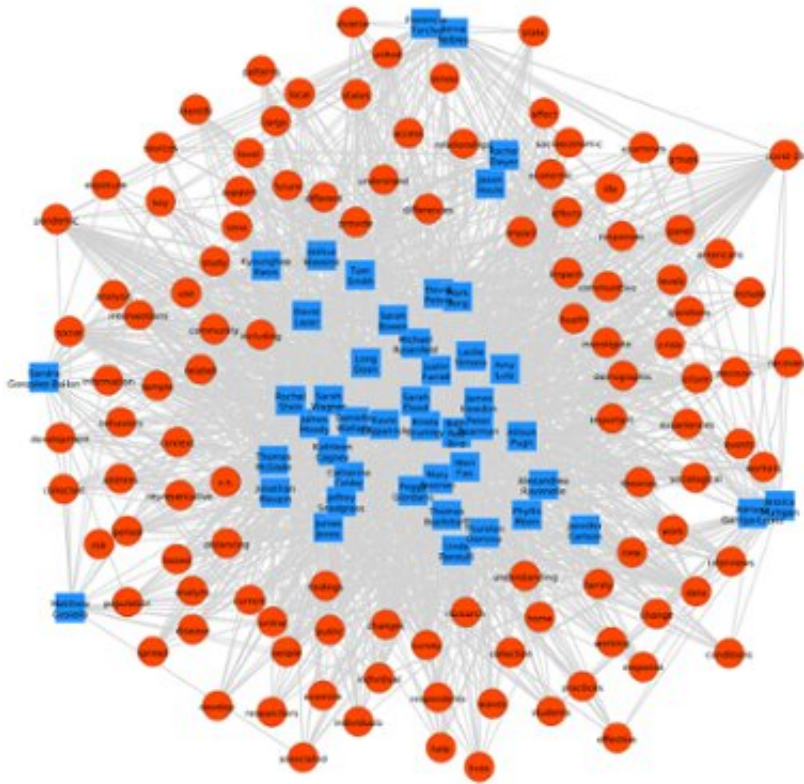


Ethnographic researchers often end up with large amounts of [text](#), especially when they conduct online research. Sociologist John Boy wondered whether, instead of using the usual methods (reading all the texts, encoding them one by one and slowly building up categories and concepts), we should use a mixed methods approach to analyze these enormous quantities of text. Such an approach is enabled by Boy's program. "I use digital technology to analyze texts. What I have developed is a way to make text analysis visually understandable. It is then left to the researcher to add interpretation and meaning."

Code and culture

Boy has been programming since he was a teenager. In the past, he used his own code for his dissertation research. During his postdoc, he developed [software](#) called "Kijkeens", a program that was able to analyze Instagram data and store it in a database. Boy became intrigued by the potential of automated text analysis, but not in the way that the techniques are usually used. "I think most of the computational work is done by people who mainly ask questions based on quantity and causal inference. That's not the kind of background I have. I'm mainly interested in what you can do with software with the purpose of being able to ask qualitative questions."



Grants made by the American National Science Foundation (NSF) to researchers in the fields of Sociology and Cultural Anthropology for projects relating to COVID-19. Credit: Leiden University

Textnets

The aim of Textnets is simple: analyzing collections of texts at a much higher level. Instead of immersing yourself in individual texts, Textnets provides a visualized overview of text documents. Important words or phrases are highlighted. Textnets analyses the documents and breaks them down into words and phrases. If two documents contain the same word or sentence, they are linked. In this way, a web or network is generated that provides an insight into which documents are connected and why.

Visualizing large volumes of text

"Especially when you have a lot of texts, Textnets is useful. If, for example, you have 70,000 tweets, 40,000 online posts on a forum and 20,000 short stories, you cannot read them all and recognize cultural patterns. You need a computer program to support you," says Boy. "The program doesn't do all the work, it doesn't tell you what the connections represent. It only visualizes how the different documents are clustered. Researchers have to interpret the results. Textnets can be seen as a tool that helps you do that, the visualization helps with the interpretation. Not just because it looks nice, but it makes it easier to convey a sense of what is going on."

