features derived from the picture description tasks were strongly correlated with psychopathological dimensions of psychosis, with the dynamic features outperforming the static features. Our findings suggest that incorporating semantic information in graph modeling of speech can increase the performance of such models. In the picture description tasks, only the semantic graph features were connected to psychopathological dimensions other than language disturbances, including dimensions such as affect, abolition, and thought content. There are no other existing reports of the use of semantic graph methodology in studying psychosis. Therefore, this finding remains to be replicated.

Speech graphs of PS+ were smaller, less connected, and more randomly organized compared to those of PS-. This was true for both semantic and structural graphs in our study, and the finding is consistent with previous studies using structural speech graphs^{11–13}. For example, Mota et al. found that schizophrenia was related to decreased size in terms of the number of nodes¹¹ and measures of internal distances¹². The association between psychosis and decreased speech graph size may reflect a similar

phenomenon as the decreased semantic density found by Rezai et al.⁵ and decreased idea density reported by Moe et al.¹⁹. With regard to measures of connectedness and organization, recent studies have relied on the LSCC, the largest strongly connected component, as an absolute and relative quantity with respect to randomly generated graphs^{8,20,21}. Our findings suggest that additional information can be captured with other non-redundant features describing size, connectedness, and organization. For example, we found that diameter and average short path length convey non-redundant information about the expansiveness of the network in addition to number of nodes and edges, network density and average degree can be used along the connectedness measure of LSCC, and z-score analysis of ASPL can be considered as a measure of speech graph organization. Each of these graph metric domains should be included in future efforts to quantify clinical speech characteristics.

We found that graph features are related to psychosis in a task-dependent manner. Previous studies reported that dream reports and picture description tasks are more informative about psychosis compared to narrations pertaining to everyday life, as reflected in larger effect sizes for association with psychosis^{12,13}. Since not all participants are able to recall and report dreams¹³, picture descriptions have developed into the preferred source of speech samples for graph analysis^{8,20}. Accordingly, our findings

Table 3. Relationships between structural and semantic graph features and psychosis.

Graph feature	Task	Static action-predication graph (S_AP)	Dynamic action-predication graph (D_AP)	Static sequential graph (S_Seq)	Dynamic sequential graph (D_Seq)
<i>A. Size</i>					
Number of nodes (NN)	Picture	0.17*	0.27***	0.20*	0.34***
	Open-ended	0.13	0.21*	0.14	0.33***
Number of edges (NE)	Picture	0.19*	0.31***	0.17*	0.27**
	Open-ended	0.12	0.19*	0.11	0.37***
Diameter	Picture	0.16	0.23**	0.24**	0.30***
	Open-ended	0.00	0.03	0.28***	0.25**
Average shortest path length (ASPL)	Picture	0.11	0.18*	0.24**	0.28**
	Open-ended	0.08	0.17*	0.27**	0.26**
<i>B. Connectedness</i>					
Average weighted degree (AWD)	Picture	0.15	0.22**	-0.05	- 0.32***
	Open-ended	-0.03	0.06	-0.13	-0.29***
Density	Picture	-0.06	-0.15	-0.26**	- 0.33***
	Open-ended	-0.20*	-0.23**	-0.21*	- 0.31***
Size of largest strongly connected component (LSCC)	Picture	0.15	0.24**	0.20*	0.18*
	Open-ended	0.08	0.13	0.14	0.29***
<i>C. Organization</i>					
LSCC z-score (LSCCZ)	Picture	-0.22**	- 0.31***	0.28***	0.25**
	Open-ended	-0.09	-0.13	0.26**	0.31***
ASPL z-score (ASPLZ)	Picture	-0.22**	- 0.30***	-0.02	0.25**
	Open-ended	-0.11	-0.14	0.01	0.24**

