

Data Science In Banking

Fatemeh Rahimian, PhD.
Senior Data Scientist

fatemeh.rahimian@swedbank.se

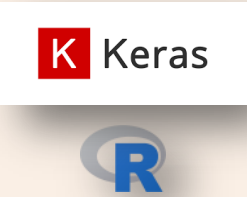






Our approach to operational data science

The data science team.



Computer Scientists

Physicists

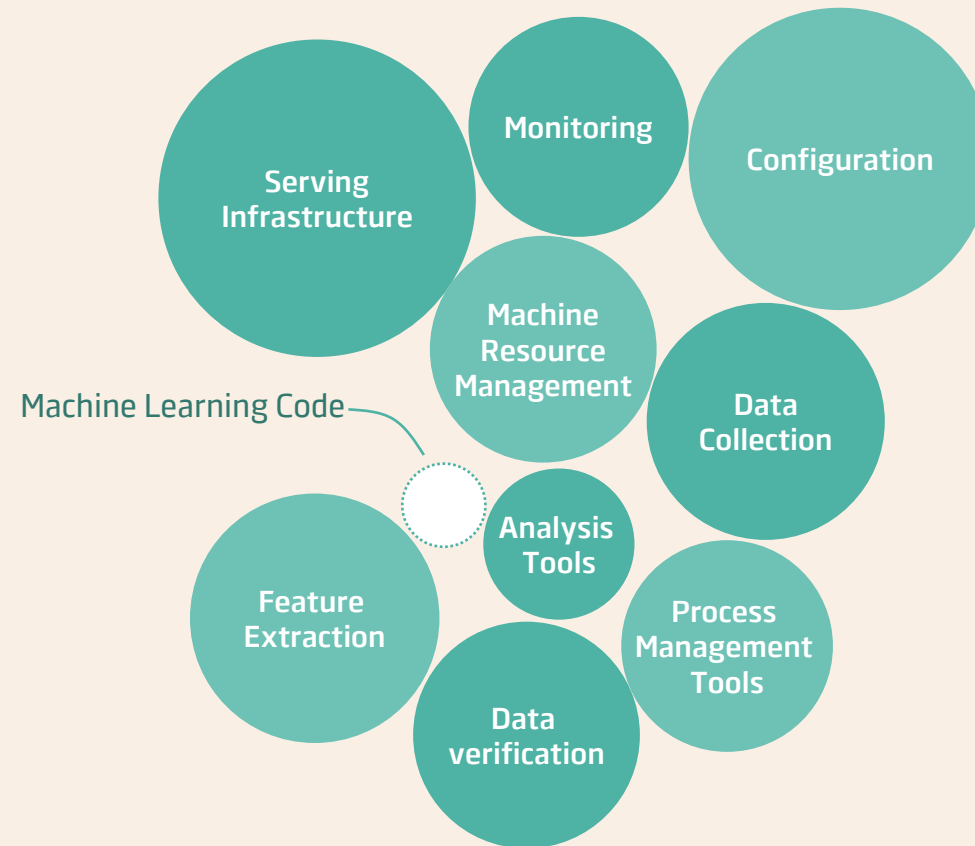
Data Analysts

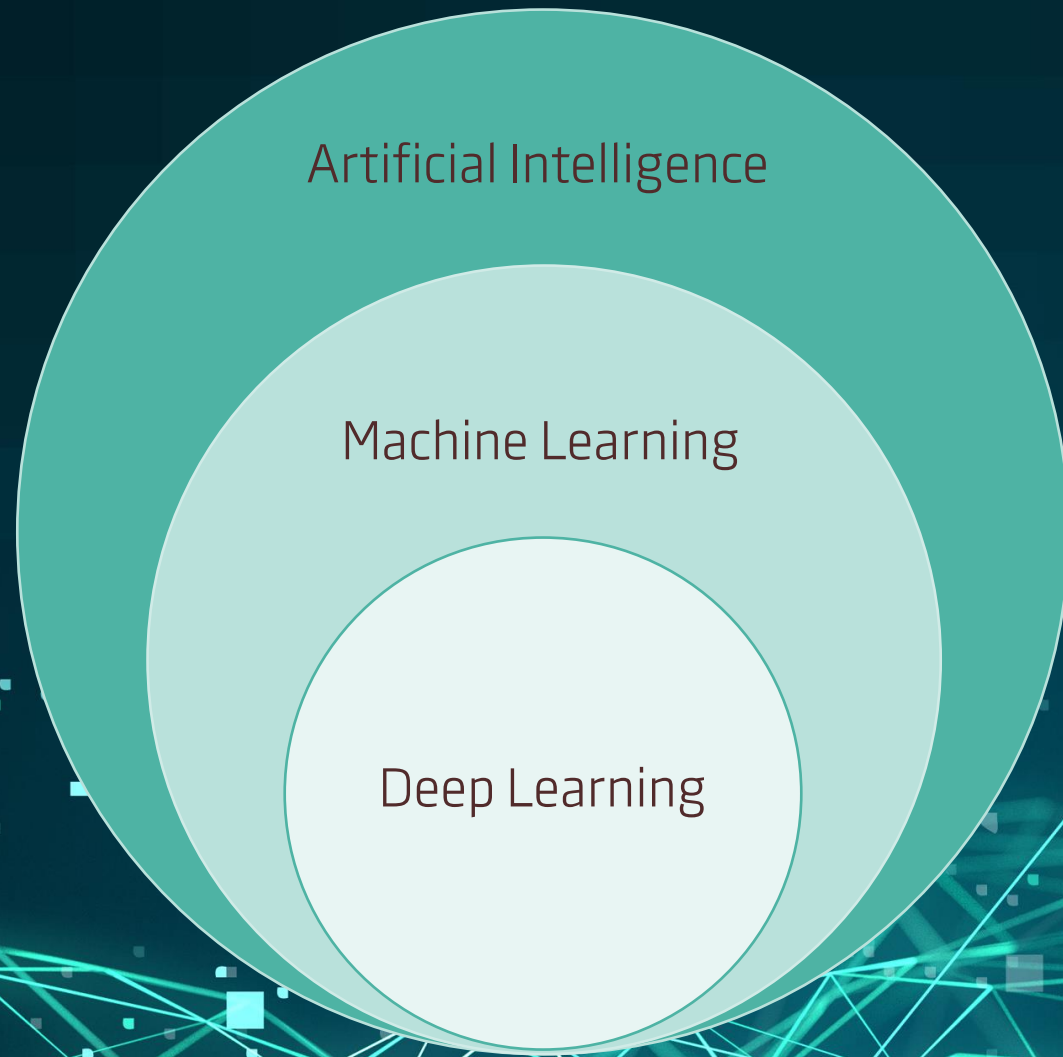
Machine Learning Engineers

Statisticians

Our approach to operational data science

Machine learning code takes us only so far.