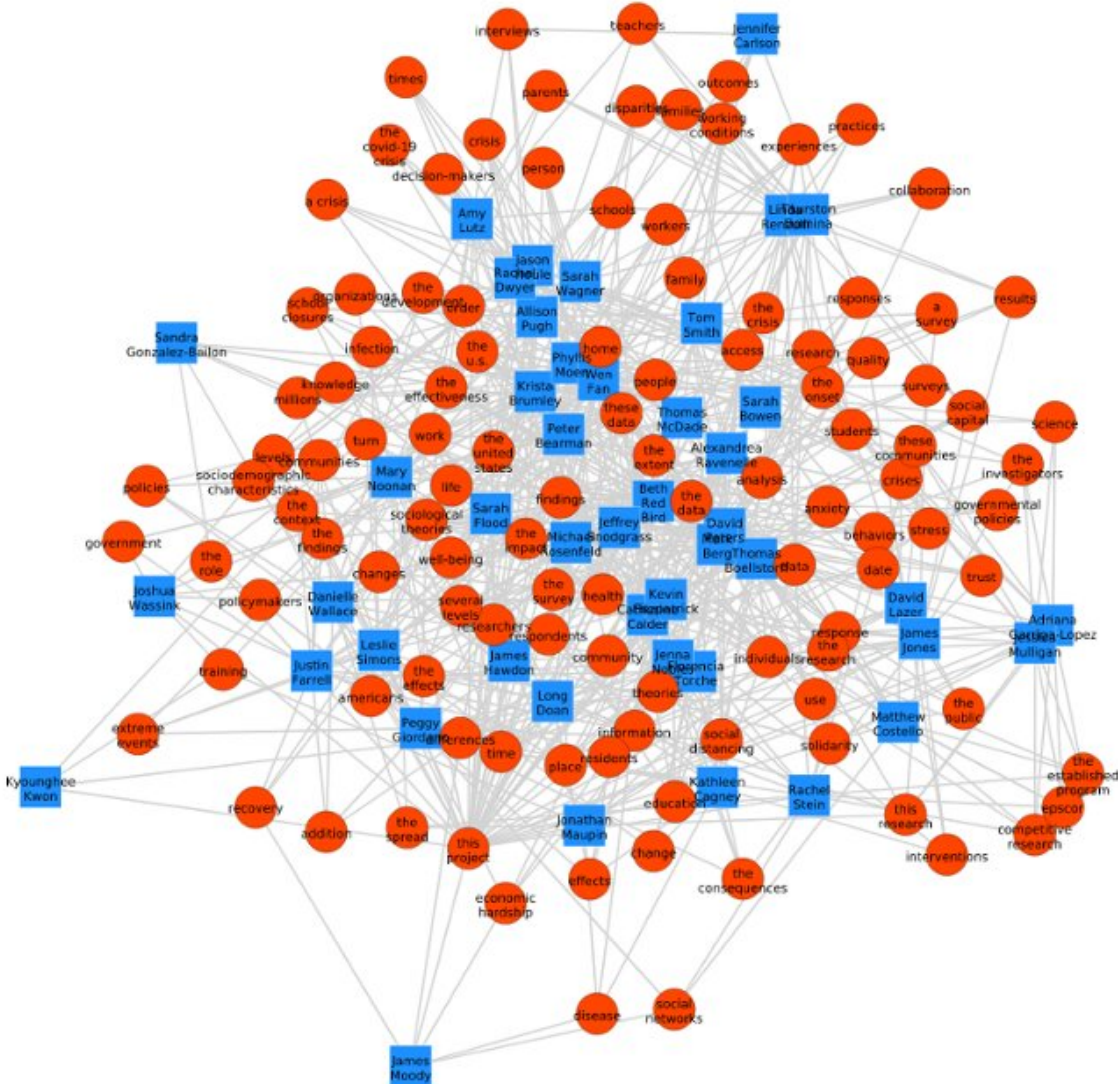


Illustration of using noun phrases instead of individual words for the connections. Credit: Leiden University

Creating connection and meaning

In addition to analyzing documents, Textnets can also link words and phrases that appear in the same document. For example, imagine a text in which someone talks about couch, Netflix and boredom and another

text in which someone talks about couch, children and coffee. The word couch is then linked to Netflix and boredom, but also to children and coffee. The program can then show that the word couch can relate to different themes. "This gives you an insight into the different phrases and expressions people use that bridge different ways of talking about the world," he says. The way you can use the software is twofold. One way is to cluster documents together and the other is to cluster words and see how those words create meaning and connection."

Free software as a way of thinking

The program that Boy has developed, like all other contributions he makes as a programmer, is [open source software](#), or as he prefers to call it, '[free software](#)'. "I see software as a way of thinking. When you consider software to be property, you are actually saying that you own that way of thinking. To me, and to people in the free software movement, that is unethical." Free' should not only be understood in the sense of gratis, but in the sense of freedom. It does not necessarily mean that you do not pay for a program, but it does mean that there are no restrictions on its usage. For Boy, it was clear that he would release his project under the GNU General Public License. This license gives the author the copyright but the programmer can keep the use of the software as open as possible. Boy uses his copyright to keep the software as free as possible. "This means that nobody is allowed to convert my software into an proprietary product. Everything that is built from it must also be 'free'."

More information: Textnets: textnets.readthedocs.io/en/stable/

Provided by Leiden University

Citation: Textnets: Software to make large amounts of text visually comprehensible (2021, March 18) retrieved 19 April 2024 from <https://techxplore.com/news/2021-03-textnets-software-large-amounts-text.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.