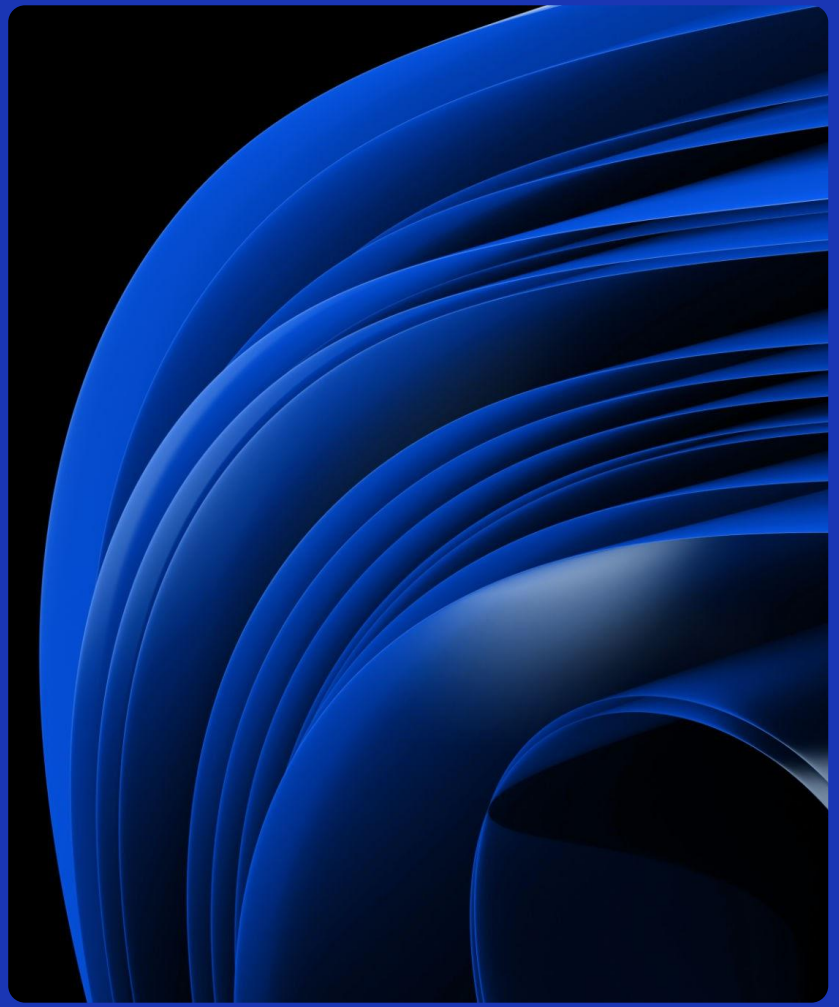
Mapping E-Commerce

A Network Approach to Logistics and Market Trends

Steven Zhu | Darren Cheng | Ayman Momin



Dataset

Amazon Product Dataset (2020):

- Data attributes: product names, corresponding categories, selling prices, development costs, quantities, and brand names.
- Additional attributes: sale regions, distribution centers, and customers.
- Data cleaning and analysis using Python libraries, (primarily Pandas and Networkx).
- Network visualization using Gephi: Open Graphical Visualization Software.

Expected size of the network

- Medium-Scale preferred: 1,000 nodes with 3,000-5,000 edges.
- Nodes: product categories, regions of demand, and distribution centers.
- Edges: distribution paths and delivering costs.
- One larger-scaled sales summary node for each given geographical location.
- One node for the distribution center: counting by regions for overall standardization.
- Subset of product type nodes under sales summary node.

Questions: Why is it important

Covers two aspects:

- Supply chain:

  - Study the paths for distribution of goods.

  - Optimize distribution paths of goods to manage costs.

- Marketing

  - Trend analysis by time to spot out the sales peaks and lows.

  - Customer segmentation: specialized offerings to improve sales.

Results: Drives revenue up, cost down

Thank you for watching!