Artificial Intelligence

Machine Learning

Deep Learning

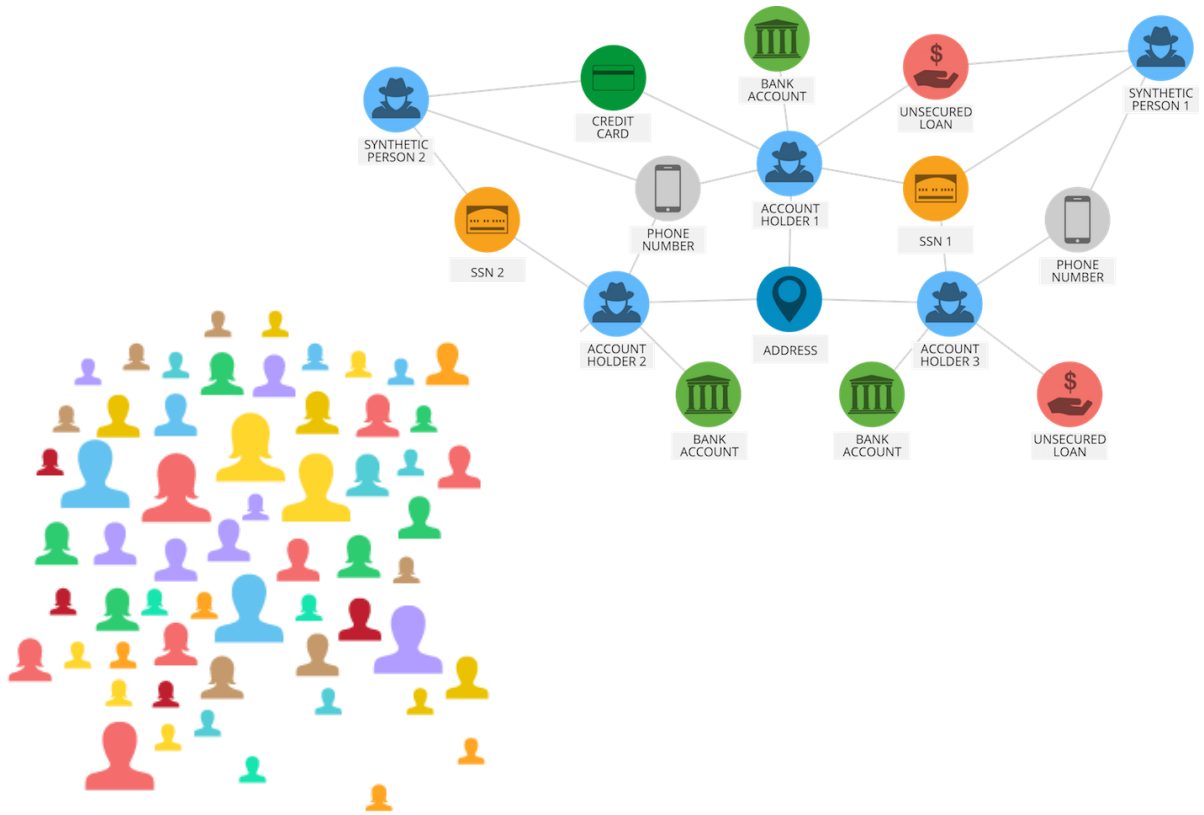
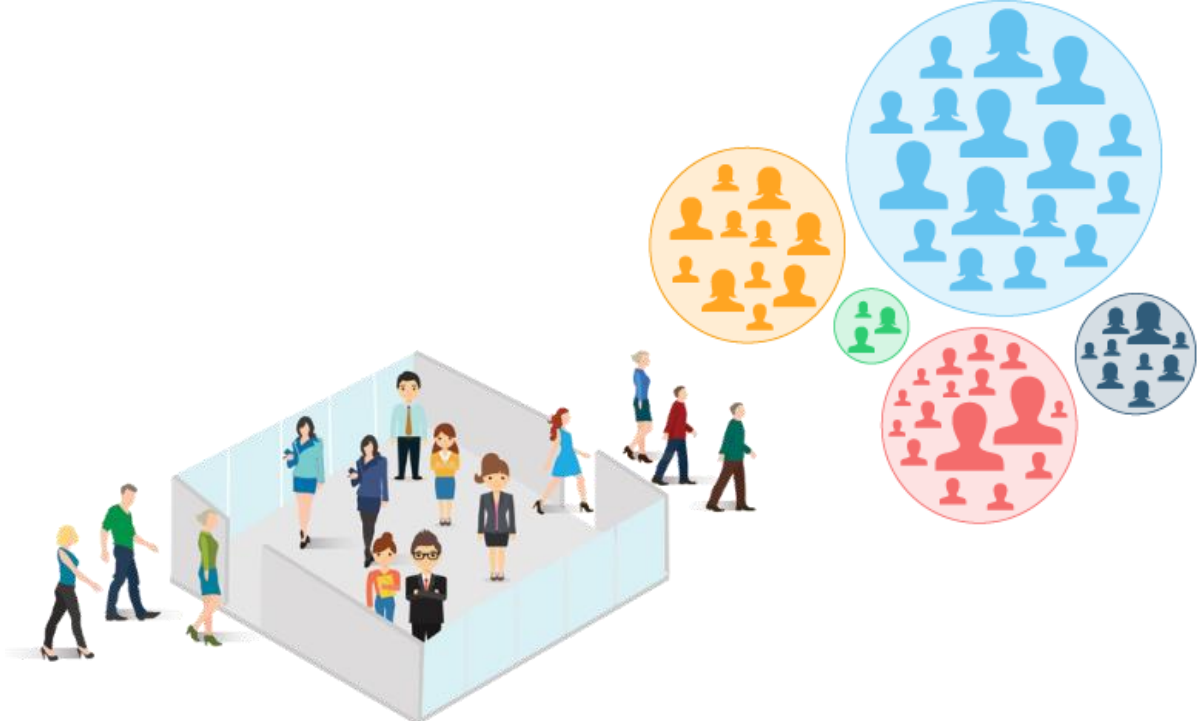
Potential Usecases

Customer Churn

Segmentation

Personalization

Fraud Detection



Fraud Detection

Fraud Detection

- Why is it hard?



A Rule Based Solution

IF

Transactions are > 3


AND

Transaction amount > 50 000 SEK

AND

Country is on red list

THEN

Flag 



A Rule Based Solution

IF
AND Furry
AND Pointy ears
AND Tail
AND Whiskers
THEN
Cat 🐱

TRAIN

Expose algorithm to millions of cats.
Algorithm learns what constitutes a cat.

TEST

Expose to pictures of all kinds of animals.
Flag as 🐱 when picture resembles a cat
from training data.



A Rule Based Solution Is not enough

IF

Transactions are > 3


AND

Transaction amount > 50 000 SEK

AND

Country is on red list

THEN


Flag 

TRAIN

Expose algorithm to millions of 'normal' cases.

