Semantic speech networks linked to formal thought disorder in early psychosis

Caroline 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 Vertes [1,3], Tom J Spencer [2], Sarah E Morgan [5,1,3,*], Philip McGuire [2,*]

1. Department of Psychiatry, School of Clinical Medicine, University of Cambridge, UK
2. Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK
3. The Alan Turing Institute, London, NW1 2DB, UK
4. Theoretical and Applied Linguistics, Faculty of Modern and Medieval Languages, University of Cambridge, UK
5. Department of Computer Science and Technology, University of Cambridge, UK

* joint last authorship

UNIVERSITY OF CAMBRIDGE

@carobellum

caroline.nettekoven@ndcn.ox.ac.uk

Background

Semantic content is altered in psychosis

Mapping a patient's speech as a network is useful to understand formal thought disorder in psychosis. However, graph theory tools have not incorporated the semantic content of speech, which is altered in psychosis.

Aim

Can *semantic speech networks* capture features of formal thought disorder in early psychosis?

Method

Netts : A toolbox for creating semantic speech networks

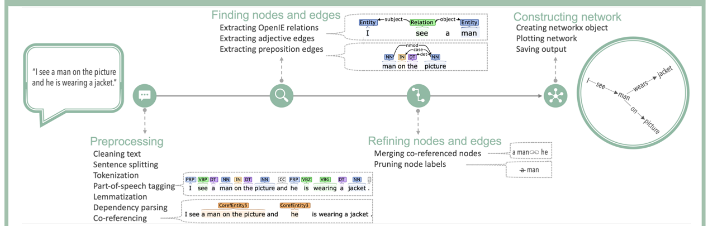


Fig. 1. *Netts* processing pipeline. *Netts* takes as input a speech transcript and outputs a network representing the semantic content of the transcript: a *semantic speech network*. *Netts* combines modern, high performance NLP techniques to preprocess the speech transcript, find nodes and edges, refine these nodes and edges and construct the final semantic speech network.

We developed an algorithm, *"netts"*, to map the semantic content of speech as a network. We applied netts to construct semantic speech networks for a general population sample (N=436) and a clinical sample (N=53). The clinical sample comprised of patients with first episode psychosis (FEP), people at clinical high risk of psychosis (CHR-P), and healthy controls.

Netts is openly available as a free Python Package:
https://pypi.org/project/netts/

Results

Example semantic speech network

Nodes in the network represent entities mentioned by the speaker ("*I*", "*man*"). Edges represent relations between nodes mentioned by the speaker ("*see*").

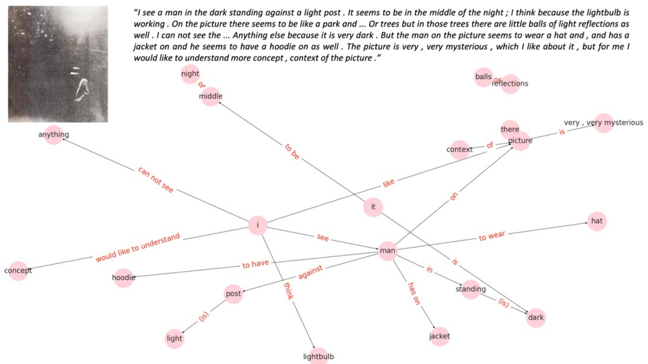


Fig. 2. Example Speech Network. Semantic speech networks map the semantic content of transcribed speech engendered by the grammatical structure. Nodes in the network represent entities mentioned by the speaker (e.g. I, man). Edges represent relations between nodes mentioned by the speaker (e.g. see).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

General Public Networks are non-random

Semantic speech networks from the general population were more connected than size-matched randomised networks, with fewer and larger connected components, reflecting the non-random nature of speech.

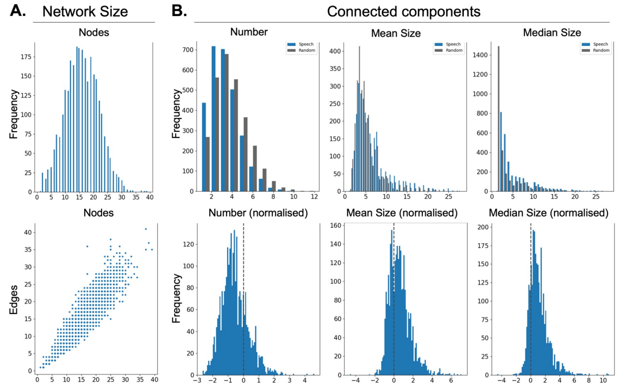


Fig. 3. General Public Networks. Semantic speech networks differ in their properties from random networks. **A.** Histogram for number of nodes and scatter plot showing the relationship between number of nodes and number of edges of semantic speech networks from the general public. Each point in the scatter plot represents a subject. **B.** Top row: 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 (grey bars). Bottom row: normalised number, mean size and median size of the connected components in speech graphs.

Clinical Networks differ between groups

Networks from FEP patients were smaller than from healthy participants. FEP networks were also more fragmented than those from controls; showing more connected components. CHR-P networks showed fragmentation values in-between.

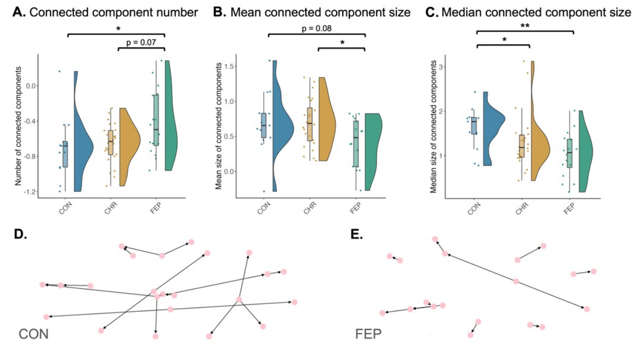
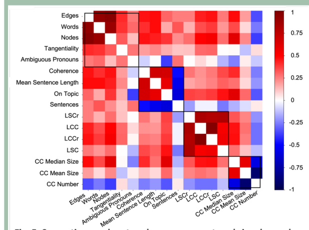


Fig. 4. **A)** Number of connected components**, B)** mean connected component size and **C)** median connected component size showed differences between the FEP patient (FEP), clinical high risk (CHR-P) and healthy control groups (CON). Network measures shown are normalised to random networks. Each point represents one subject. Values were obtained by averaging across network measures from the eight picture descriptions. *indicates significant p-values at $p < 0.05$. **indicates significant p-values at $p < 0.01$ **D** shows a typical network from a healthy control participant and **E** shows a typical network from a first episode psychosis patient.

Semantic speech networks capture novel signal



A clustering analysis suggested that semantic speech networks captured novel signal 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.

Fig. 5. Semantic speech network measures captured signal complementary to other NLP measures. Heatmap of Pearson's correlations between semantic speech network measures and NLP measures in the clinical dataset. Black lines mark communities detected by the Louvain method. Measures used in this analysis were the novel netts measures (CC Number, CC Mean Size, CC Median Size), basic transcript measures and established NLP measures (Tangentiality, Ambiguous Pronouns, Coherence, On-Topic Score and syntactic network measures proposed by Mota et al. 2017: LSC, LCC, LSCr, LCCr).

Conclusion

Semantic speech networks could enable deeper phenotyping of formal thought disorder in psychosis