Multilayer Artificial Benchmark for Community Detection (mABCD)

ABCD, Community Structure, Multilayer Networks, Random graphs, Synthetic graphs

Introduction

One of the most important features of real-world networks is their community structure. Unfortunately, there are very few datasets with ground-truth communities identified and labelled. This scarcity is particularly evident in multilayer networks [1]: a class of networks that encapsulates multiple types of relations, with separate graphs (layers) where actors are represented by corresponding nodes. To address this gap, we build on the Artificial Benchmark for Community Detection (ABCD) [2] and introduce mABCD, a generator of synthetic multilayer networks.

Operation Principle

The underlying design assumptions of mABCD were to reflect real-world multirelational data. As the exploratory study of such structures showed no consistent trends in interlayer dependencies, the model was designed to give users freedom in modelling them. In consequence, mABCD is governed by two groups of parameters: the first defines the global properties of the network (Tab. 1), and the second characterises layer-specific dependencies (Tab. 2). The subsequent graph generation process depends entirely on them and consists of six steps.

The mABCD generation process comprises the following steps. (**1**) Determine which actors are active (i.e. have a positive degree) in each network layer. (**2**) Generate degree sequences for each layer independently, following a truncated power law. (**3**) Create communities: for each layer, draw partition sizes independently from a power-law distribution and assign nodes using a latent d-dimensional ball into which actors are randomly embedded. Communities are populated by iteratively selecting the closest neighbours from the exterior inward. Reusing the same latent space across layers reflects the common real-world tendency for individuals to share the same acquaintances across different social circles. (**4**) Create edges in each layer independently, based on the established community structure. This includes a "community graph" that connects communities according to the predefined division and a "background graph" that links nodes from different communities while respecting the desired degree distribution. (**5**) Rewire edges to remove multi-edges and self-loops from each layer, using a heuristic designed to preserve the degree and community size distribution. (**6**) Achieve the desired interlayer edge correlations through a series of rewirings across pairs of layers. This phase is performed in small batches of edges using a greedy-like strategy. As it can affect the entire network, the process is carefully controlled to return a locally optimal state of the network.

Model Capabilities

As a result, mABCD can produce multilayer networks with a prescribed community structure according to a highly customisable configuration. It also offers high speed, and smooth transitions between disjoint communities and random graphs with no community structure. Experiments confirm mABCD's convergence and stability. Moreover, owing to its design and Julia implementation, the model outperforms its competition [3] in terms of efficiency (Fig. 1).

As this work introduces a synthetic graph generator, it does not involve personal or sensitive data, and thus we believe it raises no direct ethical concerns.

References

[1] Mikko Kivelä et al. “Multilayer Networks”. In: *Journal of Complex Networks* 2.3 (2014), pp. 203–271.

[2] Bogumił Kamiński et al. “Artificial Benchmark for Community Detection (ABCD) - Fast random graph model with community structure”. In: *Network Science* (2021), pp. 1–26.

[3] Marya Bazzi et al. “A framework for the construction of generative models for mesoscale structure in multilayer networks”. In: *Physical Review Research* 2.2 (2020), p. 023100.

Table 1: Global parameters of `mABCD`.

Parameter	Range	Description
n	$\mathbb{N}$	Number of actors
ℓ	$\mathbb{N}$	Number of layers
$\mathbf{R}$	$[0,1]^{\ell\times\ell}$	Correlation between edges
d	$\mathbb{N}$	Dimension of reference layer

Table 2: Layer-specific parameters of `mABCD`.

Parameter	Range	Description
q_i	$(0,1]$	Fraction of active actors
τ_i	$[-1,1]$	Correlation coefficient between degrees and labels
r_i	$[0,1]$	Correlation strength between communities and the reference layer
γ_i	$(2,3)$	Power-law degree distr. with exponent γ_i
δ_i	$\mathbb{N}$	Minimum degree as least δ_i
Δ_i	$\mathbb{N}\ (1\le\delta_i\le\Delta_i<n)$	Maximum degree at most Δ_i
β_i	$(1,2)$	Power-law community size distr. with exponent β_i
s_i	$\mathbb{N}$	Minimum community size at least s_i
S_i	$\mathbb{N}\ (\delta<s_i\le S_i\le n)$	Maximum community size at most S_i
ξ_i	$(0,1)$	Level of noise

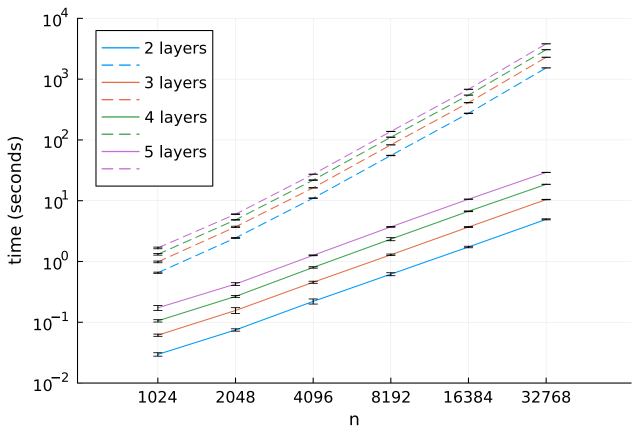

Figure 1: Average execution time (over 10 repetitions; ± 1 std) for the `multilayerGM` (dashed) and `mABCD` (solid) models, for $n=2^k$ nodes with $k\in\{10,11,\ldots,15\}$ and $\ell\in\{2,3,4,5\}$ layers.