Implicit models, latent compression, intrinsic biases, and cheap lunches in community detection in networks

Tiago de Paula Peixoto

(Central European University (CEU), Vienna, Austria)

28.06.2023, 16.00 (c.t.)

Department of Statistics, Ludwigstr. 33, Room 144
and online via Zoom (Link)
(Meeting-ID: 913-2473-4411; Password: StatsCol22)

The task of community detection, which aims to partition a network into clusters of nodes to summarize its large-scale structure, has spawned the development of many competing algorithms with varying objectives. Some community detection methods are inferential, explicitly deriving the clustering objective through a probabilistic generative model, while other methods are descriptive, dividing a network according to an objective motivated by a particular application, making it challenging to compare these methods on the same scale. In this talk, I present a solution to this problem that associates any community detection objective, inferential or descriptive, with its corresponding implicit network generative model. This allows us to compute the description length of a network and its partition under arbitrary objectives, providing a principled measure to compare the performance of different algorithms without the need for "ground truth" labels. Our approach also gives access to instances of the community detection problem that are optimal to any given algorithm, and in this way reveals intrinsic biases in popular descriptive methods, explaining their tendency to overfit. Using our framework, we compare a number of community detection methods on artificial networks, and on a corpus of over 500 structurally diverse empirical networks. We find that more expressive community detection methods exhibit consistently superior compression performance on structured data instances, without having degraded performance on a minority of situations where more specialized algorithms perform optimally. Our results undermine the implications of the "no free lunch" theorem for community detection, both conceptually and in practice, since it is confined to unstructured data instances, unlike relevant community detection problems which are structured by requirement.

Biography:
Tiago de Paula Peixoto is an Associate Professor in the Department of Network and Data Science at the Central European University (CEU), Vienna, Austria. He received his Habilitation in Theoretical Physics at the University of Bremen in 2017. Previously, he had been an Assistant Professor in Applied Mathematics at the University of Bath (2016-2019), External Researcher at the ISI Foundation (2015-2020), and post-doc researcher at the University of Bremen (2011-2016) and Technical University of Darmstadt (2008-2011). The research of his group lies at the interface between Statistical Physics, Complex Systems, Data Science, Applied Mathematics, and Machine Learning, with a special interest in the methodological foundations of Network Science.