MethodsX 5 (2018) 773–780

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

A unifying framework for fast randomization of ecological networks with fixed (node) degrees

Corrie Jacobien Carstens[a], Annabell Berger[b], Giovanni Strona[c,*]

[a] University of Amsterdam, KdV Institute for Mathematics, Amsterdam, Netherlands
[b] Martin Luther University Halle-Wittenberg, Institute of Computer Science, Halle(Saale), Germany
[c] European Commission Joint Research Centre, Directorate D – Sustainable Resources, Bio-Economy Unit, Ispra (VA), Italy

A B S T R A C T

The Curveball algorithm is an efficient and unbiased procedure for randomizing bipartite networks (or their matrix counterpart) while preserving node degrees. Here we introduce two extensions of the procedure, making it capable to randomize also unimode directed and undirected networks. We provide formal mathematical proofs that the two extensions, as the original Curveball, are fast and unbiased (i.e. they sample uniformly from the universe of possible network configurations).

A R T I C L E I N F O
Keywords: Binary matrices, Co-occurrence, Curveball algorithm, Food web, Null model
Article history: Received 2 May 2018; Accepted 25 June 2018; Available online 5 July 2018

Methods details

Background

A good algorithm to generate random networks with prescribed degree distribution (which is identical to the issue of generating random binary matrices with fixed marginal totals) should have two properties: it should be able to generate any one among all possible networks having a certain node degrees with the same probability, i.e. it should not tend towards the generation of networks having particular structural properties; and it should be able to generate truly random networks fast.

* Corresponding author.
E-mail address: giovanni.strona@ec.europa.eu (G. Strona).

https://doi.org/10.1016/j.mex.2018.06.018

Markov chains, where the randomization takes place in subsequent steps, each involving a small change in the network structure, represent a common solution to this problem. Several network randomization Markov chains have been shown to converge to the uniform distribution on their state space, that is they have been shown able to generate truly random networks with prescribed degree distribution [1–5]. By contrast, most of Markov chains exhibit an important limit, that is it is not clear how many randomization steps they require to ensure that the randomized network is truly random.

The best known Markov chain approach for randomizing network while preserving their degree sequence is the *switching model* (also known as rewiring, switching chain and swapping edges) [6,7,2,8]. It can be applied to different kinds of networks, being able to properly randomize bipartite networks, undirected networks or directed networks with given node degrees, by repeatedly switching the ends of non-adjacent edge pairs (with some additional rules required for the correct randomization of directed networks, [2]). Yet, this method has a fundamental drawback, which is requiring a very large number of switches in order to ensure an unbiased randomization, which grows very rapidly with the size of the network (see, for example, [9]).

A more recent Markov chain approach is the Curveball algorithm [10]. Experimentally, this chain has been shown to mix much faster than the corresponding switching chain [10]. Why the Curveball algorithm mixes faster than the switching model can be understood when thinking of both algorithms as games in which kids trade cards. That is, think of the Curveball algorithm as an algorithm that randomises the binary $n \times m$ bi-adjacency matrix of a bipartite network. Imagine that each row of the adjacency matrix corresponds to a kid, and the 1's in each row correspond to the cards owned by the kid. Then at each step in the Curveball algorithm, two kids are randomly selected, and trade a number of their differing cards. Using this same analogy for the switching model, in each step two cards are randomly selected and traded if firstly they are different and secondly they are owned by different kids (note that various algorithms implementing similar approaches were discovered independently by Verhelst [4]). Intuitively, the Curveball algorithm is clearly a more efficient approach to randomise the card ownership by the kids. More formally, the Curveball algorithm is also based on switches but instead of making one switch, several switches can be made in a single step, which leads to possibly exponentially many networks being reached in a single step, in contrast with the switching model where at most n^4 (the maximum number of possible edge pairs) networks can be reached in a single step.

Designed to randomize only bipartite networks, the Curveball permits the randomization of both species $\times$ locality matrices, and bipartite ecological networks such as host-parasite and plant-pollinator ones. There is, however, an important reason why such design requires an urgent upgrade. Notably, bipartite ecological networks have often been studied separately from food webs, even though all those networks belong to the same broader ecological class of 'resource-consumer' networks [11].

Food webs belong to a different class of networks, that is directed networks. In such networks, nodes cannot be attributed unambiguously to two different classes, since the same node can be simultaneously a consumer and a resource (for example, a predator can be eaten by another predator of a higher trophic level) [12]. A third class of networks is that of undirected networks, which has importance in various fields, such as social sciences and epidemiology, with networks of those kinds being well suited, to represent contacts between persons, and that is also becoming increasingly relevant in the ecological context. In fact, there is a growing interest in the study of co-occurrence networks, that is networks obtained by linking species found together more often than random expectation, and hence considered as potentially interacting (see, for example, [13–15]).

Although some attempts has been recently made to provide measures of network structure applicable to different kinds of ecological networks (see, for example, [16,17]), we are still very far from having a unifying analytical toolbox. Here we take a step further in this direction, by showing how the efficient Curveball algorithm can be extended to work also with unimode directed networks, and undirected networks. Besides providing ecologists with a common procedure to analyze different ecological entities, this constitutes an important advance for network science in general, with the potential of bringing benefit to various disciplines.