Inferring the Modular Structure of Networks with Weighted Stochastic Blockmodeling

Simone Santoni

2024-11-26

Synopsis   This notebook shows how to infer the modular structure of weighted networks using the Weighted Stochastic Blockmodeling (WSBM) approach. WSBM extends the concept of the Stochastic Blockmodel (SBM) to incorporate edge weights into the modeling process, allowing for a more detailed representation of the network's structure. We will use a synthetic network dataset to illustrate the key steps involved in inferring the modular structure of a weighted network using the WSBM approach.

1 Overview: inferring the modular structure of networks

The notebook `community_detection.qmd` shows how to infer the modular structure of weighted networks using Louvain's community-detection algorithm. The current notebook focuses on a different approach to the same problem: the weighted stochastic blockmodeling (WSBM).

WSBM[1] approach is an extension of the stochastic blockmodel (SBM) that incorporates edge weights into the modeling process.[2] SBMs are probabilistic models used to detect community structures in networks by partitioning nodes into blocks or communities, where the probability of an edge existing between any two nodes depends only on the blocks to which the nodes belong.

WSBM extends this concept by considering not just the presence or absence of edges, but also the weights of the edges, which can represent the strength or capacity of connections between nodes. This allows for a more nuanced understanding of the network's structure, capturing variations in connection strengths within and between communities.

The WSBM approach involves defining a likelihood function that accounts for the observed edge weights and optimizing this function to find the most likely partition of nodes into communities. This is typically done using techniques such as the Expectation-Maximization (EM) algorithm or variational inference.

---

[1]Tiago P. Peixoto, "Nonparametric weighted stochastic block models", Phys. Rev. E 97, 012306 (2018), DOI: 10.1103/PhysRevE.97.012306, arXiv: 1708.01432

[2]Paul W. Holland, Kathryn Blackmond Laskey, Samuel Leinhardt, "Stochastic blockmodels: First steps", Social Networks Volume 5, Issue 2, Pages 109-137 (1983). DOI: 10.1016/0378-8733(83)90021-7

By incorporating edge weights, WSBM provides a more detailed and accurate representation of the network, making it particularly useful for analyzing weighted networks such as social networks with varying interaction strengths, biological networks with different interaction intensities, and transportation networks with different capacities. This approach enhances the ability to uncover meaningful community structures and understand the underlying processes governing the network.

In this notebook, we will learn how to implement the WSBM approach using the `graph-tool` library in Python. We will use a synthetic network dataset to illustrate the key steps involved in inferring the modular structure of a weighted network using the WSBM.

2 Notebook setup

We will use the `graph-tool` library for network analysis and visualization. If you haven't already installed `graph-tool`, you can do so using the following command:

```bash
{bash}
conda activate %YOUR_ENVIRONMENT%
conda install -c conda-forge graph-tool
```

Then, we can import the necessary libraries and set up the notebook environment:

```python
from IPython.display import Image
import numpy as np
import matplotlib
from graph_tool.all import *
```

3 Toy Dataset: 'Food Web' network

To illustrate the WSBM approach, we will use a synthetic network dataset representing a food web. The food web network consists of different species (nodes) connected by interactions (edges) representing predator-prey relationships. The edge weights in the network represent the strength of these interactions, with higher weights indicating stronger relationships.[3] It is worth noticing that the edge values are positive $[0, \infty]$.

```python
g = collection.ns["foodweb_baywet"]
graph_draw(g, pos=g.vp._pos, output_size=(300, 300), output="foodweb.png")
Image(filename="foodweb.png")
```

