

Analyzing Co-Purchase Network of CDs in Japanese Online Store

R. Nishida^{1,2}, M. Mori³ and T. Iba^{4,5}

¹ Graduate School of Media and Governance, Keio University, Endo 5322, Fujisawa, Kanagawa, Japan

² Fellow Assistant, Rakuten Institute of Technology Lab., Shinagawa Seaside Rakuten Tower, 4-12-3 Higashi-shinagawa, Shinagawa-ku, Tokyo, Japan

³ Rakuten Institute of Technology Lab., Shinagawa Seaside Rakuten Tower, 4-12-3 Higashi-shinagawa, Shinagawa-ku, Tokyo, Japan

⁴ Faculty of Policy Management, Keio University, Endo 5322, Fujisawa, Kanagawa, Japan

⁵ Research Fellow, Rakuten Institute of Technology Lab., Shinagawa Seaside Rakuten Tower, 4-12-3 Higashi-shinagawa, Shinagawa-ku, Tokyo, Japan

Theme: Economy, Communications and Management

Introduction

Every day, many customers buy many kinds of products like books, CD, and DVD etc, at online stores. Then, the companies often have giant data stock of the transactions. The data are important from the viewpoint of business, because the company can get the chance to know the hidden laws in the market and to let the product sell more. In academic viewpoint, the data is also important because it is able to become a clue to understand the complexity of the market. The main issue is what kind of order is emerged as a result of compiling the customers' actions? For the purpose, we visualized and analyzed a co-purchase network as an order of the market, with using the real market data of CDs, which is collected by the online store "Rakuten Books" (<http://books.rakuten.co.jp/>), which is one of the biggest online stores in Japan. Note that this research was done as an analysis by Rakuten Institute of Technology, and the data do not include any personal information.

Results

We visualize co-purchase networks in order to grasp the relations among CDs, which represent the order from the bottom-up. The data we use is the real market data of 30,000 customers at Rakuten Books by picking up at random. The target term is from April, 2005 to May, 2007. The co-purchase network is compiled by the following way. We describe a node A if there is the product A is purchased by the target customers. Then we describe an edge to connect node A and node B if the product A and the product B is purchased by a customer. For describing the edge, we try two types of connection method: "full connection" and "sequential connection". In the former method, all the nodes which user bought connect each other. In the latter method, nodes connect as the sequential order of user bought. It means that an undirected graph is generated by the former method and a direct graph is generated by the latter.

These networks include many nodes and edges. For instance the network has 2014 nodes and 4923 edges in the case of cutting off customers purchasing one product only from 30000 users. And our analysis contains some kind of network indicators of the networks, including node degrees, the diameter of a network, the average number of neighbors, average clustering coefficients, and shortest path lengths etc. In addition, we refer the spreading of network at the view point of degrees. Based on the findings of our study, we would like to understand dynamics of the network from the bottom-up.