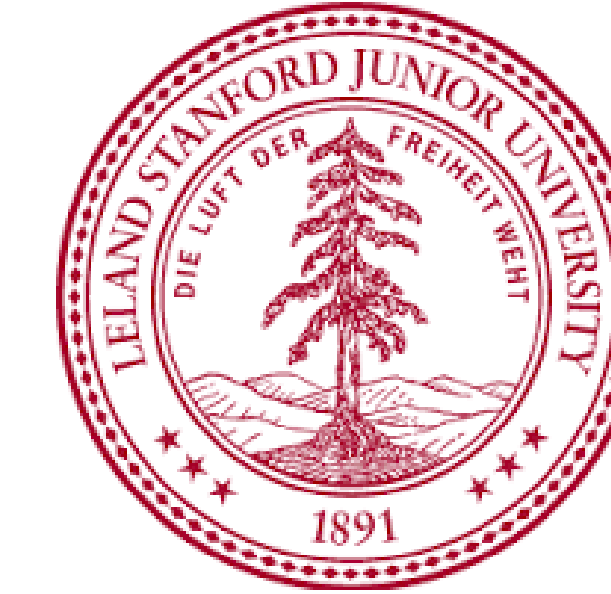
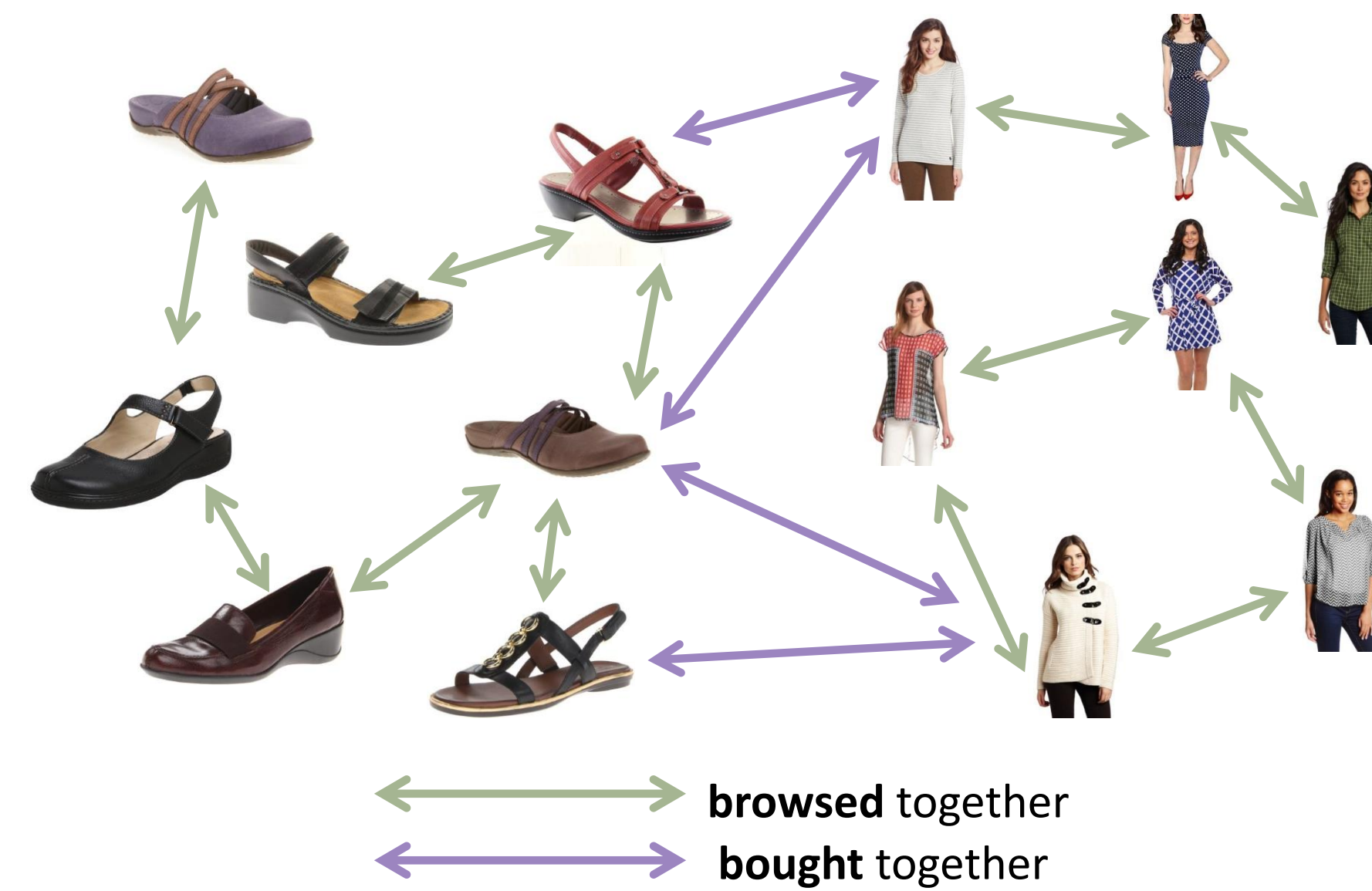


