

Week 13

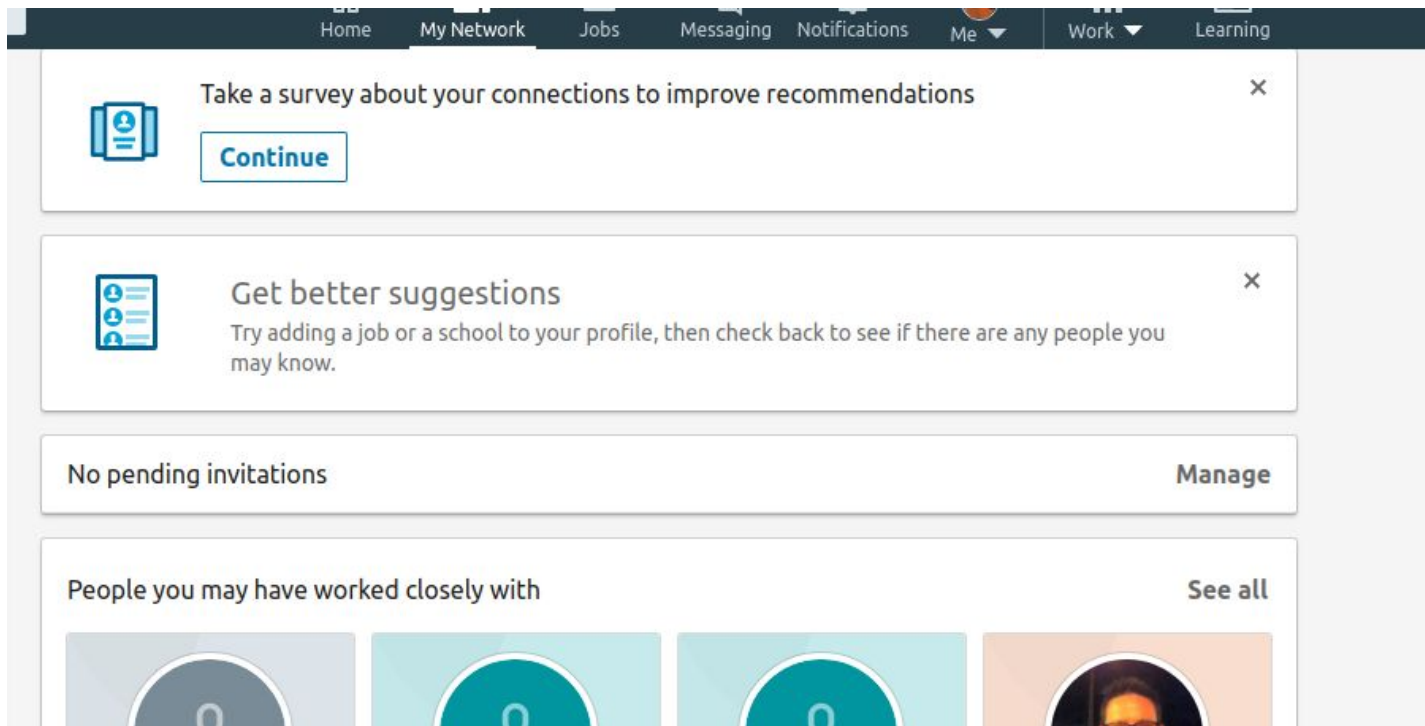
Agenda

- Project 4
- Review networks
- What is a recommender system?
- Early recommender systems: Content vs Behavior
- Collaborative Filtering
- Cold start problem
- Discussions/case studies

