

THE AGE OF INCOGNITO

Utilizing Big D in the Adult Film Industry

MSDS 2020 LT 15

Cordero | David | Gloria | Ong

this is not
about
porn

this is about

your

privacy


for adult film sites,

72%


use cookies

16%


have privacy
policies




the internet
knows *who*
you are



the internet
knows *where*
you've been



the internet
knows *what*
you like



the internet
knows *who*
you've been
talking to

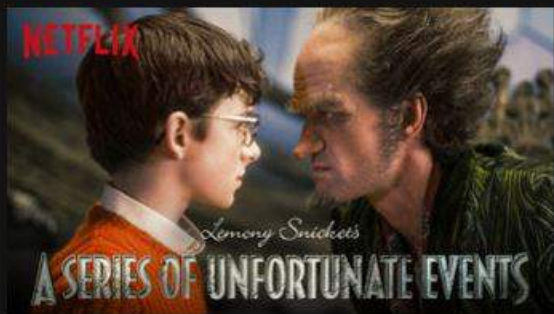
NETFLIX

Browse

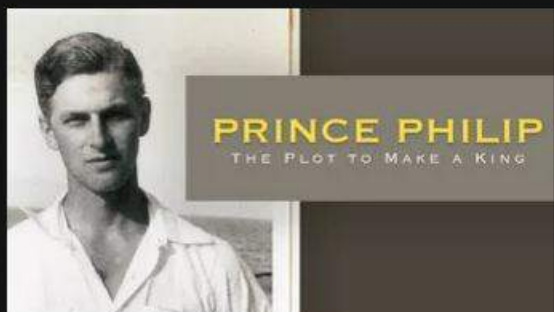
DVD

Q S

Because you watched Stranger Things



Because you watched The Crown



Because you watched American Crime Story: The People v. O.J. Simpson





Search

12153399053

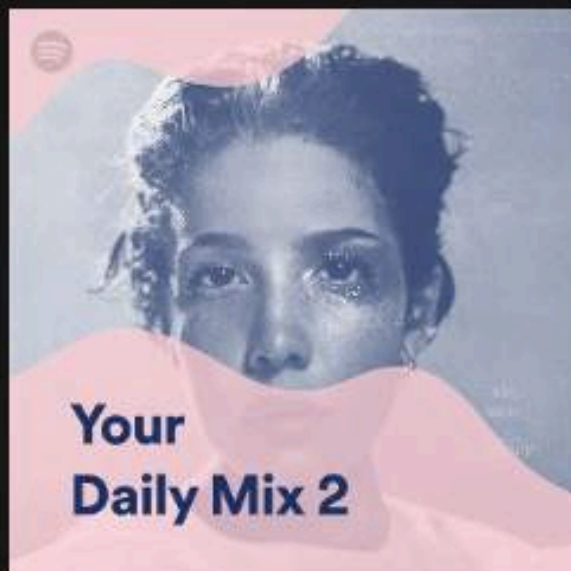
Home

Made for you

Get better recommendations the more you listen.



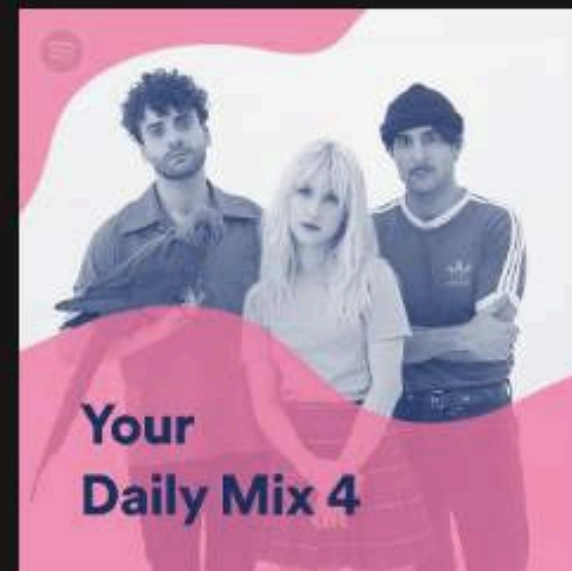
Daily Mix 1
Arctic Monkeys, The Killers, The Kooks and more



