Pangraph – A Graph Network Reading and Writing Library

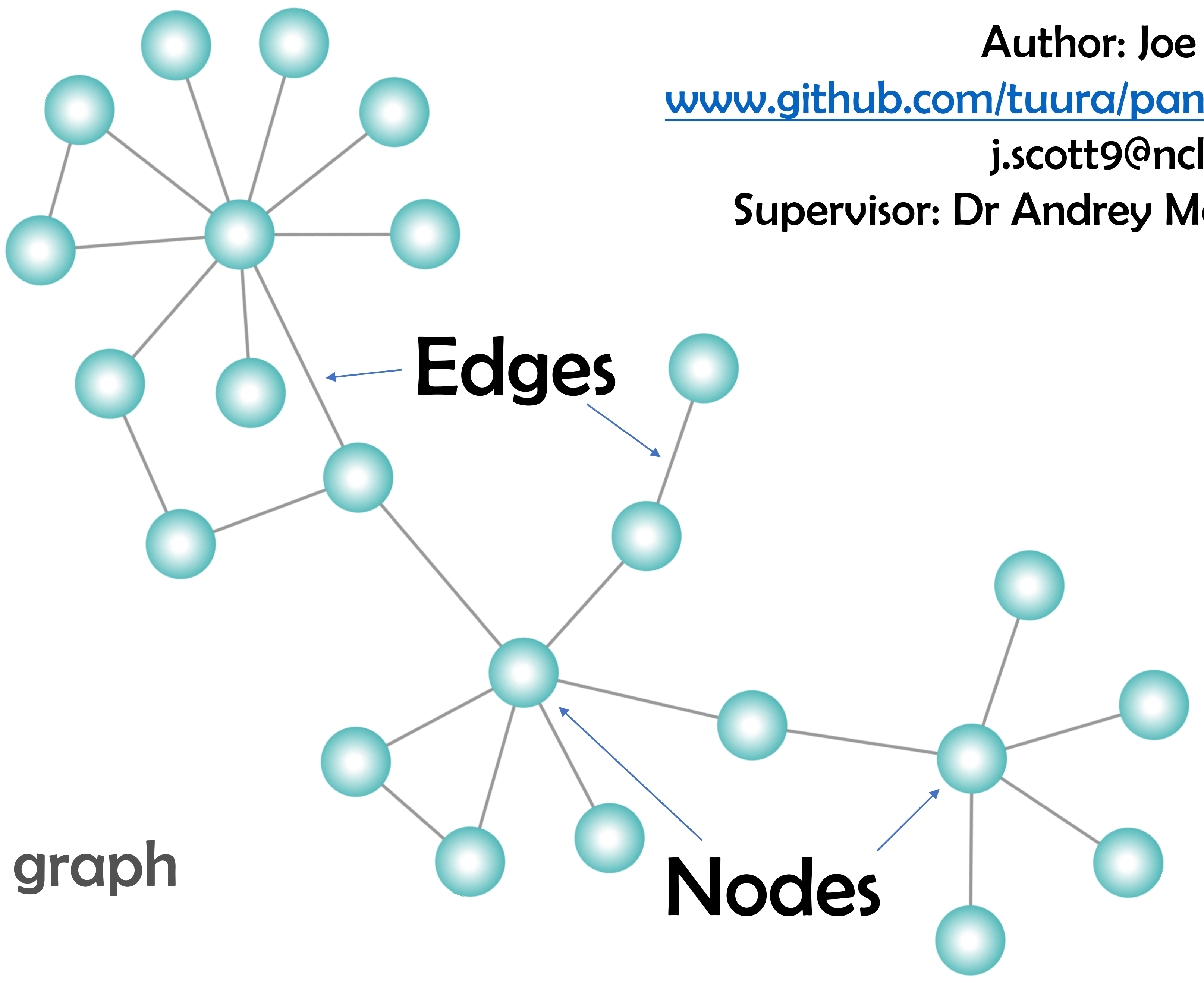
Author: Joe Scott*
www.github.com/tuura/pangraph
j.scott9@ncl.ac.uk
Supervisor: Dr Andrey Mokhov

Graph Networks

A powerful branch of maths used to solve problems in STEM and beyond. Properties of networks allow analysis of complex data sets.

Software Libraries

Libraries allow programmers to share code and expertise. Often the task to be programmed has already been completed and shared.



http://www.cytoscape.org/screenshots.html

Aims

- To implement a scalable representation of graph networks capable of handling millions of edges
- To allow users to trivially work with multiple graph libraries and file formats
- To provide documentation and pragmatic development practise for an ongoing open source project

Applications

This project is not a library for manipulating networks directly. There are already many libraries for this. Pangraph is a meta library to allow data to be more flexible.

Some applications:

- Computers; processor design, internet searching
- Biology; protein interaction, neural modelling
- Logistics; shortest path, fastest route through all stops

Design Considerations

The aims for this project are scalability with large networks and flexibility working with current software. Finding a balance in how the graphs are stored while in use and which information can be discarded. A graph topology known as a tree provides lightweight and fast lookup.

Conclusion

Currently, the project includes full support of one file format and access to three graph libraries and one file format with full conversion between them. Work on other formats is ongoing as planned.

Use Case: Protein Interaction Networks

The networks used to model biological systems are extremely large. In this case the proteins become nodes and edges their interactions. With conventional computers this represents a challenge due to the way the graph is processed. Using the Pangraph library and the included Fantasi[1] project, an electronic circuit can be generated which models each particular network.

Using these circuit designs in a big automatically configurable system (FPGA) the analysis of the graph experiences a thousand factor speed up on typical large industrial networks[1].

References

[1] A. d. G. G. T. J. W. G. L. S. M. J. S. A. Y. A. B. Andrey Mokhov, "Language and Hardware Acceleration Backend for Graph Processing," 2017. [Online]. Available: https://github.com/tuura/papers/blob/master/fdl-2017/graphs-on-fpga.pdf. [Accessed 5 10 2017].

https://www.haskell.org/ http://www.cytoscape.or/