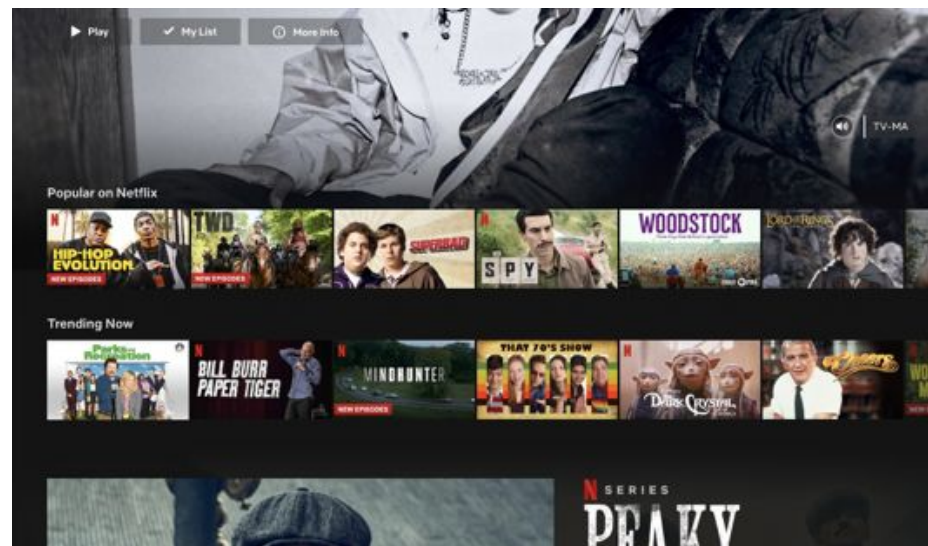
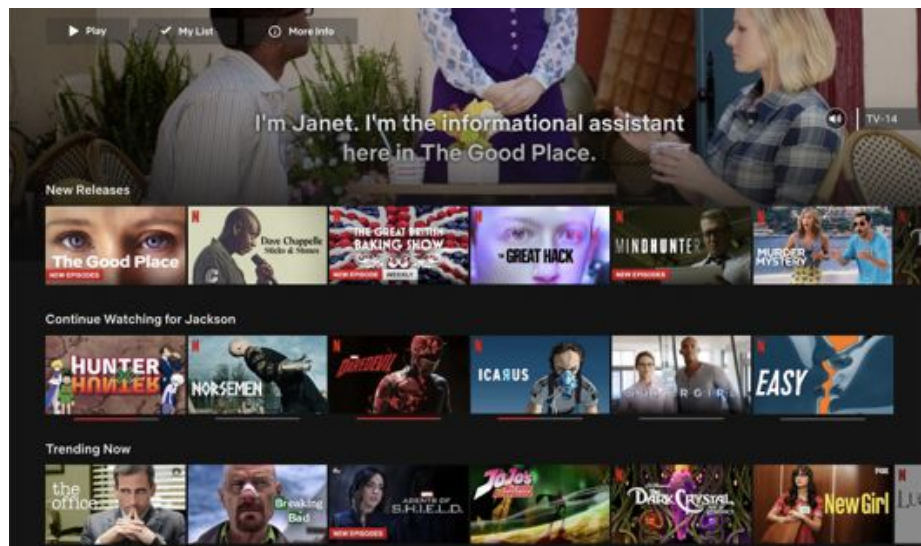
Intro: What are recommender systems?

And where does the data come from?

- Is LinkedIn's "do you know this person" a prediction or a recommender system?
- And does that make a difference?
- Or is it something else entirely?




Netflix

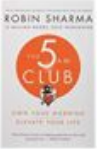


datascience@berkeley

Amazon


**Hi, Darin**
Customer since 2009

Your recent order



See your orders


Continue watching



Kitchen Nightmares

See more from Prime Video

Recently viewed




Yesterday

Edit your browsing history

Amazon Day

Free delivery on everything you've ordered each week, on the day that works for you

[Learn more](#)



COSEQUIN

Recommended for maximum strength joint health

Cosequin Joint Health Supplements

★★★★☆ (29)

\$12.99 prime


[Add to Cart](#)

[Add feedback](#)

Video: Recommended for you

ORIGINAL


The Boys Season 1



See more from Prime Video

NEW SERIES


1- or 2-hour delivery



Complete your collage checklist.

