The Amazon Referral Network

by
Rohit Hiwale
Satish Chandra Ch
Sanjeev Chandra V

under the guidance of
Joydeep Chandra

Problem Statement

Amazon.com, Inc. (Amazon.com) offers services to consumers, sellers, and developers through its retail websites; books.amazon.com is a section of this website which sells books online which has one of the largest databases of books on wide spectrjm of subjects and also includes other information about the books like reviews and ratings of books, authors' information etc; This entire database can be conceptually perceived as a complex network with books as nodes and relations between various books like same author or same genre etc; as edges.

The problem statement is to study the properties of the underlying complex network in the database and answer the following questions:

a. Will a book likely to be popular if any one of its author is popular?

b. Books co-authored by popular authors are likely to be more popular than books written by a single popular author?

c. The problem statements also requires us to discuss the assortativity properties of the co-author networks measured in terms of the author’s popularity.

Tasks Identified and Working Plan

1. Crawl the website of books.amazon.com

    1. Get information about various books like author, ratings and other books liked by the same customer etc;
    2. Build a network based on the relationship between various books based on the liking of customers.

2. Build a database which contains details about the author popularity derived from the ratings of the books written by him/her.

3. Based on the database and the complex network obtained, answer the questions mentioned in the problem statement regarding the author popularity-book poularity relationship, variations in the popularity of books with the number and popularity of respective authors;

4. Build a network based with authors as nodes and co-authorship relation as edges.

5. Analyse this network and discuss the assortativity properties of the co-author networks measured in terms of the author's popularity.

Literature Survey

1.Exploratory Data Analysis of Amazon.com Book Reviews is a review of data analysis of books from amazon.com and acts as reference on how to collect relevant data and analyse it.

2.Sun Tutorial for crawler in Java is a web tutorial on how to design an effective crawler which obtains the necessary information to be analysed and reported.

3.The Structure and Function of Complex Networks. SIAM Review 45, 167-256 (2003) by M.E.J. Newman is a documentation which gives useful insights into the concepts of complex networks.