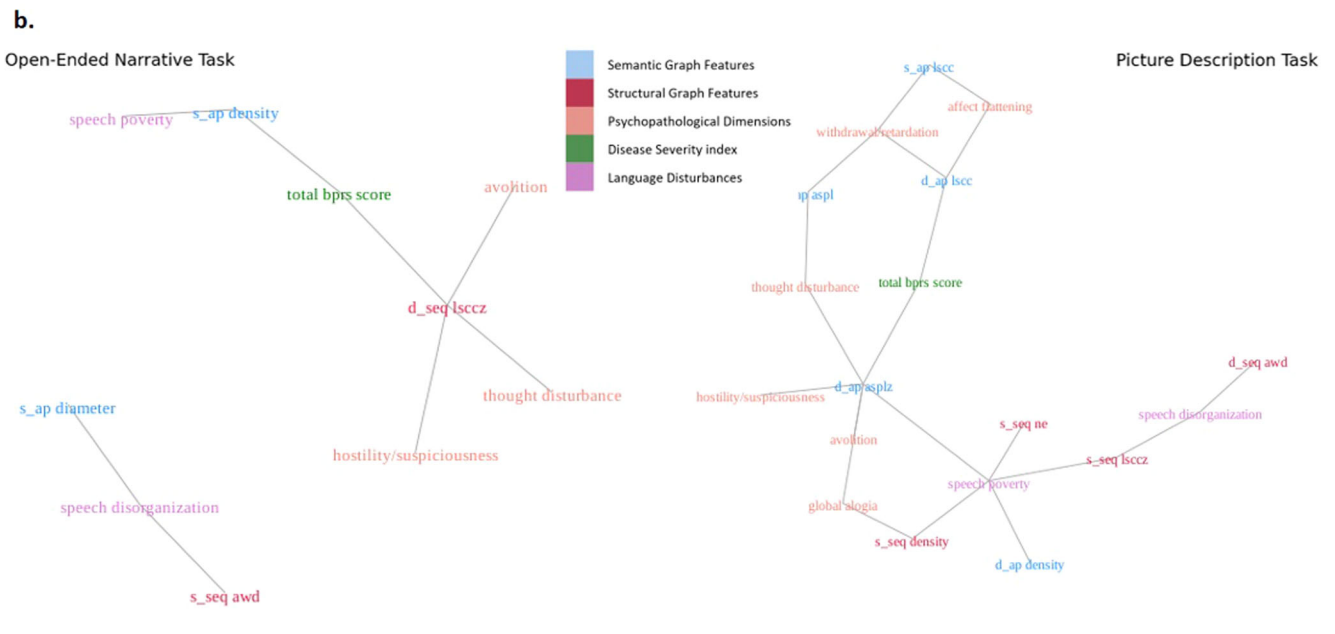
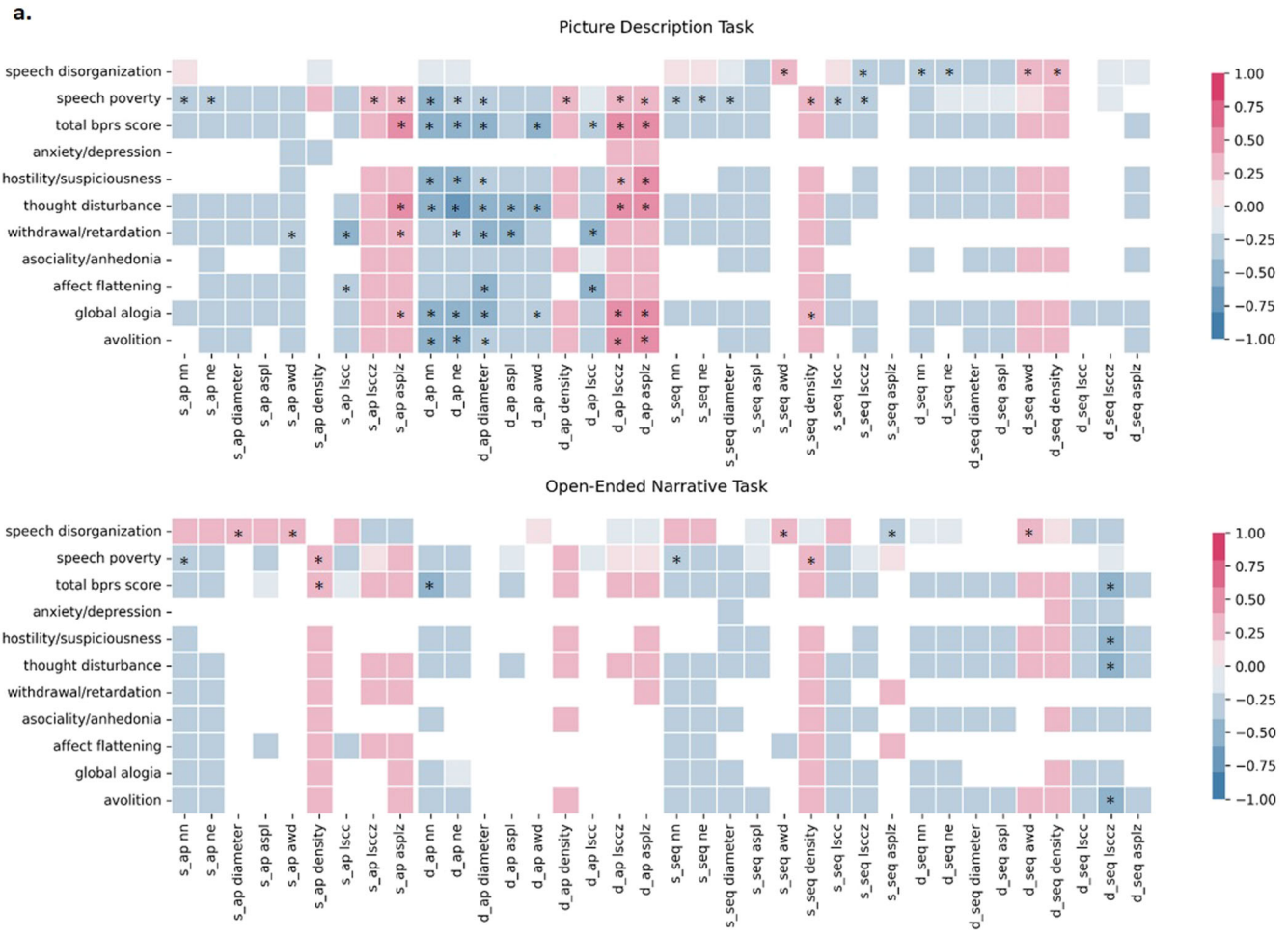
Graph features are categorized into three domains of size, connectedness, and psychosis. For each feature rank biserial correlation coefficient (RBC) is reported for picture description task (Picture) and open-ended narrative (Open-ended) as a measure of effect size. The significant associations are tagged by asterisks (* <0.05 , ** <0.01 , and *** <0.001). Associations survived in Bonferroni correction are bolded ($\alpha = 0.0003$). Semantic graphs showed a preference for picture description tasks with higher effect sizes and more significant associations. Structural graphs showed a task-independent relation with psychosis. In overall, dynamic graph features outperformed static counterparts in both graph types.

