Using Graphs for High Quality Recommendations

PyData
Nov 2015

Amit Bhattacharyya
Sr Data Scientist
Teachers Pay Teachers

@aabtzu

TpT is an online marketplace for teachers to buy, sell and share original educational resources.


50k
ACTIVE SELLERS
1M+
RESOURCES
3M+
ACTIVE MEMBERS

$150,000,000+

PAID TO TEACHERS

TpT landing page

Product page for a best seller

Teachers Pay Teachers

About Us | Blog | FAQs & Help | Gift Cards

All Categories | SEARCH | Log In | Not a member? Join for Free | Cart

2 Share 1 Tweet 4K+ Pin it 2 8+1

Spring Math and Literacy Packet NO PREP (Kindergarten)

Subjects: Reading, Math, Spring

Grade Levels: Kindergarten, 1st, Homeschool

Resource Types: Printables, Literacy Center Ideas, Math Centers

Product Rating: 4.0 4,489 ratings

PDF (Acrobat) Document File
Be sure that you have an application to open this file type before downloading and/or purchasing.
35.27 MB | 80 pages

PRODUCT DESCRIPTION

This Spring Math and Literacy Packet has it all and requires NO PREP! The resources in this packet are designed to meet Common Core Standards for Kindergarten while making learning FUN, hands-on and interactive!

Be sure to follow us on Facebook for exclusive FREEBIES!
The Moffatt Girls on Facebook

Total Pages	80
Answer Key	N/A
Teaching Duration	N/A

Report Copyright Infringement

Comments & Ratings | Product Q & A

$7.50
Digital Download

ADD ONE TO CART

BUY LICENSES TO SHARE

ADD TO WISH LIST

Made by
The Moffatt Girls
User Rating: 4.0/4.0
Follow Me (31,135 Followers)
Visit My Store

May NO PREP Packet (Kindergarten)
$6.00

March NO PREP Packet (Kindergarten)
$6.00

Sight Word Make a Match NO PREP Packet (Primer)

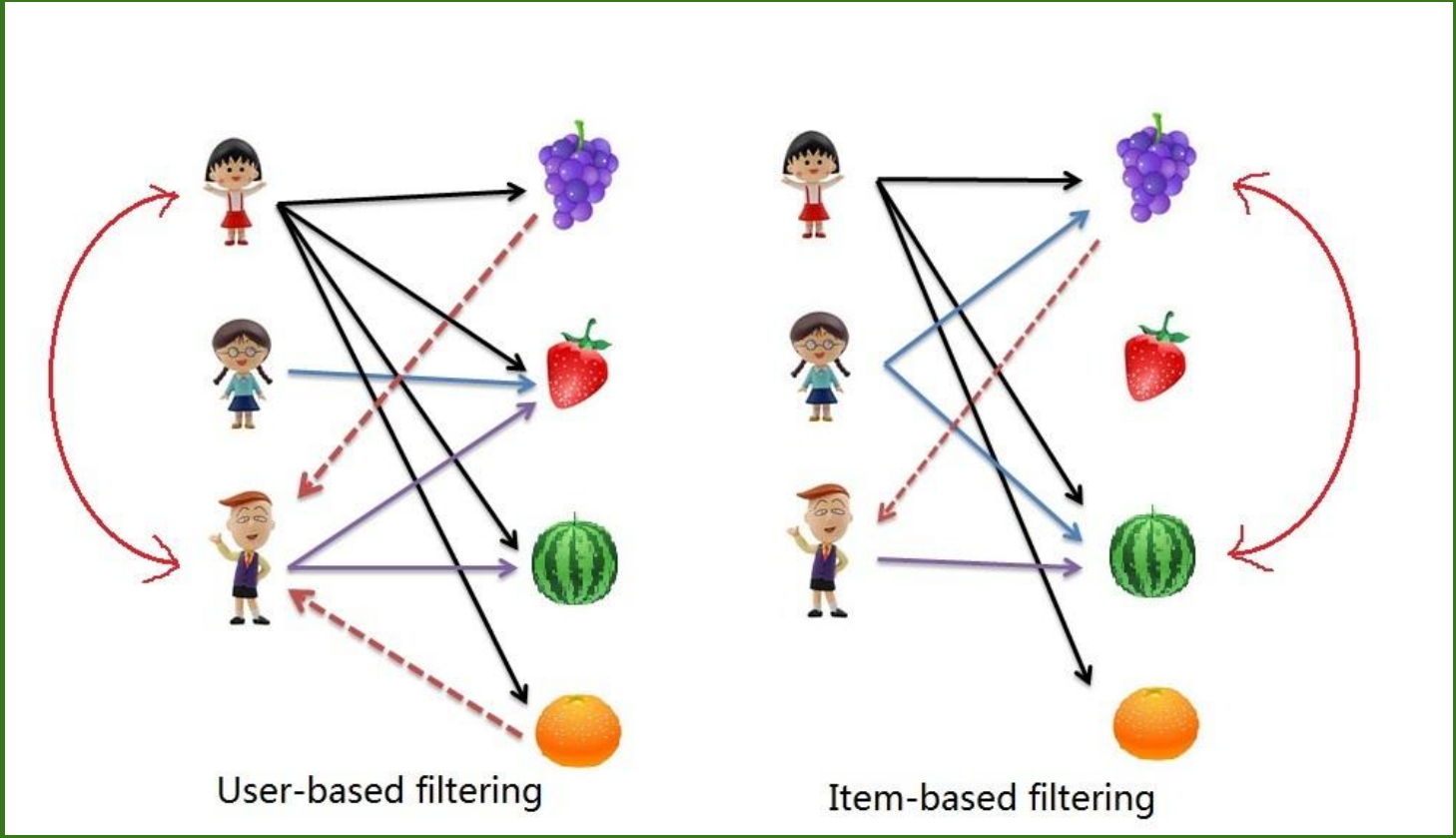
What are we trying to do?

