Distinguishing Influence and Homophily

- Shuffle Test
- Edge-Reversal Test
- Randomization Test

Distinguishing Influence and Homophily

- Which social force (influence or homophily) resulted in an assortative network?
- To distinguish between an influence-based assortativity or homophily-based one, statistical tests can be used
- In all these tests, we assume that several temporal snapshots of the dataset are available (like LTM) where we know exactly, when each node is activated, when edges are formed, or when attributes are changed

I. Shuffle Test (Influence)

IDEA:

- Influence is temporal
- If u influences v, then u should have been activated before v.
- Define a temporal assortativity measure
- If there is no influence, then *a shuffling of the activation timestamps* should not affect the temporal assortativity measurement

Shuffle Test of Influence

If influence does not play a role, the timing of activations should be independent of users.

Even if we randomly shuffle the timestamps of user activities, we should obtain a similar temporal assortativity value

User	A	B	C
Time	1	2	3

→

User	A	B	C
Time	2	3	1

Test of Influence

After we shuffle the timestamps of user activities, if the new estimate of temporal assortativity is significantly different from the original estimate based on the user's activity log, **there is evidence of influence**.

Measuring Temporal Assortativity

- Assume node activation probability depends on a, the number of already-active friends of the node.
  - Denote the probability as $\mathrm{p}(a)$

- Assume $p(a)$ can be estimated using a logistic function

$$p(a) = \frac{e^{\alpha a+\beta}}{1+e^{\alpha a+\beta}} \quad \Longrightarrow \quad \ln \frac{p(a)}{1-p(a)} = \alpha a + \beta$$

- a is the number of active friends,
- α is the temporal assortativity (**social correlation**) : **variable**
- β is a constant to explain the innate bias for activation

Activation Likelihood

Suppose at time t

- $y_{a,t}$ users with a active friends become active
- $n_{a,t}$ users with a active friends, stay inactive
- Number of users with a friends activated/not-activated at any time

$$y_a = \sum_t y_{a,t} \qquad n_a = \sum_t n_{a,t}$$

The probability of observing your data (**likelihood function**) is

$$\prod_a p(a)^{y_a}(1-p(a))^{n_a}$$

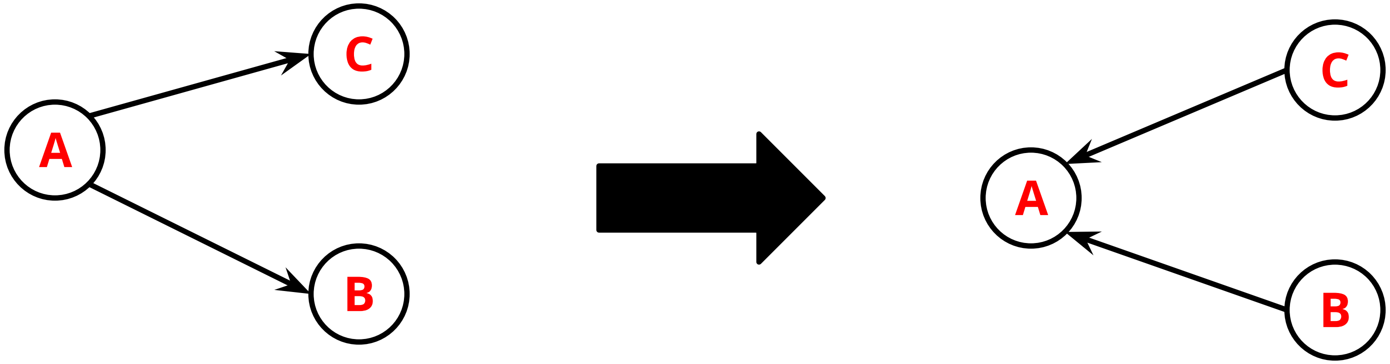
Given the user's activity log, we can compute a correlation coefficient α and bias β to maximize the above likelihood

- Using a maximum likelihood iterative method

2. The Edge-reversal Test (Influence)

If influence resulted in activation, then the direction of edges should be important (who influenced whom).

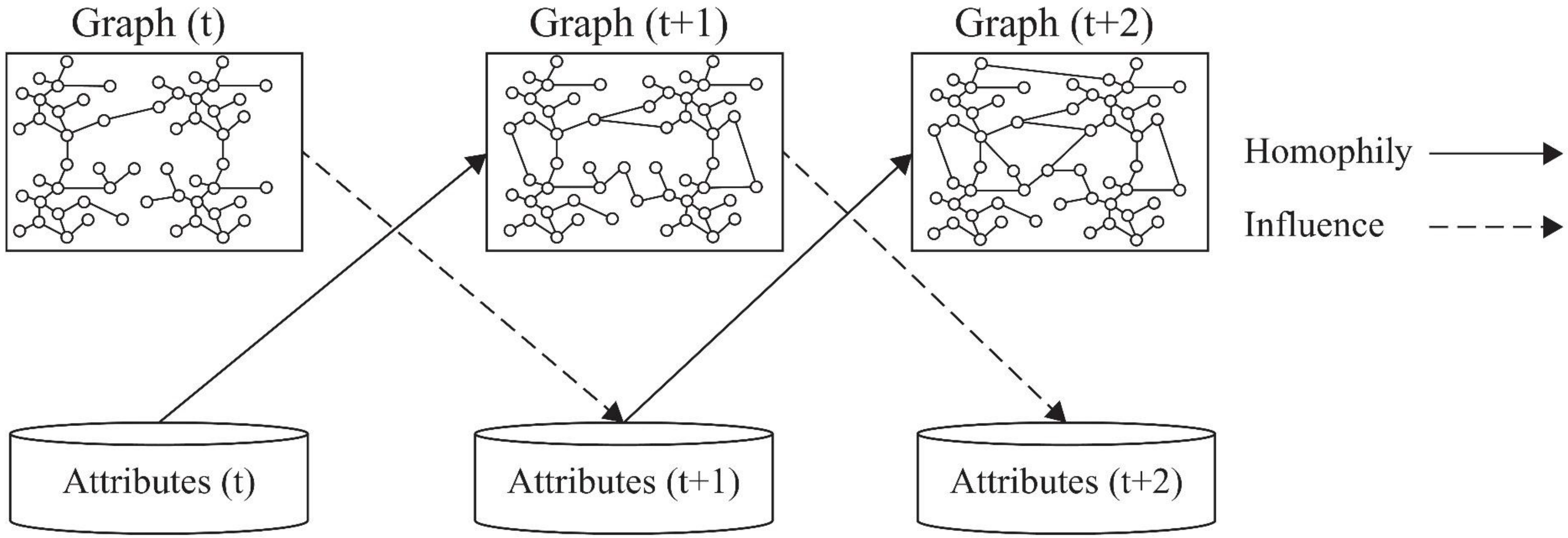
- Reverse directions of all the edges
- Run the same logistic regression on the data using the new graph
- If correlation is not due to influence, then α should not change