suggest better performance of speech graphs based on picture description tasks compared to open-ended narratives. It may be the case that the additional structure provided in these tasks is able to reduce noise—i.e., variations in speech that are not related to psychopathology, which may be more dependent on the mood or social status of individual speakers. Furthermore, relative to open-ended narratives, picture description tasks have a pre-set common ground between speaker and listener: the picture. This setting helps navigate the speaker's response, making it relevant and informative enough for graph analysis. Although everyday verbal communications are not directed by specific tasks, picture descriptions seem suitable for brief speech sampling for computational language analysis in clinical context. Further studies with larger open-ended narrative speech samples may show similar results.

Moreover, we found that controlling the amount of speech using the moving-window method enhanced the performance of speech graph models. This is in line with previous studies conducted by structural speech graphs, where different approaches attempted to control speech quantity, including using normalized features (i.e., measuring features per word count)¹¹,

setting time limits for speech recordings^{8,13,20}, and incorporating moving-window methods (i.e., using dynamic graph features)^{8,12,13}. Mota and colleagues reported that measuring graph features per word count reversed the relationship of graph features with psychopathological conditions (e.g., the number of nodes were lower in the speech graph of patients with schizophrenia compared to that of patients with manic disorder, but this was reversed when normalized for word count)¹¹. Therefore, to make the methodology more uniform and applicable to a variety of samples, we suggest using dynamic graph features for subsequent studies.

