PROJECT PROPOSAL OF NETWORK TOUR OF DATA SCIENCE

Amazon Product Social Network

Süha Kağan Köse Tao Sun Xiangzhe Meng Xingce Bao

1 Introduction

Amazon is reshaping and recording our life as the biggest online retailer and as an invisible social network. We are connected with each other because we bought or reviewed the same product. And because of our behaviors, purchasing, reviewing or just viewing, the products in Amazon are also connected together. Actually, these two networks are forming and disappearing in every supermarket every day but only in Amazon the networks are recorded and updated every day. This project would focus on these two networks. There are too many questions we want to ask:

- What do the networks look like?
- Do we have friends in the network? Am I popular?
- Do we even form groups? In what way?
- What's the star product?
- How is the friendship of product world?

2 Data acquisition

In this project, we are going to use the **Amazon Product Dataset**, provided by Julian McAuley, UCSD. This dataset includes product metadata (descriptions, category information, price, brand, and image features), and links (also viewed/ also bought graphs), and millions of reviews (ratings, text, helpfulness votes) spanning from May 1996 to July 2014.

3 Data exploration

- Use and clean one subset of the Amazon dataset: **Electronics**.
- Build the also-viewed/also-bought graphs of products using metadata.
- Connect people/reviewers together using review dataset.

4 Data exploitation

The project contains two main subtasks:

- Subtask 1
  - Analyze and compare the basic properties of each network.
  - Find the most similar generating model or build our own generating method.
- Subtask 2
  - Clustering with the help of graph Laplacian and machine learning method.
  - Intuitively visualize the networks from different perspectives.