3. Randomization Test (Influence/Homophily)

- Capable of detecting both Influence and Homophily in networks

- *Influence* changes *attributes* and *Homophily* changes *connections*



Notation and Preliminaries

- X denotes node attributes
  - X^i denotes the attributes of node v_i
  - X_t denotes the attributes of nodes at time t

- $A(G_t, X_t)$ denotes the assortativity of network G and attributes X at time t

- The network becomes more assortative at time t if

$$A(G_{t+1}, X_{t+1}) - A(G_t, X_t) > 0$$

Influence Gain and Homophily Gain

- If the assortativity is due to influence, **Influence gain** is positive

$$G_{Influence}(t) = A(G_t, X_{t+1}) - A(G_t, X_t) > 0$$

- If the assortativity is due to homophily, **Homophily gain** is positive

$$G_{Homophily}(t) = A(G_{t+1}, X_t) - A(G_t, X_t) > 0$$

- In randomization test, we check if these gains are **significant**

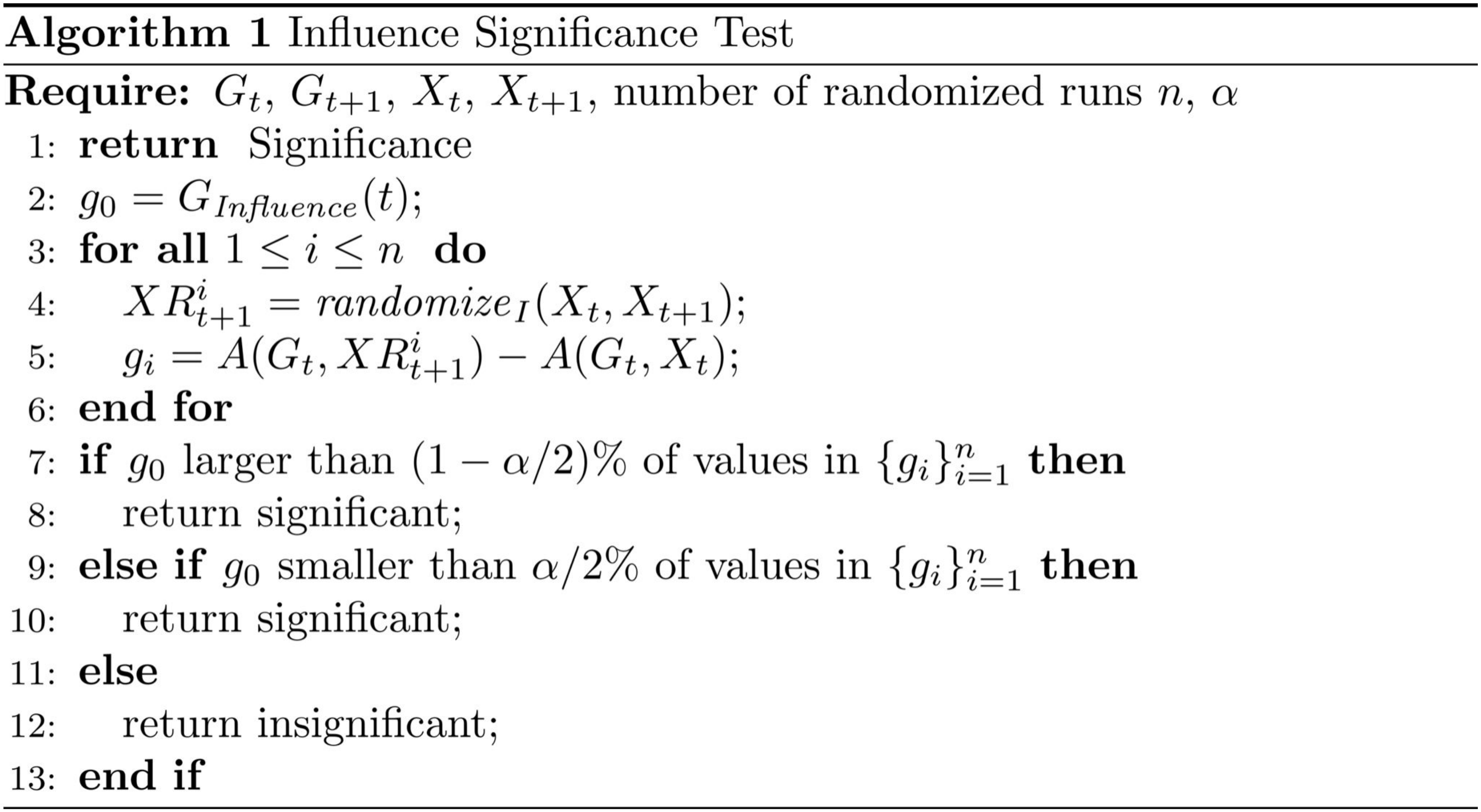
Influence Significance Test

- Compute influence gain at time t
  - Denote as g_0
- Compute n random attributes sets for time t+1
  - Denote as XR_{t+1}^{i} , $1 \leq i \leq n$