Daily Mix 2
Halsey, Billie Eilish, Echobros and more



Daily Mix 3
John Mayer, The Veronicas, Tori Kelly and more



Daily Mix 4
Paramore, Fall Out Boy, Panic! At The Disco and more



In 2018, **Cambridge Analytica** harvested millions of personal data of facebook users without consent and used it for **political advertising**.

facebook

how can we deliver *quality*
recommendations without
compromising *data privacy*?



our *solution*

content based
recommender system

6,800,000+ videos

6,800,000+ videos

380,000+ tags

6,800,000+ videos

380,000+ tags

304 billion views

6,800,000+ videos

380,000+ tags

304 billion views

collected from the top adult film site



diversity

behavior

similarity

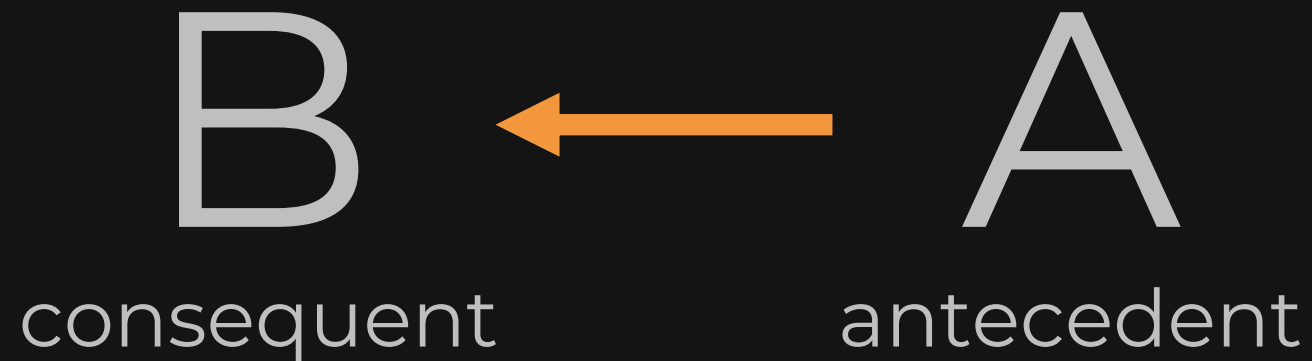
diversity behavior
association rule mining

