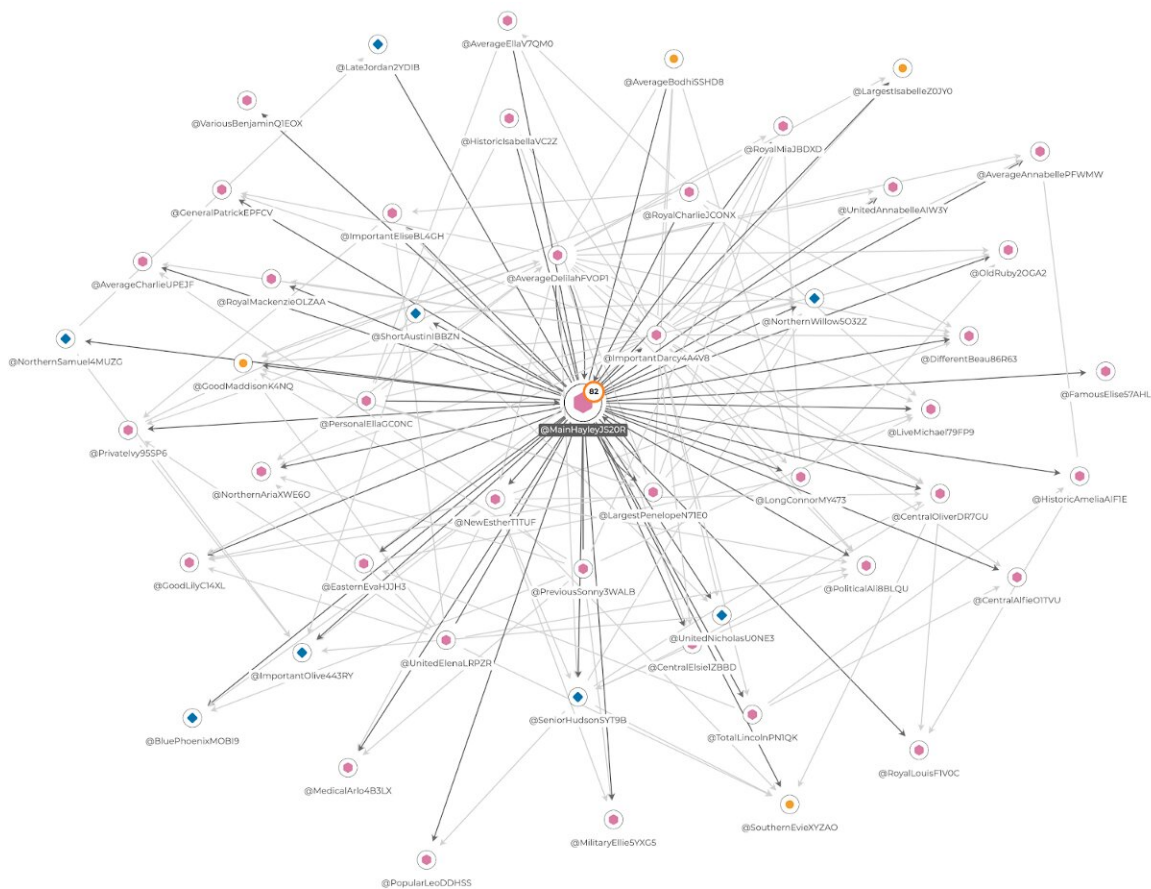


New algorithms help scientists connect data points from multiple sources to solve high risk problems

May 6 2020, by Emily Connell



Credit: StellarGraph

Open source graph machine learning library StellarGraph has today launched a series of new algorithms for network graph analysis to help discover patterns in data, work with larger data sets and speed up performance while reducing memory usage.

StellarGraph is part of Australia's national science agency, CSIRO, through its [data](#) science arm, Data61.

Problems like fraud and cybercrime are highly complex and involve densely connected data from many sources.

One of the challenges data scientists face when dealing with connected data is how to understand relationships between entities, as opposed to looking at data in silos, to provide a much deeper understanding of the problem.

Tim Pitman, Team Leader StellarGraph Library said solving great challenges required broader context than often allowed by simpler algorithms.

"Capturing data as a network graph enables organizations to understand the full context of problems they're trying to solve—whether that be law enforcement, understanding genetic diseases or fraud detection."

The StellarGraph library offers state-of-the-art algorithms for graph machine learning, equipping data scientists and engineers with tools to build, test and experiment with powerful machine learning models on their own network data, allowing them to see patterns and helping to apply their research to solve real world problems across industries.

"We've developed a powerful, intuitive graph machine learning library for data scientists—one that makes the latest research accessible to solve data-driven problems across many industry sectors."

The version 1.0 release by the team at CSIRO's Data61 delivers three new algorithms into the library, supporting graph classification and spatio-temporal data, in addition to a new graph data structure that results in significantly lower memory usage and better performance.

The discovery of patterns and knowledge from spatio-temporal data is increasingly important and has far-reaching implications for many real-world phenomena like traffic forecasting, air quality and potentially even movement and contact tracing of infectious disease—problems suited to deep learning frameworks that can learn from data collected across both space and time.

Testing of the new graph classification algorithms included experimenting with training graph neural networks to predict the chemical properties of molecules, advances which could show promise in enabling data scientists and researchers to locate antiviral molecules to fight infections, like COVID-19.

The broad capability and enhanced performance of the library is the culmination of three years' work to deliver accessible, leading-edge algorithms.

Mr Pitman said, "The new algorithms in this release open up the library to new classes of problems to solve, including fraud detection and road traffic prediction.

"We've also made the library easier to use and worked to optimize performance allowing our users to work with larger data."

StellarGraph has been used to successfully predict Alzheimer's genes , deliver advanced human resources analytics, and detect Bitcoin ransomware, and as part of a Data61 study, the technology is currently being used to predict wheat population traits based on genomic markers

which could result in improved genomic selection strategies to increase grain yield.

The technology can be applied to network datasets found across industry, government and research fields, and exploration has begun in applying StellarGraph to complex fraud, medical imagery and transport datasets.

Alex Collins, Group Leader Investigative Analytics, CSIRO's Data61 said, "The challenge for organizations is to get the most value from their data. Using network graph analytics can open new ways to inform high-risk, high-impact decisions."

StellarGraph is a Python library built in TensorFlow2 and Keras, and is freely available to the open source community on GitHub at [Stellargraph](https://github.com/CSIROData61/stellargraph).

Provided by CSIRO

Citation: New algorithms help scientists connect data points from multiple sources to solve high risk problems (2020, May 6) retrieved 9 April 2024 from <https://techxplore.com/news/2020-05-algorithms-scientists-multiple-sources-high.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.
