```
export SPARK_HOME=/home/saikat/Downloads/spark-3.4.0-bin-hadoop3
export PATH=$PATH:$SPARK_HOME/bin
spark-shell
```

# Amazon product co-purchasing network, March 02 2003

### Dataset information

Network was collected by crawling Amazon website. It is based on *Customers Who Bought This Item Also Bought* feature of the Amazon website. If a product *i* is frequently co-purchased with product *j*, the graph contains a directed edge from *i* to *j*.

The data was collected in March 02 2003.

```
:load /home/saikat/Downloads/basic.scala
```

# PageRank-

PageRank is a popular algorithm used to rank web pages based on their importance. It assigns a numerical weight to each page, indicating its relative importance, based on the number and quality of other pages that link to it.

In the context of GraphX, PageRank is computed as a property of vertices in a graph. Each vertex is assigned an initial rank, and then iteratively updated based on the rank of its neighbors, until convergence is achieved.

:load /home/saikat/Downloads/page.scala

## Profile Performance-

This type of graph is commonly referred to as a co-purchasing network and can be used for a variety of applications, such as recommendation systems and market basket analysis.

We want to identify groups of products that are frequently co-purchased together or recommend products to customers based on their purchase history.