LEVERAGING AUTOMATIC DIFFERENTIATION TO DETERMINE STATISTICAL SIGNIFICANCE OF NETWORK PATTERNS*

BART DE CLERCK[†‡], FILIP VAN UTTERBEECK[†], AND LUIS E.C. ROCHA [‡§]

Key words. network ensembles, automatic differentiation, null model

MSC codes. 65D25, 68R10, 90B15, 90C35

Abstract. Pattern detection on networks can be realised by considering maximum entropy ensembles of networks with topological constraints. One can consider different empirical patterns such as the presence of specific motifs, a large number of triangles around specific nodes, average neighbour degree etc. [4]. Exponential Random Graph Models (ERGM) [3] can be used to find statistically significant discrepancies between an observed network and a suitable null model. In the specific case of ERGMs with local constraints, it is possible to obtain an analytical expression for the expected value and variance of each pattern of the network that can be written in function of the elements of the adjacency matrix. The general analytical methodology was established by Squartini et al. [5]. Additional network ensemble models have been defined in Cimini et al. [1].

For a specific null model, it can be possible to know the exact probability distribution function and calculate the associated a p-value for a specific pattern. In the majority of the cases, the statistical significance of a pattern is expressed by its z-score. However, in practice, one often resorts to sampling an arbitrarily large number of instances from the ensemble of random networks to estimate the expected value and variance of the considered pattern instead of computing the analytical values.

In this work we combine the analytical methodology with the principle of automatic differentiation [2]. This allows us to circumvent the need for generating a (large) number of random network instances to estimate statistical significance. We provide an implementation in the Julia programming language and a number of applications on real-world networks. We also compare our method with the sampling approach, analyse the impact of the sample size and discuss the computational cost.

REFERENCES

[1] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, *The statistical physics of real-world networks*, Nature Reviews Physics, 1 (2019), pp. 58–71, https://doi.org/10.1038/s42254-018-0002-6.

[2] C. C. Margossian, *A review of automatic differentiation and its efficient implementation*, Wiley interdisciplinary reviews: data mining and knowledge discovery, 9 (2019), p. e1305.

[3] J. Park and M. E. J. Newman, *Statistical mechanics of networks*, Phys. Rev. E, 70 (2004), p. 066117, https://doi.org/10.1103/PhysRevE.70.066117.

[4] T. Squartini, *Maximum-entropy networks : pattern detection, network reconstruction and graph combinatorics*, Springer, Cham, Switzerland, 2017.

[5] T. Squartini and D. Garlaschelli, *Analytical maximum-likelihood method to detect patterns in real networks*, New Journal of Physics, 13 (2011), p. 083001, https://doi.org/10.1088/1367-2630/13/8/083001.

*Submitted to the editors July 25, 2022.

[†]Department of Mathematics, Royal Military Academy, Belgium

[‡]Department of Economics, Ghent University, Belgium (bart.declerck@ugent.be)

[§]Department of Physics and Astronomy, Ghent University, Belgium