Learn About Assortativity Coefficient in Python With Data From UK Faculty Dataset (2008)

Learn About Assortativity Coefficient in Python With Data From UK Faculty Dataset (2008)

Student Guide

Introduction

This example dataset introduces the assortativity coefficient for networks. This concept allows researchers to quantify the extent to which connected nodes share similar properties. For example, in many social networks, two people who are friends tend to have similar backgrounds, interests, jobs, etc.; such networks will have a large assortativity coefficient, and this pattern may be indicative of homophily or assortative mixing, which is edge formation due to similarity in node characteristics. There are also networks that have low or negative assortativity coefficients, such as sexual networks, where sexual relationships are often between people of different genders, and those networks are called disassortative mixing by the corresponding property.

This example describes the assortativity coefficient, discusses the assumptions underlying its measurement, and shows how to compute and interpret its measure. We illustrate this concept using a subset of data from the 2008 UK Faculty dataset. Specifically, we compute the assortativity coefficient of the friendship network of the faculty of a UK university in terms of their departmental affiliations and assess whether friends in this network tend to be from the same department. The example illustrates how the assortativity coefficient can help us understand how the network nodes, here the faculty members, interact with each other.

What Is Assortativity Coefficient?

The assortativity coefficient is analogous to the Pearson correlation coefficient that assesses the association between two continuous variables; in fact, the assortativity coefficient measures the correlation between the characteristics of every pair of nodes that are connected.

Calculating Assortativity

For a given undirected network with *N* nodes and *M* edges, let *A* be its adjacency matrix. Further, let k_i be the degree of node *i* and x_i be the value of some property of node *i* (i.e., some variable as a characteristic about *i*). Depending on whether the node property of interest is a continuous or a categorical variable, there are two definitions of the assortativity coefficient. If the node property is a continuous variable, such as age or income, then the assortativity coefficient is equivalent to the Pearson correlation for that variable across all edges in the network. Specifically, the assortativity coefficient is defined as

(1)

$$H = c\sum_{i}\sum_{j}(A_{ij} - \frac{k_i k_j}{2M})x_i x_j$$

where *c* is a normalizing constant $\sum_i\sum_j (k_i\delta(i, j) - \frac{k_i k_j}{2M})x_i x_j$, so that the value of the assortativity coefficient falls between −1 and 1, like a correlation coefficient does. The delta function, $\delta(i, j)$, takes value 1 if $i = j$ and 0 otherwise.

If the node property is a categorical variable such as race and gender, then the assortativity coefficient is equivalent to modularity (see SAGE Research Methods Datasets on Modularity). Specifically, it is defined as the fraction of edges between nodes with same properties compared to the fraction that would be expected under the configuration model:

(2)

$$H = \frac{1}{2M}\sum_{i}\sum_{j \neq i}\left(A_{ij} - \frac{k_i k_j}{2M}\right)\delta\left(x_i,\ x_j\right)$$

Interpreting Assortativity

The assortativity coefficient is a number between −1 and 1, just as are correlation coefficients. A large positive value means that connected nodes very much tend share similar properties; a large negative value means that connected nodes tend to possess very different properties; and a value close to 0 means no strong association of the property values between connected nodes (where strength is gauged in distance from what would be expected with a random null model). Note that the assortativity coefficient is always about a specific property or variable of the nodes in a network. If the nodes have multiple properties, then it is possible that nodes are assortatively mixed for one property but disassortatively mixed for another and non-assortative (unassociated, not different from random) in a third property of the nodes.

Illustrative Example: Assortativity of UK Faculty Network

This example introduces the assortativity coefficient using the friendship network of a faculty of a UK university. Specifically, we compute the assortativity coefficient of this faculty network in terms of their affiliations and assess whether friends in this network (where edges = friendship relations) are from the same department (the node characteristic).

Thus, this example aims to shed lights on the following research question:

> Are faculty members from the same department more likely to be

friends?

However, the assortativity coefficient measures correlation but not causation, so it cannot fully address the question of whether departmental affiliations cause the faculty members to be friends. Instead, it directly answers the question: Are friends more likely (than non-friends) to be from the same department?

The Data

This example uses a subset of data collected and studied by Nepusz, et al. in their paper "Fuzzy communities and the concept of bridgeness in complex networks" published in *Physical Review E* (2008) (http://hal.elte.hu/~nepusz/research/datasets/) and made publicly available by the R package "igraphdata." The data are about a social network consisting of 81 nodes with each representing a faculty member of a UK university, and the edges of the network represent friendships between the faculty members. There are 817 undirected, unweighted edges in total.

Analyzing the Data

Which department each faculty member is affiliated with is a categorical variable, and hence, we should use the second definition of the assortativity coefficient (Equation 2). The corresponding assortativity coefficient is calculated to be 0.71, which is moderately large. Therefore, we conclude that the faculty social network shows affiliation-based homophily. That is to say, the faculty members who are friends with each other tend to come from the same department.

Presenting Results

As a descriptive statistic, one simply states the calculated assortativity coefficient; there is no formally prescribed manner of reporting it. For instance, the analysis in this example could be stated as follows:

"We used a subset of data collected by Nepusz, et al. to assess the assortativity by department of a social network of friendship relations among the faculty members of a UK university. We investigate the following research question:

> Are faculty members from the same department more likely to be friends?

The network includes 81 faculty members, with each of them corresponding to a node, and the edges of the network represent friendships between the faculty members. There are 817 undirected, unweighted edges in total. The assortativity coefficient based on their department affiliations is calculated to be 0.71, which suggests homophily between the faculty members based on their affiliations. That is to say, the faculty members who are friend with each other tend to come from the same department."

Review

The assortativity coefficient quantifies the extent to which connected nodes share similar properties. It is analogous to the Pearson correlation coefficient but measures the correlation between every pair of nodes that are connected. Depending on whether the node property of interest is a continuous or a categorical variable, there are two definitions of the assortativity coefficient.

You should know:

- What is the assortativity coefficient.
- How to compute both versions of the assortativity coefficient.
- How to interpret and report the results of assortativity coefficient.

Your Turn

You can download this sample dataset along with a guide showing how to calculate the assortativity coefficient using statistical software. See whether you can reproduce the results presented here. There is another property of the nodes called degree, which is the degree of each node. Try calculating your own assortativity coefficient using this variable. Hint: Since this variable is continuous, you will want to use the definition of assortativity for continuous variables.