[Shop Prime Now](#)


Deal of the Day




\$11.99 - \$41.99

Save up to 30% on AUKEY USB Chargers, Speakers and Accessories


[See all deals](#)

**Hi, Alex**
Customer since 2016


Recommendations for you




Your Drinks



Electronics




Home & Kitchen



Health & Household


Recently viewed



Yesterday

Edit your browsing history

Deal of the Day



\$11.99 - \$41.99


Save up to 30% on AUKEY USB Chargers, Speakers and Accessories

[See all deals](#)

FREE Same-Day Delivery

Get delivery as fast as today on over three million items.

[Learn more](#)



vitaFusion

NUTRITIOUS AND DELICIOUS

prime now


[Shop now](#)

[Add feedback](#)


Video: Recommended for you

ORIGINAL


Carnival Row - Season 1




Essentials delivered fast




Grocery



Electronics




Personal care



Household

\$10 off your order of \$35+ and FREE 2-hour delivery

\$10 off Whole Foods



\$10 off \$50 at Whole Foods Market and FREE 2-hour delivery

"import your contacts"

"take this survey"

91%

Prefer Brands that Provide
Personalized Offers/
Recommendations

83%

Willing to Passively
Share Data in Exchange for
Personalized Experiences

74%

Willing to Actively
Share Data in Exchange for
Personalized Experiences

Mary Meeker, 2019

Personal Data Marketplaces



Data Management Platform (DMP): Personal data marketplaces
Discuss:

- How do you feel about them?
- Which datasets about yourself do you assume are public?

Cambridge Analytica

As the company's CEO, Alexander Nix, explains in a promotional video: “The more you know about someone, the more you can align a campaign with their requirements or their wants and needs.” He did not respond to a request for comment.

Source:

<https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data>

Personal Data Marketplaces



TiVo: A Buyer's Guide

DATA TYPES: TV Viewership

TiVo Data 101

TiVo's expertise spans multi-screen functionality and back-end services that support linear television, video on demand (VOD), mobile apps, streaming, etc. Our experience in the market puts TiVo in a unique position to offer superior data for targeting television viewers and measuring sales impact. Our data segments power any planning, modeling, or reporting you can imagine. Our viewership data is ready to be manipulated by sophisticated data science teams familiar with processing large datasets.

Description of Data Types:

Flat Files or TV Viewership data – Raw data processed by a buyer's internal data teams.

Syndicated TV Segments – Viewership segments that span across verticals and viewership attributes.

Campaign Segments – Ad-exposed segments used for an attribution ad matched to digital campaigns.

Custom Segments – Ability to create segments not readily available in data stores.

Collection Methodology:

TiVo's TV Viewership data is sourced from TiVo's retail set-top boxes and from MVPDs who use TiVo's software. For our segments, we deploy a 1:1 deterministic match to 1st and 3rd party data with our owned and licensed set-top box data. Experian is our leading partner to execute the 3rd party match process based on name/address. For our raw data, matching occurs at household level via Experian with each device viewership mapped to each household. While we don't model our raw data, our segments can be modeled in-house via a 3rd party partner or within the data stores.

Use Our Data For:

- Targeting cross-platform
- Optimizing frequency
- Extending reach
- Improving campaign KPIs
- Appending TV data to digital campaigns

Personal Data Marketplaces

Evite: A Buyer's Guide

Evite Data 101

With more than 22 million registered users and over 25,000 invitations sent each hour, Evite is the top online invitation and social planning website.

Launched in 1998, Evite is headquartered in Los Angeles.

