# Fraud Detection with AI

Lovejeet Singh - 108167414
Brett Porcelli - 109610861
Rustam Azimbaev - 110666839
Kenneth Chiguichon - 109867025
Mingchen Zhang - 110138345

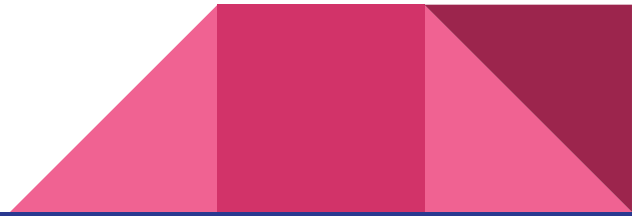
TEAM 1
CSE 352 - Artificial Intelligence
Professor Anita Wasilewska

# Bibliography


1. https://www.ayasdi.com/blog/bigdata/why-topological-data-analysis-works/
2. https://www.ayasdi.com/resources/whitepaper/topology-and-topological-data-analysis/
3. http://www.sciencedirect.com/science/article/pii/S0950705100000502
4. https://engineering.purdue.edu/~engelb/abe565/cbr.htm
5. http://www.acfe.com/rttn-summary.aspx
6. Quah JT, Sriganesh M. Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications 2008;35:1721–32.
7. Jarrod West, Maumita Bhattacharya, Intelligent financial fraud detection: A comprehensive review, Computers & Security, Volume 57, March 2016, Pages 47-66, ISSN 0167-4048, http://dx.doi.org/10.1016/j.cose.2015.09.005.
8. http://www.sciencedirect.com/science/article/pii/S0167404815001261
9. https://en.wikipedia.org/wiki/Self-organizing_map
10. https://www.cs.hmc.edu/~kpang/nn/som.html
11. http://bit.csc.lsu.edu/~jianhua/quang.pdf
12. https://www.youtube.com/watch?v=F2t_ytTLrQ4

# Roadmap

- Introduction
  - What is fraud?
  - Why does it matter?
  - What is fraud detection?
- Fraud detection techniques
  - Topological Data Analysis
  - Case Based Reasoning
  - P-RCE Neural Network
  - Self-Organizing Maps

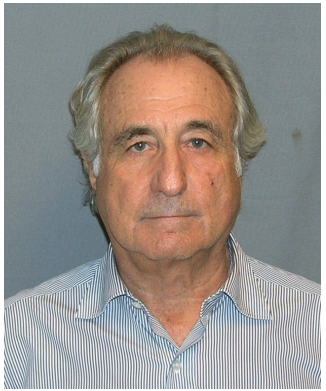


# What is fraud?

According to the People's Law Dictionary, fraud is...

> "The intentional use of deceit, a trick or some dishonest means to deprive another of his/her/its money, property or a legal right."

**Examples:** Credit card fraud, insurance fraud, etc.



Fraudster Bernie Madoff

Image: https://en.wikipedia.org/wiki/Bernard_Madoff

# Implications of Fraud Detection in A.I.

- ➔ Different types of Terrorism
- ➔ Types of Financial Crimes
- ➔ Intrusion/Spam Detection
- ➔ Healthcare Fraud
- ➔ Insurance Fraud
- ➔ Money Laundering
- ➔ Telecommunications


Image: http://www.greycloaktech.com/tag/ppc-fraud-detection/

# Why Does Fraud Matter?

- According to the Association of Certified Fraud Examiners (ACFE), **75%** of businesses globally experienced some level of enterprise fraud in 2013, with a total of over **3.7 trillion dollars in damage**
- From all parties that suffered Financial loss, only 14 % were able to recover the losses and 58% didn't recover anything at all. (ACFE)
- Various different ways of detecting and analyzing fraud
  - Algorithms used can influence how fast one can detect a fraud
  - Deploying the right technique is imperative
- The ability to efficiently detect fraud at an earlier stage has multiple benefits:
  - Businesses save money
  - Businesses experience high customer satisfaction
  - Reduces Crime in society
  - Saves taxpayer dollars in instances of governmental fraud
  - Makes society a better and more sincere environment to reside in

# Challenges in Identifying Fraud

Many old techniques for detecting fraud are becoming obsolete ***(Ayasdi)***

WHY?