  Example
  - u has influence over v
  - movies is in hobbies of u at time t, but not in hobbies of v at time t
  - At time $t + 1$ movies is added to hobbies of v
  - To remove influence effect, we remove movies from hobbies of v at time $t + 1$ and replace it with some ***random*** hobby (e.g., reading)

- Compute the [random] influence gain for all XR_{t+1}^{i} sets
  - Call them g_i
- If g_0 is greater than $\left(1 - \frac{\alpha}{2}\right)$% of all g_i's (or smaller than $\left(\frac{\alpha}{2}\right)$% of them), the influence gain is **significant**

Influence Significance Test



Algorithm 1 Influence Significance Test

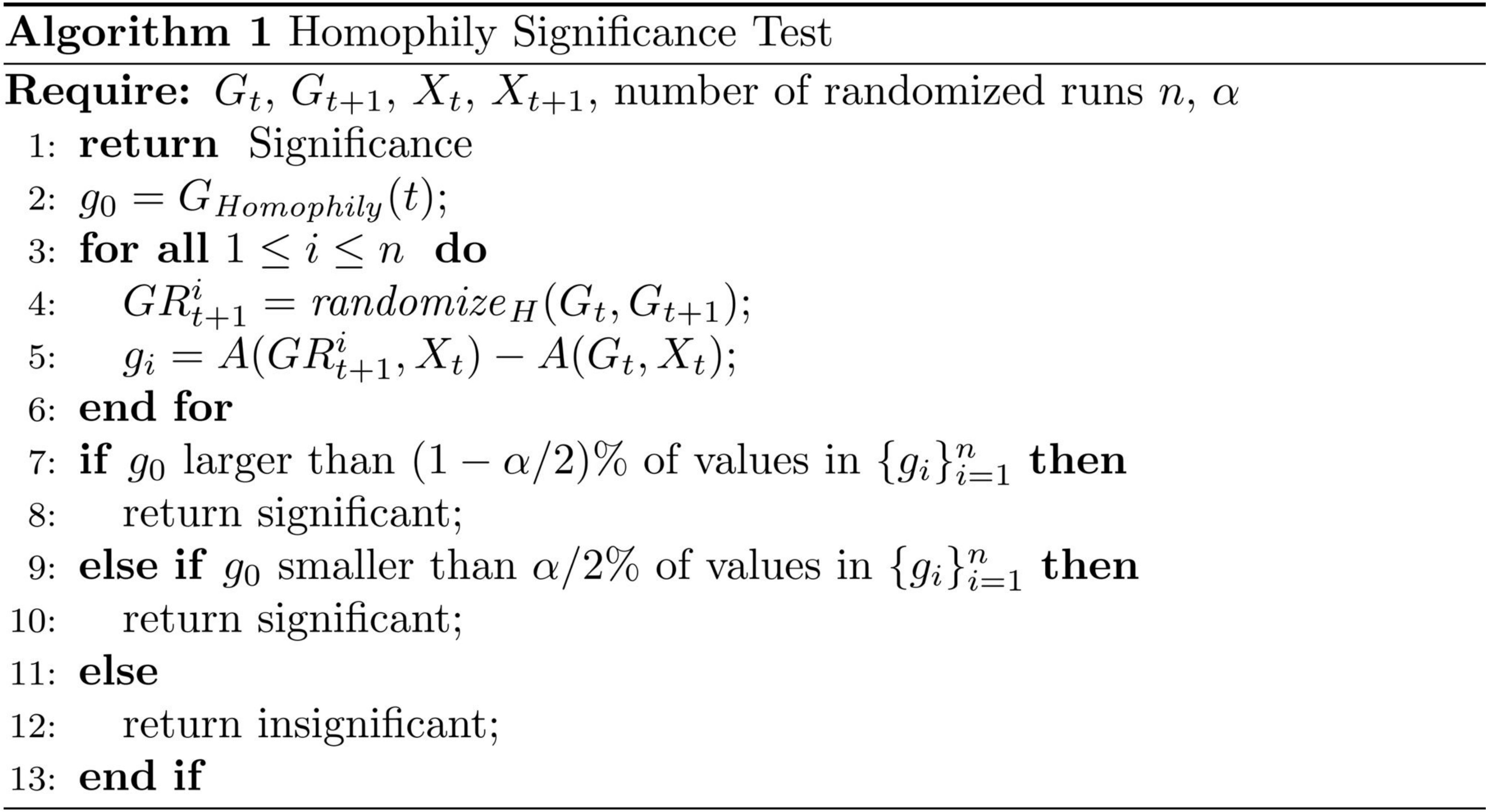
```
Require: $G_t$, $G_{t+1}$, $X_t$, $X_{t+1}$, number of randomized runs $n$, $\alpha$
return Significance
$g_0 = G_{Influence}(t)$;
for all $1 \leq i \leq n$ do
    $XR^i_{t+1} = randomize_I(X_t, X_{t+1})$;
    $g_i = A(G_t, XR^i_{t+1}) - A(G_t, X_t)$;
end for
if $g_0$ larger than $(1 - \alpha/2)\%$ of values in $\{g_i\}^n_{i=1}$ then
    return significant;
else if $g_0$ smaller than $\alpha/2\%$ of values in $\{g_i\}^n_{i=1}$ then
    return significant;
else
    return insignificant;
end if
```

Homophily Significance Test

- We construct random graphs with fixed attribute sets

- We remove the effect of homophily by generating n random graphs GR_{t+1}^{i} at time $t+1$
  - For any two (randomly selected) edges e_{ij} and e_{kl} formed in the original graph G_{t+1}
  - We form edges e_{il} and e_{kj}
  - Homophily effect removed / degrees stay the same

Homophily Significance Test



Algorithm 1 Homophily Significance Test

```
Require: $G_t$, $G_{t+1}$, $X_t$, $X_{t+1}$, number of randomized runs $n$, $\alpha$
return Significance
$g_0 = G_{Homophily}(t)$;
for all $1 \leq i \leq n$ do
    $GR^i_{t+1} = randomize_H(G_t, G_{t+1})$;
    $g_i = A(GR^i_{t+1}, X_t) - A(G_t, X_t)$;
end for
if $g_0$ larger than $(1 - \alpha/2)\%$ of values in $\{g_i\}^n_{i=1}$ then
    return significant;
else if $g_0$ smaller than $\alpha/2\%$ of values in $\{g_i\}^n_{i=1}$ then
    return significant;
else
    return insignificant;
end if
```

CS 579: Online Social Network Analysis

Recommendation in Social Media

Kai Shu

Spring 2023

Difficulties of Making Decision

- Which digital camera should I buy?
- Where should I spend my holiday?
- Which movie should I rent?
- Whom should I follow?
- Where should I find interesting news article?
- Which movie is the best for **our family**?