There are several limitations to the current study. The scope of our comparisons was focused on psychosis as a heterogeneous and multi-faceted condition. Future studies should evaluate whether there are more fine-grained relationships between semantic graph features and psychosis subtypes, as well as whether the presence of potentially comorbid conditions, differences in treatment history, and social determinants affect these measures. In addition, participants of our study were recruited from both inpatient and outpatient services and included individuals with and without formal thought disorders.



This heterogeneity might have contributed to our dimensional characterization of clinical symptoms. Future studies should evaluate the clinical correlates of semantic graph features in specific subpopulations, for example in early psychosis, acute hospitalization, chronic psychosis, and other groups. Sampling methods were not uniform across the entire sample, as further

detailed in the method and supplement. Our primary findings remained consistent when accounting for these deviations statistically. Although we utilized lemmatization of words to merge different inflected forms of lexical units, shared entities that are addressed by different words stayed detached from each other. Incorporating an algorithm to identify co-referents can help

Fig. 2 Correlations between structural and semantic graph features and dimensional clinical characteristics. **a** Heatmap representations of the Spearman's correlation coefficient for structural and semantic graph features and clinical measures in picture description and open-ended narrative tasks across all participants. Significant relationships with uncorrected p values < 0.05 are shaded based on their effect sizes (Spearman's rho). Correlations surviving Bonferroni correction are starred. Bar plots of correlation coefficients per clinical dimension are available in Supplementary Fig. 2. **b** Network representation of significant relationships between graph features and clinical measures. Multi-collinearities were separately handled for structural and semantic graph features by stepwise comparison of variance inflation factors and feature exclusion. Multiple comparisons were accounted for using Bonferroni correction. S_AP static action-predication graph feature, D_AP dynamic action-predication graph feature, S_SEQ static sequential graph feature, D_SEQ dynamic sequential graph feature, NN number of nodes, NE number of edges, diameter graph diameter, ASPL average shortest path length, AWD average weighted degree, density graph density, LSCC size of largest strongly connected component, LSCCZ z-score of LSCC compared to 1000 random graphs, ASPLZ z-score of ASPL compared to 1000 random graphs. More details on graph features are available in Supplementary Table 3.

better represent the semantic structure of discourse. Moreover, devising more sophisticated methods sensitive to the different senses of each word will be able to enhance the performance of similar models in future studies. We have limited our semantic model to the relations between actors and undergoers. However, semantic theories have identified a variety of more differentiated semantic roles such as experiencer, instrument, source, and goal that can be used to produce more fine-grained semantic representations of speech.

The ultimate goal of computational speech measures in the context of psychosis is to develop scalable quantitative methods that improve our understanding of the psychosis disease process and improve our ability to deliver the right treatment to the right person at the right time. The development of novel and informative computational methods moves the field closer to these goals. Our work suggests that graphical speech measures based on semantic relationships capture unique and clinically meaningful aspects of psychosis-related speech disturbances. This method was particularly informative when combined with a moving-windows technique and semi-structured tasks. Future efforts may further refine the semantic graph approach by incorporating more differentiated semantic categories for comprehensive characterization of speech in psychosis spectrum disorders and by relating these features to clinical outcomes like relapse risk and treatment response.

METHODS

Data acquisition and clinical assessment

Participants ($N = 205$) were recruited from the Zucker Hillside Hospital inpatient and outpatient services; healthy volunteers were recruited based on prior participation in other studies or through online advertisements. Primary diagnoses of psychotic disorders were established among 81 individuals (36 schizophrenia, 10 schizoaffective disorder, 5 schizophreniform disorder, 18 unspecified psychotic disorder, and 12 mood disorders with psychotic features). 87 had primary diagnoses of non-psychotic conditions, and 37 were healthy volunteers. We were interested in the main effect of psychosis, and so chose to compare PS+ (including schizophrenia spectrum and mood disorders with psychotic features) with PS- participants (healthy volunteers and individuals with non-psychotic disorders). All procedures were approved by the Institutional Review Board and all participants provided informed consent or assent as minors. Each participant provided speech samples in response to three picture description tasks and two open-ended narrative prompts. Open-ended narrative prompts included "Tell me about yourself.", "How have things been going recently?", and "How have you spent your time recently?". For picture description tasks, subjects were prompted to describe scenes with multiple interacting characters and abstract ink blots, as detailed in Supplementary Table 1. Participants were encouraged to talk for at least one minute, but no time limit was imposed on them. The speech samples were collected via three sampling protocols with different ascertainment goals and minor differences in procedures, as detailed in Supplementary Table 1. All assessments and rating scales were completed by the same team of trained research coordinators (SB, LB). Statistically accounting for the protocol differences did not change our primary findings. Additional details on participants and methods are provided in the Supplement.

