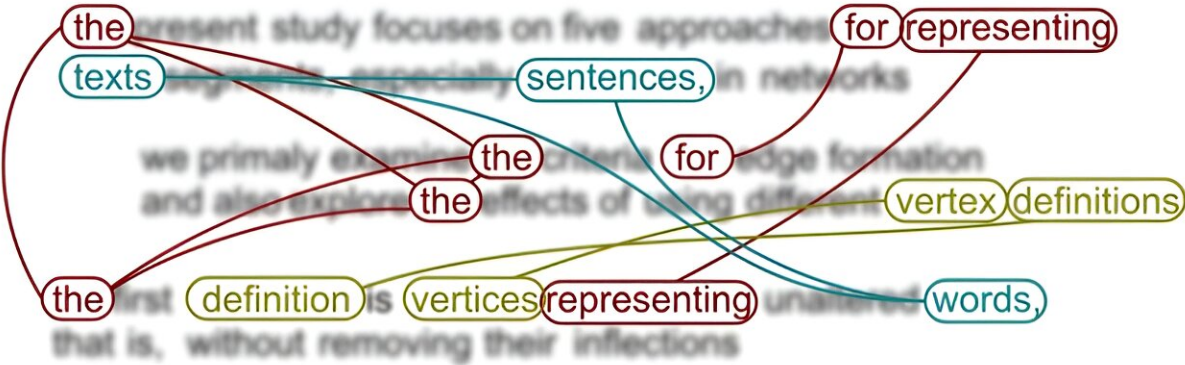# Five ways to model text using networks

August 5 2024, by Clare Sansom

Some examples of how words connect to each other in a text, forming a network. While words such as "vertex" and "vertices" are connected for their shared form, words such as "texts," "sentences" and "words" are connected because of their meanings. Credit: SciencePOD

The explosive growth of AI "chatbots" over the last few years and their ability to generate text that simulates human writing, often very accurately, has focused attention on how text is structured.

One useful way of analyzing text is to think of it as a network, and methods of network analysis that are familiar to mathematicians and computer scientists can be powerful in linguistics.

Network theory can be used in different ways to model the relationship between words in a block of text, linking analytical patterns to coherence and to some more subjective aspects of writing quality.

Davi Alves Oliveira and Hernane Borges de Barros Pereira from the University of Bahia State, Bahia, Brazil have compared five methods of representing sentences as networks, showing that each has value for specific applications. This analysis has now been published in *The European Physical Journal B.*

Their research focuses on a property of text called cohesion, which is essentially what makes a block of text work as a whole, rather than a collection of random sentences. Its cohesion is largely built up from the relationships between words. "Imagine a text as like a map, with words as cities... [and] we connect words based on how they relate to each other," explains Oliveira. "This lets us explore how language users strategically choose words to build a cohesive structure."

Network theory is based around nodes connected by edges that define the relationships between them. Oliveira and Pereira present five different ways of defining these nodes and edges in text, and then use network analysis tools to measure the strength and pattern of the connections.

In some models, individual words are replaced as nodes by lemmas, or base words (so "text" would represent both "texts" and "textual") and/or linking words like "and" or "the" removed; edges might connect consecutive words, or words in the same sentence.

"This [analysis] allows us to see how word choices influence each other and contribute to the overall meaning and structure of the text," adds Oliveira.

Coherence, and also more subjective aspects of writing quality like clarity and flow, can be linked to network patterns. This suggests that the researchers' analyses may have practical applications for language teachers, writers and translators.

More information: Davi Alves Oliveira et al, Modeling texts with networks: comparing five approaches to sentence representation, *The European Physical Journal B* (2024). [DOI: 10.1140/epjb/s10051-024-00717-0](DOI:10.1140/epjb/s10051-024-00717-0)

Provided by SciencePOD