- If interested, see some conference tutorials
  - WWW2015 Like and Recommendation in Social Media
  - SIGKDD2014, Recommendation in Social Media
  - RecSys2014, Personalized Location Recommendation

When Does This Problem Occur?

- There are too many choices
- There are no obvious advantages among them
- We do not have enough resources (time) to check all options (information overload)
- We do not have enough knowledge and experience to choose, or
  - I'm lazy, but don't want to miss out on good stuff
  - Defensive decision making

Goal of Recommendation:

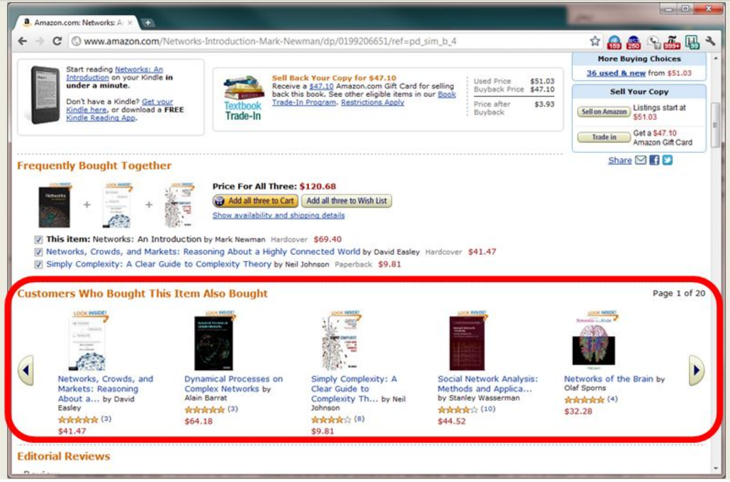
To come up with a short list of items that fits user's interests

Common Solutions to the Problem

- Consulting friends
- Obtaining information from a trusted third party
- Hiring a team of experts
- Search the Internet
- Following the crowd (pick the item from top-n lists, e.g., best sellers on Amazon)
- **Can we automate all of the above?**
  - **Using a recommendation algorithm (a.k.a. a recommender system)**

Recommender Systems - Examples

Book recommendation in Amazon



www.amazon.com/Networks-Introduction-Mark-Newman/dp/0199206651/ref=pd_sim_b_4

Start reading Networks: An Introduction on your Kindle in under a minute.

Don't have a Kindle? Get your Kindle here, or download a FREE Kindle Reading App.

Sell Back Your Copy for $47.10
Receive a $47.10 Amazon.com Gift Card for selling back this book. See other eligible items in our Book Trade-In Program. Restrictions Apply

Used Price	$51.03
Buyback Price	$47.10
Price after Buyback	$3.93

More Buying Choices
36 used & new from $51.03

Sell Your Copy
Sell on Amazon — Listings start at $51.03
Trade in — Get a $47.10 Amazon Gift Card

Frequently Bought Together

Price For All Three: $120.68

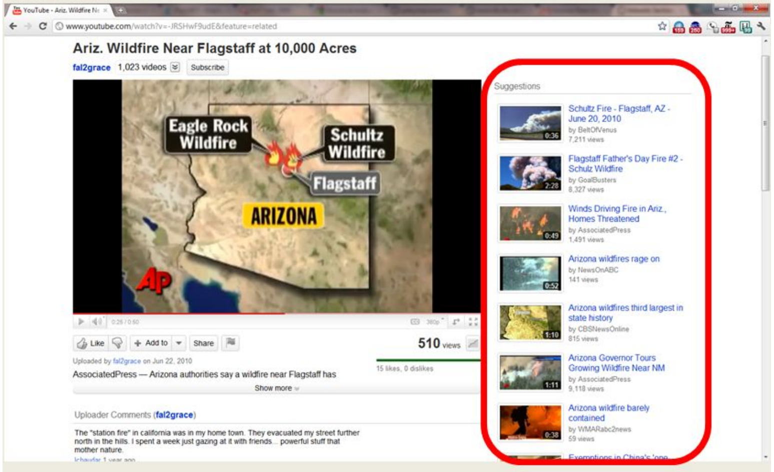
☑ This item: Networks: An Introduction by Mark Newman Hardcover $69.40
☑ Networks, Crowds, and Markets: Reasoning About a Highly Connected World by David Easley Hardcover $41.47
☑ Simply Complexity: A Clear Guide to Complexity Theory by Neil Johnson Paperback $9.81

Customers Who Bought This Item Also Bought — Page 1 of 20

- Networks, Crowds, and Markets: Reasoning About a... by David Easley ★★★★★ (3) $41.47
- Dynamical Processes on Complex Networks by Alain Barrat ★★★★★ (3) $64.18
- Simply Complexity: A Clear Guide to Complexity Th... by Neil Johnson ★★★★☆ (8) $9.81
- Social Network Analysis: Methods and Applica... by Stanley Wasserman ★★★★☆ (10) $44.52
- Networks of the Brain by Olaf Sporns ★★★★★ (4) $32.28