similarity
clustering

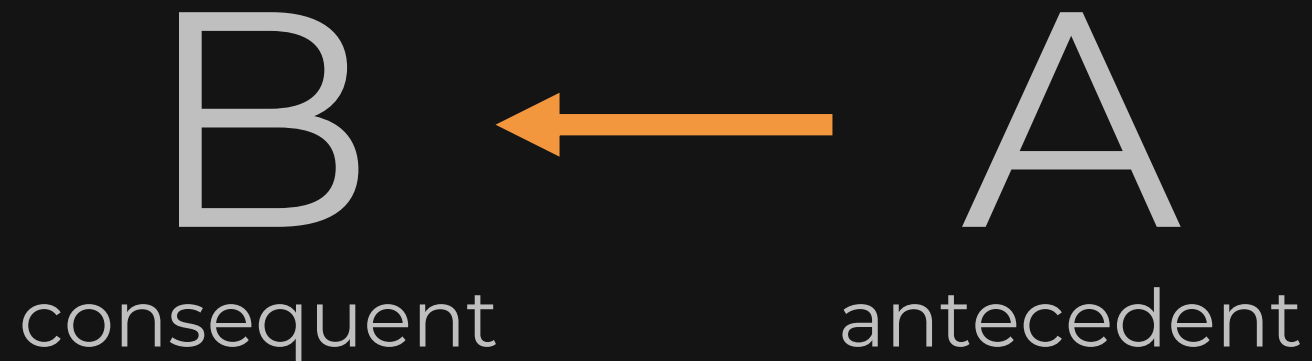
association rule mining

A

association rule mining



association rule mining



association rule mining

transaction

item

rating

association rule mining

transaction *video*

item *tag*

