

# Analyzing Co-Purchase Network of Books in Japanese Online Store

Y. Kitayama<sup>1 2</sup>, M. Yoshida<sup>3 2</sup>, S. Takami<sup>4</sup> and T. Iba<sup>3 5</sup>

<sup>1</sup> Faculty of Environment and Information Studies, Keio University, Endo 5322, Fujisawa, Kanagawa, Japan

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

<sup>3</sup> Faculty of Policy Management, Keio University, Endo 5322, Fujisawa, Kanagawa, Japan

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

<sup>5</sup> 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

In this research, we visualize a network of co-purchase books in order to clarify the hidden relationship in the book sale market. As the characteristic of the market, it is already known that the sales ranking follows power law in book sale market, which is also called as “Long-tail” distribution. In the co-purchase network, a node represents a book, and the nodes linked each other is connected if they are purchased by same person. The source data of the network is collected by “Rakuten Books” (<http://books.rakuten.co.jp/>), which is one of the biggest online book stores in Japan. The data that we use are an extraction of 30,000 people from customers of online shop at random. Note that this research was done as an analysis by Rakuten Institute of Technology, and the data do not include any personal information.

## Results

For visualizing the network, 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.

As a result, we can visualize the rough sketch of the map of the products based on the choices by the customers. For an instance in the case of “full connection”, a lot of components are found in the co-purchase network, few large components and many middle-size and small components. The similarity among the books in the components is formed by same genre books. To see the network more concretely, the large components are formed by comics, law-books, X-rated books. It is supposed that the books which are not so much circulated in the market, or the books which have some mental difficulties to buy. These large components are connected to other components by famous books. This kind of books becomes a bridge which connects the some different genres components.