Semantic Speech Networks Linked to Formal Thought Disorder in Early Psychosis

Caroline R. Nettekoven[*,1,©], **Kelly Diederen**[2], **Oscar Giles**[3], **Helen Duncan**[3], **Iain Stenson**[3], **Julianna Olah**[2], **Toni Gibbs-Dean**[2], **Nigel Collier**[4,©], **Petra E. Vértes**[1,3,©], **Tom J. Spencer**[2], **Sarah E. Morgan**[1,3,5,6], and **Philip McGuire**[2,6]

[1]Department of Psychiatry, School of Clinical Medicine, University of Cambridge, Cambridge, UK; [2]Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [3]The Alan Turing Institute, London, UK; [4]Theoretical and Applied Linguistics, Faculty of Modern and Medieval Languages, University of Cambridge, Cambridge, UK; [5]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK; [6]Joint last authorship.

*To whom correspondence should be addressed; tel: +1-(226)-582-4800; e-mail: crn29@cantab.ac.uk

Background and Hypothesis: Mapping a patient's speech as a network has proved to be a useful way of understanding formal thought disorder in psychosis. However, to date, graph theory tools have not explicitly modelled the semantic content of speech, which is altered in psychosis. *Study Design*: We developed an algorithm, "*netts*," to map the semantic content of speech as a network, then applied *netts* to construct semantic speech networks for a general population sample (*N* = 436), and a clinical sample comprising patients with first episode psychosis (FEP), people at clinical high risk of psychosis (CHR-P), and healthy controls (total *N* = 53). *Study Results*: Semantic speech networks from the general population were more connected than size-matched randomized networks, with fewer and larger connected components, reflecting the nonrandom nature of speech. Networks from FEP patients were smaller than from healthy participants, for a picture description task but not a story recall task. For the former task, FEP networks were also more fragmented than those from controls; showing more connected components, which tended to include fewer nodes on average. CHR-P networks showed fragmentation values in-between FEP patients and controls. A clustering analysis suggested that semantic speech networks captured novel signals not already described by existing NLP measures. Network features were also related to negative symptom scores and scores on the Thought and Language Index, although these relationships did not survive correcting for multiple comparisons. *Conclusions*: Overall, these data suggest that semantic networks can enable deeper phenotyping of formal thought disorder in psychosis. Whilst here we focus on network fragmentation, the semantic speech networks created by *Netts* also contain other, rich information which could be extracted to shed further light on formal thought disorder. We are releasing *Netts* as an open Python package alongside this manuscript.

Key words: natural language processing/formal thought disorder/disorganized speech/graph theory/psychosis

Introduction

One of the great challenges facing psychiatry is mapping the complexity and heterogeneity of mental health conditions, including psychotic disorders. Although psychotic disorders are recognized for their multifaceted and diverse phenotypes, objectively describing and measuring these remains difficult. Generating new approaches for the automated analysis and quantification of psychotic symptoms could not only significantly diversify the tools available to probe this debilitating condition, but also aid disease monitoring and early diagnosis.

A core symptom of psychosis is formal thought disorder (FTD), manifesting as changes in the patient's speech which can appear incoherent and disorganized. FTD has several dimensions, as described by the Thought and Language Index,[1] including poverty of speech and loosening of associations, where the connection between ideas is tenuous or extraneous ideas intrude into the train of thought.[1] Several studies have highlighted the importance of FTD. A large longitudinal study showed loosening of associations in speech appeared early in the disorder, was present in the majority of cases, and did not resolve over 30 years of investigation.[2] More generally, FTD has been related to more severe forms of

psychosis,[3,4] lower quality of life[5] and has been suggested to predict functional outcomes and relapse.[6]

The advent of natural language processing (NLP) techniques has provided new computational approaches to measure disorganized speech and FTD.[7–17] These approaches can be applied automatically by computers and are therefore substantially less time-consuming than manual assessments using the TLI. For example,[13] created networks to measure disorganized speech, in which nodes were represented as words, and edges connected words in the order in which they were spoken. These word-trajectory networks showed patient-control differences in speech and predicted diagnosis 6 months in advance.[11]

However, these networks focus on the syntactic characteristics of speech and largely ignore semantic content. Meanwhile, semantic abnormalities are a key feature of FTD.[18–21] Measuring semantic coherence has been found to predict psychotic disorders,[8,22] but this provides only a few summary measures of semantic speech content. Speech excerpts likely include additional semantic information that could be used to build a deeper understanding of how the semantic content of language is altered in schizophrenia; and provide extra power to monitor outcome or relapse for individual patients. Networks provide a natural way to represent the semantic content of a speech transcript in more detail, building on the idea that "reality is knowable as a set of informational units and relations among them."[23,24] Hence, it is plausible that representing the semantic content of speech as a network could shed fresh light on the nature of speech in psychosis.

We, therefore, developed a novel speech network algorithm that maps the semantic content of transcribed speech, creating *semantic speech networks*. These semantic speech networks provide a natural framework to capture the information conveyed by the speaker by representing the entities (eg, a person, object, or color) the speaker mentions as nodes and the relationships between entities as edges in the network. While we term these networks semantic speech networks, we note that grammar plays a key role in generating these networks: Our novel algorithm uses the grammatical structure of the text along with the semantic content of the words to represent the semantic relations conveyed through grammar and semantics as a network. We can then test whether the properties of these networks and relationships are related to particular properties of FTD, such as speech fragmentation and loosening of associations. Because the semantic speech networks represent how ideas are grammatically connected in the shape of referential and predicative expressions, fragmentation of the semantic speech networks might reflect abnormalities in this connecting mechanism. Here we focus on speech fragmentation, although we note that the networks created by *Netts* could also be used to study other aspects of FTD, for example, anaphoric referential structure over time. We are releasing

the tool as an openly available Python package named **Networks of Transcribed Speech in Python (*netts*)** alongside this article. The tool can be installed from PyPI[25] and used to construct a semantic speech network from a text file with a single command.

In the following, we introduce *netts* and outline the processing steps the algorithm takes to construct a semantic speech network from transcribed speech. We then describe results from applying *netts* to speech transcripts from a general population sample, where we explored the general properties of semantic speech networks and compared them with random networks. We also used the tool to test for group differences between semantic speech networks of first episode psychosis patients, individuals at clinical high risk for psychosis, and healthy controls. Finally, we explore the relationships of the semantic speech network measures with symptom severity, manual Thought and Language Index scores and with other established NLP markers.

Methods

Network Algorithm

To construct semantic speech networks we created *netts*. *Netts* takes as input a speech transcript (eg, *I see a man*) and outputs a semantic speech network, where nodes are entities (eg, *I, man*) and edges are relations between nodes (eg, *see*).

In the following, we describe the *netts* pipeline, illustrated in figure 1.

