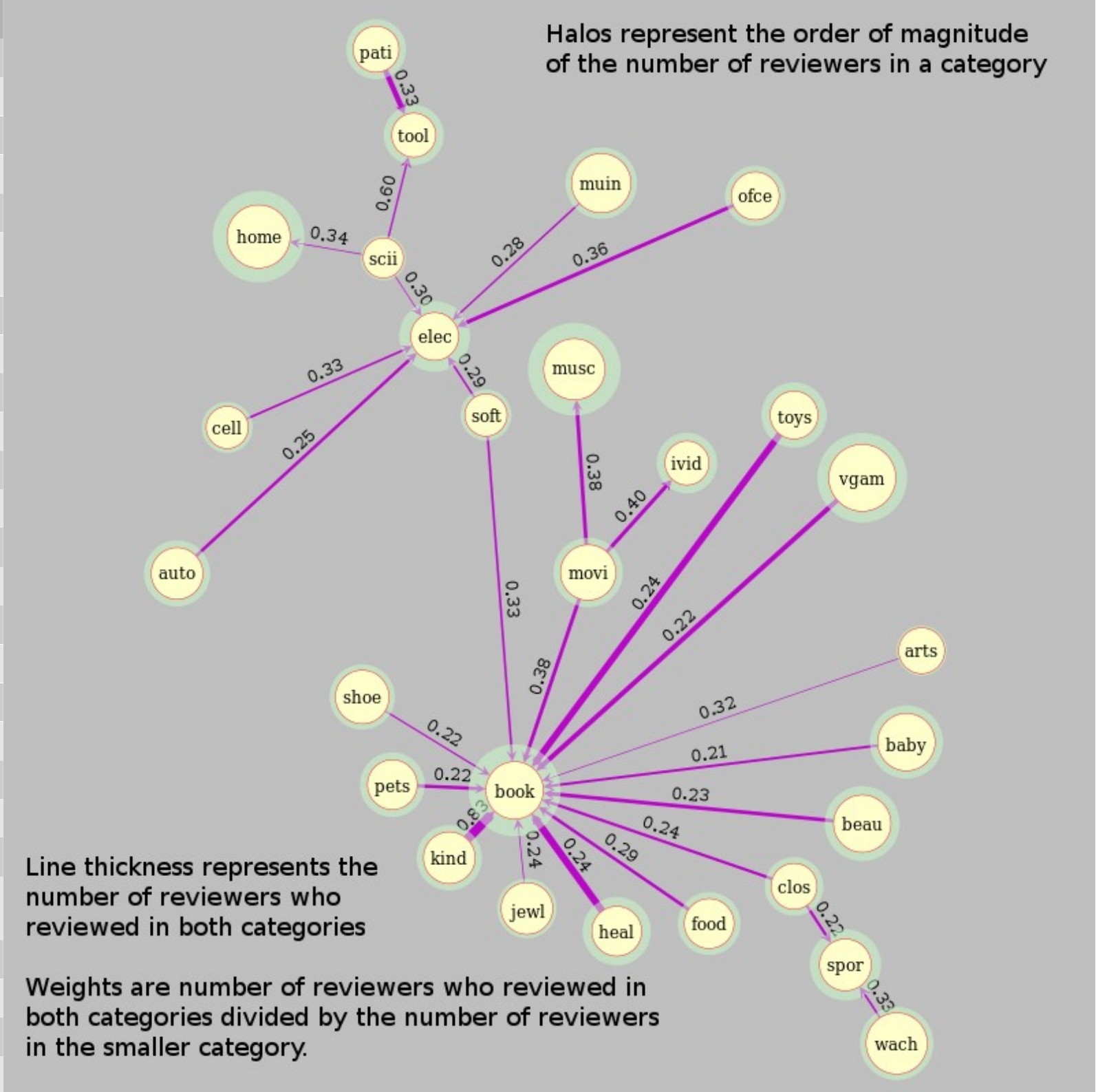
ivid	Amazon Instant Video
arts	Arts
auto	Automotive
baby	Baby
beau	Beauty
book	Books
cell	Cell Phones & Acc.
clos	Clothing & Acc.
elec	Electronics
food	Gourmet Foods
heal	Health
home	Home & Kitchen
scil	Industrial & Scientific
jewl	Jewelry
muIn	Musical Instruments
musc	Music
ofce	Office Products
pati	Patio
pet	Pet Supplies
shoe	Shoes
soft	Software
spor	Sports
tool	Tools & Home Improv.
toy	Toys and Games
vGame	Video Games
wach	Watches
kind	Kindle Store



Next Steps: Set up baseline for data by building recommender Systems for individual categories using Apache Mahout.

What is Apache Mahout?

The Apache Mahout™ project's goal is to build a sc

With scalable we mean:

Scalable to large data sets. Our core algorithms for clustering, classfication and collaborative filtering are implemented on top of scalable, distributed systems. However, contributions that run on a single machine are welcome as well.

Scalable to support your business case. Mahout is distributed under a commercially friendly Apache Software license.

Scalable community. The goal of Mahout is to build a vibrant, responsive, diverse community to facilitate discussions not only on the project itself but also on potential use cases. Come to the mailing lists to find out more.

Currently Mahout supports mainly three use cases: Recommendation mining takes users' behavior and from that tries to find items users might like. Clustering takes e.g. text documents and groups them into groups of topically related documents. Classification learns from exisiting categorized documents what documents of a specific category look like and is able to assign unlabelled documents to the (hopefully) correct category.

Collaborative Filtering *with CLI Drivers*

User-Based Collaborative Filtering

Item-Based Collaborative Filtering

Matrix Factorization with ALS

Matrix Factorization with ALS on Implicit Feedback
Weighted Matrix Factorization, SVD++

Classification *with CLI Drivers*

Logistic Regression - trained via SGD

Naive Bayes / Complementary Naive Bayes

Random Forest

Hidden Markov Models

Multilayer Perceptron

Clustering *with CLI Drivers*

Canopy Clustering

k-Means Clustering

Fuzzy k-Means

Streaming k-Means

Spectral Clustering