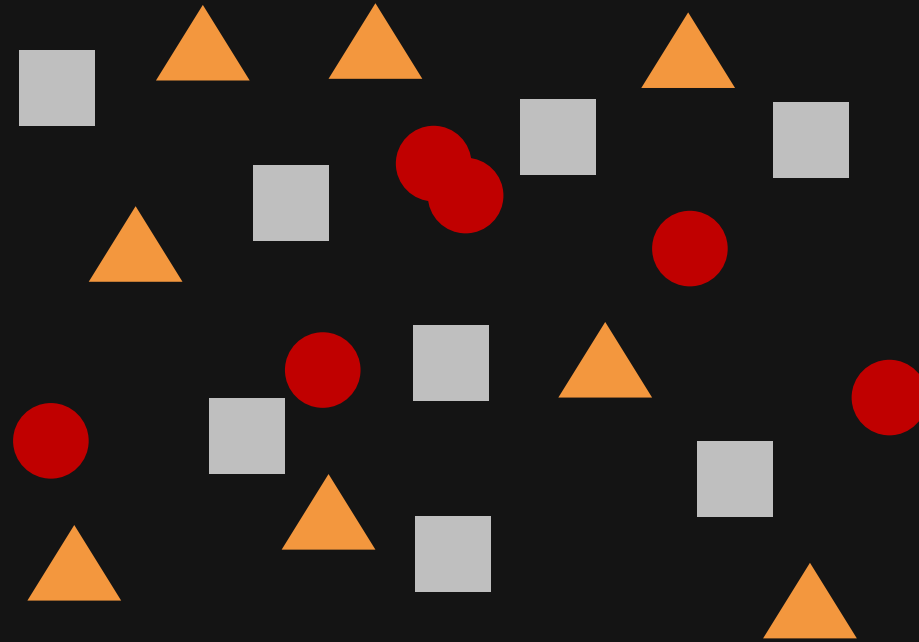
rating *views*

association rule mining

tag  tag

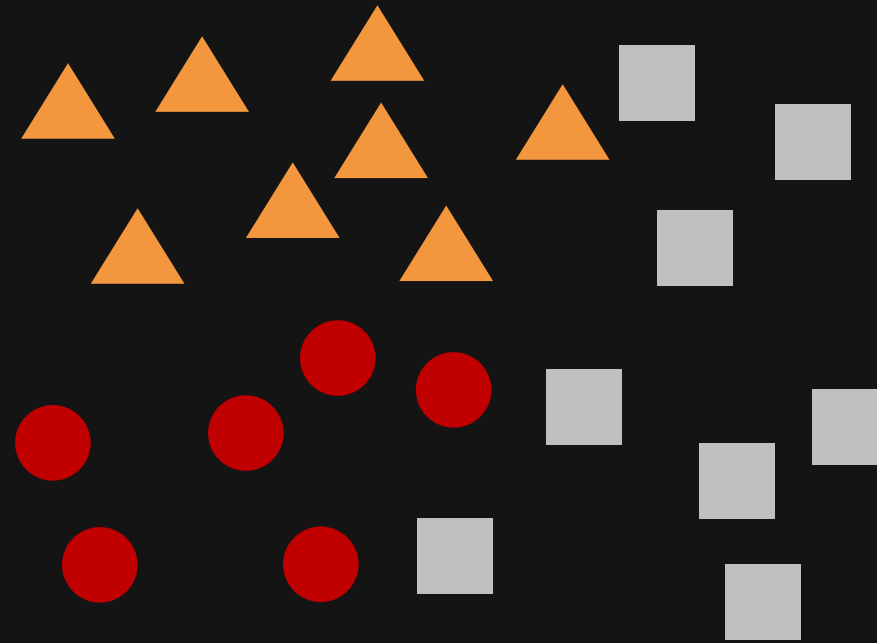
unsupervised *clustering*

grouping videos based on inherent similarity



unsupervised *clustering*