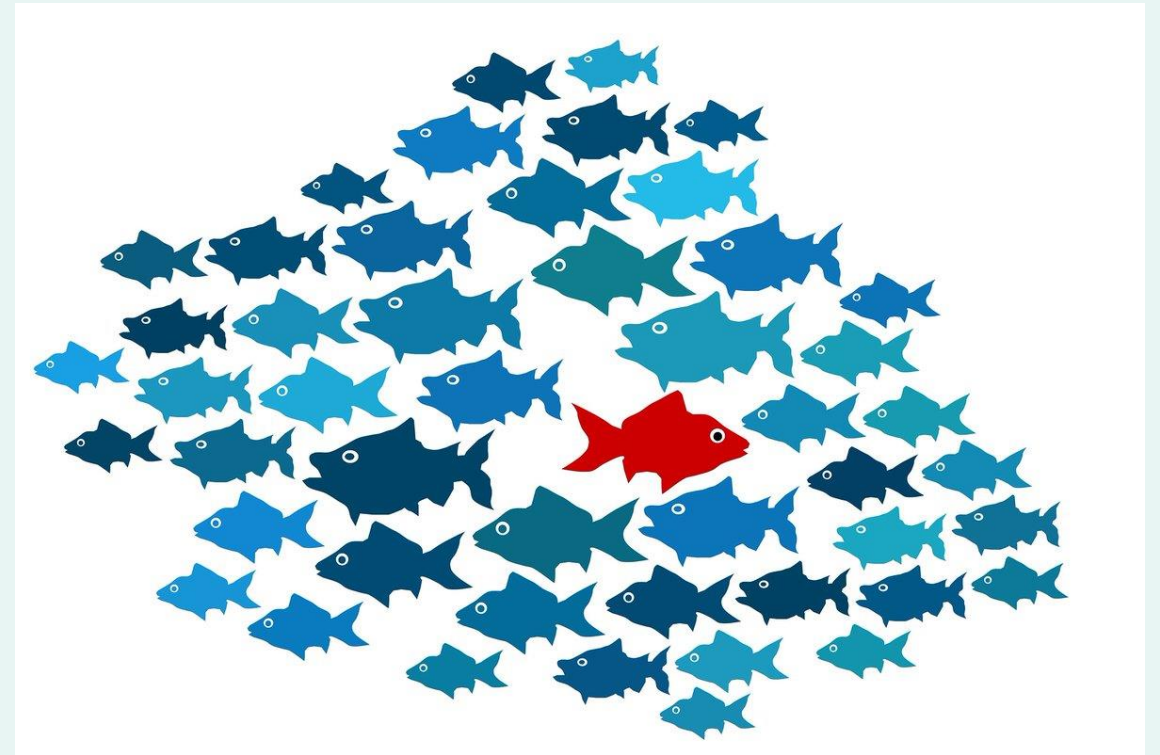
Teach model what constitutes 'normality'.

TEST

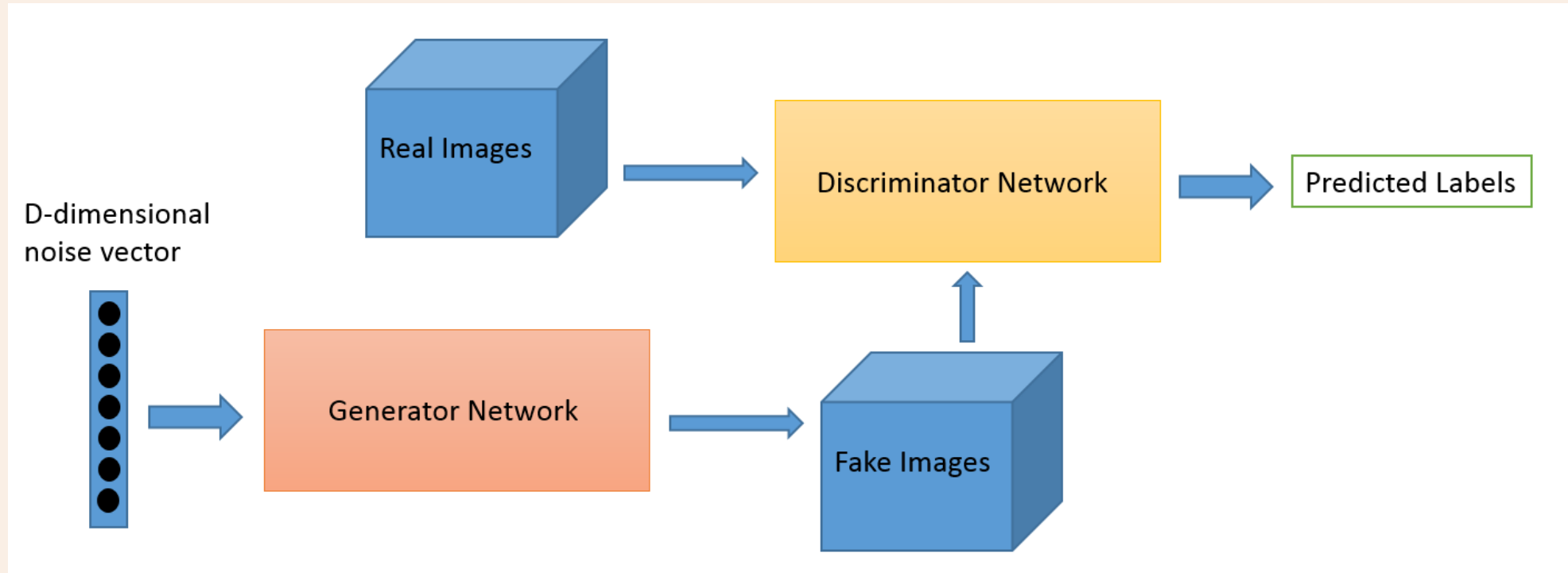
Run on real-time data. Flag  when patterns deviate from the normal. Continuously learn and adapt.