Editorial Reviews

Video clip recommendation in YouTube



www.youtube.com/watch?v=-JRSHwF9udE&feature=related

Ariz. Wildfire Near Flagstaff at 10,000 Acres

fal2grace 1,023 videos Subscribe

510 views — 15 likes, 0 dislikes

Uploaded by fal2grace on Jun 22, 2010

AssociatedPress — Arizona authorities say a wildfire near Flagstaff has

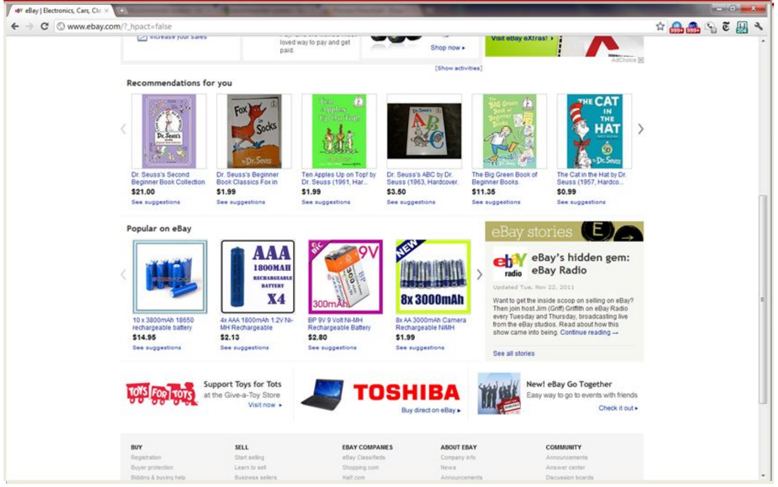
Uploader Comments (fal2grace)

The "station fire" in california was in my home town. They evacuated my street further north in the hills. I spent a week just gazing at it with friends... powerful stuff that mother nature.

Suggestions

- Schultz Fire - Flagstaff, AZ - June 20, 2010 by BeltOfVenus 7,211 views (0:36)
- Flagstaff Father's Day Fire #2 - Schulz Wildfire by GoalBusters 8,327 views (2:28)
- Winds Driving Fire in Ariz., Homes Threatened by AssociatedPress 1,491 views (0:49)
- Arizona wildfires rage on by NewsOnABC 141 views (0:52)
- Arizona wildfires third largest in state history by CBSNewsOnline 815 views (1:10)
- Arizona Governor Tours Growing Wildfire Near NM by AssociatedPress 9,118 views (1:11)
- Arizona wildfire barely contained by WMARabc2news 59 views (0:38)

Product Recommendation in ebay



www.ebay.com/?_hpact=false

Recommendations for you

Dr. Seuss's Second Beginner Book Collection	$21.00
Dr. Seuss's Beginner Book Classics Fox in	$1.99
Ten Apples Up on Top! by Dr. Seuss (1961, Har...	$1.99
Dr. Seuss's ABC by Dr. Seuss (1963, Hardcover	$3.50
The Big Green Book of Beginner Books	$11.35
The Cat in the Hat by Dr. Seuss (1957, Hardco...	$0.99

Popular on eBay

10 x 3800mAh 18650 rechargeable battery	$14.95
4x AAA 1800mAh 1.2V Ni-MH Rechargeable	$2.13
BP 9V 9 Volt Ni-MH Rechargeable Battery	$2.80
8x AA 3000mAh Camera Rechargeable NiMH	$1.99

eBay stories

eBay's hidden gem: eBay Radio

Updated Tue, Nov 22, 2011

Want to get the inside scoop on selling on eBay? Then join host Jim (Griff) Griffith on eBay Radio every Tuesday and Thursday, broadcasting live from the eBay studios. Read about how this show came into being. Continue reading →

See all stories

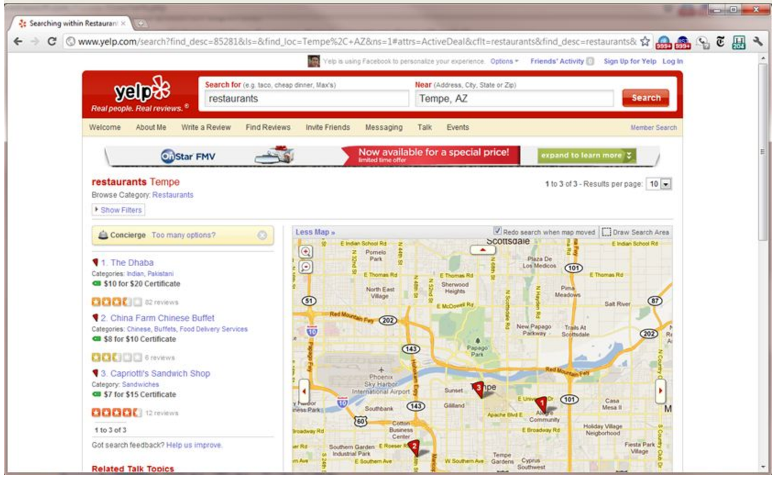
Support Toys for Tots at the Give-a-Toy Store Visit now ▸

TOSHIBA Buy direct on eBay ▸

New! eBay Go Together Easy way to go to events with friends Check it out ▸

BUY — Registration, Buyer protection, Bidding & buying help
SELL — Start selling, Learn to sell, Business sellers
EBAY COMPANIES — eBay Classifieds, Shopping.com, Half.com
ABOUT EBAY — Company info, News, Announcements
COMMUNITY — Announcements, Answer center, Discussion boards

Restaurant Recommendation in Yelp



www.yelp.com/search?find_desc=85281&ls=&find_loc=Tempe%2C+AZ&ns=1#attrs=ActiveDeal&cflt=restaurants&find_desc=restaurants&

Yelp is using Facebook to personalize your experience. Options ▾ Friends' Activity Sign Up for Yelp Log In

Search for (e.g. taco, cheap dinner, Max's): restaurants
Near (Address, City, State or Zip): Tempe, AZ

restaurants Tempe
Browse Category: Restaurants

1 to 3 of 3 - Results per page: 10

1. The Dhaba
Categories: Indian, Pakistani
$10 for $20 Certificate
82 reviews

2. China Farm Chinese Buffet
Categories: Chinese, Buffets, Food Delivery Services
$8 for $10 Certificate
6 reviews

3. Capriotti's Sandwich Shop
Category: Sandwiches
$7 for $15 Certificate
12 reviews

1 to 3 of 3

Got search feedback? Help us improve.

