The University of Sydney
School of Computer Science

THE UNIVERSITY OF SYDNEY

COMP5313—Large Scale Networks S1 2022
Assignment 2

This is an **individual** assignment.
This assignment is worth **30%** of the final mark of the course.
Submit your presentation recording and your final report in Canvas.

Project Proposal Due (Optional): Week 8, Thursday the 14th of April, 23:59 AEST
Presentation Recording Due: Week 12, Wednesday the 18th of May, 23:59 AEST
Final Report Due: Week 12, Friday the 20th of May, 23:59 AEST

Task

You can select one of the following tasks.

1. Writing a short research paper exploring a research topic related to the course presenting the related work and an analysis of this topic. Surveyed papers must be published in 2010 or later. Example topics include, but are not limited to, the following.
   - Google's PageRank and beyond (e.g. [1, 2, 3])
   - Social network analysis (e.g., [4, 5, 6])
   - Link prediction in social networks (e.g. [7, 8])
   - Graph embeddings: Theory and practice (e.g. [9, 10, 11])
   - See other related scientific articles https://canvas.sydney.edu.au/courses/39698/pages/related-scientific-articles

2. Programming an algorithm related to the course in C/C++, Java or Python and making a demo of it. Write a report on your findings. For example,
   - Take a well known graph algorithm and study its performance. For example you can take several implementations of the algorithm (i.e. in different libraries NetworkX and iGraph or in different programming languages) and benchmark its performance using various datasets. Also, you can compare this with its Big-O complexity and comment whether the implementations match the expected performance.

3. Analyses a real word graph dataset and identify interesting properties of the structure and the dynamics of the graph. Write a report on your findings. For example,
   - Crawl a graph (cf. Twitter lab), select a large graph of your everyday life, or
   - Take a graph dataset online (e.g., https://canvas.sydney.edu.au/courses/39698/modules/318197)
   - Extract properties of the graph, analyse, characterise, visualise and conclude
   - Make sure you observe something new (not mentioned by someone else) or choose a novel dataset that was not analysed by someone else

Project Proposal (Optional, NOT marked, but feedback will be given)

Submit one (short) paragraph. Project proposal can contain

- The project option you are choosing and an overall summary of the project
- The datasets you are planning to use or if you are planning to collect your own data, methodology of collecting data
- The tools, libraries, and programming languages you are planning to use
- Any involved algorithms and graph metrics

Recorded Presentation(6 marks, late submission will get zero mark)

Whether your choice is to do one of the programming assignments or the literature review assignment, you have to do a 5-minute presentation. You only need to submit a recording of your presentation as an mp4 file in Canvas. Presentation video should contain two parts at the same time that one part shows the speaker and second part shows the slides.

Final Report (24 marks)

Whether your choice is to do one of the programming assignments or the literature review assignment, you have to write a report of ≥ 4 pages (maximum 6 pages). You only need to submit your report in Canvas as a PDF. You can include your codes as appendices of the report. Cover page and appendices will not be counted towards the page limit.

Marking scheme

Both the report and the presentation should include (it could be short):

1. The context of the study/paper (e.g., social networks, programming languages) and novelty (e.g. dataset that is not been analysed in the past)
2. The problem addressed or the question answered by the study/paper including the motivations for it (why this question/problem is relevant)
3. The methodology (steps taken in terms of experiments of analysis to answer this question or to solve this problem / depth of the methods used)
4. The result (what is the result about and why does it address the problem or answer the question)
5. The conclusion (what this implies, what the limitations, any recent developments are and what the next step would be)

All these 5 points above have the same weight in the mark and should be clearly stated in both the presentation and the report.

References

[1] D. F. Gleich, "Pagerank beyond the web," *SIAM Rev.*, vol. 57, no. 3, pp. 321–363, 2015.

[2] I. M. Kloumann, J. Ugander, and J. M. Kleinberg, "Block models and personalized pagerank," *Proc. Natl. Acad. Sci. USA*, vol. 114, no. 1, pp. 33–38, 2017.

[3] P. Lofgren, S. Banerjee, and A. Goel, "Personalized pagerank estimation and search: A bidirectional approach," in *Proc. of WSDM'16*, 2016, pp. 163–172.

[4] P. Rozenshtein, N. Tatti, and A. Gionis, "Inferring the strength of social ties: A community-driven approach," in *Proc. of KDD'17*, 2017, pp. 1017–1025.

[5] J. Ugander, L. Backstrom, C. Marlow, and J. M. Kleinberg, "Structural diversity in social contagion," *Proc. Natl. Acad. Sci. USA*, vol. 109, no. 16, pp. 5962–5966, 2012.

[6] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, "Predicting positive and negative links in online social networks," in *Proc. of WWW'10*, 2010, pp. 641–650.

[7] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," in *Proc. of KDD'11*, 2011, pp. 1100–1108.

[8] M. Al Hasan and M. J. Zaki, "A survey of link prediction in social networks," in *Social network data analytics*. Springer, 2011, pp. 243–275.

[9] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. of KDD'16*, 2016.

[10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[11] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, 2017, pp. 1024–1034.