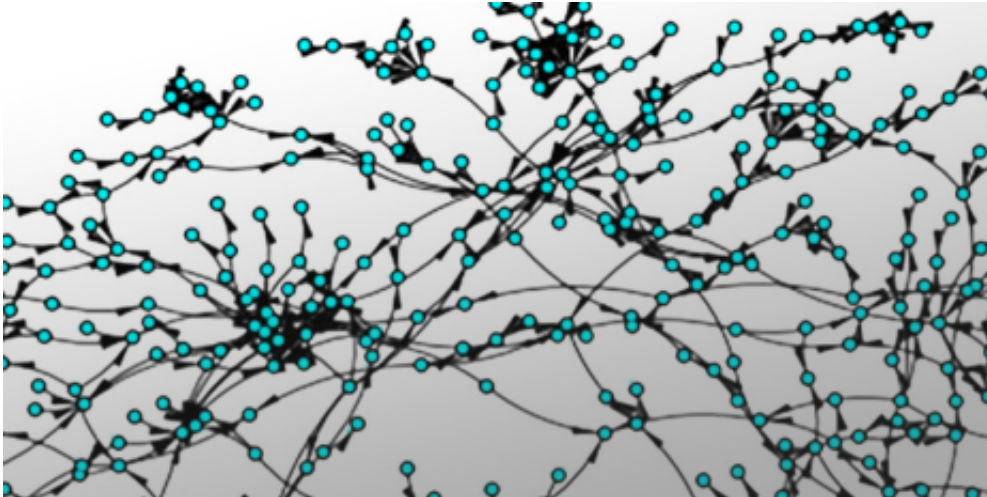# The right order

August 14 2017, by Inken De Wit

How can relevant information be found in a complex network? The results delivered by the usual methods are often too imprecise. Credit: ETH Zurich

Ingo Scholtes from the Chair of Systems Design has developed an analytical method that takes account of the chronological order of connections within networks. This not only makes it possible to more accurately identify links between topics on the internet, but also makes it easier to predict the spread of epidemics, for example.

Why change something when it seems to work well already? Networks have been analysed according to more or less the same pattern for many decades. For example, the more paths leading to a certain article on the web, the higher the relevance attributed to it will be. Accordingly, an article to which many other important articles refer will be listed at the

top of search engine results. The key factor in this is therefore the complex structure of the links or, in other words, which articles are linked to which other articles.

The chronological order is very important

"Current network analysis methods disregard one very important dimension, namely the chronological order of the connections," says Ingo Scholtes, Senior Assistant at the Chair of Systems Design. This means that the algorithms analysing links between articles on the web fail to take account of the chronological order in which users access articles. However, that is precisely what's important, because the order in which we navigate information networks contains valuable information about which articles are most closely linked in terms of topic and are therefore relevant to users.
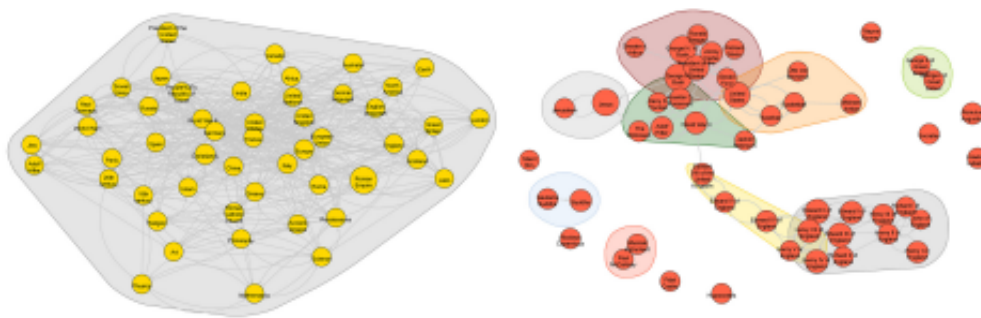
Which Wikipedia articles are particularly relevant when searching for famous people? Unlike the traditional method (left), the new algorithm (right) ranks articles about famous people as most relevant. Credit: ETH Zurich

Take Wikipedia, for example: if a user clicks on a link to "ETH Zurich" in an article about "Albert Einstein", they are more likely to subsequently access an article relating to physics than to earth sciences.

"If we take the temporal dimension into account in the network analysis, we can make much better predictions about user behaviour and, based on this, obtain more-relevant search results and provide better recommendations," says Scholtes.

A model applicable to all networks

Ingo Scholtes' findings have important implications that go beyond internet search algorithms. His analytical method can also be applied to any kind of network. "Networks are abstract models for things that are interconnected," the data science specialist explains. "For example, this also includes transport infrastructures, the global financial system, energy grids, the cells in our bodies and, in addition to virtual networks like Facebook, the connections between people in the real world."

The importance of chronological order becomes clear when Scholtes' new network analysis method is applied to the prediction of epidemics: in order to predict how intensely a wave of influenza will spread and which sections of the population will be affected, it is not enough simply to know which people know each other – that is, who is connected to whom in a network. The sequence of their encounters is also critical to predicting, for example, whether Alice will pass on her flu infection to Bob and also to Carol. Only if Alice meets Bob first and Bob then meets Carol can Carol become infected. If Bob meets Carol first and does not meet Alice until later, Carol will not catch the flu.

Likewise, in an analysis of the London underground network, Scholtes established that the frequently used network models of the "Tube" are often incorrect. For instance, just because a certain underground station has many connections does not mean that the passengers are equally likely to use all of the routes departing from this point. If the station is closer to the outskirts of the city, the passengers will favour connections that take them further into the city centre. Line A is therefore followed

by line B much more often than by lines C or D – even if all of them seem equally likely from a network perspective. "Our results in relation to the London Tube cast a critical light on the application of network-based methods, which are also used, for example, to analyse risks in the Swiss rail network," says Scholtes.

A departure from previous analysis methods

Scholtes emphasises that it is important to take account of both the structure and the temporal relationships when analysing and modelling networks in future: "It is fundamentally important for our society that we use the right methods to analyse networks, as this has direct consequences for matters as wide-ranging as the resilience of critical infrastructures or the spread of epidemics."

More information: The free software is available at Github: github.com/IngoScholtes/pathpy

Provided by ETH Zurich