We would like to identify a cluster of buyers and send a product recommendations email to the cluster

The Problem

Collaborative filtering (CF) methods are great for individual recommendations but hard for clusters.

Collaborative filtering for recommendations



I use GraphLab Create for easy to implement recommender systems

Core version is open source
https://dato.com/products/create/

```
m = graphlab.recommender.create(data_frame,
                                user_id='user',
                                item_id='movie')
recs = m.recommend()
```

It is harder to use CF methods when making recommendations for a group

- not optimized for finding clusters
- coefficients are hard to interpret
- lack of distance measurement for k-means clustering
- ratings on TpT are useless

Why want clusters anyways?

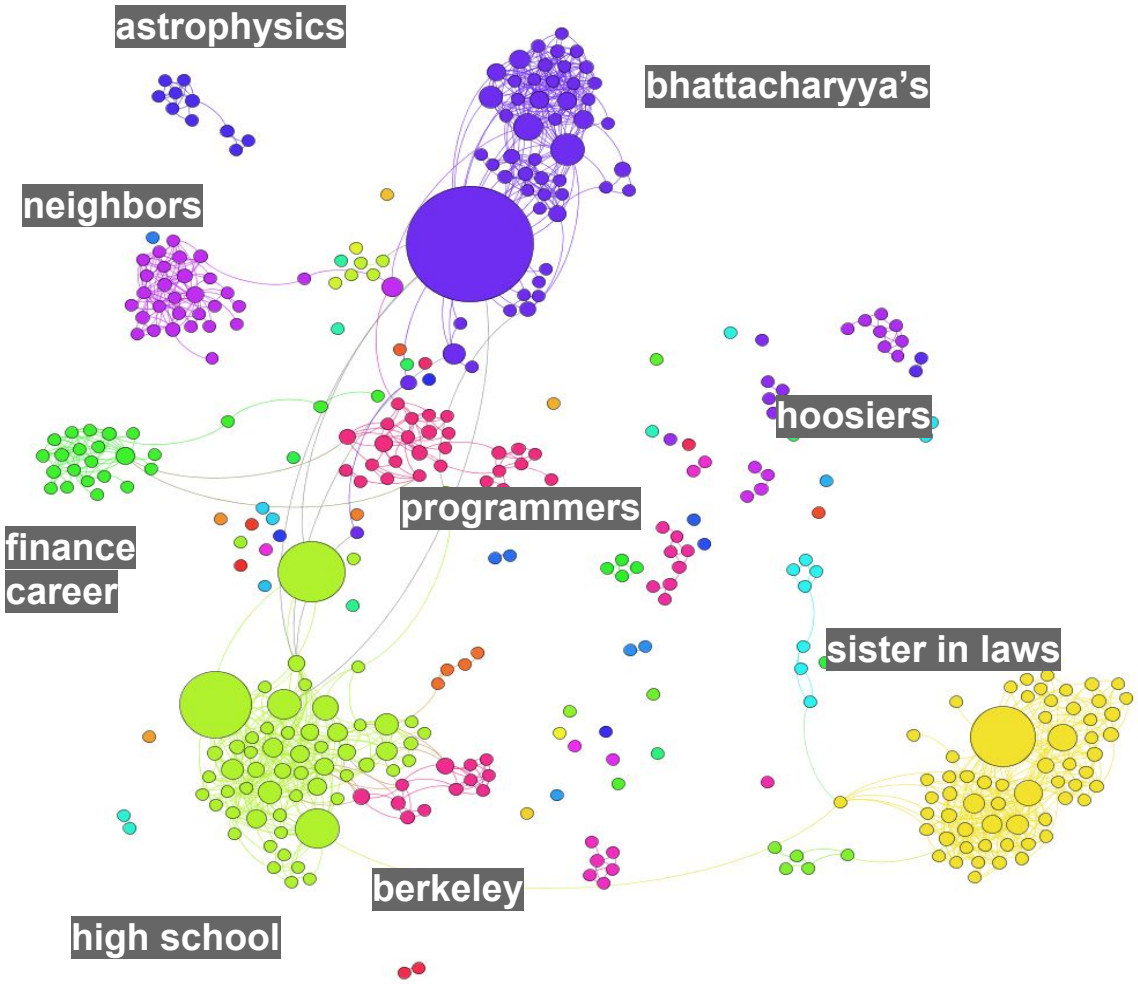
Clusters at TpT

- natural groupings inform buyer behavior
- teachers just getting started may not have long purchase history but are likely to be very similar to others
- can reduce complexity of CF recommendations by pre-clustering users

The Solution

Cluster Users by Creating a Graph

My LinkedIn Network



All we are given is a list of users and items they have purchased

user_id	item_id	item_name
206067	1020240	Rotation and Revolution Model: Sun, Earth and ...
3927028	1498533	Naming Compounds Puzzle - A Fun Chemical Nomen...
3927028	821435	Ionic Bonding Task Cards
3927028	355690	Make Your Own Color By Number Clipart Collecti...
268012	1100472	Introduction to Meso-America Vocabulary
268012	882788	Aztec, Mayan, & Incan Graphic Organizer
268012	1417449	Rise of Empires (Maya, Inca, Aztec)

How to construct the graph:

- each user is a node
- there is an edge between two users if they have made a common purchase
- multiple common purchases will get a higher edge weight