Preprocessing. *Netts* first cleans the text by expanding common contractions (eg, *I'm* becomes *I am*), and removing interjections (eg, *Mh*, *Uhm*) and any transcription notes (eg, timestamps). Netts then uses the CoreNLP library to perform sentence splitting, tokenization, part of speech tagging, lemmatization, dependency parsing, and coreferencing on the transcript,[26] using the default CoreNLP language model (for more details, see supplementary note 1).

Finding Nodes and Edges. Netts submits each sentence to OpenIE5 to extract relationships between entities in the sentence using OpenIE5 default settings.[27,28] For example, from the sentence I see a man we identify the relation see between the entities I and a man. From these extracted relations, *netts* creates an initial list of network edges.

Next, *netts* uses the previously identified part of speech tags and dependency structure to extract additional edges defined by prepositions or adjectives: For instance, *a man on the picture* contains a preposition edge where the entity *a man* and *the picture* are linked by an edge labeled *on*. An example of an adjective edge is *dark background*, where *dark* and *background* are linked by an implicit *is*.
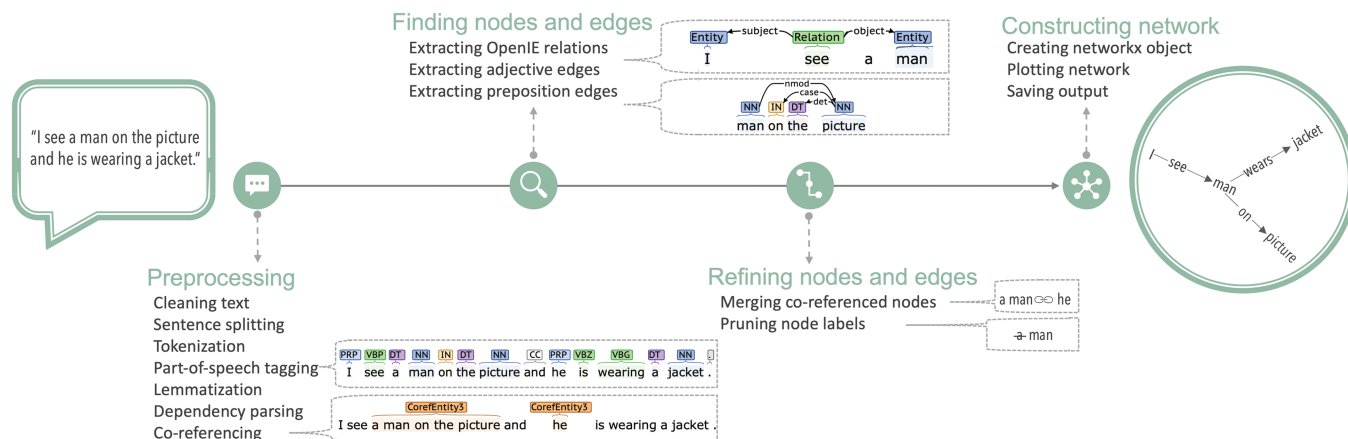
Fig. 1. *Netts* processing pipeline. *Netts* takes as input a speech transcript and outputs a network representing the semantic content of the transcript: a *semantic speech network*. *Netts* combines modern, high performance NLP techniques to preprocess the speech transcript, find nodes and edges, refine these nodes and edges and construct the final semantic speech network.

Refining Nodes and Edges.   After creating the edge list, *netts* uses the previously identified co-referencing structure to merge nodes that refer to the same entity. For example, *a man* might be referred to by the pronoun *he*, or the synonym *the guy*. To ensure every entity is represented by a unique node, nodes referring to the same entity are merged by replacing them with the most representative node label (first mention of the entity that is a noun). In our example, *he* and *the guy* would be replaced by *a man*.

Node labels are then cleaned of superfluous words such as determiners, eg, replacing *a man* with *man*.

Constructing Network.   Finally, *netts* constructs a semantic speech network from the edge list using the networkX library[29] and the network is plotted. The networkX object and the network image are saved for later analysis. The resulting graphs are directed and unweighted.

An example semantic speech network is shown in figure 2.

Data

We applied *netts* to transcribed speech data from two studies: (1) a general population cohort ($N = 436$), collected online, and (2) a clinical dataset including $N = 16$ first episode psychosis (FEP) patients, $N = 13$ healthy control subjects and $N = 24$ subjects at clinical high risk of psychosis (CHR-P), collected in person (total $N = 53$).[30] CHR-P participants met the ultra high risk criteria from the Comprehensive Assessment of At-Risk Mental States (CAARMS[31]). Controls had no history of psychiatric illness. Further recruitment and demographic details are given in supplementary note 2. This dataset has previously been used to investigate patient-control differences in established NLP measures, including the speech graph measures proposed by Mota et al,[12,13] the results of which are reported in Morgan et al[22] and Spencer et al.[32] Here,

we used this clinical dataset to investigate differences in our novel semantic speech network measures. We used the established NLP measures solely to contextualize the signal captured by the semantic speech networks.

In both studies, participants described ambiguous pictures from the Thematic Apperception Test (TAT[33]). Participants were presented with 8 pictures from the TAT and instructed to talk about each picture for 1 minute while audio was recorded.

In the clinical dataset, participants additionally performed a story recall task, in which they were read 6 stories from the Discourse Comprehension Test[34] and asked to retell them with as many details as possible. In the following, we focus primarily on the picture description transcripts, which were available for both datasets, then assess whether we see similar clinical group differences in semantic networks with the story recall task.

The positive and negative syndrome scale (PANSS[35]) was used to measure symptom severity. The Thought and Language Index (TLI[1]) was used to assess FTD.

The speech recordings were transcribed manually by a trained assessor, who was blinded to participant group status in the case of the clinical dataset.

All participants were fluent in English and gave written informed consent after receiving a complete description of the study. Ethical approval for the clinical study was obtained from the Institute of Psychiatry Research Ethics Committee, whilst ethical approval for the general population study was obtained from the King's College London Research Ethics Committee.