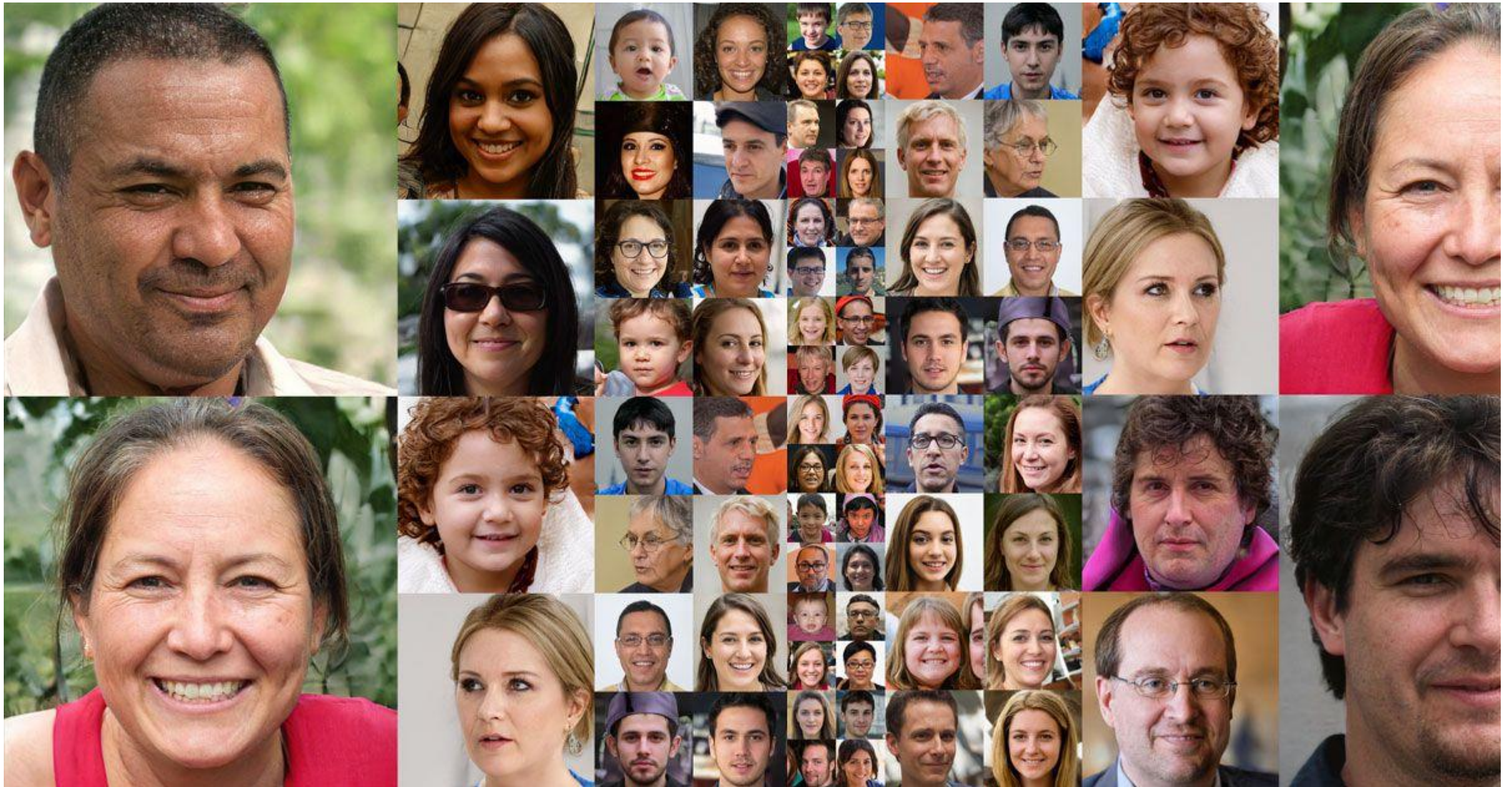
Still hard! Why?

- What is a normal behavior?
- Not all normal behaviors are the same
- Not enough fraudulent cases
- Fraudulent behavior changes over time
- Fraudulent guys do not always behave suspicious



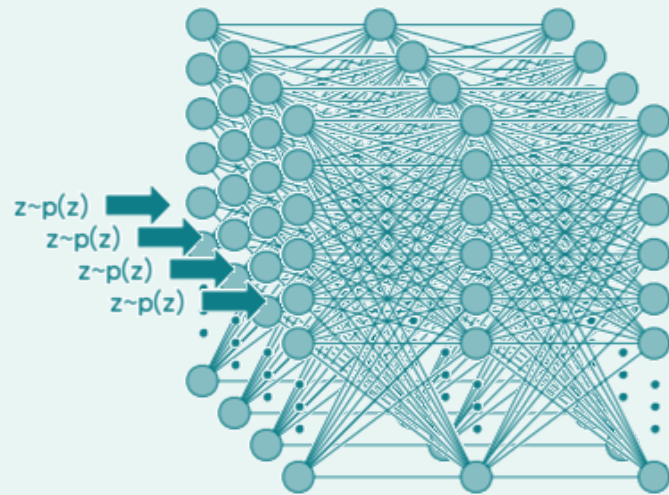
GANs for Anomaly Detection



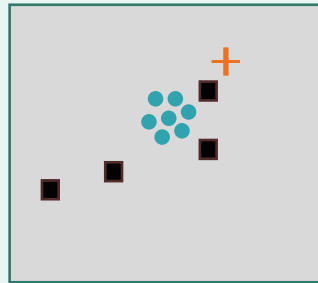




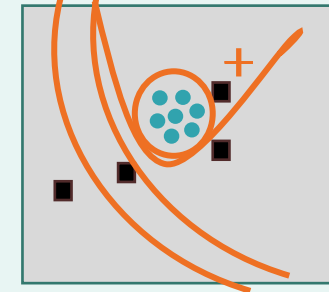
How to Use GANs for AML?