Description of Data Types

Segment/Definition

- Presence of Child in Household
 - » Host of Kids Birthday
 - » Babys first Bday
 - » Kids Corner
 - » Kids themes
 - » Halloween for Kids
- Age of Child in Household
 - » Title scrape for numbers (1st, first, First) in Host of Birthday for Kids
 - » Age assumed to be 1 for Host of Baby's first
- Recent Movers/Furnishers
 - » Host of Housewarming party
- Bride
 - » Host of Save the Date
- Wedding Attendee
 - » Host and Guest of Wedding/Engagement
 - » Bridal Shower
 - » Bachelor
 - » Bachelorette Party
 - » Save the Date

Evite

- Pre-Natal/Expecting
 - » Host or Guest of Baby Shower
- Upcoming Birthday
 - » Host or Guest of Birthday for Her
 - » Birthday for Him
- Recent Graduate
 - » Host or Guest of Graduation
- Sports Enthusiast
 - » Host or Guest of Sports/Leagues
- Home Entertainers
 - » Host or Guest of Hostess Party
 - » Dinner Party
 - » Cocktail Party
 - » House Party
 - » BBQ/Pool Party
 - » Pot Luck
 - » Game Night
- Travel intenders
 - » Hosts or Guest of Trips/Getaways
- Halloween
 - » Host or Guest of Halloween Party
- Winter Holidays
 - » Host or Guest of Winter Holiday Party
- Super Bowl
 - » Host or Guest of The Big Game Party
- Thanksgiving
 - » Host or Guest of Thanksgiving
- Religious
 - » Host or Guest of Religious Event

Personal Data Marketplaces

TECH

Vizio nears \$17 million settlement for TV data-tracking lawsuit

The deal is still subject to final approval

By Chaim Gartenberg | @cgartenberg | Oct 4, 2018, 1:30pm EDT

f t  SHARE



Photo by James Bareham / The Verge

Personal Data Marketplaces

“One of the biggest challenges in protecting privacy is that many of the violations are invisible,” Cook writes. “For example, you might have bought a product from an online retailer—something most of us have done. But what the retailer doesn’t tell you is that it then turned around and sold or transferred information about your purchase to a ‘data broker’—a company that exists purely to collect your information, package it, and sell it to yet another buyer.

“The trail disappears before you even know there is a trail. Right now, all of these secondary markets for your information exist in a shadow economy that’s largely unchecked—out of sight of consumers, regulators, and lawmakers. Let’s be clear: You never signed up for that. We think every user should have the chance to say, ‘Wait a minute. That’s my information that you’re selling, and I didn’t consent.’”



Tim Cook - Apple

Personal Data Marketplaces



Apple Sold Its Customers' iTunes Data Without Consent, Lawsuit Says

Suit says that third-parties could buy music habits and user home addresses from tech giant for \$136 per 1,000 people

[Sean Burch](#) | May 28, 2019 @ 1:32 PM

Last Updated: May 28, 2019 @ 3:34 PM

Taste Domains



Taste Domains

- Early research focused on 'taste' domains, particularly movies, music, and books.
- In these domains, 'finding' often involved suggestions from friends or tastemakers
- Researchers created collaborative filtering and other approaches as means of emulating this process
- Usually cast as ratings prediction problem in part because its relatively easy to collect ratings data

1996 MovieLens (Minnesota)

- <http://en.wikipedia.org/wiki/MovieLens>
- Research project collected ratings on movies, etc
- Very early Amazon and Netflix strongly influenced by this
- User-Based Collaborative Filtering (UB-CF): Find k-nearest users and use their ratings
- 2001 Item based CF: Find k nearest items to those items a user prefers.
http://files.grouplens.org/papers/www10_sarwar.pdf
- Still popular recommendation algorithm

CF - Collaborative Filtering

- What's that?
- Two types:
 - Memory based - similarity between users or items
 - Model based - predict ratings of unrated items
 - hybrid/deep learning approaches
- Problems:
 - Cold start
 - Scalability
 - Sparsity

Movielens (Late 90's)

movielens helping you find the right movies

Welcome aib@...
You've rated 15 movies.
You're the 21st visitor in the past hour.