Related Talk Topics

Main Idea behind Recommender Systems

Use historical data such as the user's past preferences or similar users' past preferences to predict future likes

- Users' preferences are likely to remain stable, and change smoothly over time.
  - By watching the past users' or groups' preferences, we try to predict their future likes
  - Then we can recommend items of interest to them
- Formally, a recommender system takes a set of users U and a set of items I and learns a function f such that:

$$f : U \times I \rightarrow \mathbb{R}$$

Recommendation vs. Search

- One way to get answers is using search engines
- Search engines find results that match the query provided by the user
  - You have to search!
- The results are generally provided as a list ordered with respect to the relevance of the item to the given query
- Consider the query “best 2014 movie to watch”
  - The same results for an 8 year old and an adult

Search engines’ results are not personalized

Challenges (I)

- The Cold Start Problem
  - Many recommendation systems use historical data or information provided by the user to recommend items, products, etc. However, when individuals join sites, they haven’t bought any product, or they have **no history**. This makes it hard to infer what they are going to like when they start on a site. The problem is referred to as the *cold-start* problem.
- Data Sparsity
  - Similar to the cold-start problem, *data sparsity* is when not enough historical or prior information is available. Unlike the cold start problem, this is in the **system as a whole** and is not specific to an individual.

Challenges (II)

- Attacks
  - Push Attack: pushing the ratings up by creating fake users
  - Nuke attack: DDoS attacks, stop the whole recommendation systems
- Privacy
  - Employing user’s private information to recommend to others.
- Explanation
  - Recommendation systems often recommend items without any explanation of why recommending them

Classical Recommendation Algorithms

- Content-based algorithms
- Collaborative filtering

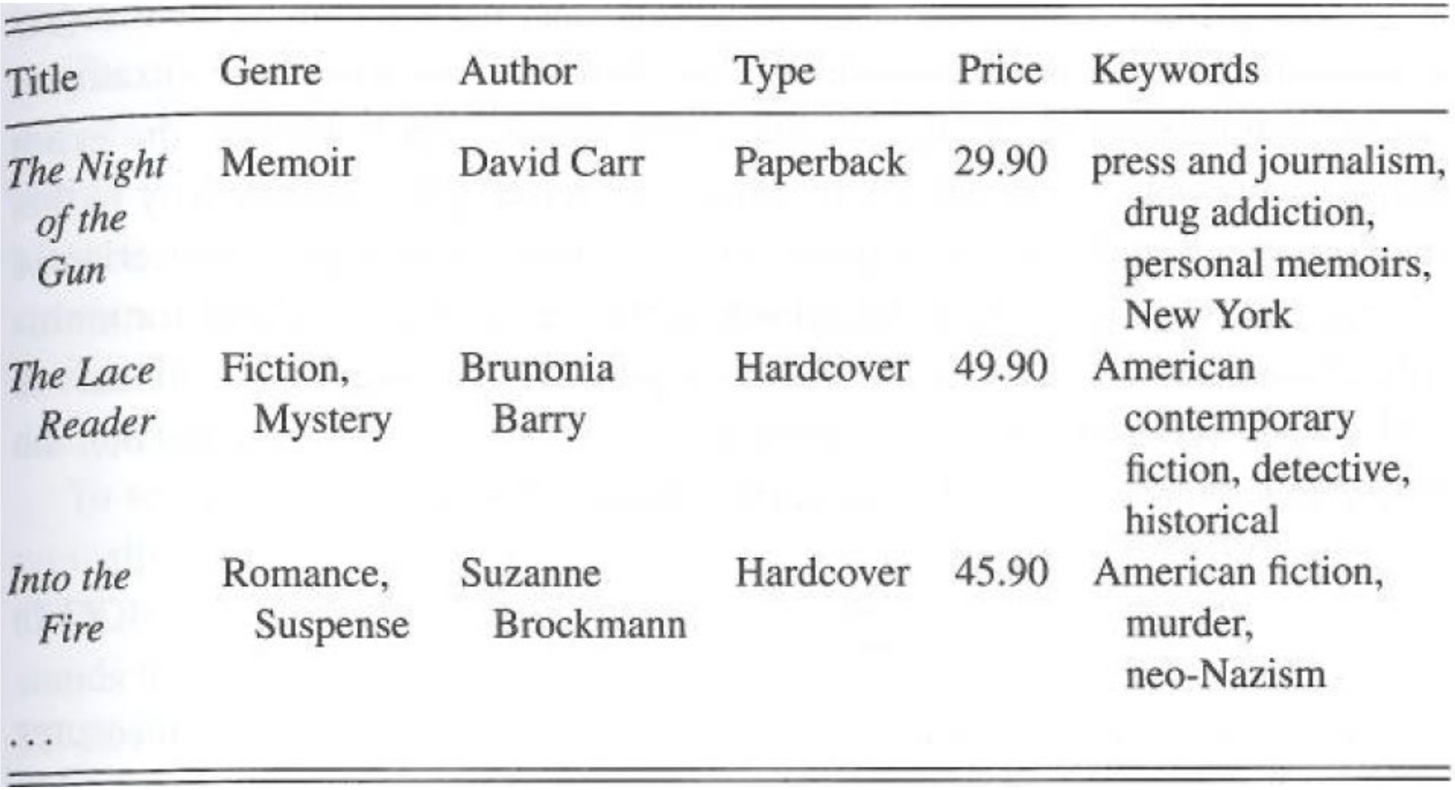
Content-Based Methods

- Content-based recommendation systems are based on the fact that a **user's interest** should match the **description of the items** that she should be recommended by the system.
- In other words, the more similar the item's description to that of the user's interest, the more likely that the user finds the item's recommendation interesting.

