

TOPOLOGY OF COMPLEX NETWORKS: MODELS AND ANALYSIS

CORRIE JACOBIE CARSTENS

(Received 29 August 2016; first published online 5 January 2017)

2010 Mathematics subject classification: primary 05C82; secondary 05C80, 55U15, 68R10, 90B15.

Keywords and phrases: random graphs, complex networks, network models, chain complexes.

There is a large variety of real-world phenomena that can be modelled and analysed as networks. Part of this variety is reflected in the diversity of *network classes* that is used to model these phenomena. However, the sometimes subtle and sometimes obvious differences between network classes are not always taken into account in their analysis. This thesis carefully addresses how to deal with distinct classes of networks in two different contexts.

First, we address the problem of randomising specific classes of networks while fixing their degree sequence. The literature on this topic is extensive but disparate. The research presented here addresses the need for both a more comprehensive overview and a greater variety of random network models by reformulating existing algorithms and introducing new algorithms.

We study two models in detail: the well-known but ill-defined switching model and the recently introduced curveball algorithm. We reformulate both models in terms of their underlying Markov chains. This provides us with a framework which simplifies making adjustments to these models and we introduce several new variations of both algorithms. Furthermore, we use the well-established mathematical theory of Markov chains to establish under which conditions these algorithms sample without bias.

We compare both models to several other random network models. For multi-graphs with self-loops we recommend using the configuration model, for networks without multiple edges or without self-loops we recommend using the curveball algorithm and for directed acyclic networks we recommend using the ordered

Thesis submitted to RMIT University in January 2016; degree awarded on 9 March 2016; senior supervisor Kathy Horadam, associate supervisor Asha Rao.

© 2017 Australian Mathematical Publishing Association Inc. 0004-9727/2017 \$16.00

switching model. Our work on Markov chain models for network randomisation was published in [2, 3].

Second, we extend the theory of motif analysis to the class of directed acyclic networks. Motif analysis is a popular network measure that has been used in many applications, usually for simple directed networks or simple undirected networks. Directed acyclic networks are an important subclass of simple directed networks. We establish experimentally that there is no difference in the motifs detected by existing motif analysis methods and our customised method. As far as motifs are concerned, our experiments show that it is not necessary to take the acyclic nature of networks into account. However, we show that there *are* differences in the detected anti-motifs. Hence, we recommend taking into account the acyclic nature of directed acyclic networks when running motif finding experiments. These results were published in [1].

Intuitively, it is clear that the neighbourhood of a node can provide us with valuable information about the node itself. We are interested in the identification of individuals across dark social networks hidden in recorded networks. We investigate the potential for using neighbourhood attributes alone to match unidentified entities across networks. We test an Enron email database and the CiteSeer citation database, and show that neighbourhoods are highly distinctive in both networks. This work resulted in two publications [5, 6].

Network science is a young and active field of research. Most existing network measures originate in statistical mechanics and focus on statistics of local network properties. Such statistics have proven very useful. However, they do not capture the complete structure of a network. Recently, several papers have been published in which ideas from topological data analysis are used in the analysis of complex networks. We discuss these works in detail and present results from an experimental study on using persistent homology to analyse collaboration networks. We show that persistent homology is a versatile tool for the analysis of several classes of networks. This work was published in [4].

References

- [1] C. J. Carstens, 'Motifs in directed acyclic networks', in: *SITIS 2013, Ninth International Conference on Signal-Image Technology Internet-Based Systems* (eds. K. Yetongnon, A. Dipanda and R. Chbeir) (IEEE Computer Society, Tokyo, 2013), 605–611.
- [2] C. J. Carstens, 'A uniform random graph model for directed acyclic networks and its effect on motif-finding', *J. Complex Networks* **2** (2014), 419–430.
- [3] C. J. Carstens, 'Proof of uniform sampling of binary matrices with fixed row sums and column sums for the fast curveball algorithm', *Phys. Rev. E* **91** (2015), article ID 042812.
- [4] C. J. Carstens and K. J. Horadam, 'Persistent homology of collaboration networks', *Math. Probl. Eng.* **2013** (2013), article ID 815035.
- [5] A. Hecker, C. J. Carstens and K. J. Horadam, 'Neighbourhood distinctiveness: an initial study', in: *Complex Networks VI, Studies in Computational Intelligence*, 597 (Springer International, Switzerland, 2015), 99–110.
- [6] J. Jeffers, K. J. Horadam, C. J. Carstens, A. Rao and S. Boztas, 'Influence neighbourhoods in CiteSeer: a case study', in: *SITIS 2013, Ninth International Conference on Signal-Image Technology Internet-Based Systems* (eds. K. Yetongnon, A. Dipanda and R. Chbeir) (IEEE Computer Society, Tokyo, 2013), 612–618.

CORRIE JACOBIE CARSTENS, Korteweg-de Vries Institute for Mathematics,
University of Amsterdam, Science Park 105, 1098 XG Amsterdam,
The Netherlands
e-mail: c.j.carstens@uva.nl