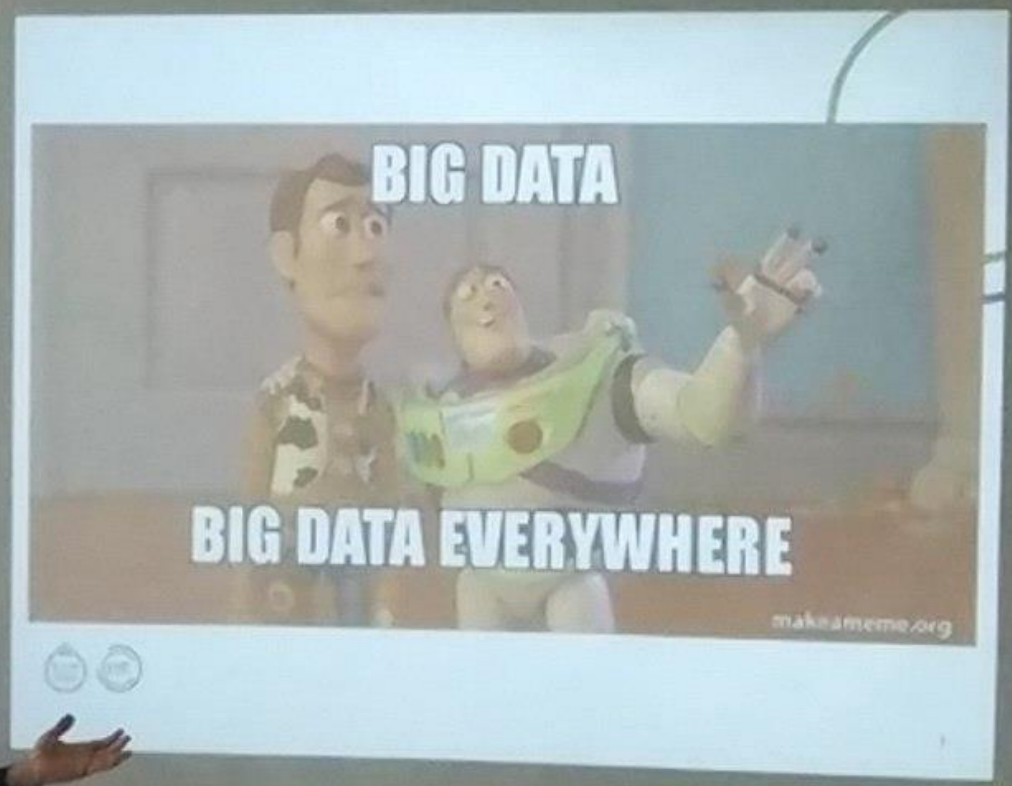
Generator G



Discriminator D



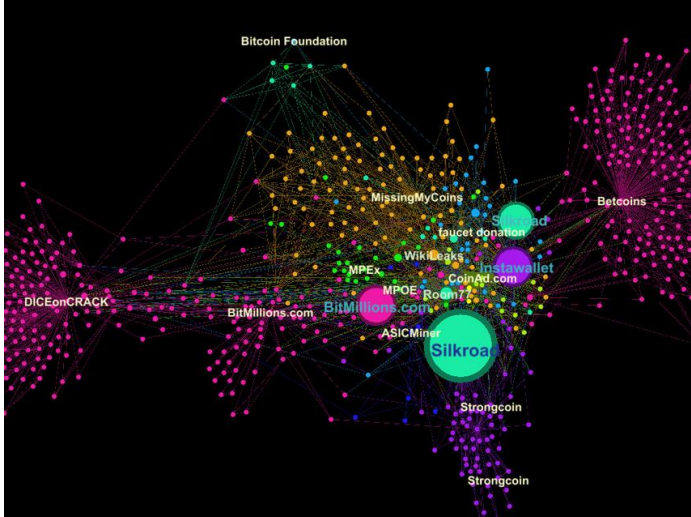
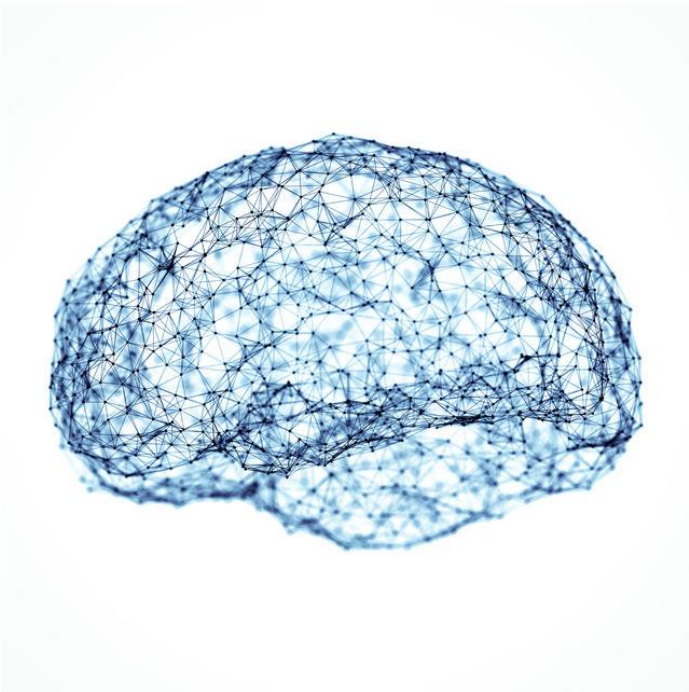
A Few Applications of Graph Analytics



with
Bosnia and
Herzegovina

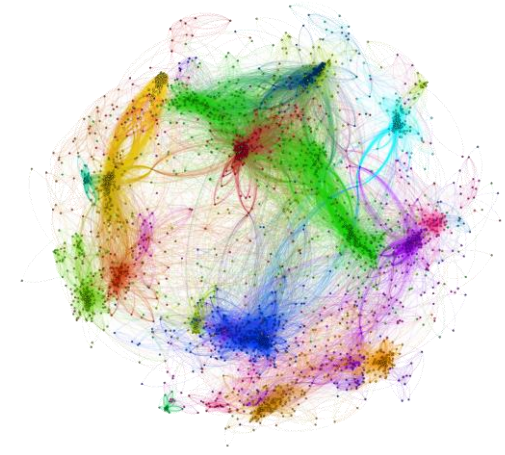


Big ``Linked`` data



Challenge of working with graphs

- Effective graph modeling
 - What should be the nodes and what should be the edges?
- Choosing the right algorithm
 - That solves the problem
 - That suits the size of the problem
- Efficient parallel processing
 - Each node or edge should be processed in the context of its neighborhood
 - The load should be uniformly distributed.
- Select and use appropriate visualization tools

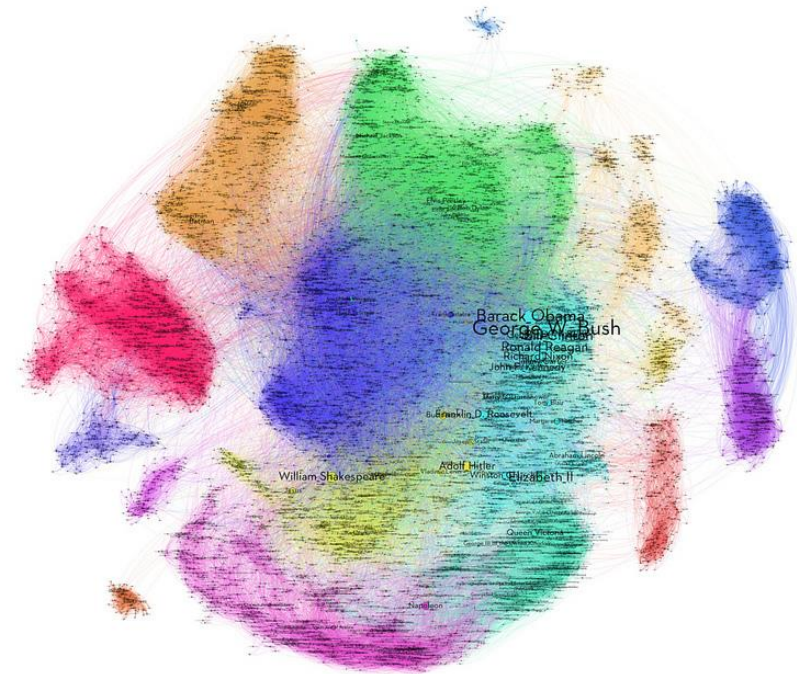


Clustering

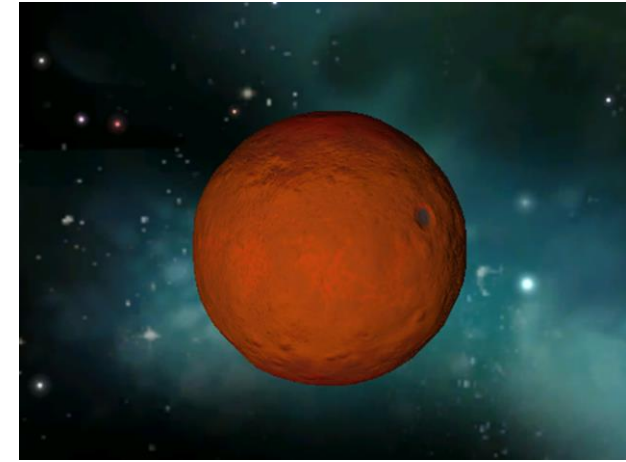
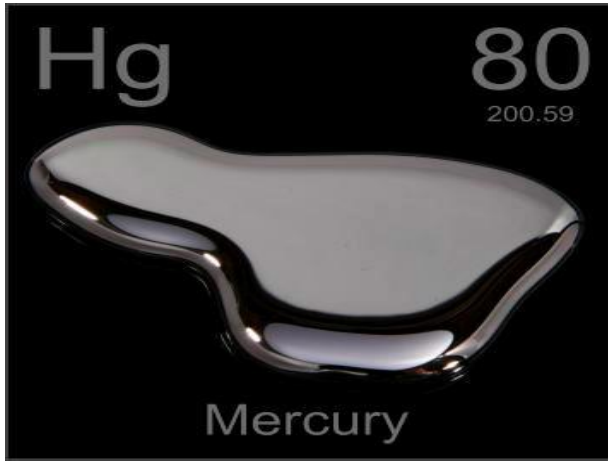
Clustering (Community Detection)