Speech Measures

Netts Measures.   We constructed a semantic speech network for each transcript using netts. Overall, netts took ~40 seconds processing time per 1 minute speech transcript. We then calculated the number of nodes and edges.
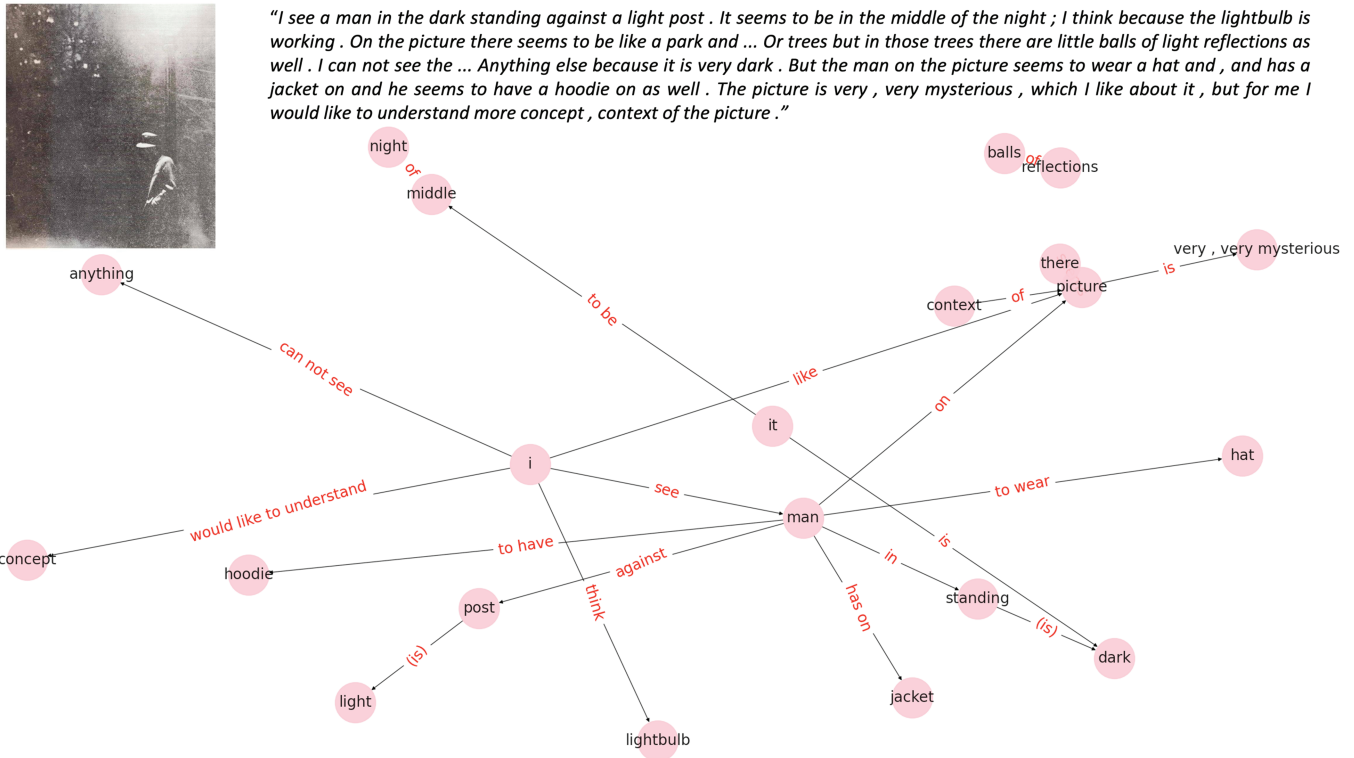
"I see a man in the dark standing against a light post . It seems to be in the middle of the night ; I think because the lightbulb is working . On the picture there seems to be like a park and ... Or trees but in those trees there are little balls of light reflections as well . I can not see the ... Anything else because it is very dark . But the man on the picture seems to wear a hat and , and has a jacket on and he seems to have a hoodie on as well . The picture is very , very mysterious , which I like about it , but for me I would like to understand more concept , context of the picture ."

Fig. 2. Example speech network. Semantic speech networks map the semantic content of transcribed speech engendered by the grammatical structure. Nodes in the network represent entities mentioned by the speaker (eg, *I*, *man*). Edges represent relations between nodes mentioned by the speaker (eg, *see*). Top left figure inset shows the stimulus picture that the participant described. Top right figure inset is the speech transcript.

We also calculated the number of connected components in each network, alongside their mean and median size. Here, a connected component is a subgraph of the original graph where all nodes are connected to each other by a path, ignoring edge directions. We decided to focus our analysis on connected components because when visualizing networks from the general population dataset it was clear the networks often contained multiple connected components. Prior work on syntactic speech graphs has also shown size of connected components can be informative for psychosis.[12]

All *netts* measures were compared to 1000 random Erdős–Rényi networks, matched for number of nodes and edges. The *netts* measures were then normalized to the random networks by z-scoring.

Additional NLP Measures. We compared our semantic speech graph measures with other NLP approaches that previously showed significant group differences in the clinical dataset.[22,32] Specifically, to contextualize the semantic speech network measures, we calculated semantic coherence, tangentiality, on-topic score and ambiguous pronoun count, and connectivity measures from the syntactic speech graphs proposed by Mota et al.[11,12] In these syntactic speech graphs proposed by Mota et al, words are

nodes and words spoken consecutively are connected by an edge. Supplementary figure 6 illustrates the difference between syntactic speech graphs and semantic speech networks derived from the same speech transcript: While in syntactic speech graphs[11,12] each node is a word (eg, *I*, *see*, *a*, *man*) and words spoken in succession are connected, semantic speech networks represent the content of speech where nodes are entities (eg, *I*, *man*) and edges are relations between nodes (eg, *see*, *wears*). Connectivity of the syntactic speech graphs is measured by calculating the total number of nodes in the largest connected component (LCC) and the largest strongly connected component (LSC)[11,12] (for a detailed description of the syntactic graph measures as well as the other additional NLP measures, see supplementary note 3).

Statistical Analysis

All measures were calculated for each transcript from each participant. To obtain a single value per participant, we calculated the mean average across the eight TAT picture descriptions.

We assessed the normality of the measures using the Shapiro-Wilk test. Because some measures were not normally distributed and to mitigate the presence of potential outliers, we used the Mann-Whitney U-test to test for group differences (see supplementary table 1).

To explore the relationship between the *netts* measures and other speech measures, we calculated Pearson's correlation coefficients between all measures. The resulting correlation matrix was clustered using the popular Louvain method for community detection.[36]

Finally, we investigated the relationships between the *netts* measures and symptom and TLI scores, using linear regression, controlling for group membership as covariates.

Results

General Public Networks

We first constructed semantic speech networks from the 2861 speech transcripts obtained from 436 members of the general public. 16 of the resulting networks were empty and therefore excluded from further analysis. Our final general public sample, therefore, consisted of 2845 networks. The resulting semantic speech networks consisted of 15.77 ± 5.03 nodes and 15.04 ± 5.61 edges on average; see figure 3.

Speech Networks Had Fewer and Larger Components Than Random Networks

Compared to size-matched random networks, the speech networks had fewer connected components ($t = -28.51$, $P = < .00001$, Mean ± Std: $-0.55 ± 0.41$); see figure 3. These connected components were also significantly larger than the connected components of size-matched random networks (Mean: $t = 23.19$, $P = < .00001$; Median: $t = 35.0$, $P = < .00001$; figure 3).

Networks in the Clinical Dataset

The clinical dataset consisted of 415 transcripts. FEP patients spoke trendwise fewer words and more sentences than controls, resulting in shorter sentences (words: $z = 1.73$, $P = .08$; sentences: $z = -2.39$, $P = .02$, mean sentence length: $z = 2.83$, $P < .01$).