So far you have rated **15** movies.
MovieLens needs at least **15** ratings from you to generate predictions.
Please rate as many movies as you can from the list below.

Your Rating	Movie Information
??? Notseen	Beneath the Planet of the Apes (1970) Action, Sci-Fi
??? Notseen	Gift, The (2000) Thriller
??? Notseen	Great Muppet Caper, The (1981) Children, Comedy
??? Notseen	Heaven Can Wait (1978) Comedy
★★★★ 4.0 stars	Hitch (2005) Comedy, Romance
??? Notseen	Kate & Leopold (2001) Comedy, Romance
??? Notseen	Muppets Take Manhattan, The (1984) Children, Comedy, Musical
??? Notseen	Police Academy 4: Citizens on Patrol (1987) Comedy
??? Notseen	Saturday Night Fever (1977) Comedy, Drama, Romance
??? Notseen	Teenage Mutant Ninja Turtles II: The Secret of the Ooze (1991) Action, Children, Fantasy

To get a new set of movies click the [next](#)>

movielens helping you find the right movies

Welcome dus@infovis.net
You've rated 48 movies.
You're the 24th visitor in the past hour.

★★★★ = Must See
★★★★☆ = Will Enjoy
★★★★☆ = It's OK
★★★☆☆ = Fairly Bad
★☆☆☆☆ = Awful

[Home](#) | [Manage Buddies](#) | [Your Account](#) | [Help](#) | [Logout](#)

You've searched for **all titles**.
Found 7233 movies, sorted by **Prediction**
Genres: **All** | Exclude Genres: **None**
Dates: **All** | Domain: **All** | Format: **All** | Language: **All**
[Show Printer-Friendly Page](#) | [Download Results](#) | [Suggest a Title](#)

Page 1 of 483 | Go to page: 1...96...192...288...384...480...last page 2>

Predictions for you	Your Ratings	Movie Information	Wish List
★★★★★	Notseen	Tainted (1998) info imdb Comedy, Thriller	<input type="checkbox"/>
★★★★★	Notseen	Friday Night Lights (2004) info imdb Action, Drama	<input type="checkbox"/>
★★★★★	Notseen	Harry Potter and the Prisoner of Azkaban (2004) info imdb Adventure, Children, Fantasy	<input type="checkbox"/>
★★★★★	Notseen	Spider-Man 2 (a.k.a. Spiderman 2) (2004) info imdb Action, Fantasy, Sci-Fi, Thriller	<input type="checkbox"/>
★★★★★	Notseen	Finding Nemo (2003) DVD, VHS info imdb Adventure, Animation, Children, Comedy	<input type="checkbox"/>
★★★★★	Notseen	X-Men 2 (a.k.a. X2: X-Men United) (2003) DVD, VHS info imdb Action, Adventure, Sci-Fi	<input type="checkbox"/>
★★★★★	Notseen	Oliver Twist (1948) info imdb Adventure, Crime, Drama	<input type="checkbox"/>
★★★★★	Notseen	Raiders of the Lost Ark (1981) DVD info imdb Action, Adventure	<input type="checkbox"/>
★★★★★	Notseen	Indiana Jones and the Last Crusade (1989) DVD info imdb Action, Adventure	<input type="checkbox"/>

Search

Search Titles
[]

☐ Use selected buddies!

Search by Genre/Date

All Genres [v] All Dates [v]
Domain: All movies [v]
☐ Use selected buddies!