Partition a graph into multiple components with **natural cut**

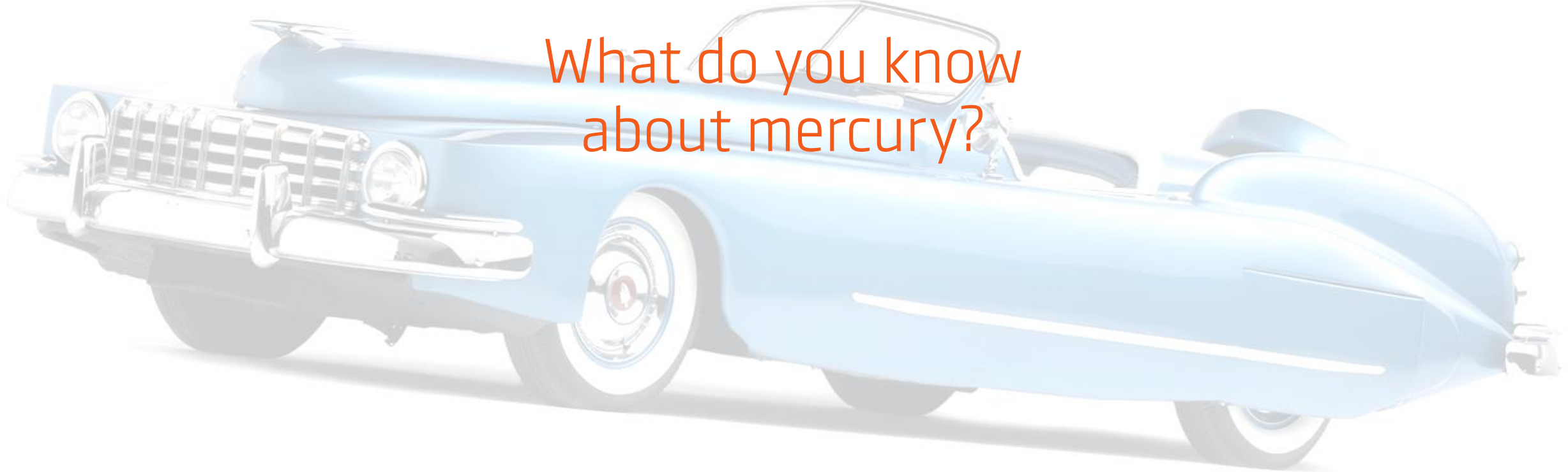
- > The number of components is unknown.
- > The size of the components could be very different.



Graphs for Text Analytics



What do you know
about mercury?





What do you know
about mercury?

Coreference Resolution

Given a set of documents with an **ambiguous word**, classify the ambiguous words into different groups, where each group contains all those words that are co-referent.

	Mercury easily forms <u>alloys</u> with other <u>metals</u> , such as <u>gold</u> , <u>silver</u> , and <u>aluminum</u> .
	The <u>heat</u> encountered once in Mercury 's <u>orbit</u> will be the equivalent of 11 <u>suns</u> beating down on <u>Earth</u> , about 700 degrees.
	Mercury probably acquired much of its <u>water</u> and organic <u>material</u> the same way <u>Earth</u> did, researchers said.
	Mercury is a relatively poor <u>conductor</u> of <u>heat</u> . Most <u>metals</u> are excellent thermal conductors.
	Mercury generally is not allowed on <u>aircraft</u> because it combines so readily with <u>aluminum</u> , a <u>metal</u> that is common on aircraft.
	Mercury <u>orbits</u> the <u>sun</u> every 88 <u>Earth</u> days, zipping around at a faster pace than any other <u>planet</u> .

Coreference Resolution

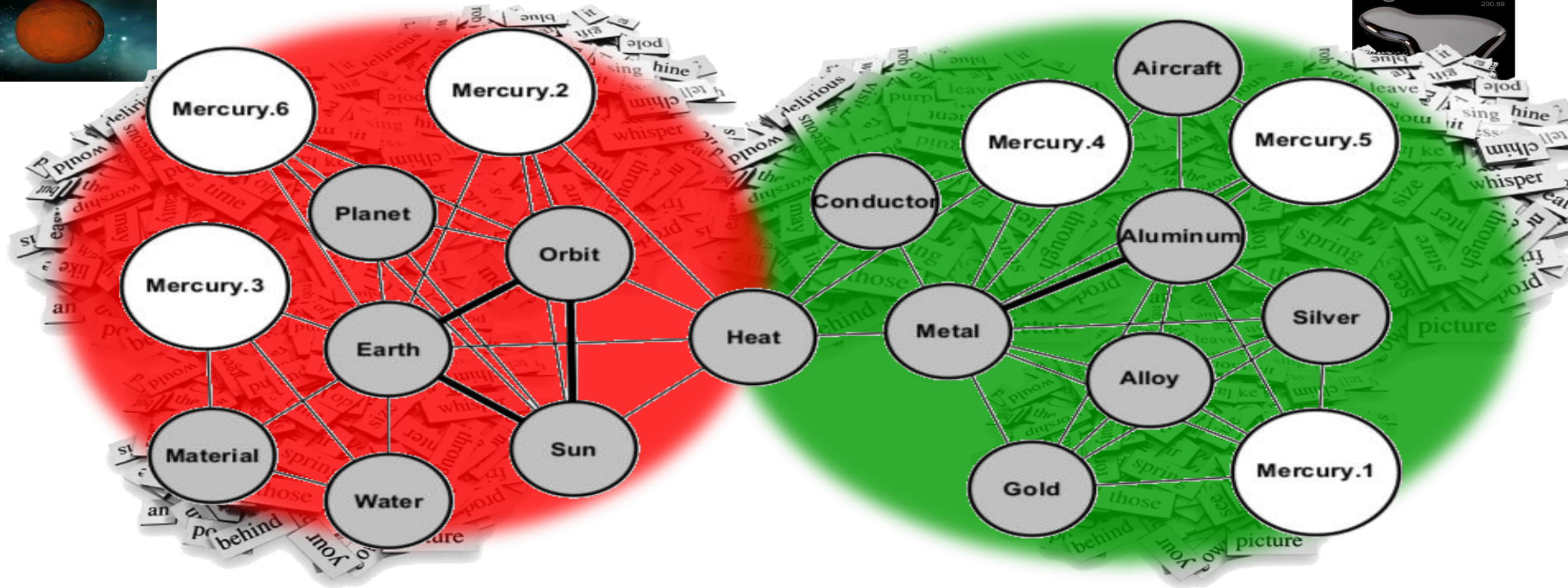
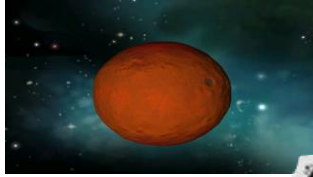
Given a set of documents with an **ambiguous word**, classify the ambiguous words into different groups, where each group contains all those words that are co-referent.