Inferring networks of substitutable and complementary products

Julian McAuley, UCSD; Rahul Pandey, Pinterest; Jure Leskovec, Stanford



Goal: to infer networks of **substitutable** and **complementary** items

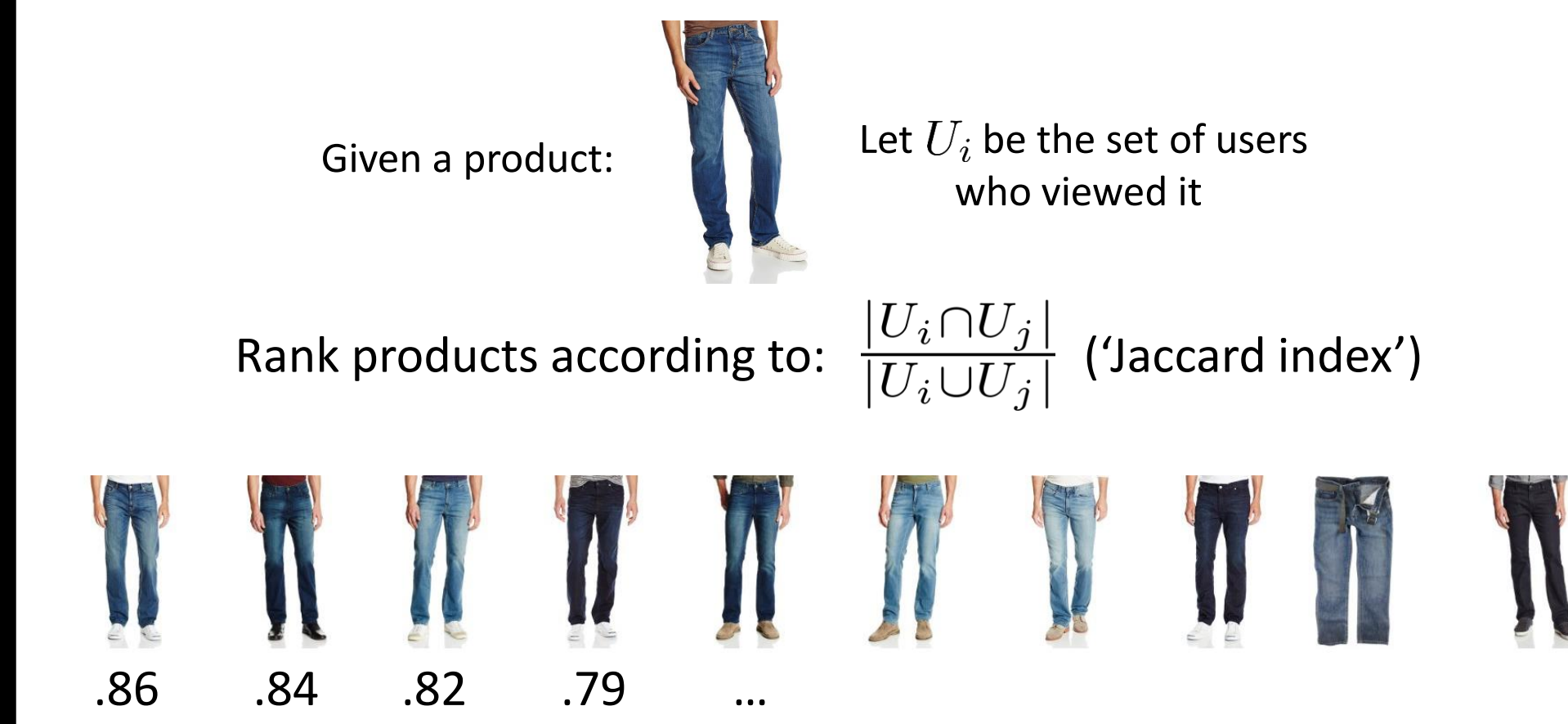


Data:
<http://cseweb.ucsd.edu/~jmcauley/>



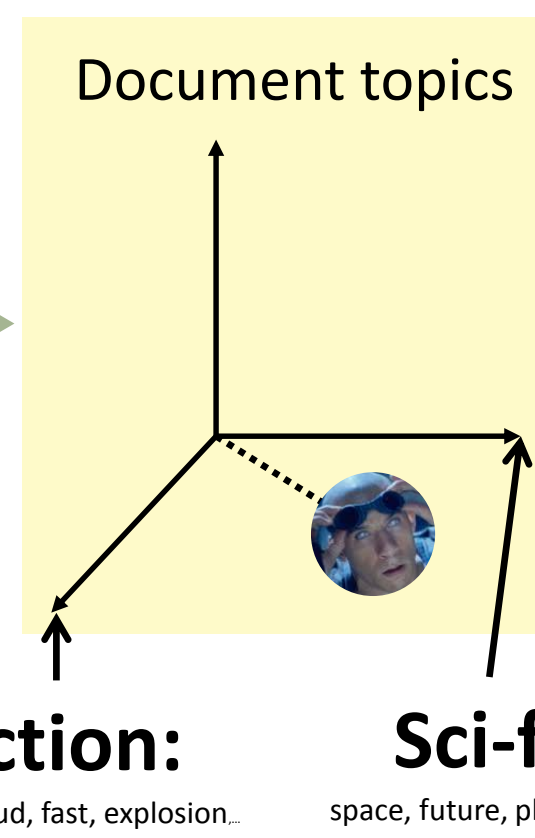
- Amazon product network:
- thousands of **categories**
 - 9 million **products**
 - 21 million **users**
 - 140 million **reviews & ratings**
 - 300 million **relationships** (co-purchased & co-browsed)

Currently, this is solved using **collaborative filtering** approaches, but this works poorly for new items, and is not interpretable