Perpetrators of fraud are constantly updating their strategies to exploit vulnerable points within the system.

Identifying an extremely small sample from immense big data

WHY?

Provided the size of corporations, immense amount of data processed (stored in databases), and analyzing it has become a major modern day issue in Computer Science.

Identifying Fraud is a very labor intensive, time consuming and expensive endeavor.

WHY?

Analyzing Big Data requires the deploying of complicated data mining/machine learning techniques.

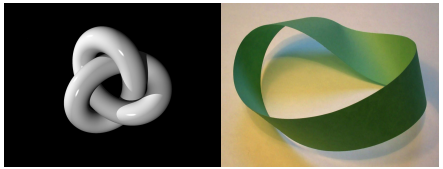
# Fraud Detection Techniques

- ★ Various techniques and methods are utilized modern day to combat, detect, and analyze fraudulent transaction in an enterprise across all industries.
- ★ Many companies use “Specific Techniques” or “Methods”
- ★ **Goal: To detect a potential fraudulent transaction as early, accurately, and precisely (compared to other Fraud Transactions) as possible.**
- ★ We will introduce different techniques for the next 20 minutes

# Roadmap

- ✓ **Introduction**
  - ✓ What is fraud?
  - ✓ Why does it matter?
  - ✓ What is fraud detection?
- ❑ Fraud detection techniques
  - ❑ Topological Data Analysis
  - ❑ Case Based Reasoning
  - ❑ P-RCE Neural Network
  - ❑ Self-Organizing Maps

# What is Topological Data Analysis?

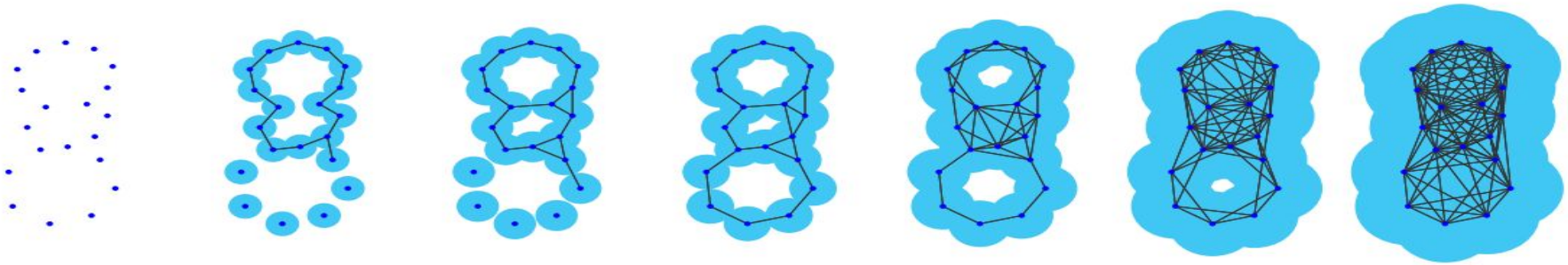


- ➔ General technique for discovering information in large, complex data sets
- ➔ Involves analyzing, categorizing, manipulating data and vital information in 3-D views (or Multi-Dimensional) depicting topological structures. Essentially, datasets are viewed as complex and distorted 3-D shapes or figures.
- ➔ Data is viewed in a form of distorted or complicated “Shape”
- ➔ Take the concept of a shape and envision it as a form of Node based presentation with connected lines which show relationships between respective nodes.

**Definition:** Topology is the study and analysis of space in complex deformations (Stretching + twisting)

Image: https://en.wikipedia.org/wiki/Topology

# Representation and Rendering of TDA.



- ➔ The progressive images above represent nodes in a complicated graphical view.
- ➔ **Goal: To analyze clusters of nodes categorized using a classification mechanism of defined properties or specific information.**
- ➔ The earlier stage consists of nodes.
- ➔ Clusters and relationships start to form as time goes on and data sets become larger.
- ➔ The graphical presentation is 3-Dimensional because each node has many different data values
- ➔ Each node has a particular set of Data Information categorized on a 3 - Dimensional grid point
- ➔ A topological network consists of Nodes and Edges (connective lines jointing 2 nodes)
  - ◆ Edges are used **IF AND ONLY IF** 2 nodes have a common data (some type of relation) in between.

