Random graph models with fixed constraints : proving the validity of generation methods

supervision : Lionel Tabourier (LIP6 – CNRS / SU)
location : Sorbonne Université – 4 place Jussieu, Paris

stages@complexnetworks.fr, http://complexnetworks.fr

1 Context

Complex network analysis (CNA), is the field that has been developed to describe real-world interactions or relations using graphs. The field has rapidly grown during the last 20 years since it was discovered that graphs which seem unrelated in terms of what they represent (social relationships, connected computers, protein interactions, trophic relations in an ecosystem, etc.), exhibit striking similarities. Therefore, the problems of describing these data and their evolution, understanding processes occurring within these networks call to similar formalizations and thus to the same kind of mathematical or numerical solutions. The need for reference graph models has been identified since the inception of the field. Indeed, specificities of a graph can only be described against a well-founded benchmark. Here, we mean by reference model a synthetic graph with specific structural properties, which aim at mimicking real network characteristics.

Graph models are widely used for **simulation purposes**. Much of the work in CNA has focused on dynamical processes that the network supports : message passing on computer networks, transactions among firms in economic networks, information spreading in social networks, etc. Researchers rely on past observations to infer realistic properties and then simulate the process under study on a reference graph model. Fundamentally, null models are also **essential for describing and understanding** structural properties of a network. Important landmarks in network science are related to observations where real data significantly deviate from reference graph models. For example, the early measurements of the scale-free degree distributions of the Internet structure [FFF99] in the late '90s and the properties of robustness that ensue were surprising in comparison to properties of the Erdős-Rényi model.

For more than a decade, progress made in describing properties of real networks indicates that it is necessary to go beyond standard reference models. For instance, random graphs with a given degree sequence are very popular reference models, however they do not account for the high local density observed in most real world graphs, which plays an essential role in spreading phenomena. The main problem that the CNA community faces is that there are no available method to properly build more elaborate, realistic null models.

2 Internship goals

The purpose of this internship is to prove the validity of the k-edge switching method [TRC11] to generate specific reference models. This method is a Markov Chain Monte-Carlo (MCMC) sampling method, derived from the edge switching method which has been proven to be a uniform sampling method in some very interesting cases [Tay81, EGM$^+$19]. The principle of MCMC methods is to generate a random graph from an existing graph by making slight changes (e.g. edge switches) of its structure iteratively. This procedure can also be described as a random walk in the set of graphs that is targeted. Unfortunately, it is not always possible to do so, as it assumes that any graph of the target set is reachable by the random walk process. In this desired case, the random walk is said to be **ergodic**.

The k-edge switching method is parametrized by k, which represents the number of edges involved in a switch. It is known that a large range of models can be properly generated (that means that the random walk is ergodic) using this method if k is "large enough". However, the method cannot be used practically if k is larger than a few units. The purpose of this internship is to prove in several cases that a low value of k is sufficient to have an ergodic process. The first target clearly identified are random graphs with a given number of triangles : it is believed that $k = 4$ is sufficient to obtain ergodicity, but it remains to be proven. . .Other reference graph models can also be considered during the internship.

This work is essentially theoretical : the intern will mostly work on graph theory proofs. The internship does not demand advanced knowledge in graph theory, but it requires a taste for this field, and some ease with this sort of mathematical reasoning. While not mandatory, some knowledge of Markovian processes is also certainly helpful.

Références

[EGM+19] Péter L Erdős, Catherine Greenhill, Tamás Róbert Mezei, István Miklós, Dániel Soltész, and Lajos Soukup. The mixing time of the swap (switch) markov chains : a unified approach. *arXiv preprint arXiv :1903.06600*, 2019.

[FFF99] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review*, volume 29, pages 251–262. ACM, 1999.

[Tay81] Richard Taylor. Contrained switchings in graphs. In *Combinatorial Mathematics VIII*, pages 314–336. Springer, 1981.

[TRC11] Lionel Tabourier, Camille Roth, and Jean-Philippe Cointet. Generating constrained random graphs using multiple edge switches. *Journal of Experimental Algorithmics (JEA)*, 16 :1–7, 2011.