Name: YelpNYC
Description:
Yelp NYC dataset contains Yelp reviews for restaurants located in New York City. It has 160,255 users, 923 products and 359,052 reviews. So a User-Review-Product graph can be constructed which has 520,230 nodes and 718,144 edges. The goal is to classify the reviews is fake or not.

Task:
We can randomly choose 1000 fake reviews and 1000 genuine reviews as the training dataset and test the remaining reviews. Not all features are need to make the classification. For example, the review content is too long which need to be removed or adjusted.

File Descriptions:

metadata.txt:
The metadata file contains meta information in the following order:
1. user_id
2. prod_id
3. rating:0-5
4. label:benign,1; fake,-1
5. date

productIdMapping.txt: Product id.

reviewGraph.txt: User-Review-Product Graph, (user_id(1 .. N), prod_id(1 ... M), rating).

userIdMapping.txt: User id.

reviewContent.txt: the review text.

Reference:
@inproceedings{DBLP:conf/sigkdd/Akoglu15,
author = {Shebuti Rayana and Leman Akoglu},
title = {Collective Opinion Spam Detection: Bridging Review Networks and metadata},
booktitle = {Proceeding of the 21st ACM SIGKDD international conference
on Knowledge discovery and data mining, {KDD’15}},
year = {2015},
}