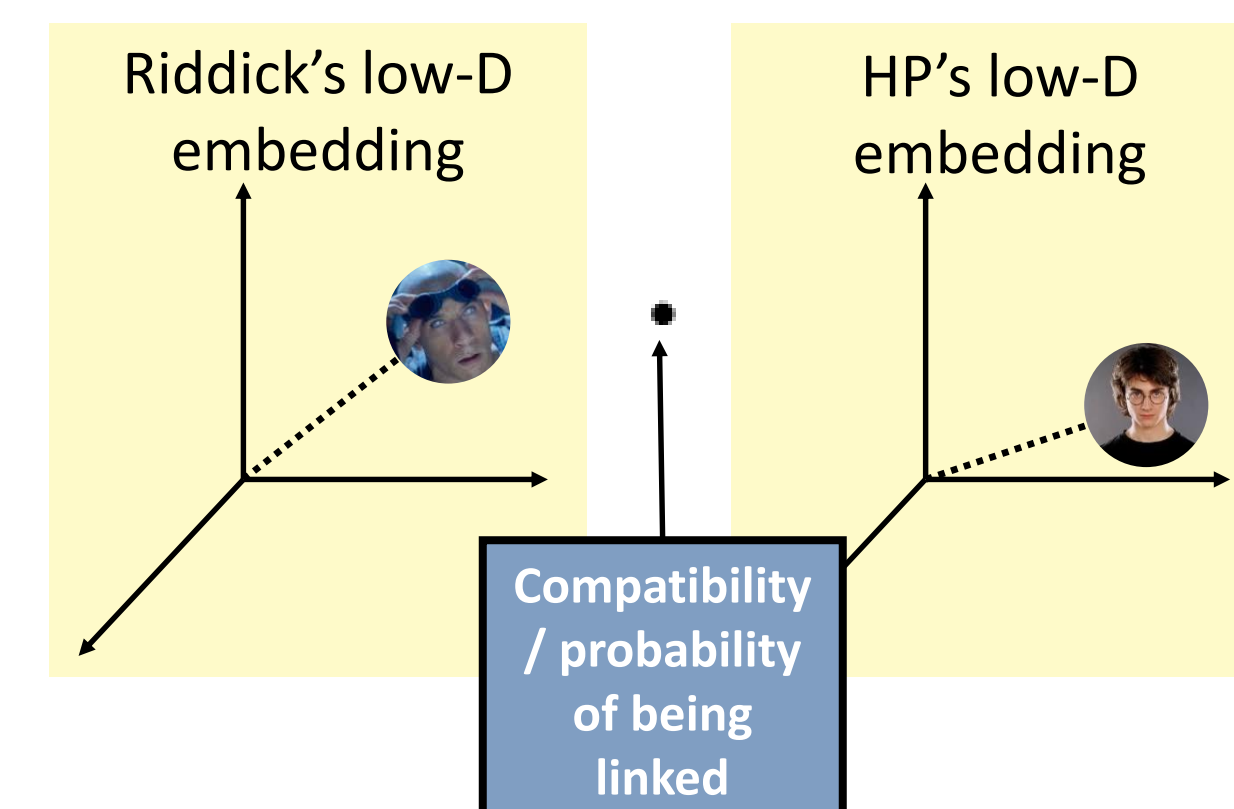


A standard way to model **documents** is to project them into a low-dimensional space, where dimensions encode "topics"

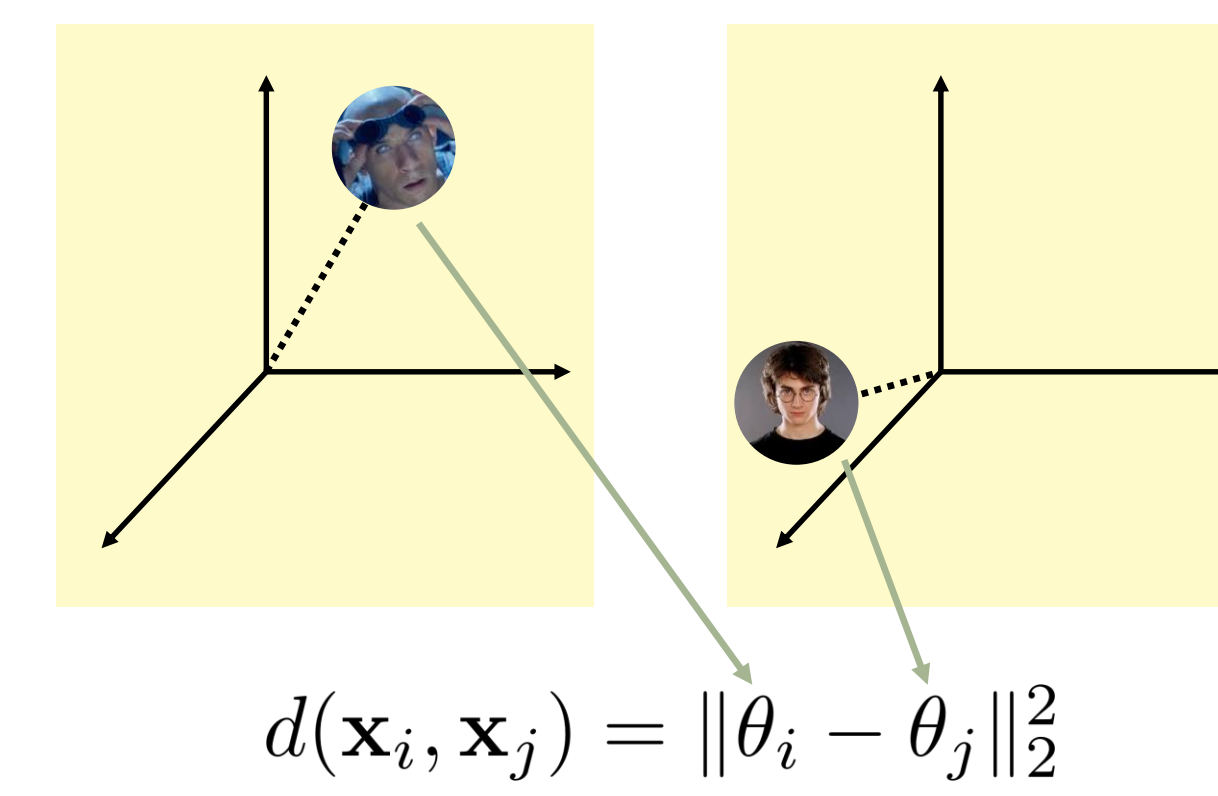
87 of 102 people found the following review helpful
 ★★★★★ **You keep what you kill**, December 27, 2004
 By **Schinsky "Schinsky"** (Washington State) - See all my reviews
 ...
 This review is from: **The Chronicles of Riddick: Vengeance** (Directed by James Cameron)
 Even if I have to apologize to my friends and family, and my family, I have to admit that I really liked this movie. It's a Sci-Fi movie with a "Mad Max" appeal that, while changing many things, left Riddick from "Pitch Black" to be just Riddick. They did not change his attitude or soften him up or bring him out of his original character, which was very pleasing to "Pitch Black" fans like myself!