Clinical measures included the Scale for Assessment of Thought, Language, and communication (TLC) for speech and language

disturbances²², the Scale for Assessment of Negative Symptoms (SANS)²³, and the Brief Psychiatric Rating Scale (BPRS) for overall psychosis symptoms²⁴. Two-factor scores were calculated from the TLC based on the factor model by Peralta et al. speech disorganization (pressure of speech, tangentiality, derailment, incoherence, illogicality, circumstantiality, and loss of goal) and speech poverty (poverty of speech and poverty of content)^{25,26}. Global scores were taken from the SANS for affective flattening, avolition, and asociality/anhedonia domains of negative symptoms. BPRS total scores were used as a measure for general psychopathology, in addition to its four factors describing anxiety/depression, hostility/suspiciousness, thought disturbance and withdrawal/psychomotor retardation²⁴.

Language pre-processing

Utterance boundaries were determined manually based on syntactic completeness and the presence of pauses. Within each utterance, tokens were identified by NLTK word-tokenizer²⁷. Each token was tagged for its part-of-speech (POS) and lemmatized using spaCy modules²⁸. Semantic role labeling (SRL) was performed using transformer-srl (<https://github.com/Riccorl/transformer-srl>), a BERT-based model built as an extension to AllenNLP and pre-trained on CoNLL 2012 dataset derived from the OntoNotes v5.0 corpus²⁹⁻³². Semantic roles in OntoNotes are span-based and follow the PropBank formalism where a predicate is annotated with predicate specific, core, or numbered, arguments such as A0, A1, etc., and adjunct arguments such AM-TMP (temporal), AM-LOC (location), etc., which are shared across all predicates. Although core arguments are predicate specific, A0 typically marks the Proto-agent and A1 (and sometimes A2) mark the Proto-patient. For this study we only used a subset of core arguments: A0, A1 and A2, where the relationships between verbs and their arguments captured the predication relationships, and the relationships between A0 and A1 or A2 captured the action relationships³³.

For structural graphs, we included nouns, pronouns, non-auxiliary verbs, adjectives, and adverbs. Interjections, filled pauses, articles, and conjunctions (e.g., "yes", "um", and "then") were excluded so as to capture tokens that contribute to verbal exchanges. The lemmatized forms of the included tokens were passed to the graph formation step. We chose to lemmatize the tokens to avoid dissociation of different morphological forms of the same lexeme in network representations (e.g., "talk", "talked", and "talking") all merged into the same node labeled as "talk". For action-predication graphs we tagged verbs, actor-arguments, and undergoer-arguments using SRL, and the lemmatized forms of them were passed to the graph formation step further detailed in the Results section.

Statistical analysis

The associations between graph features and psychosis diagnosis as a dichotomous variable were evaluated using the Mann-Whitney U test. Rank biserial correlation coefficients (RBC) were used as measures of effect size³⁴ and the significance level was adjusted for multiple comparisons by Bonferroni correction (adjusted $\alpha = 0.0003$). Graph features were correlated with dimensional clinical measures using Spearman's rank-order test. The significance threshold was adjusted by Bonferroni correction (adjusted $\alpha = 0.0001$). The correlations between graph feature corresponding to each task and clinical measures were plotted as a heatmap.