There was no difference in the number of words or sentences between CHR-P participants and either FEP patients or controls (all $P > .12$).
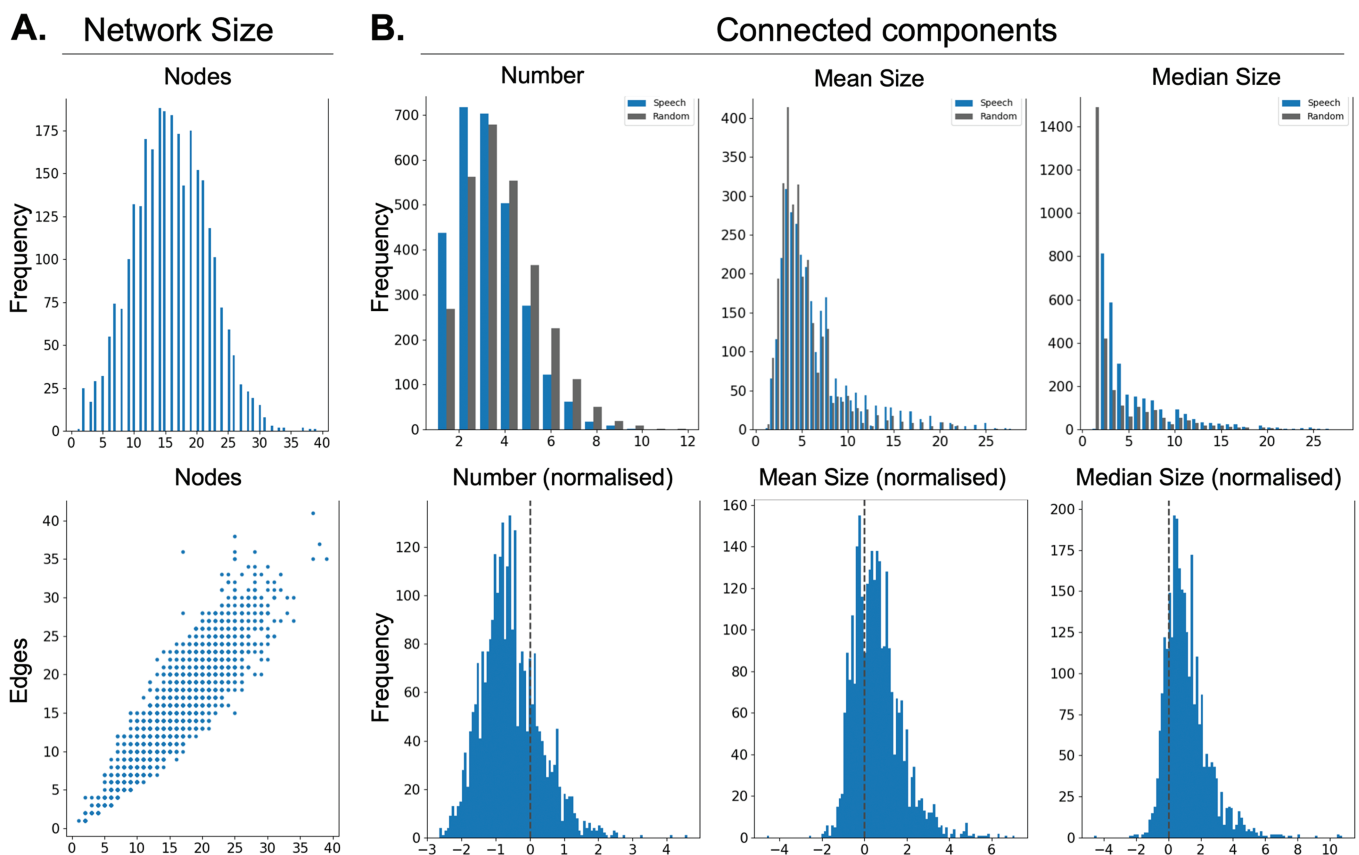

Fig. 3. General public networks. Semantic speech networks differ in their properties from random networks. (A) Histogram for number of nodes and scatter plot showing the relationship between number of nodes and number of edges of semantic speech networks from the general public. Each datapoint in the scatter plot represents one subject. Values were obtained by averaging across network measures from the eight TAT picture descriptions. (B) Top row shows number, mean size, and median size of the connected components in the speech graphs (blue bars) and a randomly chosen subset of the size-matched random graphs (gray bars). Bottom row shows normalized number, mean size, and median size of the connected components in speech graphs.

Patient Networks Were Smaller

Differences in the amount of speech were reflected in the semantic speech networks. FEP networks had fewer nodes ($z = 2.33$, $P = .02$) and fewer edges ($z = 1.97$, $P = .048$) than healthy control networks. Numbers of nodes and edges did not differ between CHR-P participants and either FEP patients or controls (all $P > .13$). Representative FEP patient and control networks are shown in figures 4D and 4E.

Patient Networks Had More and Smaller Components

The normalized number of connected components was higher for FEP patient networks than healthy control networks ($z = -2.00$, $P = .046$; figure 4A). FEP patients also had a trendwise higher number of connected components than CHR-P participants ($z = 1.81$, $P = .07$). There was no significant difference in the normalized number of connected components between CHR-P participants and controls ($z = -0.79$, $P = .43$)).

Mean connected component size was trendwise smaller for FEP patients than healthy controls ($z = 1.73$, $P = .08$; figure 4B) and smaller for FEP patients than CHR-P

participants ($z = -2.01$, $P = .04$). There was no significant difference in mean connected component size between CHR-P participants and healthy controls ($z = 0.13$, $P = .90$).

We also calculated the median connected component size, which should be less influenced than mean size by the 1 or 2 very large connected components speech networks often contain. The median connected component size should hence give a better indication of the size distribution of connected components. Both FEP patients and CHR-P participants had significantly smaller median connected component size compared to controls ($z = 2.70$, $P < .01$ for FEPs, $z = 2.14$, $P = .03$ for CHR-P subjects; figure 4C). There was no difference in median connected component size between CHR-P participants and FEP patients ($z = -1.16$, $P = .25$).