Advanced Search

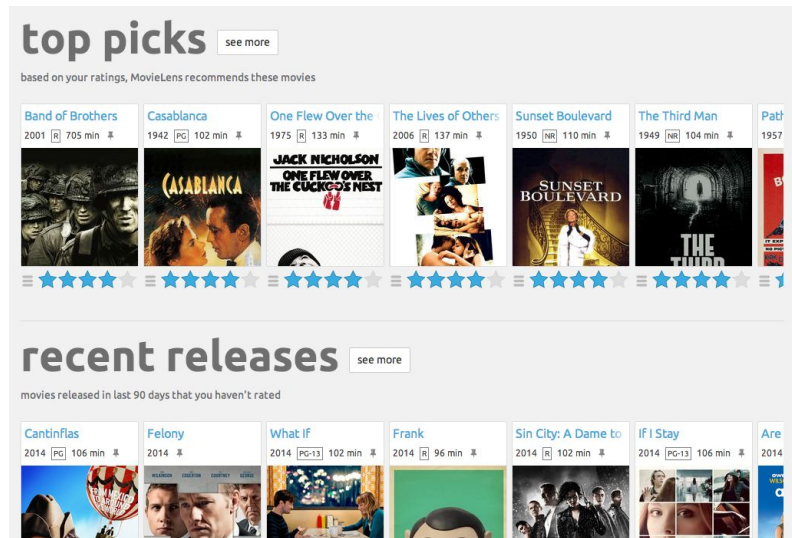
Select Buddies
☐ Test Buddy
[What are buddies?](#)

Movielens (Today)

Stable benchmark dataset. 10 million ratings and 100,000 tag applications applied to 10,000 movies by 72,000 users.
Released 1/2009.

- [README.html](#)
- [ml-10m.zip](#) (size: 63 MB, [checksum](#))

Permalink: <http://grouplens.org/datasets/movielens/10m/>



Netflix DVD Recommendations

Movies You've Rated

Based on your 745 movie ratings, this is the list of movies you've seen. As you discover movies on the website that you've seen, rate them and they will show up on this list. On this page you may change the rating for any movie you've seen, and you may remove a movie from this list by clicking the 'Clear Rating' button.

	TITLE	MPAA	GENRE
Add	12 Angry Men (1957)	UR	Classics
Add	The 39 Steps (1935)	UR	Classics
Add	An American in Paris (1951)	UR	Classics
Add	The Andromeda Strain (1971)	G	Sci-Fi & Fantasy
Add	Apollo 13 (1995)	PG	Drama
Add	The Battle of Algiers (1965) La Battaglia di Algeri	UR	Foreign
Add	Being There (1979)	PG	Drama
Add	Big Deal on Madonna Street (1958) I soliti ignoti	UR	Foreign
Add	The Birds (1963)	PG-13	Thrillers
Add	Blade Runner (1982)	R	Sci-Fi & Fantasy

★ Rating Activity [View All](#)

of Ratings: 745 [Rate More](#)
Favorite Genres: 0 [Edit Favorites](#)
Recommendations: 428 [View All](#)
of Reviews Written: 5 [View](#)

Below the table, each movie has a rating bar with 5 stars and a 'Clear Rating' button.

2002 Netflix DVDs

- Mailed 1-3 DVDs to users
- Brand based in part on ratings and recommendations
- Users could rate any movie, not just those recently watched
- Users were told recommendations get better with more ratings

Search and Discovery

- Amazon historically incorporated recommendations into search results, including to extend results
- Recommendations used at different stages of transaction (E.g. Product page, cart)

Amazon Recommendations (2003)

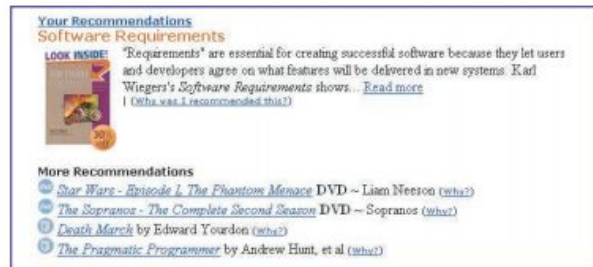


Figure 1. The “Your Recommendations” feature on the Amazon.com homepage. Using this feature, customers can sort recommendations and add their own product ratings.