Finding the similarity between the user and all of the existing items is the core of this type of recommender systems

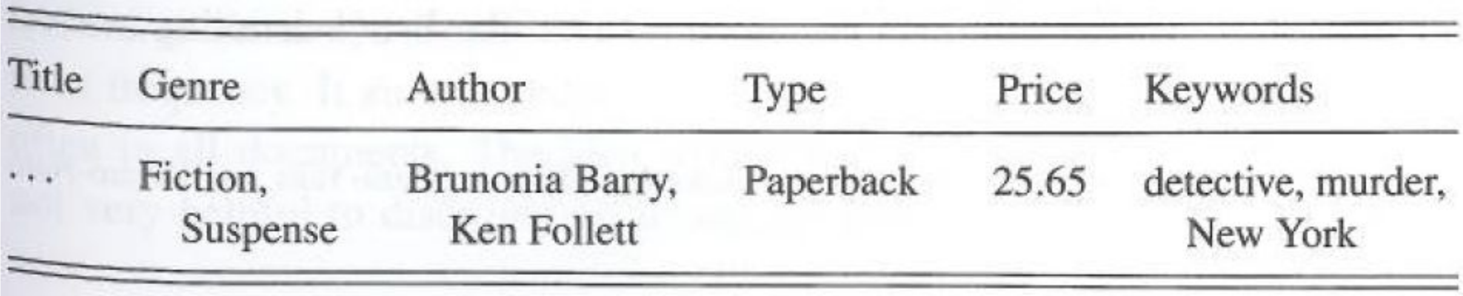
Content-based Recommendation: An Example

Book knowledge base



Title	Genre	Author	Type	Price	Keywords
The Night of the Gun	Memoir	David Carr	Paperback	29.90	press and journalism, drug addiction, personal memoirs, New York
The Lace Reader	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
Into the Fire	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism
. . .					

User Profile



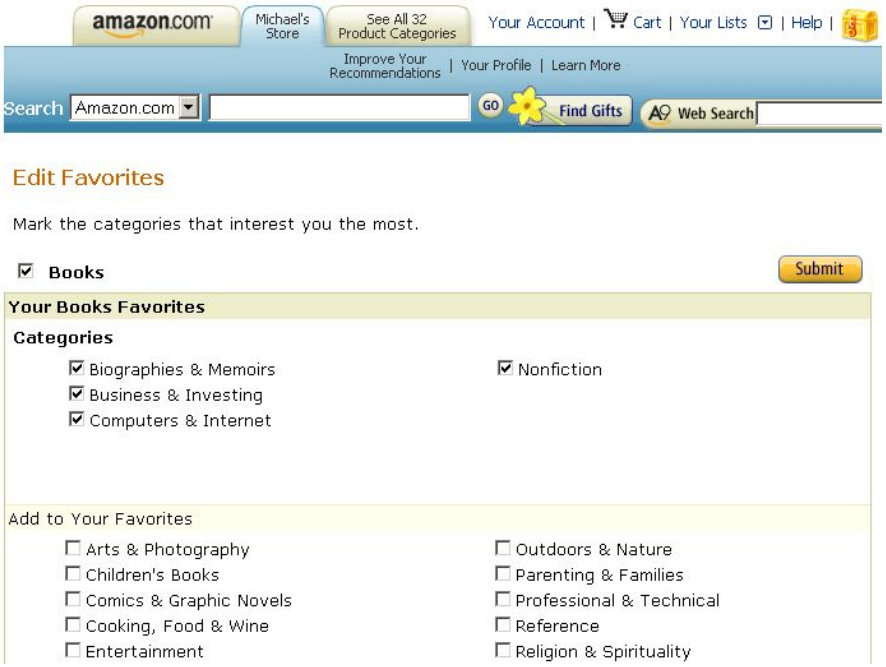
Title	Genre	Author	Type	Price	Keywords
. . .	Fiction, Suspense	Brunonia Barry, Ken Follett	Paperback	25.65	detective, murder, New York

Content-based Recommendation Algorithm

1. *Describe the items to be recommended*
2. *Create a profile of the user that describes the types of items the user likes*
3. *Compare items with the user profile to determine what to recommend*

The profile is often created, and updated automatically in response to feedback on the desirability of items that are presented to the user

Content-based Recommendation: Example


amazon.com
Michael's Store
See All 32 Product Categories
Your Account | Cart | Your Lists | Help |
Improve Your Recommendations | Your Profile | Learn More
Search Amazon.com
GO
Find Gifts
Web Search

Edit Favorites

Mark the categories that interest you the most.

☑ **Books** Submit

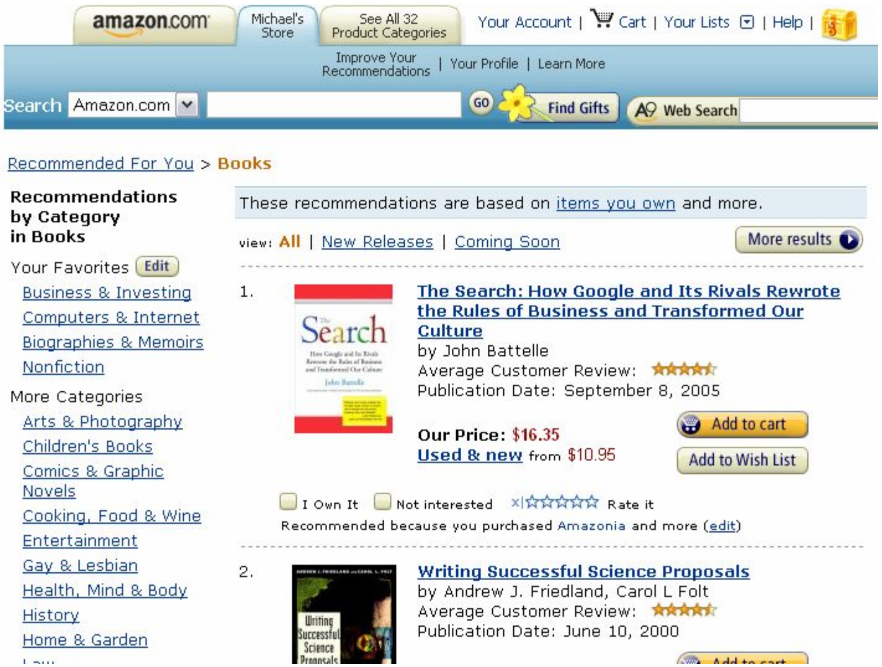
Your Books Favorites

Categories

☑ Biographies & Memoirs
☑ Business & Investing
☑ Computers & Internet
☑ Nonfiction