Sensitivity and Reliability Analyses

Our normalization procedure controls for network size. However, to ensure that semantic speech networks capture signal over and above simple differences in the amount of speech, we performed an additional sensitivity analysis testing whether the group differences in
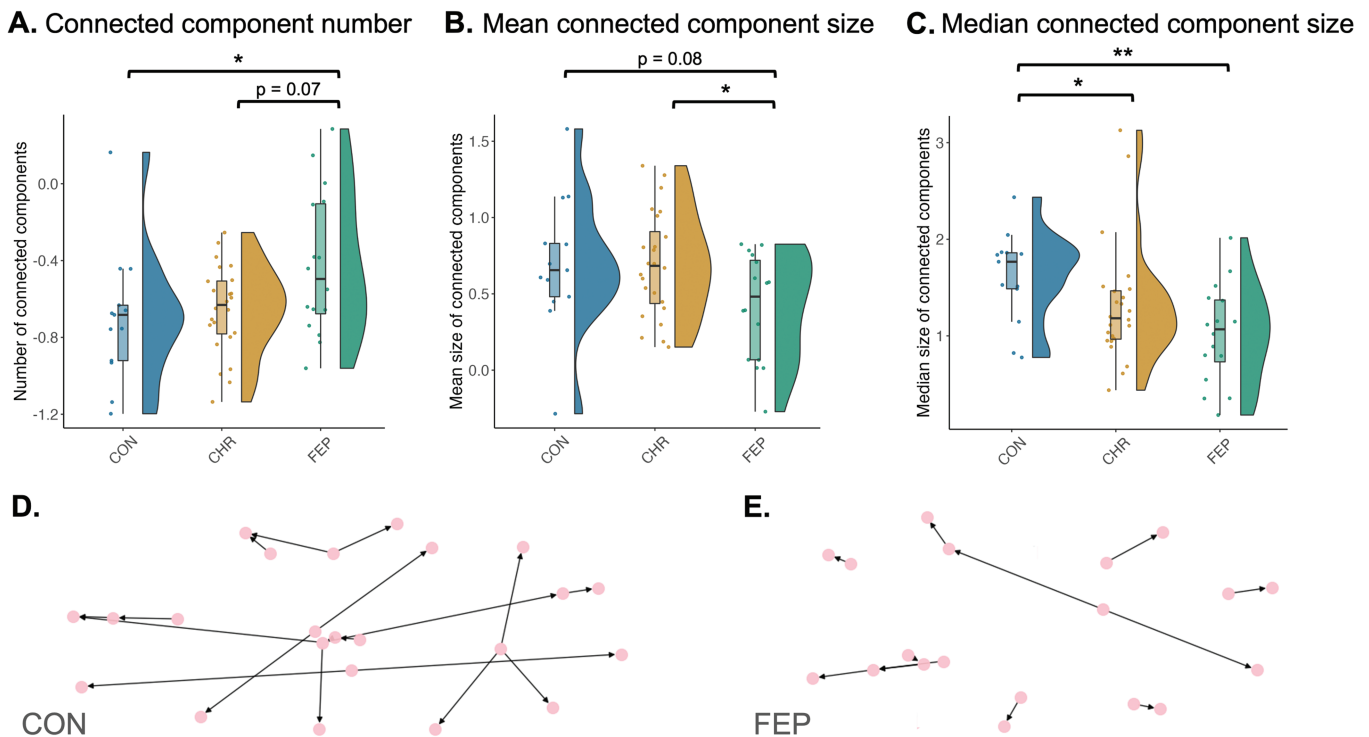

Fig. 4. Clinical networks differ between groups. (A) Number of connected components, (B) mean connected component size, and (C) median connected component size showed differences between the FEP patient (FEP), clinical high risk (CHR-P), and healthy control groups (CON). Network measures shown are normalized to random networks. Each datapoint represents one subject. Values were obtained by averaging across network measures from the eight TAT picture descriptions. * indicates significant p-values at $P < .05$. ** indicates significant p-values at $P < .01$. (D, E) Semantic speech networks of patients had more and smaller connected components than the networks of healthy controls. D shows a typical network from a healthy control participant and E shows a typical network from a first episode psychosis patient. Plots A–C were produced using the Raincloud package[37] and D–E using networkX.[29]

semantic speech network measures remained significant when controlling for number of words. To that end, we constructed a linear model of the network measure as a function of group and included number of words as a covariate.

The FEP patient-control difference in number of connected components remained significant after controlling for number of words (Significant overall regression model: $F(3, 48) = 3.33$, $P = .03$, with significant effect of patient: $\beta = 0.24$, $P = .04$), as did the FEP patient-control difference in the median size of connected components (Significant overall regression model: $F(3, 48) = 4.868$, $P < .01$, with significant effect of patient: $\beta = -0.51$, $P = .01$). The difference in median size of connected components between CHR-P participants and healthy controls did not survive controlling for number of words (no significant effect of CHR-P participant: $\beta = -0.29$, $P = .12$). There was no significant group difference between FEP, CON and CHR-P networks in mean size of connected components after controlling for number of words ($F(3, 48) = 1.97$, $P = .13$).

We also performed additional sensitivity analyses controlling for mean sentence length which showed a similar pattern to the analyses controlling for number of words and can be found in supplementary note 6.

Since these NLP measures are novel and clinical implementation hinges on appropriate reliability, we probed the reliability of our network measures within subjects. For this, we calculated the intraclass correlation coefficient (ICC) of the main variables of interest across all subjects in the clinical dataset, excluding 1 participant who had no speech excerpt for TAT picture 8. Both number of connected components and mean size of connected components had a significant intraclass correlation (CC Number: ICC = 0.14, $F(50,350) = 2.29$; $P < .00001$; CC Mean Size: ICC = 0.08, $F(50,350) = 1.69$; $P < .01$), while the median number of connected components showed a trendwise significant intraclass correlation of ICC = 0.04, $F(50,350) = 1.37$; $P = .058$. These results indicate that the semantic speech network measures assessed were relatively consistent within subjects of the clinical sample. We also calculated the ICC for the general public networks and found that all variables of interest had a highly significant intraclass correlation (CC Number: ICC = 0.10, $F(274,1918) = 1.93$; $P < .0001$; CC Mean Size: ICC = 0.07, $F(274,1918) = 1.57$; $P < .0001$; CC Median Size: ICC = 0.12, $F(274,1918) = 2.05$; $P < .0001$).

Semantic Speech Networks May Capture Novel Information

Figure 5 shows a heatmap of the correlations between the semantic speech network measures and other NLP measures, calculated using the clinical dataset. The black lines denote communities detected using the Louvain method.[36]

As expected, the number of nodes and number of edges of the semantic speech networks clustered with the number of words in the transcript. The connected component measures (CC Number, CC Median Size, CC Mean Size) formed their own community and did not cluster with number of words, number of sentences, measures of network size, or any other NLP measures.

Component Size Was Associated With TLI Scores

Supplementary table 2 shows the relationships of the network measures with the PANSS symptom scores and the TLI. Median connected component size was negatively related to TLI Negative scores ($t = -2.02$, $P = .049$, standardized $\beta = -0.29$). The number of connected components and the mean size of connected components predicted PANSS Negative scores (CC Number: $t = -2.55$, $P = .02$, standardized $\beta = 0.54$; CC Mean Size: $t = -2.12$, $P = .046$, standardized $\beta = -0.47$), with participants scoring higher on the PANSS Negative scale having more and smaller connected components. However, none of these relationships survived Bonferroni-correction (significance threshold corrected for 6×3 comparisons: $P = .0028$).