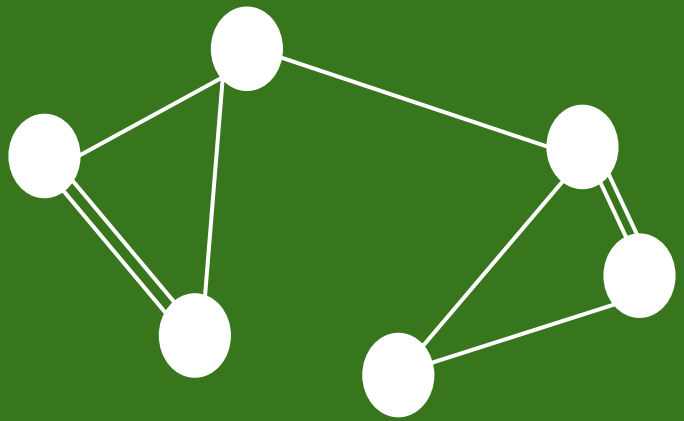
Graphs in Python

- NetworkX
- GraphLab Create (now called Dato)
- igraph

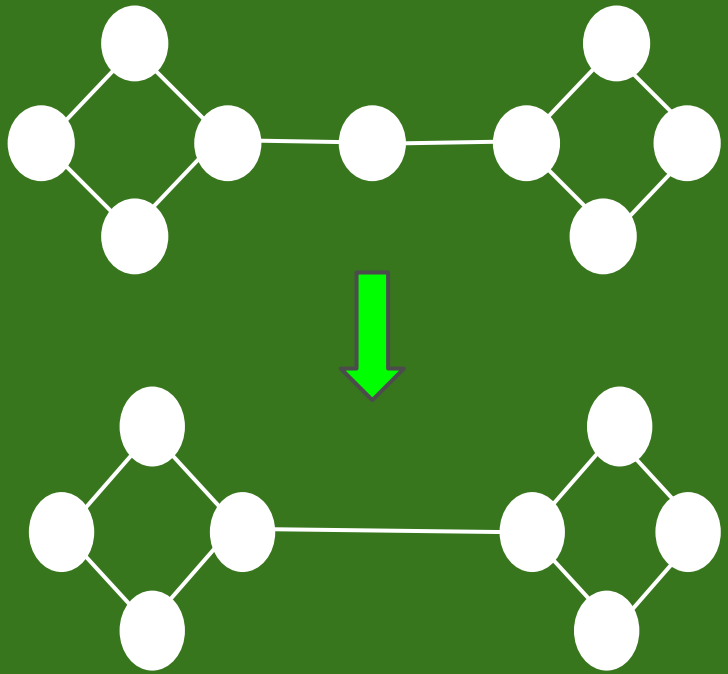
Graphs (not in Python)

- Gephi

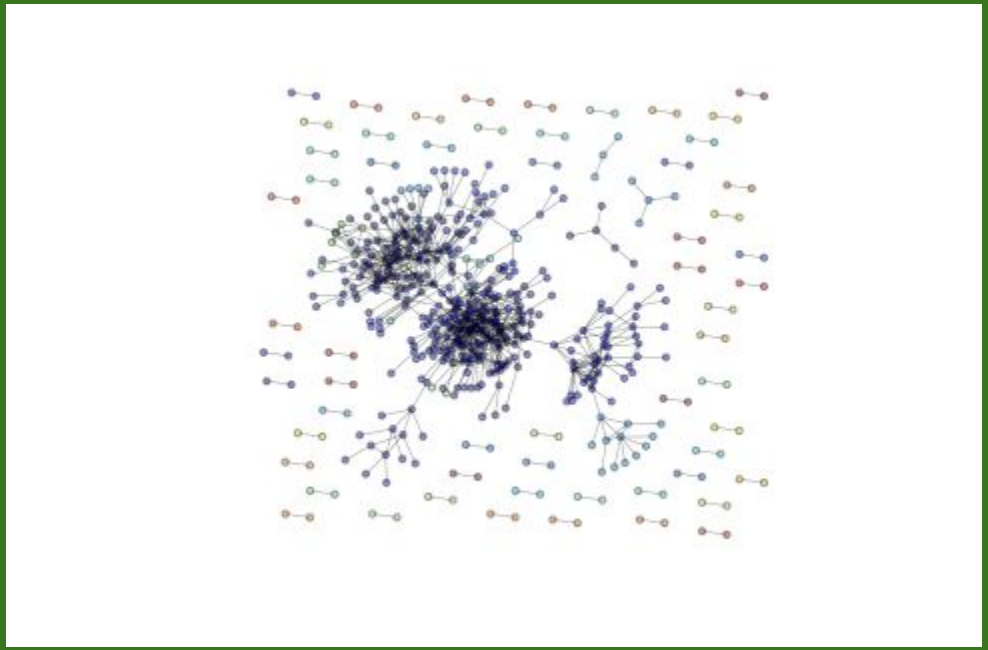
example of a simple graph
users = nodes
common purchases = edges



reduce complexity by removing all nodes with degree = 2



run community detection algorithm and clusters begin to emerge ...