Clustering is required
(NP-Hard)

	Orange	Red	Purple	Green	Yellow	Brown
Orange		X	✓	✓	X	X
Red			✓	X	X	X
Purple				X	✓	X
Green					✓	✓
Yellow						✓
Brown						

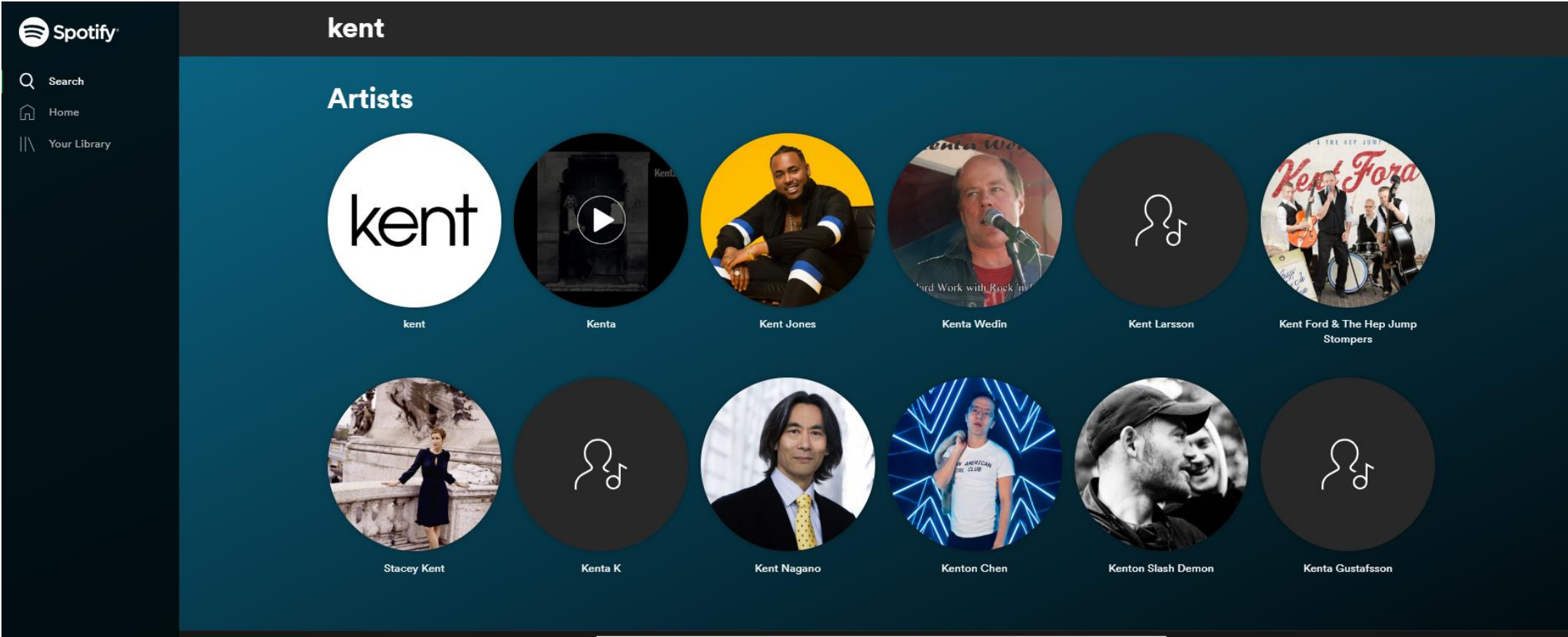
$O(N^2)$ Comparisons

Coreference Resolution -> A Graph Problem



Graphs for Artist Disambiguation

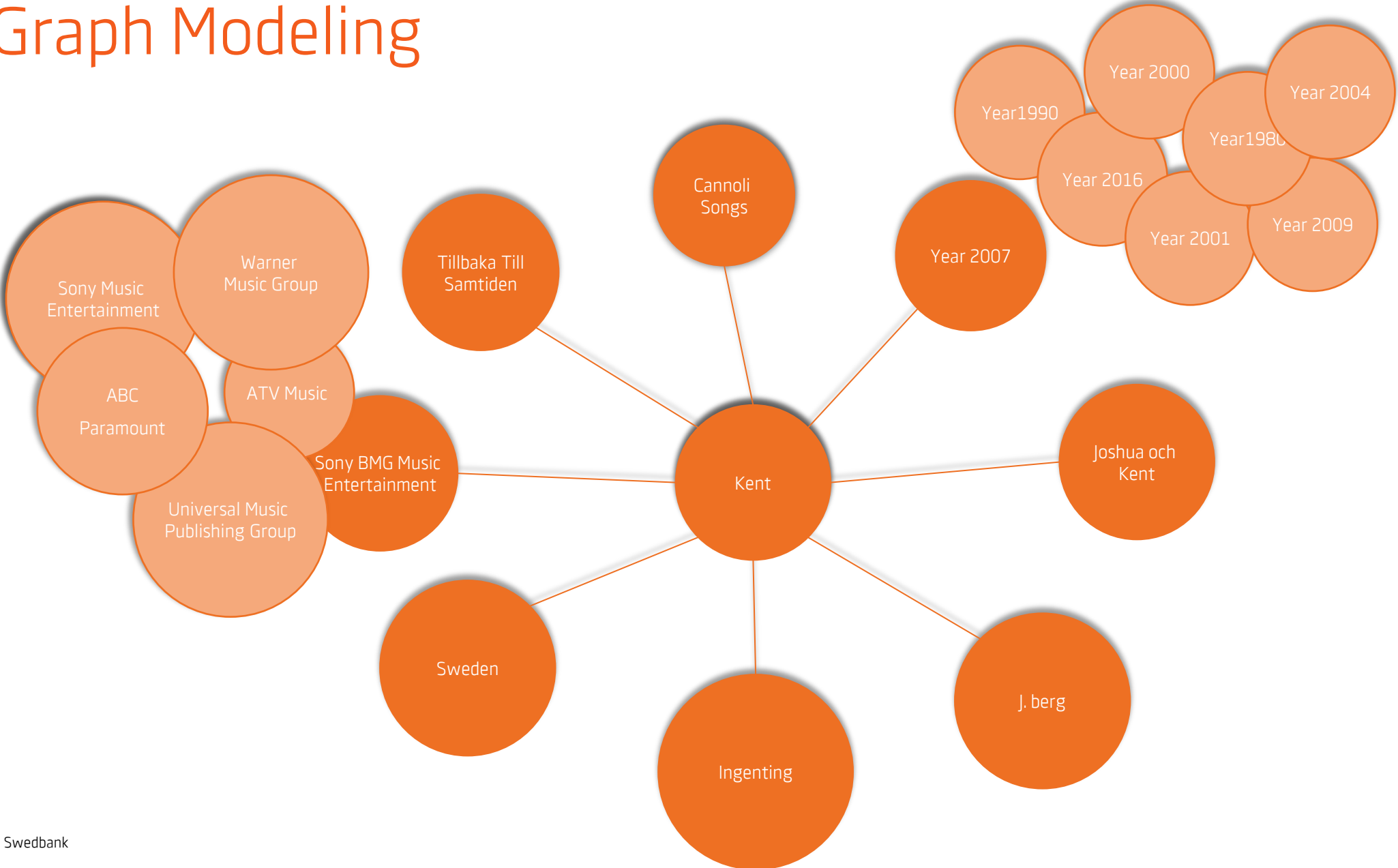
Spotify and the Kent Problem



Artist disambiguation

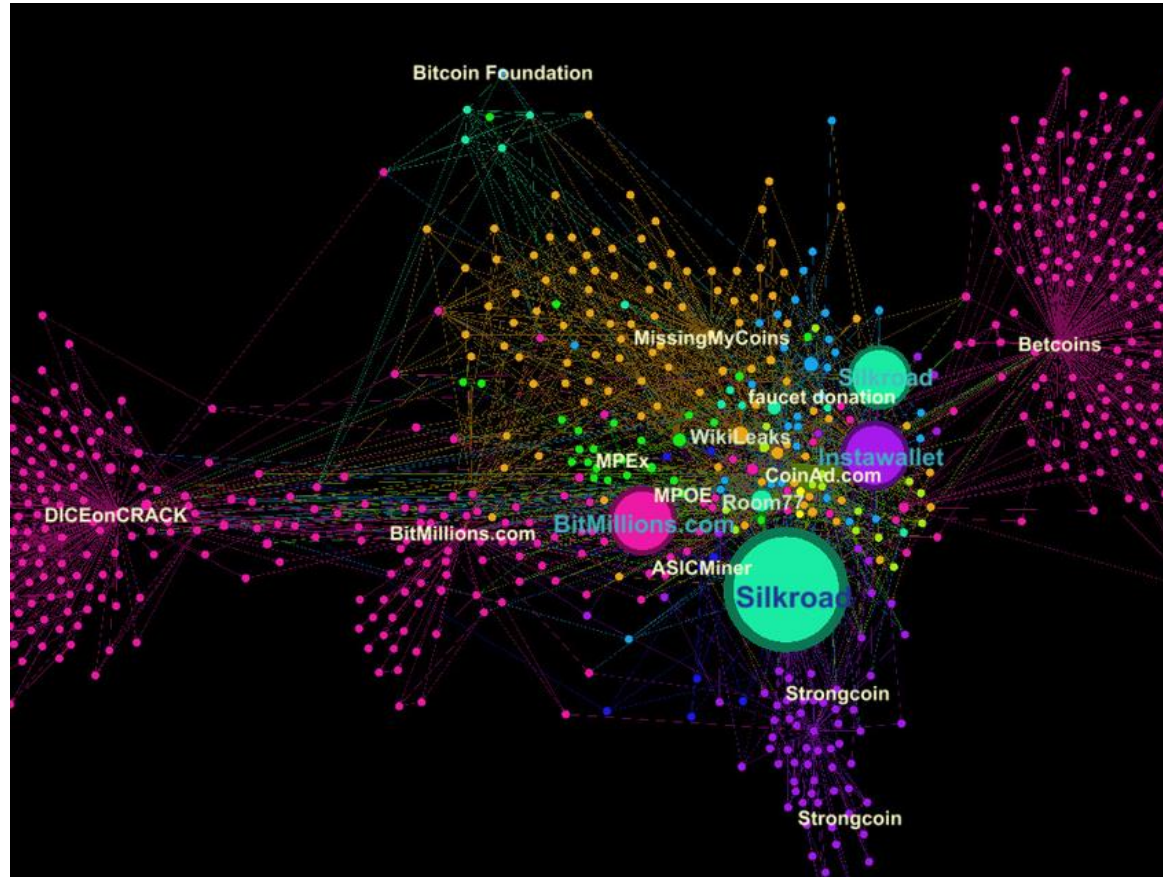
- For
 - Search
 - Deduplication
 - Recommendation
- Computational methods for classification, clustering