Inspecting the TLI subscales, we observed a relationship between median connected component size and poverty of speech ($t = -2.33$, $P = .02$, standardized $\beta = -0.34$; supplementary table 3). Number of connected components and mean connected component size were related to peculiar logic (CC Number: $t = 2.6$, $P = .01$, standardized $\beta = 0.37$; CC Mean Size: $t = -2.09$, $P = .04$, standardized $\beta = -0.30$). However, none of these relationships survived Bonferroni-correction (significance threshold corrected for 7×3 comparisons: $P = .0024$). After poverty of speech, median component size was most strongly associated with looseness of associations, although not significantly ($t = -1.45$, $P = .15$; standardized $\beta = -0.22$).

Story Networks Show Differences in Size between CHR-P participants and Healthy Controls, But Not in Connected Components

Finally, we assessed the task dependency of our results using speech generated from a story recall task.

Although number of words and number of sentences did not differ between FEP patients and healthy controls (words: $z = 1.46$, $P = .15$, sentences: -1.18, $P = .24$), FEP patients showed reduced sentence length ($z = 2.56$, $P = .01$). FEP patients also showed reduced mean sentence length compared to CHR-P participants ($z = -2.58$, $P < .01$). FEP patients spoke fewer words than CHR-P participants ($z = -2.16$, $P = .03$), but did not differ in number of sentences ($z = 0.7$, $P = .49$). There was no difference in the number of words and sentences or mean sentence length between CHR-P participants and controls (all $P > .37$).
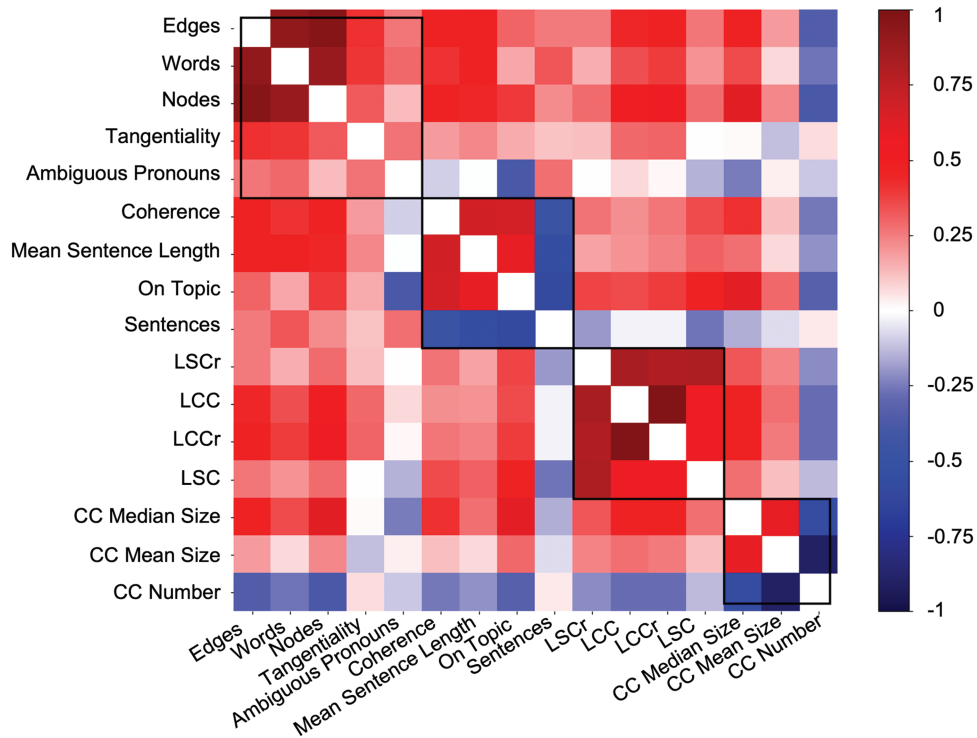
Fig. 5. Clustered speech measures. Semantic speech network measures captured signal complementary to other NLP measures. Shown is a heatmap of Pearson's correlations between semantic speech network measures and NLP measures in the clinical dataset. Black lines indicate communities detected using the Louvain method. The measures used in this analysis were the novel *netts* measures, as well as basic transcript measures and established NLP measures. *Netts* measures were number of connected components (CC Number), mean connected component size (CC Mean Size), and median connected component size (CC Median Size). Basic transcript measures were number of words, number of sentences, and mean sentence length. Established NLP measures included Tangentiality, Ambiguous Pronouns, Semantic Coherence (Coherence), On-Topic Score (On Topic) taken from Morgan et al.[22] Additionally, syntactic network measures based on the method proposed by Mota et al,[11,12] were taken from Morgan et al[22] and included number of nodes in the largest strongly connected component of syntactic networks (LSC), number of nodes in the largest weakly connected component of syntactic networks (LCC), as well as the LSC and LCC normalized to random networks (LSCr, LCCr).[12] Pearson's correlations were calculated from each subject's average NLP value. Values were obtained for each measure by averaging across values calculated from the eight TAT picture descriptions.

This difference between FEP patients and CHR-P participants in the amount of speech was reflected in the semantic speech networks. FEP patient networks consisted of fewer nodes ($z = -2.44$, $P = .01$) and fewer edges ($z = -2.9$, $P < .01$) than CHR-P networks. Network size did not differ between healthy controls and either FEP patients or CHR-P participants (all $P > .1$), although there was a trendwise difference in the number of edges between controls and FEP patients ($z = 1.87$, $P = .06$).

The normalized number of connected components did not differ between groups (all $P > .18$). We also found no significant group differences in the mean and median size of the connected components (all $P > .25$).

Finally, in a post-hoc analysis to compare the characteristics of the story recall and picture description networks, we calculated the ratio of edges to nodes for all networks. The semantic networks from the story re-telling task had significantly more edges per node than those from the picture description task ($t = -11.87$, $P < .0001$).

Discussion

We developed a novel algorithm, and software package, "*netts*" to map the semantic content of a speech transcript as a network, and investigate the nature of semantic speech networks in psychosis.

Netts maps the content of speech as semantic speech networks by (1) identifying the entities described in a speech transcript (the nodes, frequently nouns), (2) extracting the semantic relationships between entities (the edges), and (3) plotting the resulting graph. This approach has several advantages. First, relationships between entities can be extracted from relatively distant parts of the transcript. The tool is also computationally fast, and robust against artifacts typical for transcribed speech, such as interjections (Um, Ah, Err) and word repetitions (I think the the man). The algorithm, therefore, lends itself to the automated construction of speech networks from large datasets.

Having developed the tool, we initially applied it to speech transcripts from the general population. The

networks had fewer and larger connected components than size-matched random Erdős–Rényi networks. This distinction from random networks reflects the nature of semantic speech networks, and the fact that when describing the picture people tend to link different aspects of the picture together in a way that is intrinsically nonrandom. These results imply that *netts* is capable of extracting a substantial number of nodes and edges from a short transcript based on identifying entities and semantic relations in the speech transcript - a non-trivial task for current NLP tools that are primarily trained on text that is very unlike speech (eg, newspaper articles).