What do the clusters mean?

Need to look incorporate item info to determine physical intuition of cluster

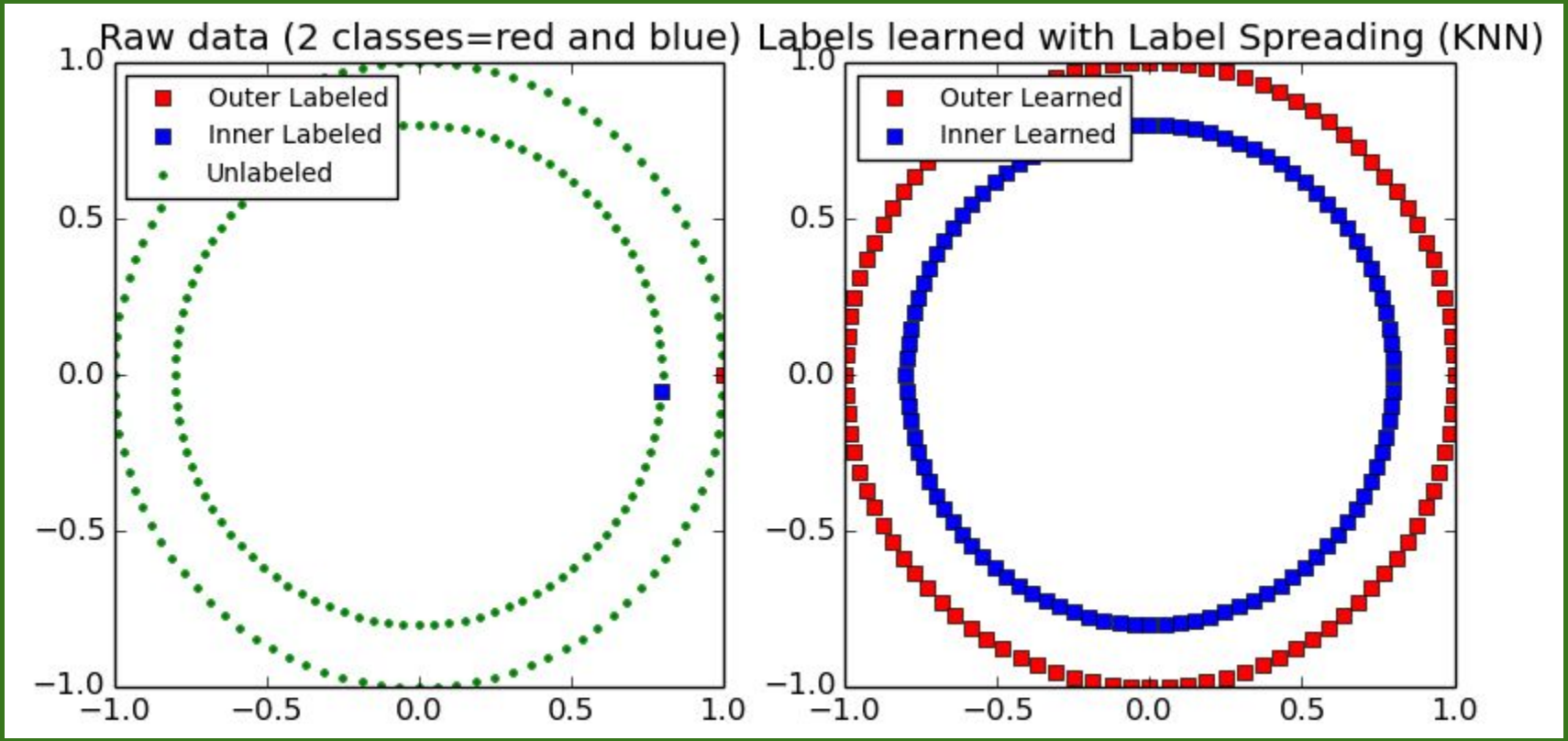
item_id	name
506116	Solving Systems of Equations by Graphing and S...
1080423	Exponent Rules Valentine's Day Coloring Activity
1415508	Trigonometry {SOH CAH TOA} Coloring Activity
590706	Factoring a Greatest Common Factor (GCF) Drag ...
1196123	Algebra: Graphing and Writing Compound Inequal...
967719	Multiplying Polynomials {FOIL} Coloring Activity
1013496	Systems of Equations Relay Races
426830	Algebra 1: Quadratic Equations (Unit 8) - Unit...