grouping videos based on inherent similarity





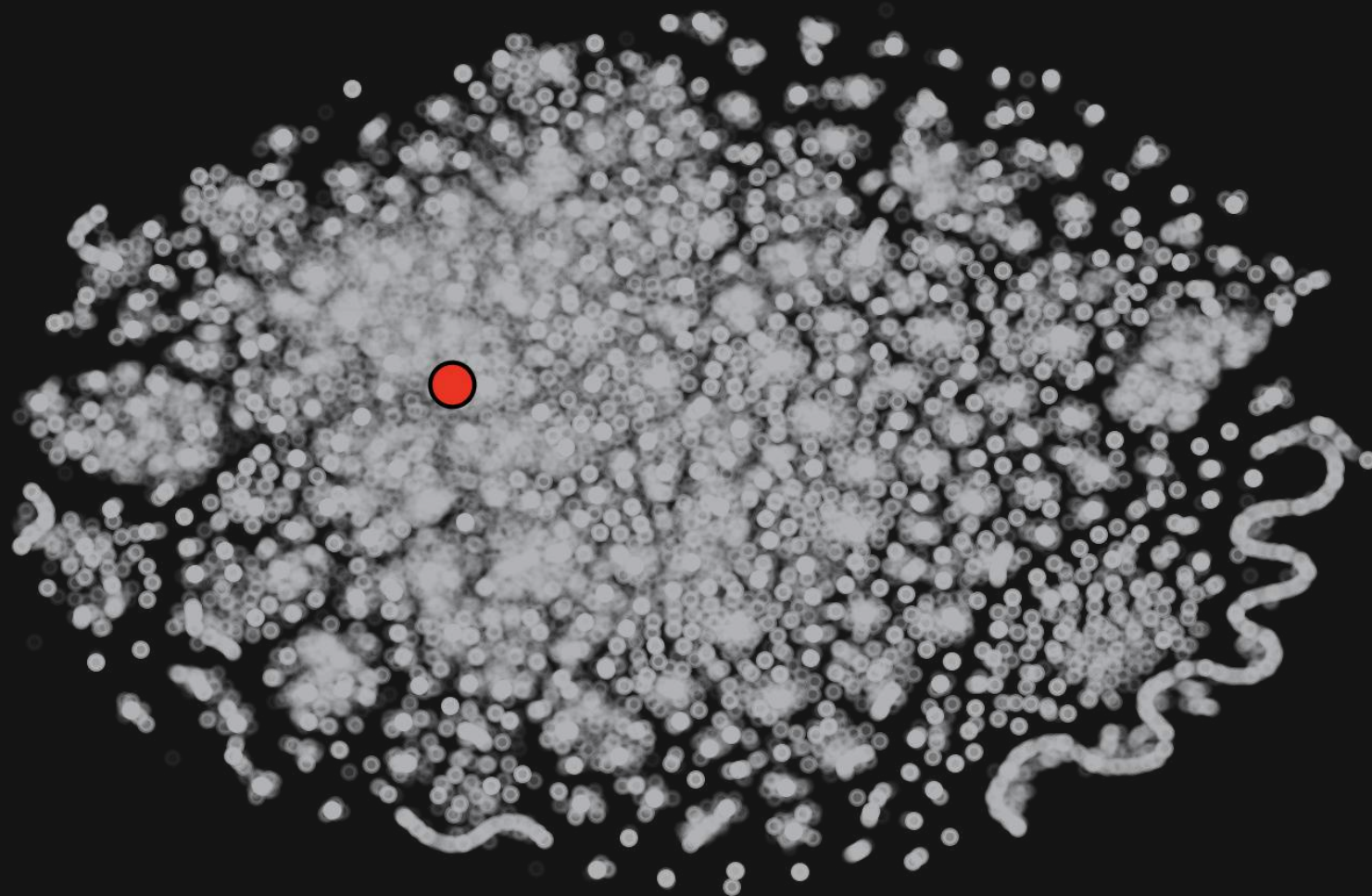
the *results*

*intra*cluster recommendations

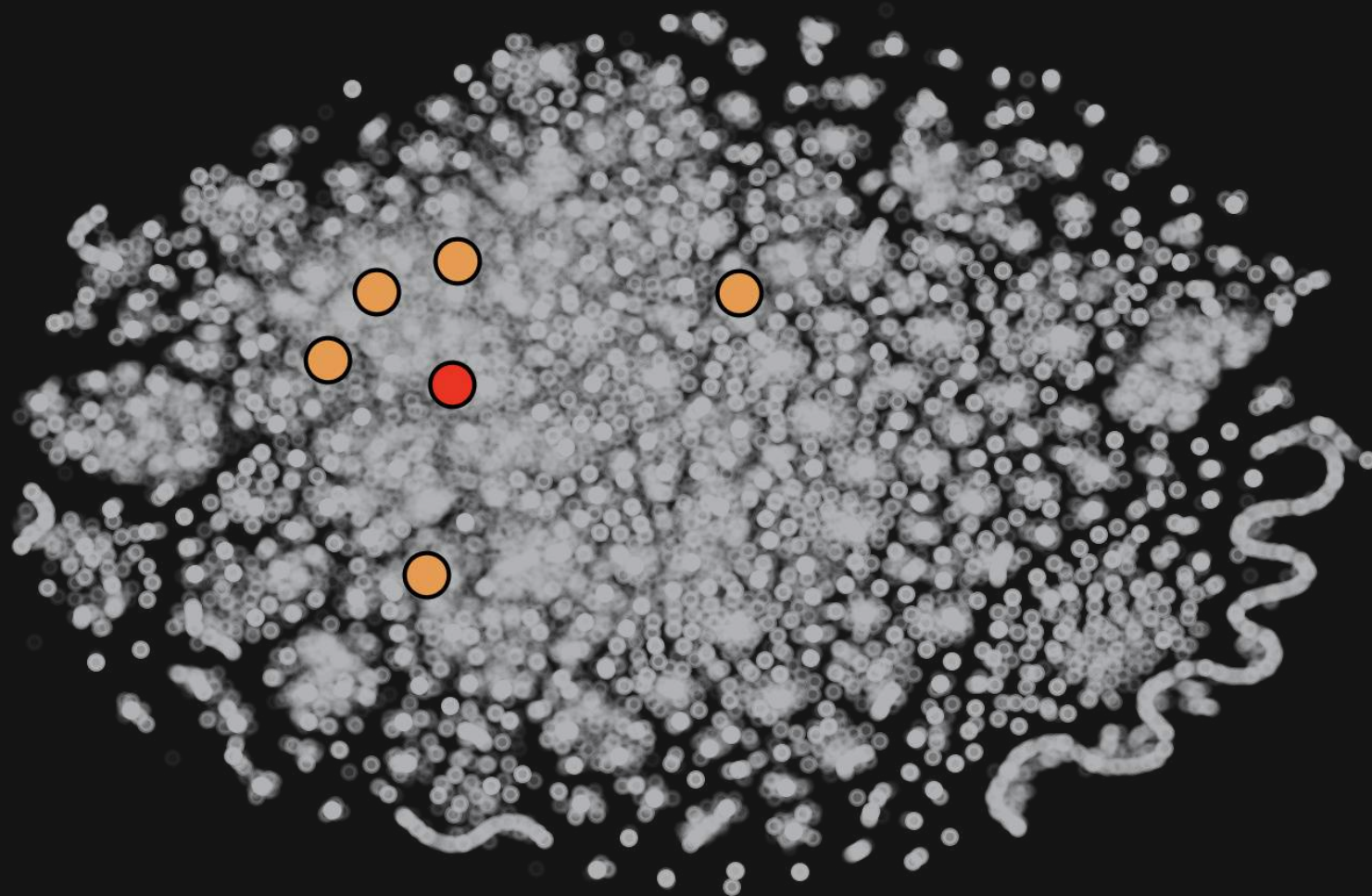
recommended videos **stay
within the same cluster.**

The persistence of the video
themes are strong.