A standard way to model **networks** is to project items into a low-dimensional space, such that related items are projected nearby



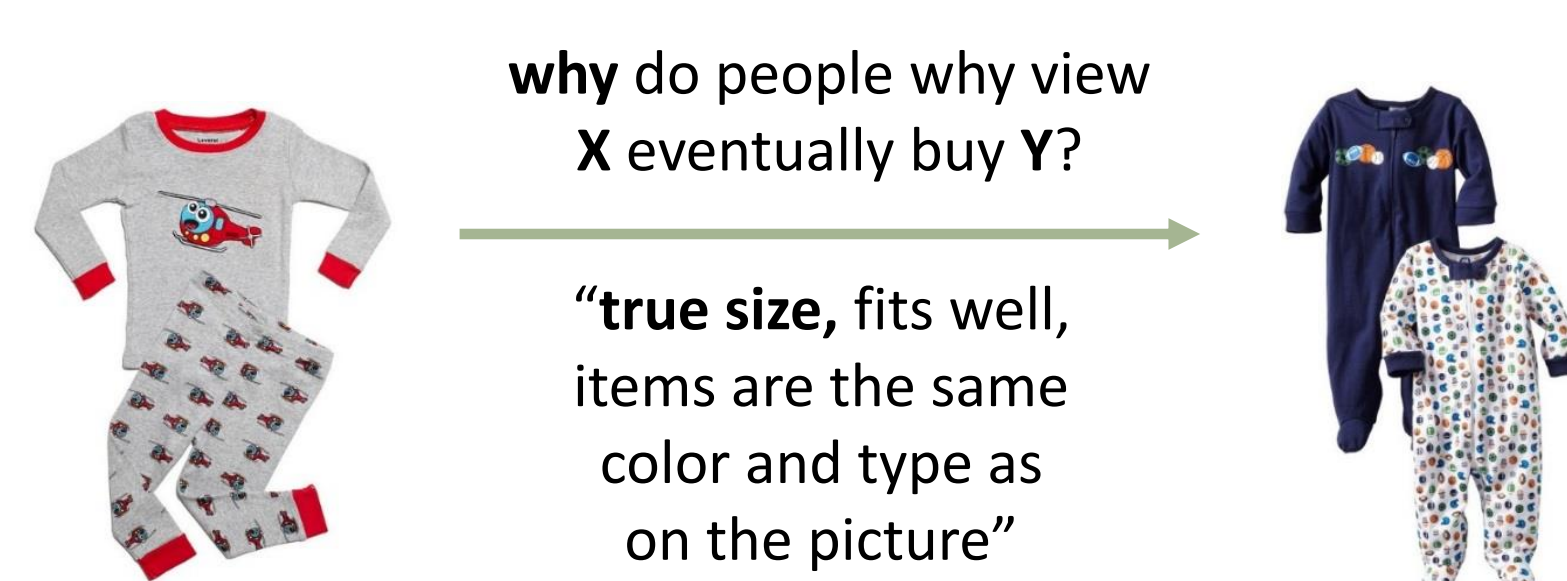
Our idea is to combine the two approaches, by learning topic representations such that related items have similar topics



