HW1 – Amazon Product Data Network

In HW1, we will build on the social network analysis to understand how products get sold on Amazon.com

Please download the data from http://snap.stanford.edu/data/index.html#amazon where you will find data on Amazon products and which products are bought together.

In particular, you will find five datasets

Product co-purchasing networks

Name	Type	Nodes	Edges	Description
amazon0302	Directed	262,111	1,234,877	Amazon product co-purchasing network from March 2 2003
amazon0312	Directed	400,727	3,200,440	Amazon product co-purchasing network from March 12 2003
amazon0505	Directed	410,236	3,356,824	Amazon product co-purchasing network from May 5 2003
amazon0601	Directed	403,394	3,387,388	Amazon product co-purchasing network from June 1 2003
amazon-meta	Metadata	548,552	1,788,725	Amazon product metadata: product info and all reviews on around 548,552 products.

The first four represent co-purchasing networks on four different days. The fifth one consists of information on each of the products.

You have the following tasks:

1) Pre-process the data to create the data in the format used in class.
2) Compute Degree centrality, closeness centrality, between-ness centrality and pagerank for all the products.
3) Find the top 100 products based on each category.

Please submit well-commented running R code and a 2 page report explaining the key findings. Compare and contrast the buying behaviour on the four different days.