In our clinical sample, semantic speech networks from FEP patients had fewer nodes and number of edges than those from healthy controls for the picture description, but only trendwise differences in edges for the story recall task. CHR-P subjects had more nodes and edges than FEP patients for the story recall task, but not the picture description task. These observations are in-line with the group differences in the number of words spoken, and poverty of speech in psychosis.

We also explored whether there were group differences in the number and size of the networks' connected components, motivated by the loosening of associations often observed in FEP. In the picture description task, FEP patient networks were more fragmented than control networks, with more, smaller connected components. The semantic speech networks generated by *netts* map referential elements to entities mentioned by the speaker and derives links between the entities based on grammatical structure. Fragmentation of the networks of patients could be the result of an abnormal connecting mechanism in patient speech compared with controls. The precise nature of this connecting mechanism and how it is altered in patients could be investigated in future research using the semantic speech networks generated by netts.

Relationships between the network connectivity measures and the TLI and PANSS scores were strongest for the negative TLI and negative PANSS scores, but did not survive correction for multiple comparisons. We note that a large majority of participants scored zero for TLI Negative (71%), peculiar logic (86%) and poverty of speech (75%), suggesting that a flooring effect could be obscuring significant relationships. Interestingly, after poverty of speech, median component size was most strongly associated with looseness of associations. This relationship did not reach significance, though here again a flooring effect might be obscuring a significant relationship because 80% of participants in the clinical dataset scored zero for looseness of associations. We note, however, that there were significant group differences in the TLI scores in this sample, specifically in the TLI Total ($P = .0029$, table 1 in Morgan et al[22]) and the TLI Positive scale ($P < .001$; table 1 in Morgan et al[22])

The increased fragmentation in FEP patient networks compared to controls was not observed with the story recall task. This could be a result of the differing cognitive demands of the two tasks and the probing of distinct mental processes. For example, story re-telling relies on working memory, but does not require the participants to spontaneously connect entities with each other, unlike the picture description task. We note that semantic speech graphs from the story recall task also had more edges per node than those from the picture description task, again likely reflecting the more structured nature of the recalled stories compared to spontaneous speech. Hence, the picture description task may be much more suited to bringing about "disconnected" or "fragmented" descriptions than the story recall task. However, we cannot exclude the possibility that the picture description task showing group differences and the story recall task not showing differences could be due to a type I or type II error. In light of this, these results should be interpreted cautiously when generalizing to other tasks and further work is required to fully understand these task dependencies, ideally with larger sample sizes. Nonetheless, our results again suggest the need for careful consideration of the choice of tasks to elicit speech responses from participants.[12,22]

As expected, in a clustering analysis, the number of nodes and edges in the semantic speech networks are clustered with the number of words. The number and size of connected components formed their own community, independent of previously reported NLP measures[22] including syntactic graph measures[11] and semantic coherence. The connected components in semantic speech networks might therefore capture signal beyond the information contained in established NLP measures or relating to the amount of speech.

Overall, *netts* allows us to map the semantic content of speech as a network, opening the door to more detailed analyses of the semantic content of speech, and how this is altered in FTD. Measures of network size and connectivity showed significant group differences between FEP patients, CHR-P subjects and healthy controls and provided complementary information to established NLP measures. We hope that the open availability of *netts* will allow other researchers to explore this new perspective on how disorganized speech is manifest in psychosis.

Limitations

We examined relatively basic topological properties of semantic speech networks, eg, size and number of connected components. Nonetheless, semantic speech networks also include additional information. For example, *netts* records temporal information about the order in which edges were formed in the networks, which could be used to give an indication of network coherence or staying on topic. Furthermore, the edges include context information on epistemic uncertainty (eg, "She does not seem," "I would imagine," "It could be," "You could say") and about whether relationships between entities are negated

(eg, "She is not tall" would result in an edge labeled as negated between "She" and "tall") which could be utilized. The words associated with nodes and edges could also be represented as vectors using word embedding methods, to provide rich node labels. These more advanced network features could provide additional power for deep phenotyping of psychotic disorders.

Preprocessing in *netts* includes sentence splitting based on punctuation. The text excerpts used in this study were transcribed manually, and therefore sentence splitting is based on punctuation by human annotators. We cannot exclude the possibility that fully automated sentence splitting would affect the generated semantic speech networks. *Netts* is currently only available for the English language and would need to be adapted for use with other languages. The term "formal thought disorder" infers alterations in speech are a result of altered thinking. We used the term "formal thought disorder" here to refer to alterations in speech that are evident in patients with psychosis, as it is widely used in this manner, but we note the distinction between thought and language. We note also that the TAT involves subjects speaking in response to a series of pictures. As such, it does not elicit true discourse, as in a free speech paradigm. Furthermore, although our sensitivity analyses controlling for potential confounds (number of words and mean sentence length) concluded that the overall pattern of patient-control differences was robust, we cannot rule out the possibility that group differences in the semantic speech network measures might be driven by other potentially confounding factors.

Finally, the modest sample size of this study, though similar to those used in previous studies is unlikely to fully capture the diverse phenotype of psychosis[38] and the heterogeneity of CHR-P subjects. It is therefore crucial for future studies to test the generalizability and the clinical validity of these results. Generating semantic speech networks for larger clinical samples will also be important to assess whether netts provides additional predictive power beyond existing NLP measures and to fully validate the relationship between semantic speech network measures and FTD, as measured by the TLI for example. We believe the main contribution of *netts* is that it provides a new framework in which semantic relationships can be mapped as a network. Although we show that features extracted from these networks can differentiate patients from controls, it will take time to develop clinical translation. To support further research into the clinical relevance of *netts*, we are releasing *netts* as a free and open-source Python package alongside this article. The *netts* toolbox is available at: https://pypi.org/project/netts/.

Supplementary Material

Supplementary material is available at *Schizophrenia Bulletin*.

Author Contributions

C.R.N. wrote the network algorithm, packaged and documented the *netts* toolbox, conceived, designed, and performed the analysis, and wrote the original draft of the manuscript. O.G., I.S., and H.D. packaged and documented the *netts* toolbox. N.C. contributed to the NLP analysis. P.E.V. contributed to the network analysis. K.D., J.O., T.G.-D., T.J.S., and P.M. collected the data. K.D. and T.J.S. acquired the funding for the general population data collection and conceived the population cohort study design. P.M. acquired the funding for the clinical data collection and conceived the original clinical study design. S.E.M. conceived and designed the network algorithm, conceived and performed the analysis, and acquired the funding for the development of the network algorithm and toolbox. All authors discussed the data interpretation and reviewed and revised the manuscript.