We achieve this by fitting a joint objective that combines a logistic regressor (to model the network) with a topic model (to model reviews)

$$\theta = \arg \max_{\theta} \underbrace{\prod_{\text{edges } (x,y)} p_{\theta}(x \text{ and } y \text{ are related})}_{\text{logistic regression}} \underbrace{\prod_{\text{non-edges } (x,y)} (1 - p_{\theta}(x \text{ and } y \text{ are related}))}_{\text{topic model}} \times p(\text{review corpus}|\theta)$$

Outcome 1: We can **explain** product relationships in terms of interpretable textual dimensions



e.g. (above) select a representative sentence from a review that best explains the predicted link

Outcome 2: We discover topics and 'microcategories' that are associated with product relationships

Electronics										
e111 cameras	e92 portable speakers	e75 cases	e79 Samsung cases	e78 Onebox cases	e50 styli	e69 batteries	e85 portable radios	e96 car radios	e89 high-end headphones	e99 budget headphones
camera zoom	little speaker	leather case	Galaxy elastic	Defender protection	pen tip	batteries	radio weather	Pioneer crank	Bose sound	Skullcandy head
pictures	portable speaker	soft magnets	Samsung velcro	leather kids	Bamboo rechargable	Wacom stylus	Lenmar solar	Motorola dash	Shure Beats	another pair
Kodak Canon	small speaker	Roccase closed	leather auto	Survivor protected	friendly Swede	Survivor pencil	Sanyo rechargeable	Headset	Alk	comfortable gym
flash digital	Home	material	stung closing	safe elastic strap	digital	auto	Lenmar alkaline	Jeep music	Beats	head
optical	wireless speaker	great speaker	mini speaker	standing	taken	mini speaker	full charge	fm	classical	Beats
picture										

Men's clothing										
e44 dress shirts	e107 dress shirts	e75 dress pants	e49 three-wolf shirt	e52 polo shirts	e110 boots	e156 minimalist running	e134 athletic running	e133 sports shoes	e24 generic clothing	e9 generic clothing
sleeves	leather	expandable waist	wolf moon	Polo Lauren	Bates Red Wing	running trail	Balance New	court play	dry cold	same durable
arms	saie	Duckers	three	Lauren Ralph	leather	wide	Nike	working	store	different
neck	dress	iron	power	Beene	good boot	casual boot	running series	running shoe	short	two
shoulders	brown	khaki	trailer	Geoffrey	dress boot	motorcycle	minimalist	feet	brand	weather
dress shirt	dress shoe	stretch waist	hair	great shirt	quality shirt	right boot	zero drop	usa	light shoe	right
dress	polish	hidden	man	short-sleeve	white shirt	motorcycle	road	cross training	great shoe	down
long sleeve	be	ironed	magic	athletic shoe	Wings	glove	run	cross	support	fine
lacket	brown pair	be	magic	athletic shoe	Wings	glove	run	cross	support	fine
iron	looking shoe	formal	elastic waist	powerful						another pair

Outcome 3: We construct a browseable product graph of related products



see:
<http://jmcauley.ucsd.edu/amazon/interface.php>