# Computing moments for arbitrary microcanonical distributions over graphs

**Alexandre René**[1], Michael Schaub[1]

[1]RWTH Aachen University, Aachen, Germany

Graphs are a natural language for expressing interactions within a complex system. Complex systems, however, are rarely static: elements move, change affinities, get created, etc.

Therefore their description typically requires defining a distribution over graphs. Defining a distribution is one thing; if we want to test our theories, or put them into practice, we also need to be able to use the distribution in computations.

From the point of view of statistical physics, one challenge with graphs is that most are not realized in continuous space: constraints on what counts as "in" or "out" of the ensemble can be arbitrarily complex, so that many of the symmetries we are used to exploiting in our calculations no longer apply. Those symmetries that do remain can be difficult to formalize. On the other hand, they might also be a gateway to the study of new, interesting phenoma.

In this work, we develop a computational foundation for studying observables (i.e. statistics) over graph distributions. Our only requirement is a boolean function which implicitly defines which graphs are in the desired ensemble. From there, we construct a Markov chain Monte Carlo (MCMC) sampler which tunes itself to maximize sampling efficiency and thus avoids the exponentially slow mixing time typical of MCMC on graphs [1]. There are two key ideas underlying this approach. First, the focus on computing observables, and more specifically on their expectations: this enables the use of importance sampling to focus on the graph subspaces which contribute the most to the expectation. Second, the partitioning of the graph space into the classes of the simplicial configuration model (SCM) [2]. This allows us to combine the fast SCM sampler with a potential over those classes, which guides the sampler towards those high-contributing subspaces.

Many common models of random graphs, such as the Erdős-Rényi-Gilbert or stochastic block model, are almost entirely unstructured. In contrast, many graphs we might want to study—be they economic, social, ecological—have some sort of local structure, even if it is hard to define. For example, while trade between countries is strongly influenced by geography, other considerations (historical, political) can also create strong links between geographically distant countries. Similarly, while most synapses in the brain connect nearby neurons, long range connections do exist, and are crucial to its function. By grouping graphs according to the prevalence of higher-order structures (namely fully-connected cliques), the SCM classes allow us to use this local structure to optimize sampling without precluding longer connections.

We hope that this work emboldens researchers to study systems with new and more complex topologies, by providing the tools to test theories or put them into practice.

References:

[1] S. Bhamidi, G. Bresler, and A. Sly, 'Mixing time of exponential random graphs', Dec. 11, 2008, arXiv: arXiv:0812.2265. http://arxiv.org/abs/0812.2265

[2] Young, J.-G., Petri, G., Vaccarino, F. Patania, A. 'Construction of and efficient sampling from the simplicial configuration model'. Phys. Rev. E 96, 032312 (2017).