The University of Sydney
School of Mathematics and Statistics

Student Project

DATA5441: Networks and high-dimensional inference — Semester 1, 2020

Web Page: http://www.maths.usyd.edu.au/u/PG/DATA5441/index.html
Lecturer: Eduardo G. Altmann

Aims (to be evaluated):

(i) to apply ideas and computational methods discussed in class to a new setting.

(ii) to learn a more advanced topic in networks

(iii) to present to your colleagues the key new concept of your project (e.g., from a paper) and your results.

Outputs:

- Oral presentations: $5 + 2$ minutes, maximum 5 slides. You can either record a 5 minute video or present it live via Zoom. Questions will be live.
- Written document (2 to 5 pages, excluding appendices and codes).

Both presentations **must** contain:

- A summary of the **main point** of the project (paper). What is the problem? What is the main result? You can quote and use figures from the papers, but the text should be written in your own words.
- A numerical part that tests, supports, or reproduces one of the results of the paper. This could be a numerical simulation and/or the data analysis of real world networks. Include the code as an appendix.

Each project can be performed by **at most 3 students**, who can work together and have a combined presentation. Each student must focus on a different aspect of the project (e.g., reproducing a different figure) and submit individual outputs.

Dates:

From April 8 to April 24: Think about the project you are interested in. Discuss with the Lecturer and colleagues about your idea.

By Wednesday April 24: choose or propose one project. Post the project number or name in the Edstem discussion forum (first come, first serve).

By Monday May 25: Submit the outputs of your projects (written document and either video or PDF of presentation).

Tuesday 26/5 and Wednesday 27/5: Oral presentation during Lectures.

PROJECT 1: Motifs

Questions

What is a motif? Compare the number of motifs in real networks to those in random graphs.

References

1. "Network Motifs: Simple Building Blocks of Complex Networks", R. Milo et al. Science (2002) Vol. 298, Issue 5594, pp. 824-827
http://www.sciencemag.org/content/298/5594/824.short

2. "Superfamilies of Evolved and Designed Networks", R. Milo et al. Science (2004) Vol. 303, Issue 5663, pp. 1538-1542,
http://science.sciencemag.org/content/303/5663/1538

Possible numerical part

Reproduce as much as possible Fig. 1 of Ref. 2 for one (or a few) directed network.

PROJECT 2: Degree Correlation

Questions

What is degree correlation? How do you quantify it? What is the interpretation in social networks? Is the degree correlation of a given network higher or smaller than a corresponding random network (i.e., a suitable null model)?

References

1. Chapter 7 of Barabasi Book, http://networksciencebook.com/chapter/7

Possible numerical part

Choose a few real world networks, compute the degree correlations in these networks and in null models of these networks (e.g., a Poisson Random Graph with the same number of links and nodes or, even better, a configuration model with same degree sequence). Compare the results.

PROJECT 3: Preferential attachment and fat-tailed distributions

Questions

Understand how modifications of preferential attachment are needed to describe fat-tailed distributions in real networks. What is the degree distribution of the World Wide Web and what mechanisms shape it? How can one modify Barabasi-Albert algorithm to create power-law degree distributions with exponents different from 3?

References

1. Power-Law Distribution of the World Wide Web, Lada A. Adamic, Bernardo A. Huberman, Science 24 Mar 2000: Vol. 287.
   http://science.sciencemag.org/content/287/5461/2115

2. Connectivity of Growing Random Networks, P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. 85, 4629 (2000).
   https://doi.org/10.1103/PhysRevLett.85.4629

Possible numerical part

Implement the algorithm of Ref. [2] and show how the degree distribution of the obtained networks depends on γ.

PROJECT 4: networks with fat-tailed distribution and clustering

Questions

Both preferential attachment and the Barabasi-Albert model generate networks with vanishing clustering coefficient. Can we have scale-free degree distribution and positive clustering coefficient?