Figure 2. Amazon.com shopping cart recommendations. The recommendations are based on the items in the customer's cart: The Pragmatic Programmer and Physics for Game Developers.

2003 Amazon

- Item-Item CF
(<http://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf>)
- Compares to User-Item based CF (as in GroupLens)
- Pointed to scalability: compute similarities between items offline, usually nightly. This is preferred as item similarities change less frequently than user similarities
- Use item similarities to recommend similar items to those a user has shown preference towards
- Item preference a weighted combination of views, ratings, cart adds, and purchases
- Item similarities can also be used in non-personalized fashion (“users who considered this, also considered that...”)
- Not fully real-time and suffers from cold-start problem
- Flawed problem setup: not directly optimizing sales

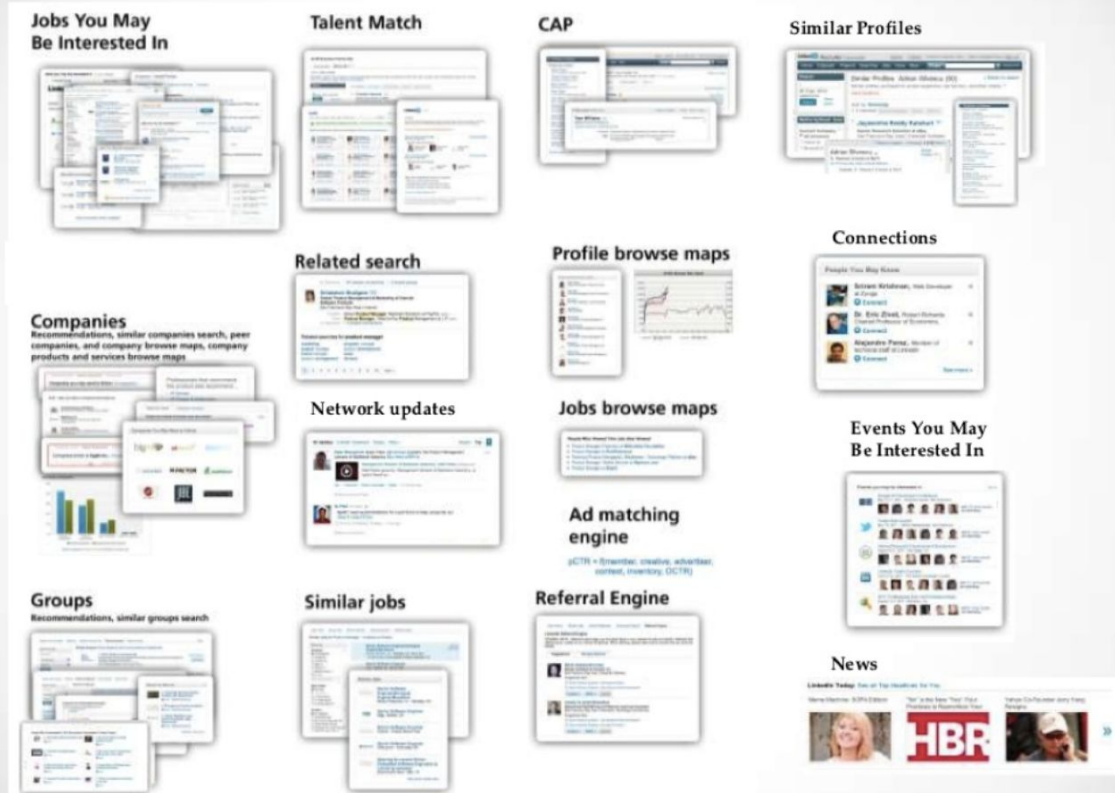
Cold Start Problem

What cases of cold start exist?

Job Recommendation - How would you do it?

LinkedIn

The Recommender Ecosystem



News Recommendation - How would you do it?

News Recommendation