 - Musical feature extraction
 - Year, Place, Record label, Lyrics
 - Mono- or polyphonic, Beat, tempo, and rhythm, etc.

Graph Modeling



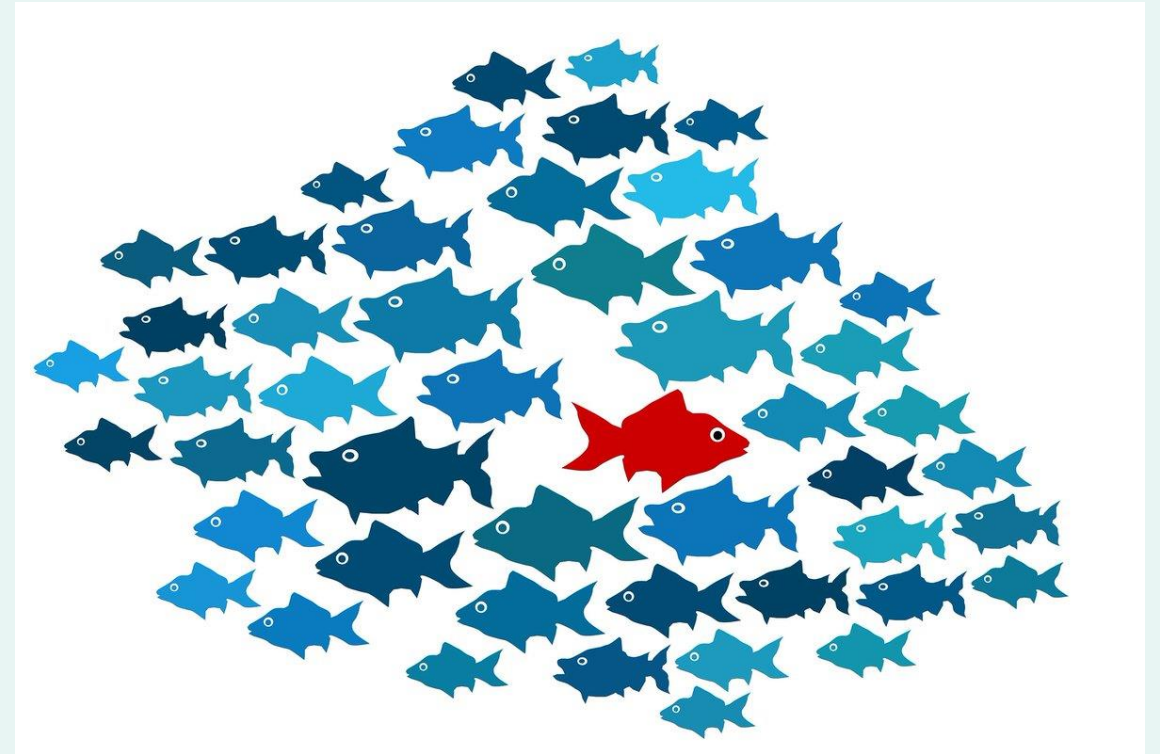
Graphs for AML and Fraud Detection

Discovery of the Silk Road network



Challenges

- What is a normal behavior?
- Not all normal behaviors are the same
- Not enough fraudulent cases
- Fraudulent behavior changes over time
- Fraudulent guys do not always behave suspicious
- We do not have complete data
 - e.g., cross-bank transactions



Come and join us 😊

