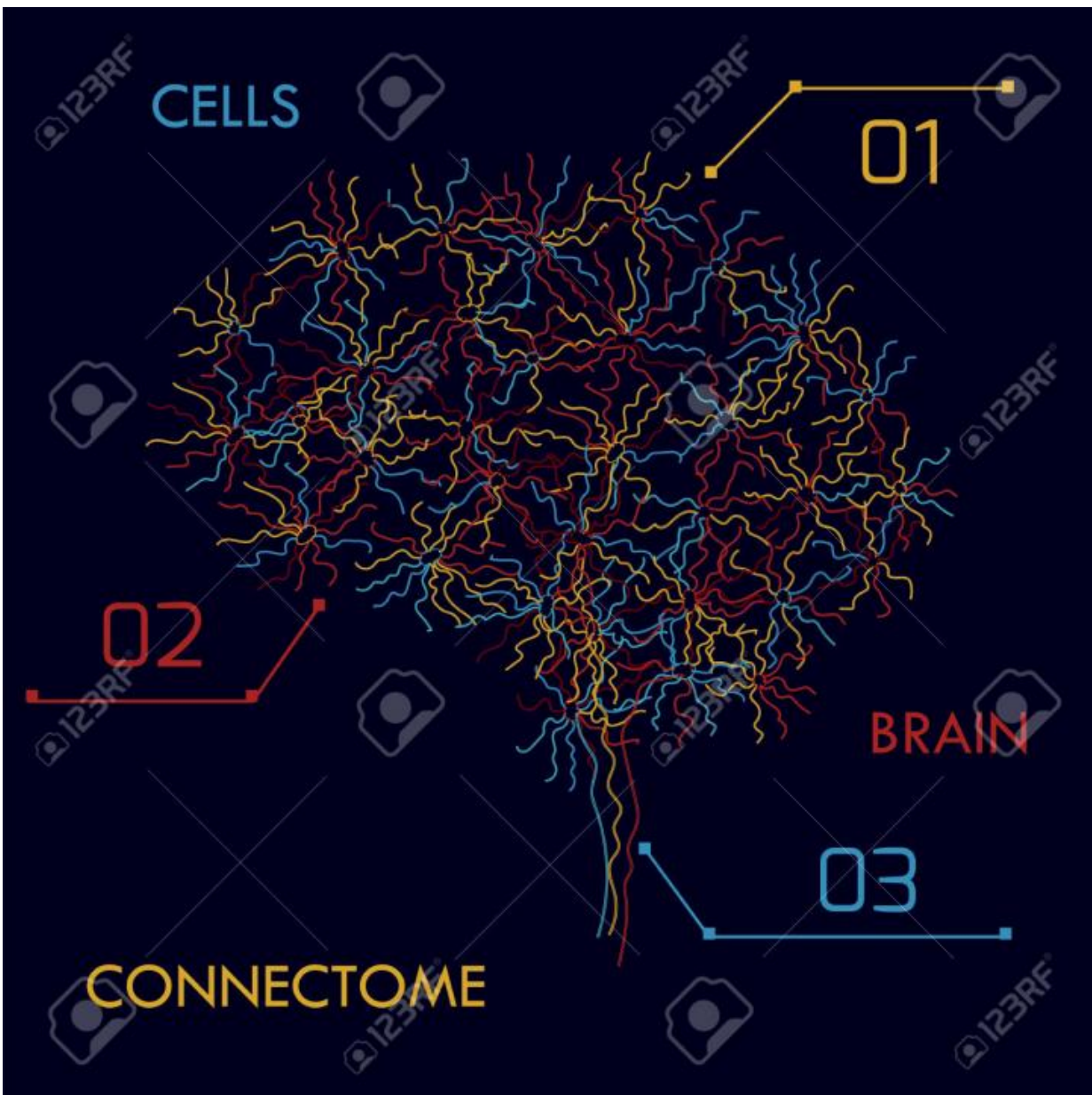


Faster performance evaluation of super-graphs

June 19 2017



Faster performance evaluation of 'super graphs' is shown. Credit: DGIST

Himchan Park and Min-Soo Kim of DGIST have developed TrillionG, a computer model that generates synthetic data for simulating real-world applications that use giant graphs. TrillionG is faster than currently available synthetic graph generators and uses fewer computer resources, such as memory and network bandwidth.

Graphs are used to model data in a way that clarifies the relationships between entries. Each individual entry into a [graph](#) is known as a node, while the relationships, or connections, between these entries are known as edges.

The need for super-graphs to process huge amounts of data is greater than ever before. Google's PageRank algorithm, which ranks websites in Google's search engine, is a good example. It represents the web as a giant graph, with nodes representing each individual web page and edges representing the links from one page to another. Facebook's Apache Giraph graph processor maps all users of the social media site with more than one billion nodes. Their connections with each other reach more than 1 trillion edges.

The performance of giant graph algorithms and systems needs to be tested, but this requires the availability of data. Real data can't be used due to privacy laws. So fabricated, or synthetic, data is required. But synthetic data does not always follow the same relational rules as real data. Also, currently available synthetic graph generators require the use of supercomputers, using several thousand server computers connected via a high speed network due to the exceptionally large amount of data being analyzed.

Park and Kim have proposed a new model for graph generation. It is a compromise between two other currently available models that require significant computational time and memory space. The new [model](#) reuses data that is kept in a very compact form and in a very fast [computer](#) cache memory during graph generation, making it more efficient and effective than existing models.

TrillionG generates more realistic synthetic data than both previous models and can also generate larger graphs. In addition, it can generate similar-sized trillion-edge graphs in a shorter period of time (two hours) using fewer computer resources (10 standard personal computers).

"Through extensive experiments, we have demonstrated that TrillionG outperforms the state-of-the-art graph generators by up to orders of magnitude," write the researchers in their study published in the Proceedings of the 2017 ACM International Conference on Management of Data.

The team expects that TrillionG could generate synthetic graphs the size of the human brain connectome, which consists of 100 trillion connections between neurons, using 240 standard personal computers. IT companies and universities could also use large-scale synthetic graphs as an essential tool for developing and evaluating new graph algorithms and systems.

More information: [DOI: 10.1145/3035918.3064014](https://doi.org/10.1145/3035918.3064014)

Provided by DGIST (Daegu Gyeongbuk Institute of Science and Technology)

Citation: Faster performance evaluation of super-graphs (2017, June 19) retrieved 19 April 2024

from <https://phys.org/news/2017-06-faster-super-graphs.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.