References

1. Liddle PF, Ngan ET, Caissie SL, *et al*. Thought and language index: an instrument for assessing thought and language in schizophrenia. *Br J Psychiatry.* 2002;181:326–330.

2. Pfohl B, Winokur G. The evolution of symptoms in institutionalized hebephrenic/catatonic schizophrenics. *Br J Psychiatry.* 1982;141:567–72. doi:10.1192/bjp.141.6.567.

3. Roche E, Creed L, Macmahon D, Brennan D, Clarke M. The epidemiology and associated phenomenology of formal thought disorder: a systematic review. *Schizophr Bull*. 2015;41(4):951–62. doi:10.1093/schbul/sbu12

4. Oeztuerk OF, Pigoni A, Wenzel J, *et al.* The clinical relevance of formal thought disorder in the early stages of psychosis: results from the PRONIA study. *Eur Arch Psychiatry Clin Neurosci*. 2022; 272:403–13. doi:10.1007/s00406-021-01327-y

5. Tan EJ, Thomas N, Rossell SL. Speech disturbances and quality of life in schizophrenia: differential impacts on functioning and life satisfaction. *Compr Psychiatry.* 2014;55(3):693–698.

6. Wilcox J, Winokur G, Tsuang M. Predictive value of thought disorder in new-onset psychosis. *Compr Psychiatry.* 2012;53(6):674–8. doi:10.1016/j.comppsych.2011.12.002

7. Bedi G, Carrillo F, Cecchi GA, *et al*. Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophr.* 2015;1(1):1–7.

8. Corcoran CM, Carrillo F, Fernández-Slezak D, *et al*. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry.* 2018;17(1):67–75.

9. Elvevåg B, Foltz PW, Weinberger DR, Goldberg TE. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. *Schizophr Res.* 2007; 93(1–3):304–16. doi:10.1016/j.schres.2007.03.001

10. Iter D, Yoon J, Jurafsky D. Automatic detection of incoherent speech for diagnosing schizophrenia. In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, 136–146, New Orleans, LA. Association for Computational Linguistics.

11. Mota NB, Copelli M, Ribeiro S. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *npj Schizophr.* 2017;3(1):110.

12. Mota NB, Furtado R, Maia PP, Copelli M, Ribeiro S. Graph analysis of dream reports is especially informative about psychosis. *Sci Rep.* 2014;4:3691. doi:10.1038/srep03691

13. Mota NB, Vasconcelos NA, Lemos N, *et al*. Speech graphs provide a quantitative measure of thought disorder in psychosis. *PLoS One.* 2012;7(4):1–9.

14. Rezaii N, Walker E, Wolff P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. *npj Schizophr.* 2019;5(1):1–7.

15. Ratana R, Sharifzadeh H, Krishnan J, Pang S. A comprehensive review of computational methods for automatic prediction of schizophrenia with insight into indigenous populations. *Front Psychiatry.* 2019;10(September):1–15.

16. Tang SX, Kriz R, Cho S. Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. *npj Schizophr.* 2021;7(1):1–8.

17. Gutiérrez ED, Corlett PR, Corcoran CM, Cecchi GA. Using automated metaphor identification to aid in detection and prediction of first-episode schizophrenia. *Proc Conf Empir Methods Nat Lang Process.* 2017;2012:2923–2930.

18. Covington MA, He C, Brown C. Schizophrenia and the structure of language: the linguist's view. *Schizophr Res.* 2005;77(1):85–98.

19. Kuperberg GR. Language in schizophrenia part 1: an introduction. *Lang Linguist Compass.* 2010;4(8):576–589.

20. Ditman T, Kuperberg GR. Building coherence: a framework for exploring the breakdown of links across clause boundaries in schizophrenia. *J. Neurolinguistics.* 2010;23(3):254–269.

21. Alonso-Sánchez MF, Ford SD, MacKinley M, Silva A, Limongi R, Palaniyappan L. Progressive changes in descriptive discourse in first episode schizophrenia: a longitudinal computational semantics study. *Schizophrenia.* 2022;8(1):1–9.

22. Morgan SE, Diederen K, Vértes PE, *et al*. Natural Language Processing markers in first episode psychosis and people at clinical high-risk. *Transl Psychiatry.* 2021; 11:630. doi:10.1038/s41398-021-01722-y

23. Zurn P, Bassett DS. Network architectures supporting learnability. *Philos Trans R Soc Lond B Biol Sci,* 2020;375(1796):1–13.

24. Lynn CW, Bassett DS. How humans learn and represent networks. *Proc Natl Acad Sci USA.* 2020;117(47):29407–29415.

25. Python Package Index - PyPI. Python Software Foundation, 2021. http://citebay.com/how-to-cite/python-package-index/

26. Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2014; 55–60. https://stanfordnlp.github.io/CoreNLP/index.html

27. Mausam. Open information extraction systems and downstream applications. New York, NY: International Joint Conference on Artificial Intelligence (IJCAI), 2016.

28. Angeli G, Premkumar MJ, Manning CD. Leveraging linguistic structure for open domain information extraction. *Proc Conf Assoc Comput Linguist Meet.* 2015;1(1):344–354.

29. Hagberg A, Swart P, and Chult DS. *Exploring network structure, dynamics, and function using NetworkX*. Technical report. Los Alamos, NM: Los Alamos National Lab.(LANL); 2008.

30. Demjaha A, Weinstein S, Stahl D, *et al*. Formal thought disorder in people at ultra-high risk of psychosis. *BJPsych Open.* 2017; 3(4):165–170. doi:10.1192/bjpo.bp.116.004408

31. Yung AR, Yuen HP, McGorry PD, *et al*. Mapping the onset of psychosis: the comprehensive assessment of at-risk mental states. *Aust N Z J Psychiatry.* 2005; 39(11–12):964–71. doi:10.1080/j.1440-1614.2005.01714.x

32. Spencer TJ, Thompson B, Oliver D, *et al*. Lower speech connectedness linked to incidence of psychosis in people at clinical high risk. *Schizophr Res.* 2021;228:493–501.

33. Murray HA, Harvard University., and Harvard Psychological Clinic. *Thematic Apperception Test Manual*. Cambridge, MA: Harvard University Press; 1943.

34. Brookshire RH, Nicholas LE. *The Discourse Comprehension Test.* Minneapolis, MN: BRK Publishers; 1993.

35. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987; 13(2):261–76. doi:10.1093/schbul/13.2.261

36. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp.* 2008;2008(10):P10008.

37. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization [version 1; peer review: 2 approved]. *Wellcome Open Res.* 2019;4:1–51.

38. Fusar-Poli P, Cappucciati M, Borgwardt S, *et al*. Heterogeneity of psychosis risk within individuals at clinical high risk: a meta-analytical stratification. *JAMA Psychiatry.* 2016;73(2):113–20. doi:10.1001/jamapsychiatry.2015.2324