Add to Your Favorites

□ Arts & Photography
□ Children's Books
□ Comics & Graphic Novels
□ Cooking, Food & Wine
□ Entertainment
□ Outdoors & Nature
□ Parenting & Families
□ Professional & Technical
□ Reference
□ Religion & Spirituality


amazon.com
Michael's Store
See All 32 Product Categories
Your Account | Cart | Your Lists | Help |
Improve Your Recommendations | Your Profile | Learn More
Search Amazon.com
GO
Find Gifts
Web Search

Recommended For You > **Books**

Recommendations by Category in Books

Your Favorites Edit

Business & Investing
Computers & Internet
Biographies & Memoirs
Nonfiction

More Categories

Arts & Photography
Children's Books
Comics & Graphic Novels
Cooking, Food & Wine
Entertainment
Gay & Lesbian
Health, Mind & Body
History
Home & Garden

These recommendations are based on items you own and more.

view: **All** | New Releases | Coming Soon More results

1. **The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture**
by John Battelle
Average Customer Review: ★★★★☆
Publication Date: September 8, 2005
Our Price: $16.35
Used & new from $10.95
Add to cart
Add to Wish List
□ I Own It □ Not interested ☆☆☆☆☆ Rate it
Recommended because you purchased Amazonia and more (edit)

2. **Writing Successful Science Proposals**
by Andrew J. Friedland, Carol L Folt
Average Customer Review: ★★★★☆
Publication Date: June 10, 2000

Content-Based Methods

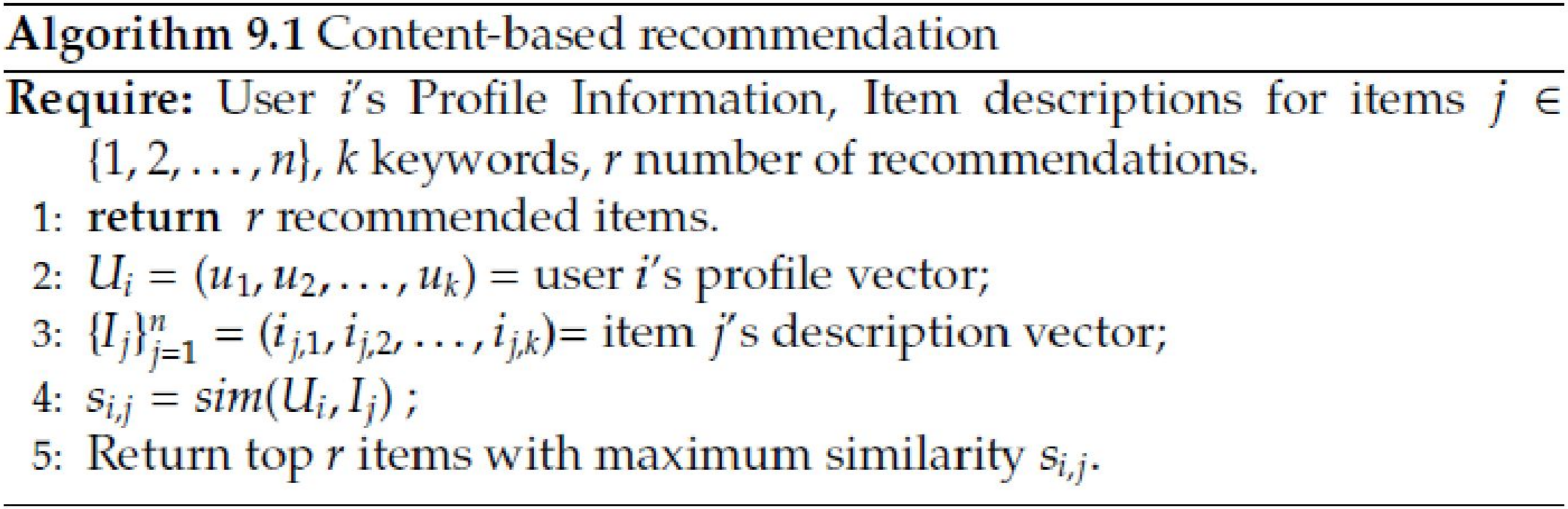
- To formalize a content-based method, we first represent both user profiles and item descriptions by vectorizing them using a set of *k* keywords
- We can vectorize (e.g., using TF-IDF) both users and items and compute their similarity

$$I_j = (i_{j,1}, i_{j,2}, \ldots, i_{j,k}) \qquad U_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,k}).$$

$$sim(U_i, I_j) = cos(U_i, I_j) = \frac{\sum_{l=1}^{k} u_{i,l} i_{j,l}}{\sqrt{\sum_{l=1}^{k} u_{i,l}^2} \sqrt{\sum_{l=1}^{k} i_{j,l}^2}}$$

- We can recommend the top most similar items to the user

Content-Based Recommendation Algorithm



Algorithm 9.1 Content-based recommendation

```
Require: User $i$'s Profile Information, Item descriptions for items $j \in \{1, 2, \ldots, n\}$, $k$ keywords, $r$ number of recommendations.
return $r$ recommended items.
$U_i = (u_1, u_2, \ldots, u_k)$ = user $i$'s profile vector;
$\{I_j\}_{j=1}^{n} = (i_{j,1}, i_{j,2}, \ldots, i_{j,k})$= item $j$'s description vector;
$s_{i,j} = sim(U_i, I_j)$ ;
Return top $r$ items with maximum similarity $s_{i,j}$.
```

- In content-based recommendation, we compute the topmost similar items to a user *i* and then recommend these items in the order of similarity