References

1. Growing scale-free networks with tunable clustering, Petter Holme and Beom Jun Kim Phys. Rev. E 65, 026107 https://journals.aps.org/pre/abstract/10.1103/PhysRevE.65.026107

Possible numerical part

Implement the network generation model proposed in Ref. 1 and show degree distribution and clustering coefficient.

PROJECT 5: Community detection

Questions

Compare different ideas and algorithms to perform community detection.

References

1. Content of lectures in Week 11.
2. Fortunato, "Community detection in graphs"
   https://doi.org/10.1016/j.physrep.2009.11.002

Possible numerical part

Apply and compare community detection algorithms in real-world networks.

PROJECT 6: Weak Ties

Questions

What are weak ties? Why are they important? How weak and strong ties affect the robusteness of networks?

References

1. Content of lectures in week 5.
2. MS Granovetter, 1973
   http://www.jstor.org/stable/10.2307/2776392
3. Onnela et al. "Structure and tie strengths in mobile communication networks"
   http://www.pnas.org/content/104/18/7332.short

Possible numerical part

In one or more weighted networks, rank links according to their weight. Delete links in order and measure the size of the largest component, reproducing Fig. 3A (or Fig. 3D) of Ref. 3 [Ref. 2 is mostly a historical reference, just to put the project in perspective].

PROJECT 7: Influentials

Questions

What are influentials? What features of the network affect the spreading of innovations?

References

1. Influentials, Networks, and Public Opinion Formation, Duncan J. Watts and Peter Sheridan Dodds Journal of Consumer Research, Vol. 34, 2007, p. 441–458, https://doi.org/10.1086/518527

Possible numerical part

Reproduce some of the results in the paper, or, show how the degree distribution of the network (power law vs. Poisson) affects one model of influence spreading.

PROJECT 8: Explosive Percolation

What is explosive percolation? How can they appear in networks?

References

1. Explosive Percolation in Random Networks, Dimitris Achlioptas, Raissa M. D'Souza, Joel Spencer, Science 13 Mar 2009. https://science.sciencemag.org/content/323/5920/1453

2. Explosive Percolation Is Continuous, Oliver Riordan, Lutz Warnke, Science 15 Jul 2011. https://science.sciencemag.org/content/333/6040/322

Possible numerical part

Reproduce Fig. 1C for the ER and PR cases (possibly other curves as well).

PROJECT 9: Phase transition in ERGMs

What are phase transitions in the ERGMs? Why are they important?

References

1. Content of lectures in week 10.
2. Sec. 15.2.5 “Two-star model” in Newman’s book
3. R. Fischer, J. C. Leitao, T. P. Peixoto, and E. G. Altmann ” Sampling motif-constrained ensembles of networks” , Phys. Rev. Lett. 115, 188701 (2015) https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.115.188701

Possible numerical part

Reproduce Fig. 1 (main panel) from Ref. 3.

PROJECT 10: Epidemic spreading in Networks 1: role of network topology

Investigate how SIR models depend on network topology.

References

1. Content of lectures in week 7.
2. Edstem Discussion, post on “Python packages to simulate epidemic models in networks”

Possible numerical part

Perform SIR simulations in a network and show how epidemic threshold changes depending on (i) degree variability or (ii) clustering coefficient of the network.

PROJECT 11: Epidemic spreading in Networks 2: flattening the curve

Investigate how social distancing can flatten the curve in a SIR models.

References

1. Content of lectures in week 7.
2. Edstem Discussion, post on “Python packages to simulate epidemic models in networks”

Possible numerical part

Perform SIR simulations in a network and show how varying the connectivity changes the shape of the infectious curve (reducing the peak of infection, as being done during the current COVID-19 pandemic).

Project X: your project

Propose a question in line with the constraints above (must relate to the content of the lectures and must contain a numerical part). Projects related to epidemic spreading and COVID-19 are particularly encouraged, they can be different from the epidemic spreading projects proposed above.

Projects related to the research question you are working in your Honours are acceptable only if:

- you indicate that this is the case at the time of the selection;
- the project contains a new numerical analysis, different from your original Honours project (explain what is it);
- you connect your project to what has been discussed in class.