Variance inflation factor (VIF) comparison was used to identify redundant graph features in order to simplify our analysis and generate a more streamlined set of features for future applications. First, we applied the method to different features within each domain (size, connectedness, and organization) for static and dynamic types separately. Features with the highest VIF were removed successively until all features showed the

VIF of $<5^{18}$. The remaining static and dynamic graph features were integrated for each graph type and the same analyses were conducted. These remaining graph features for structural and semantic graphs were then combined to produce non-redundant graph feature sets for each graph type. A final set of graph features were attained by merging all previously survived features and re-doing VIF analysis over them. The significant relationships between the ultimately survived features and clinical measures were demonstrated as a network for each task.

The potentially confounding effect of different sampling protocols on our results was evaluated by covarying for protocol type in multiple linear regressions predicting clinical measures with graph features of interest (survived VIF and Bonferroni tests). All initial findings remained significant.

All statistical analyses and visualizations were done in python using the Pandas³⁵, NumPy³⁶, Pingouin³⁷, SciPy³⁸, Statsmodels³⁹, Seaborn⁴⁰, igraph¹⁶, and Matplotlib⁴¹ libraries.

DATA AVAILABILITY

The extracted features and clinical ratings for the participants is provided at <https://github.com/STANG-lab/Semantic-Graphs>. De-identified raw transcripts can be made available to interested scientists upon request.

CODE AVAILABILITY

The codes used for language processing, graph formation, and the statistical analyses are available in a public repository at <https://github.com/STANG-lab/Semantic-Graphs>.

Received: 22 March 2022; Accepted: 1 June 2022;

Published online: 05 July 2022

REFERENCES

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.): 87–122 (2013).
- Cohen, A. S. & Elvevag, B. Automated computerized analysis of speech in psychiatric disorders. *Curr. Opin. Psychiatry* **27**, 203–209 (2014).
- Bedi, G. et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophr.* **1**, 1–7 (2015).
- Corcoran, C. M. et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry* **17**, 67–75 (2018).
- Rezaei, N., Walker, E. & Wolff, P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. *npj Schizophr.* **5**, 9 (2019).
- Corcoran, C. M. & Cecchi, G. A. Using language processing and speech analysis for the identification of psychosis and other disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **5**, 770–779 (2020).
- Tang, S. X. et al. Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. *npj Schizophr.* **7**, 25 (2021).
- Palaniyappan, L. et al. Speech structure links the neural and socio-behavioural correlates of psychotic disorders. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **88**, 112–120 (2019).
- Ferrer I Cancho, R. & Solé, R. V. The small world of human language. *Proceedings. Biological sciences* **268**, 2261–2265 (2001).
- Al-Taie, M. Z. A. & Kadry, S. A. Python for graph and network analysis. Springer Cham. 1–36 <https://doi.org/10.1007/978-3-319-53004-8> (2017).
- Mota, N. B. et al. Speech graphs provide a quantitative measure of thought disorder in psychosis. *PLoS ONE* **7**, e34928 (2012).
- Mota, N. B., Furtado, R., Maia, P. P. C., Copelli, M. & Ribeiro, S. Graph analysis of dream reports is especially informative about psychosis. *Sci. Rep.* **4**, 3691 (2014).
- Mota, N. B., Copelli, M. & Ribeiro, S. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *npj Schizophr.* **3**, 18 (2017).
- Kircher, T., Bröhl, H., Meier, F. & Engelen, J. Formal thought disorders: from phenomenology to neurobiology. *Lancet Psychiatry* **5**, 515–526 (2018).
- Van Valin, R. D. *Exploring the Syntax-semantics Interface*. (Cambridge University Press, 2005).
- Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJ. Complex Syst.* **1695**, 1–9 (2006).
- Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–60 (1960).