Another Problem

Community detection is hard to do at scale

Another Solution

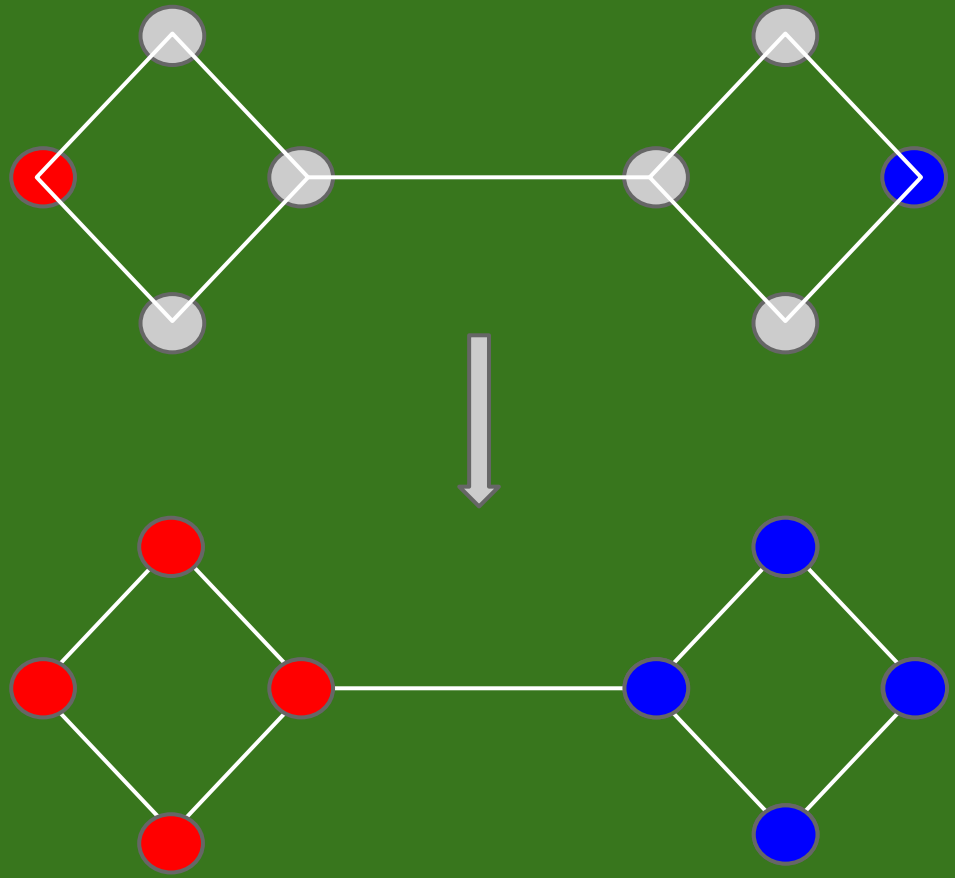
Label propagation



Two step process:

- **Run community detection on a fraction of the users to determine clusters**
- **Use *label propagation* to fill in the rest**

Label propagation in clusters



Recommendation algorithm can be as simple as choosing the most purchased by a cluster

Clustering also helps restrict number of user-item pairs in CF recommender systems

Let’s do some real Python ...

We’re Hiring!

engineering mostly

TeachersPayTeachers.com/careers