Construction of and efficient sampling from the simplicial configuration model

Jean-Gabriel Young,[1,*] Giovanni Petri,[2] Francesco Vaccarino,[3,2] and Alice Patania[3,2,†]

[1] *Département de Physique, de Génie Physique, et d'Optique, Université Laval, Québec (Québec), Canada*
[2] *Dipartimento di Scienze Matematiche, Politecnico di Torino, Torino, Italy*
[3] *ISI Foundation, Torino, Italy*

(Dated: May 18, 2017)

It has been shown recently that the structure of complex systems is not always correctly represented by networks, due to the presence of many-body interactions. An increasingly popular alternative is to instead encode these interactions explicitly, using simplicial complexes (a generalization of graphs). With this new solution comes the need for principled null models. Drawing inspiration from the network literature, we propose a natural candidate: the simplicial configuration model. The core of our contribution is an efficient and uniform Markov chain Monte Carlo sampler for this model. In a short case study, we demonstrate its usefulness by investigating the relationship between the actual and randomized Betti numbers of a few real systems. This allows us to conclude—based on sound statistical arguments—that the structure of some systems is essentially random, while large-scale organizational principle intervene in others.

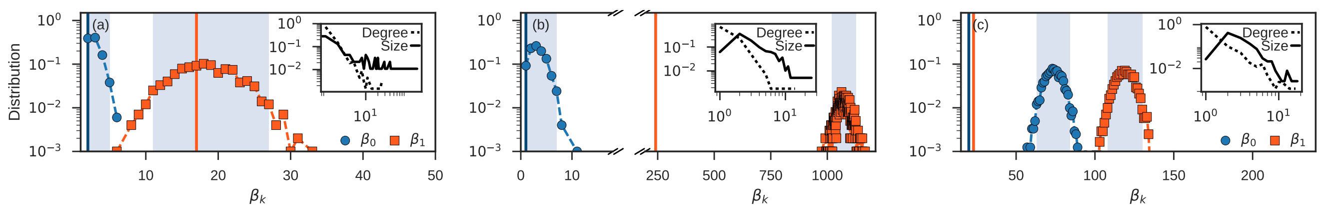

FIG. 1. **The SCM as a null model**. Significance of the topological invariants of reals systems, as quantified by their Betti numbers $\{\beta_k\}$. β_0 counts the number of connected components in the simplicial complex, and β_1 the number of cycles (dimension 1 hole). We investigate 3 datasets; they map the relation between: (a) flower-visiting insects and plants in Kyoto (b) human disease and genes linked by known disorder–gene associations (c) crimes and suspects / witnesses in St.-Louis. Solid symbols show the distributions of the Betti numbers is the randomized version of the real systems (computed from 1000 instances of the model). The Betti numbers of these real systems appear as solid vertical lines, and are (a) $\beta_0 = 2$, $\beta_1 = 17$ (b) $\beta_0 = 1$, $\beta_1 = 241$, (c) $\beta_0 = 20$, $\beta_1 = 23$. The shaded regions contain 95% of the samples. We show the parameters of the model—extracted from real systems—in insets. Our results show that the topology of system (a) is fully explained by the local structure (degree of nodes and size of interactions), while deeper organizational principles are at play for system (b) and (c).

* jean-gabriel.young.1@ulaval.ca
† alice.patania@isi.it