Amazon Product Co-purchasing Network

Fall 2020 SNAP Proposal

Northwestern

Samuel Hamilton
Shirley Liu
Aria Wang
Joe Zhang

Data source and description

- **Dataset name**: Amazon product co-purchasing network
- **Dataset source:** Leskovec, J. “Large Network Dataset Collection.” Stanford University. http://snap.stanford.edu/data/.
- **Collection date:** March 2, 2003
- **Data information:** Network was collected by crawling Amazon website. It is based on the “Customers Who Bought This Item Also Bought” feature of the Amazon website. Each node represents a product. If a product i is frequently co-purchased with product j, the graph contains a directed edge from i to j.

Problem statement

Client

Amazon.com, Inc. is an American multinational technology company based in Seattle, Washington, which focuses on e-commerce, cloud computing, digital streaming, and artificial intelligence. For the online retailing service, Amazon allows for individuals and business to sell and display products for sale online. Currently, It is the largest internet retailer in the world by revenue.

Questions

- Can co-purchased products or cluster of products be sold in bundle deals in a business perspective?
- Would the network analysis be helpful for recommendations on products for customers to buy together?
- What popular products could be advertised to drive purchasing of other products?

Product Review Metadata

TABLE I. Sample Amazon Product Metadata

Property	Value
Id	1
ASIN	0827229534
Title	Patterns of Preaching: A Sermon Sampler
Group	Book
SalesRank	396585
Similar Products	5 0804215715 156101074X 0687023955 0687074231 082721619X
Categories	2 —Books[283155]—Subjects[1000]— Religion & Spirituality[22]— Christianity[12290] —Clergy[12360]— sPreaching[12368]
Reviews	Total: 2 downloaded: 2 avg rating: 5 2000-7-28 cutomer: A2JW67OY8U6HHK rating: 5 votes: 10 helpful: 9 2003-12-14 cutomer: A2VE83MZF98ITY rating: 5 votes: 6 helpful: 5

Srivastava, A. Motif Analysis in the Amazon Product Co-Purchasing Network. *arXiv [cs.SI]* (2010)

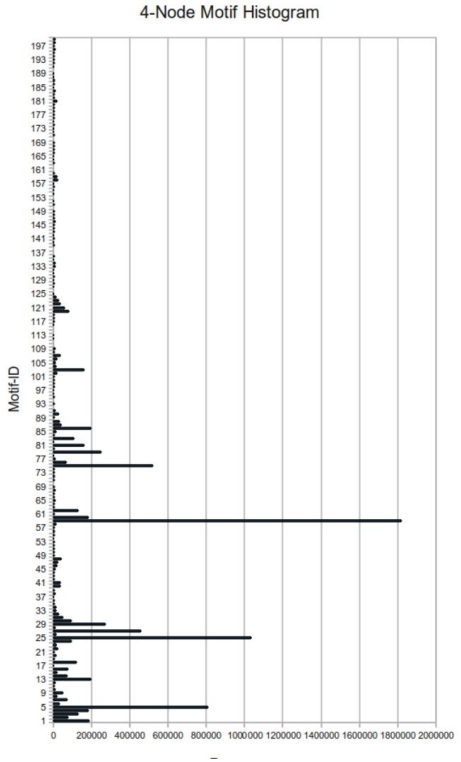
EDA results

Nodes	262,111	**Average clustering coefficient**	0.4198
Edges	1,234,877	**Number of triangles**	717,719
Nodes in largest WCC	262,111	**Fraction of closed triangles**	0.09339
Edges in largest WCC	1,234,877	**Diameter (longest shortest path)**	31
Nodes in largest SCC	241,761	**90-percentile effective diameter**	11
Edges in largest SCC	1,131,217	**Density (undirected)**	3.59e-05

Previous Work

Motif Analysis in the Amazon Product Co-Purchasing Network

Abhishek Srivastava
*Computer Science Department, Columbia University**
(Dated: December 21, 2010)



The Dynamics of Viral Marketing

JURE LESKOVEC
Carnegie Mellon University
LADA A. ADAMIC
University of Michigan
and
BERNARDO A. HUBERMAN
HP Labs

Table I. Product Group Recommendation Statistics
p: number of products, n: number of nodes, r: number of recommendations,
e: number of edges, b_b: number of buy bits, b_e: number of buy edges.

Group	p	n	r	e	b_b	b_e
Book	103,161	2,863,977	5,741,611	2,097,809	65,344	17,769
DVD	19,829	805,285	8,180,393	962,341	17,232	58,189
Music	393,598	794,148	1,443,847	585,738	7,837	2,739
Video	26,131	239,583	280,270	160,683	909	467
Full network	542,719	3,943,084	15,646,121	3,153,676	91,322	79,164

Visualization & Centrality analysis

Visualization

- Visualize macro-level structure of Amazon purchases network
- Visualize giant graph through different types of graph layouts

Individual-level

- Identify key products in the network. Find out the products that are “gateways” to other products

Global

- Use k-core decomposition to identify the group of popular products
- Detect communities and understand the dynamics within and across clusters

Question 1: Co-purchased products in bundles

Motivation:

- Exclusive deals add a feeling of value.
- The “Long tail” is a huge portion of online revenue

Strategy:

- Bundle additional products that are bought together with other frequently bought products into bundles.

Methods:

- Motif analysis
- Burt’s constraint
- Community identification
- Visualization

Question 2: Advertise products that bridge network gaps

Motivation:

- An ideal co-purchasing network has high connectivity
- By advertising nodes with high betweenness, we can increase connectivity

Strategy:

- Identify nodes that connect modules

Methods:

- Community identification
- Betweenness centrality.

Task Assignment

	Items	Team
Network Analysis	Data Preparation	All
	EDA & Visualization	Shirley, Aria
	Statistical Models	All
	Business Insights	Joe
Final Deliverables	Research Report (Background)	Sam
	Research Report (Results/Discussion/Conclusion)	Aria, Joe
	Research Report (Methods)	Shirley
	Final presentation (Slide preparation)	Sam, Joe
	Final presentation (Speaking)	All

Thank You!