---

[3] Robert E. Ulanowicz, and Donald L. DeAngelis. "Network analysis of trophic dynamics in south florida ecosystems." US Geological Survey Program on the South Florida Ecosystem 114 (2005).
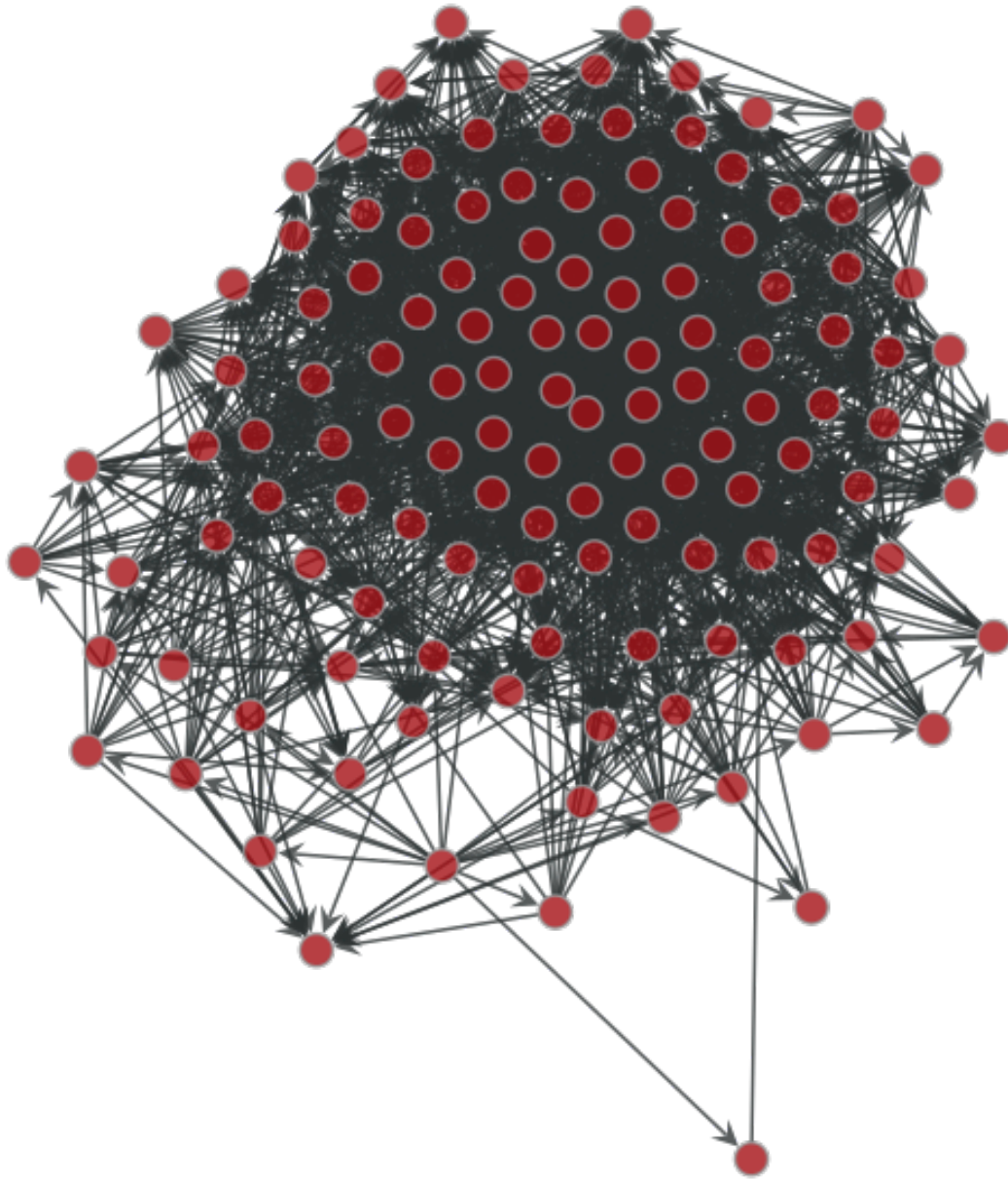
Figure 1: Food Web Network

Figure 1 does not show a clear pattern. Let us see if we can uncover the underlying modular structure of this network using the WSBM approach.

4 Weighted Stochastic Blockmodeling (WSBM)

The WSBM approach involves defining a likelihood function that accounts for the observed edge weights and optimizing this function to find the most likely partition of nodes into communities.

As the following code cell shows, we can use the `graph-tool` library to implement the WSBM approach for inferring the modular structure of the food web network. Specifically, we will use the `minimize_nested_blockmodel_dl` function to fit the WSBM model to the network data. This function takes the following arguments in our example:

- `g`: the input graph object representing the food web network
- `state_args`: a dictionary containing the edge weights (`recs`) and their types (`rec_types`) for the model fitting process
- `rec_types`: the type of edge weights, in this case, real-exponential – remember that the edge values are positive‼

```python
# model fit
state = minimize_nested_blockmodel_dl(
    g, state_args=dict(recs=[g.ep.weight], rec_types=["real-exponential"])
)
# improve solution with merge-split
for i in range(100):
    ret = state.multiflip_mcmc_sweep(niter=10, beta=np.inf)
state.draw(
    edge_color=prop_to_size(g.ep.weight, power=1, log=True),
    ecmap=(matplotlib.cm.inferno, 0.6),
    eorder=g.ep.weight,
    edge_pen_width=prop_to_size(g.ep.weight, 1, 4, power=1, log=True),
    edge_gradient=[],
    output_size=(300, 300),
    output="foodweb-wsbm.png",
)
# show the plot
Image(filename="foodweb-wsbm.png")
```
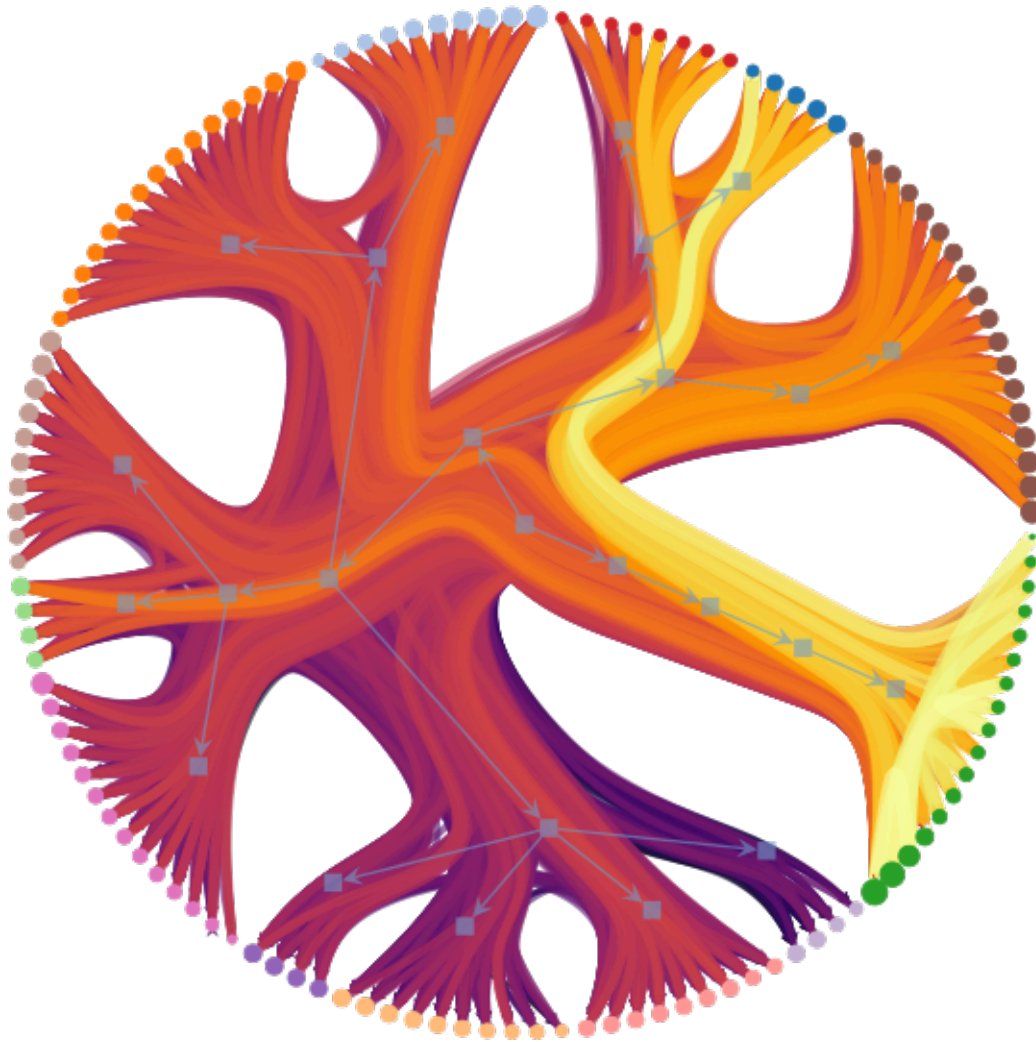
Figure 2: Modular Structure of the Food Web Network Inferred Using WSBM

Figure 2 shows the modular structure of the food web network inferred using the WSBM approach. Quite evidently, this is a complex visualization:

- The nodes are colored according to the community they belong to, with nodes of the same color representing the same community (that is, species dependent on the same resources)
- The size of nodes is proportional to their degree, with larger nodes having more connections
- The layout of the network is determined by the WSBM algorithm, which arranges nodes based on their community assignments and interaction strengths
- The color gradient represents the edge weights, with lighter colors indicating higher weights
- The set of square nodes in the center of the network represents the reduced form (i.e., simplified) version of the network

- The square nodes are arranged in a hierarchical structure, with each level representing a different level of community organization. In other words the square nodes in the outer layer are nested in the square nodes in the inner layer.