- James, G., Witten, D., Hastie, T. & Tibshirani, R. *Springer Texts in Statistics* (Springer, 2021).
- Moe, A. M., Breitborde, N. J., Shakeel, M. K., Gallagher, C. J. & Docherty, N. M. Idea density in the life-stories of people with schizophrenia: associations with narrative qualities and psychiatric symptoms. *Schizophr. Res.* **172**, 201–205 (2016).
- Spencer, T. J. et al. Lower speech connectedness linked to incidence of psychosis in people at clinical high risk. *Schizophr. Res.* **228**, 493–501 (2021).
- Malcorra, B. L. C. et al. Low speech connectedness in Alzheimer's disease is associated with poorer semantic memory performance. *J. Alzheimer's Dis.* **82**, 905–912 (2021).
- Andreasen, N. C. & Grove, W. M. Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophr. Bull.* **12**, 348–359 (1986).
- Andreasen, N. C. The scale for the assessment of negative symptoms (SANS): conceptual and theoretical foundations. *Br. J. Psychiatry Suppl.* **7**, 49–58 (1989).
- Overall, J. E. & Gorham, D. R. The brief psychiatric rating scale. *Psychol. Rep.* **10**, 799–812 (1962).
- Peralta, V., Cuesta, M. J. & de Leon, J. Formal thought disorder in schizophrenia: a factor analytic study. *Compr. Psychiatry* **33**, 105–110 (1992).
- Cuesta, M. J. & Peralta, V. Thought disorder in schizophrenia. Testing models through confirmatory factor analysis. *Eur. Arch. Psychiatry Clin. Neurosci.* **249**, 55–61 (1999).
- Bird, S., Klein, E. & Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language toolkit*. (O'Reilly Media, Inc., 2009).
- Honnibal, M., Montani, I., Landeghem, S. & Boyd, A. spaCy: Industrial-strength Natural Language Processing in Python. <https://doi.org/10.5281/zenodo.1212303> (2020).
- Shi, P. & Lin, J. Simple bert models for relation extraction and semantic role labeling. Preprint at <https://arxiv.org/abs/1904.05255> (2019).
- Gardner, M. et al. Allennlp: A deep semantic natural language processing platform. Preprint at <https://aclanthology.org/W18-2501/> (2018).
- Pradhan, S. et al. In *Proc. Seventeenth Conference on Computational Natural Language Learning*. 143–152 (2013).
- Marcus, R. W. E. H. M., Palmer, M., Ramshaw, R. B. S. P. L. & Xue, N. in *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation* (eds. Olive, J., Christianson, C. & McCary, J.). (Springer, 2011).
- Palmer, M., Gildea, D. & Kingsbury, P. The proposition bank: an annotated corpus of semantic roles. *Comput. Linguistics* **31**, 71–106 (2005).
- Kerby, D. S. The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychol.* **3**, 11 (2014). IT. 13.11.
- McKinney, W. In *Proc. 9th Python in Science Conference*. 51–56 (Austin, TX).
- Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
- Vallat, R. Pingouin: statistics in Python. *J. Open Source Softw.* **3**, 1026 (2018).
- Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**, 261–272. <https://doi.org/10.1038/s41592-019-0686-2> (2020).
- Seabold, S. & Perktold, J. In *Proc. 9th Python in Science Conference*. 61 (Austin, TX).
- Waskom, M. L. Seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
- Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

ACKNOWLEDGEMENTS

This project was supported by the Brain and Behavior Research Foundation Young Investigator Award (SXT) and the American Society of Clinical Psychopharmacology Early Career Research Award (SXT). Data for a portion of the participants ($n = 132$) was collected in partnership with, and with financial support from, Winterlight Labs, Inc. We thank the participants for their contributions. We are also grateful to the clinicians and administrators at Zucker Hillside Hospital for their support of our work, including Drs. Michael Birnbaum, Anna Costakis, and Ema Saito. We thank Bill Simpson, Jessica Robin, and Liam Kaufman of Winterlight Labs for their contributions to the ongoing collaborations.

AUTHOR CONTRIBUTIONS

A.H.N. designed the study, carried out the data processing and analyses, and drafted the manuscript. S.X.T., Y.C., K.H., S.C., S.P., D.D.D., and M.Y.L. contributed to the study design and contributed to the manuscript. S.B. and Ms. L.B. collected the data and contributed to the manuscript.

COMPETING INTERESTS

S.X.T. is a paid scientific advisor for Winterlight Labs, Inc., and receives research funding from them. S.X.T. is also a co-founder of North Shore Therapeutics and holds equity to this company. D.D.D. is a paid scientist at Winterlight Labs, Inc. The other authors have no financial conflicts of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41537-022-00263-7>.

Correspondence and requests for materials should be addressed to Amir H. Nikzad or Sunny X. Tang.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022