*intra*cluster recommendations



*intra*cluster recommendations

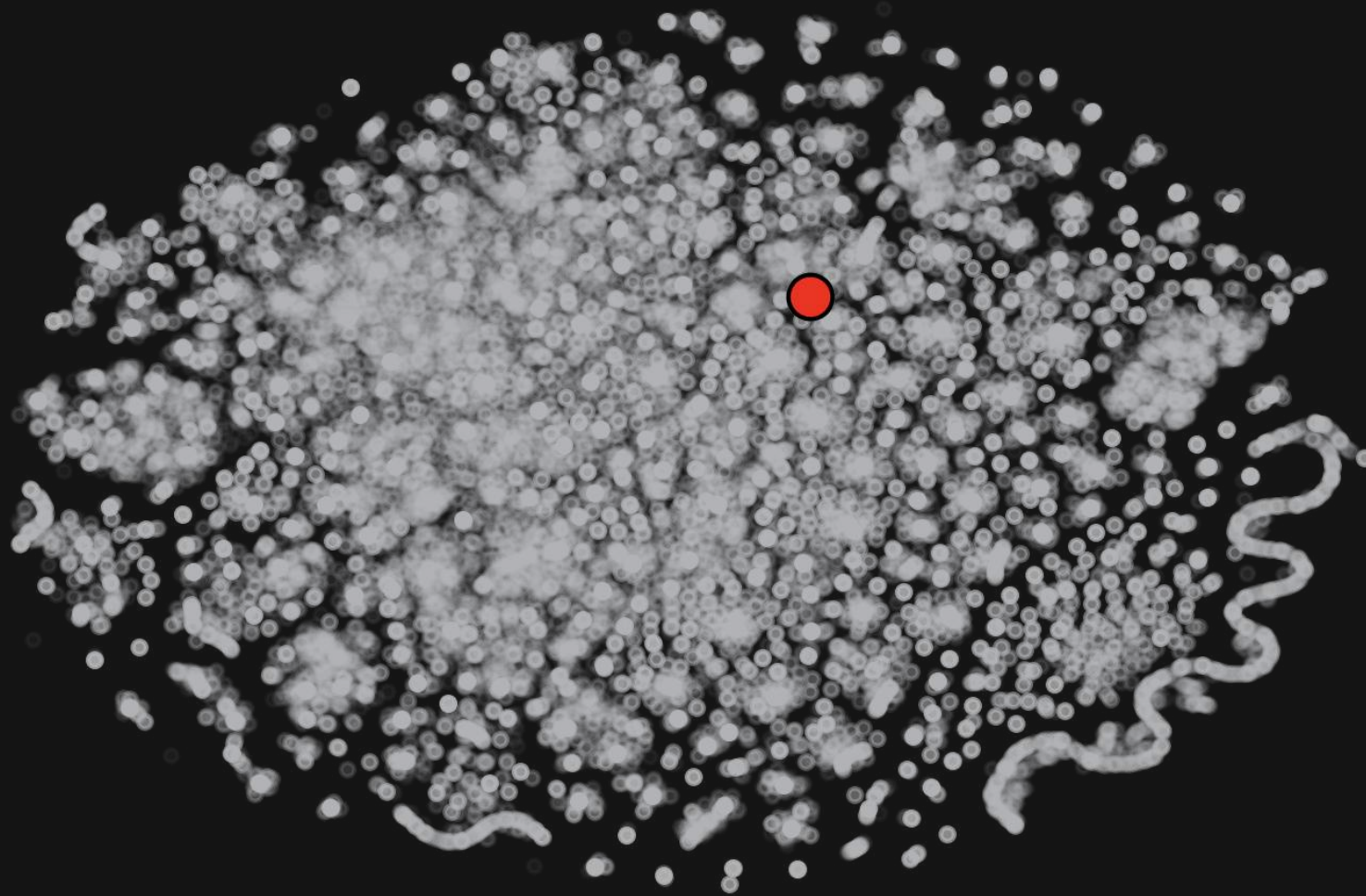


*intra*cluster recommendations

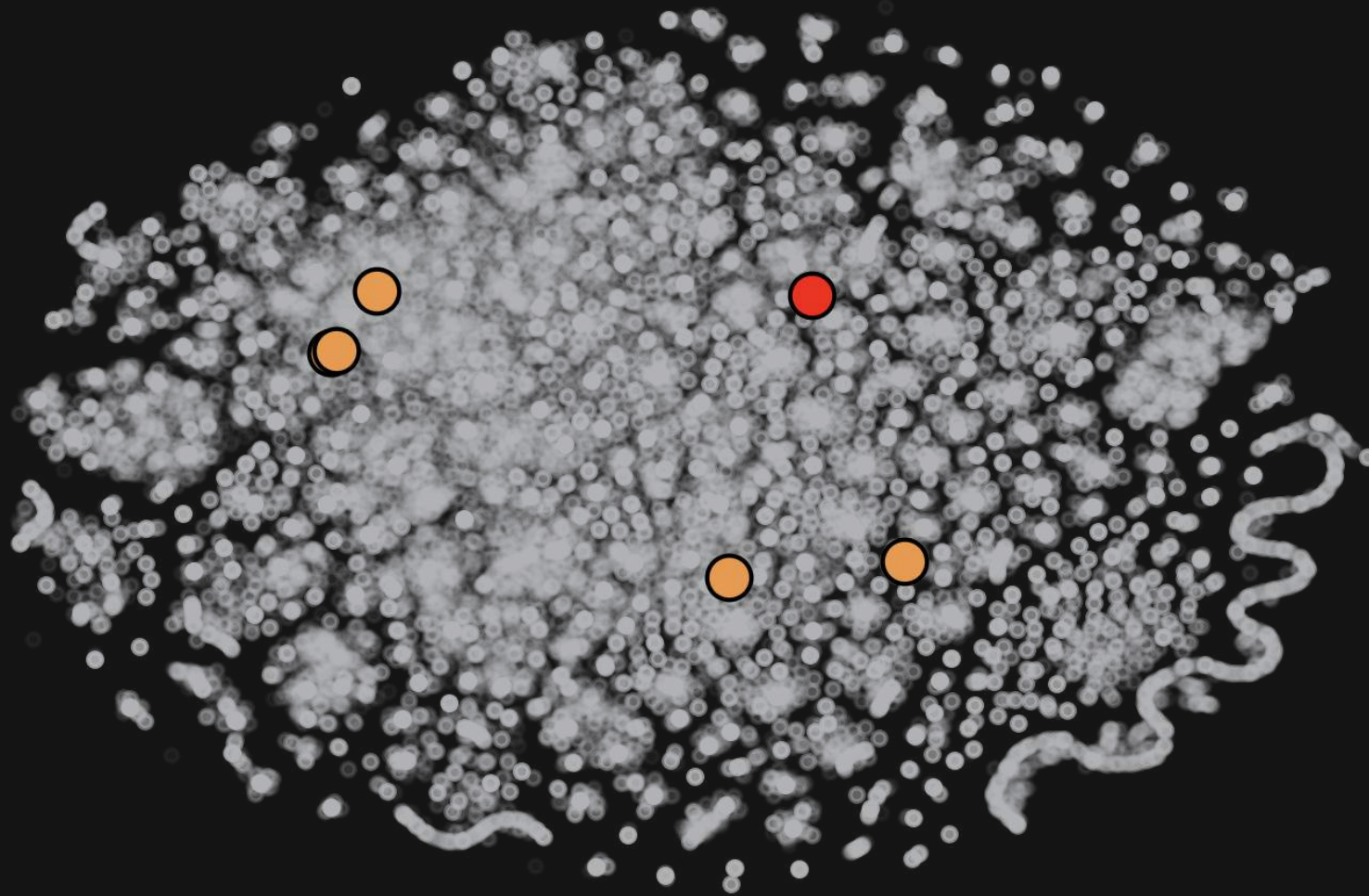


*inter*cluster recommendations

recommended videos **jump
to different clusters.** The
themes tend to lead to
other types of videos.



*inter*cluster
recommendations



*inter*cluster
recommendations



*inter*cluster
recommendations



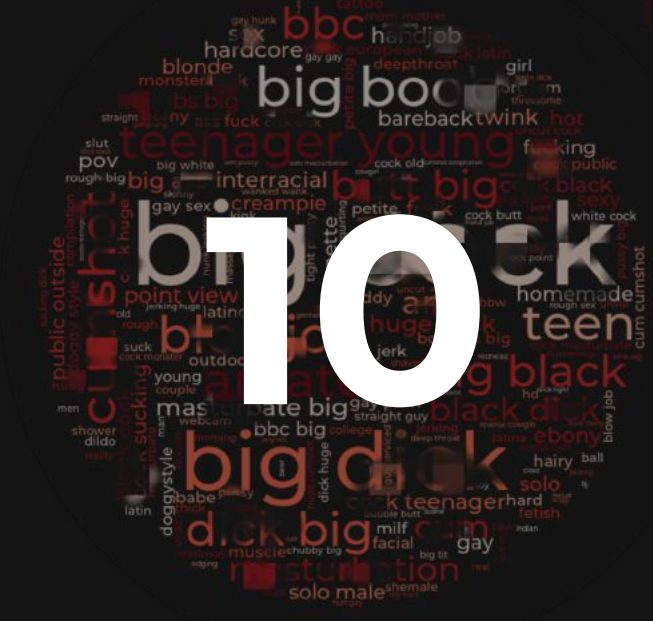
2



4



7



10

*inter*cluster
recommendations

what *now*?

As data privacy becomes a bigger concern, companies could possibly *lose access to user-specific data.*

content based recommender system

association rule mining unsupervised
clustering

But in the digital world, whether it's digital cameras or satellites or just what you click on, we need to have more explicit rules - not just for governments but for private companies.

-Bill Gates

THE AGE OF INCOGNITO

Utilizing Big D in the Adult Film Industry

MSDS 2020 LT 15

Cordero | David | Gloria | Ong