Image: http://www.jmlr.org/papers/volume16/bubenik15a/bubenik15a.pdf

# Why does TDA work? Why does it Matter?

Data and vital information from businesses/organizations is transformed into a complicated representation of Nodes, edges, and clusters.


Data transformed in a topological network reveals insights and hidden patterns which allows an enormous possibility of predicting futuristic behavior of consumers or fraudulent transactions accurately and precisely.

Ayasdi conducted a Case Study where one technology payment firm explored the use of the automated software; the technology firm tested its fraud detection model against the software to testify. The following was concluded:

- ➔ Software quickly visualized the most fraudulent transactions and produced a statically ranking of these figures.
- ➔ The approach:
  - ◆ Increased fraud detection accuracy
  - ◆ Reduced revenue loss
  - ◆ Improved customer satisfaction

# Result from Case Study

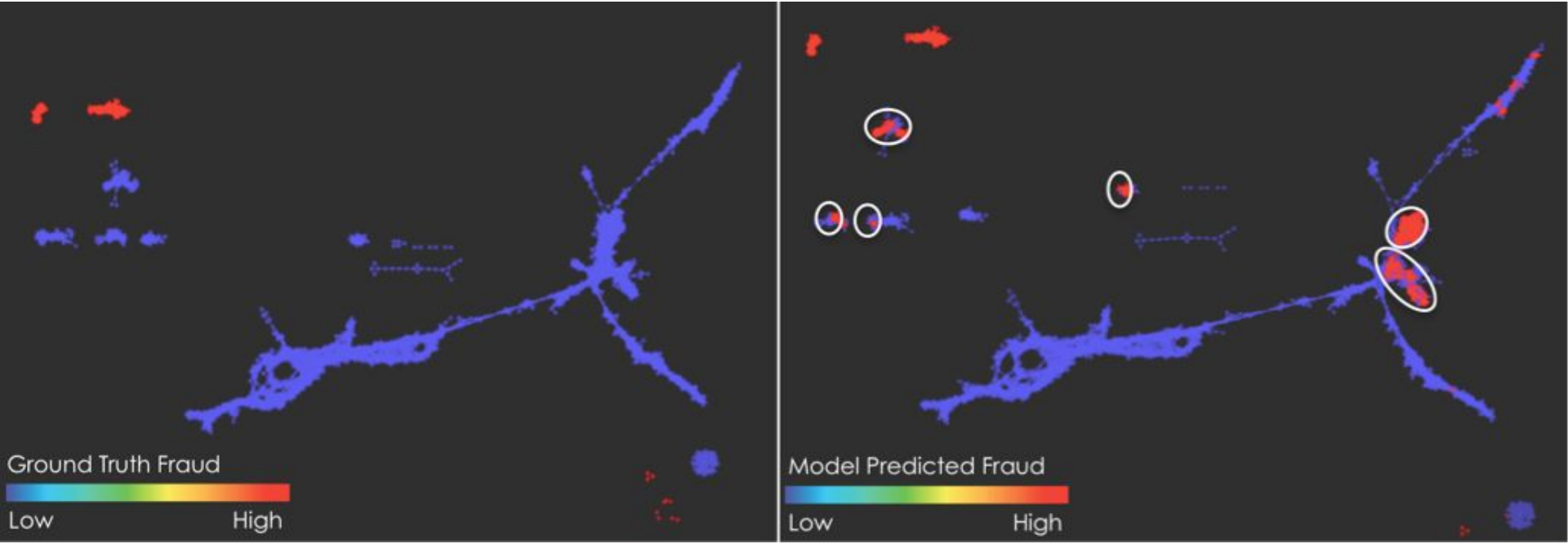
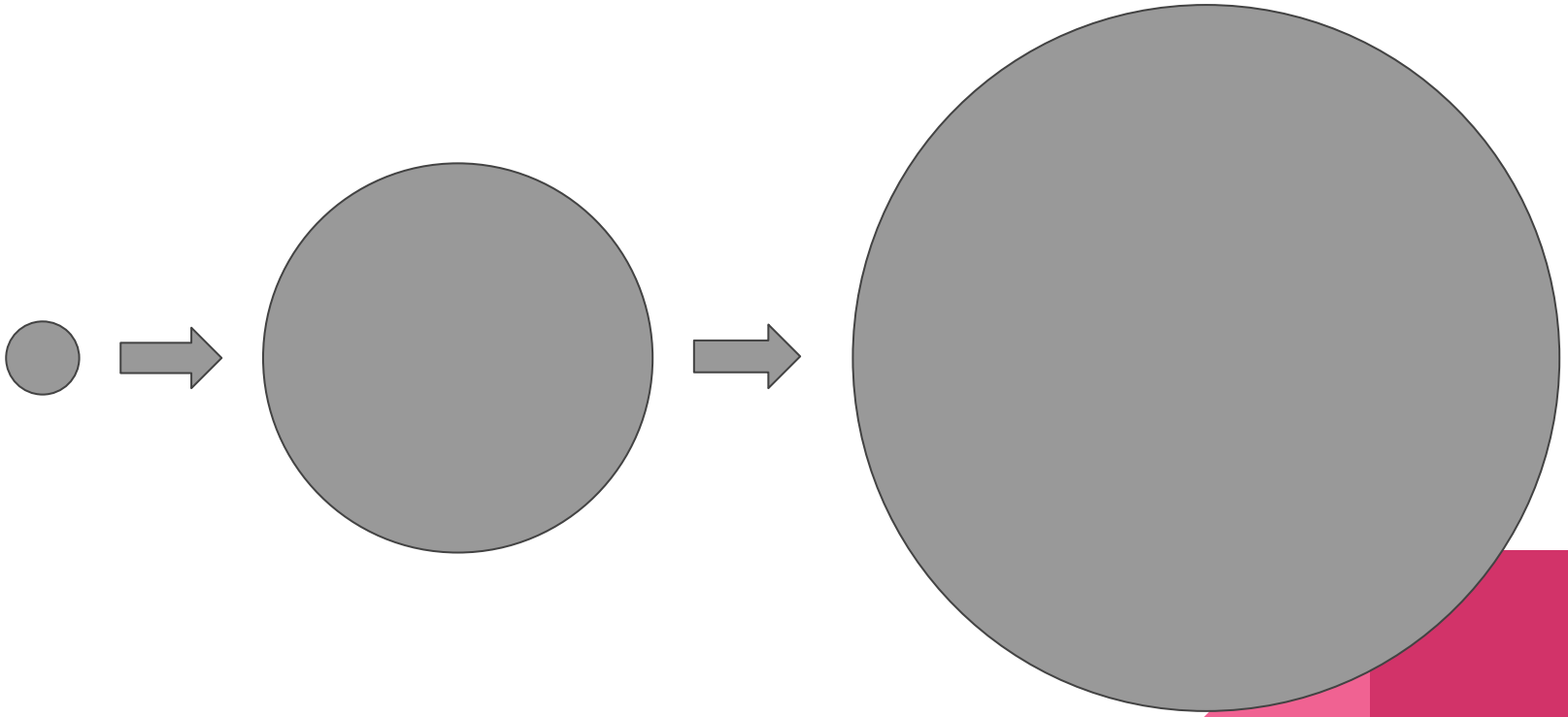
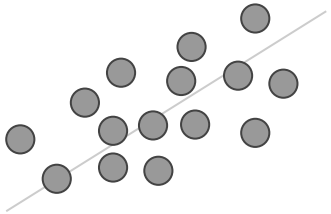



Image: https://www.ayasdi.com/industries/financial-services/fraud-detection/

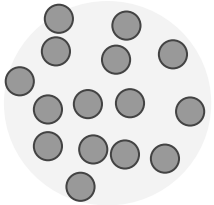
# Complexity grows exponentially



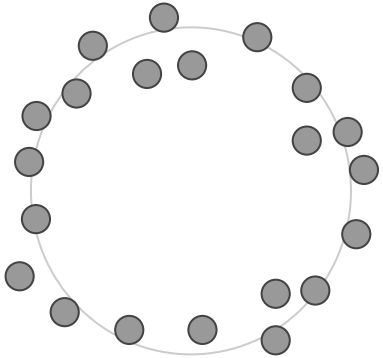
# Using TDA to discover hidden patterns



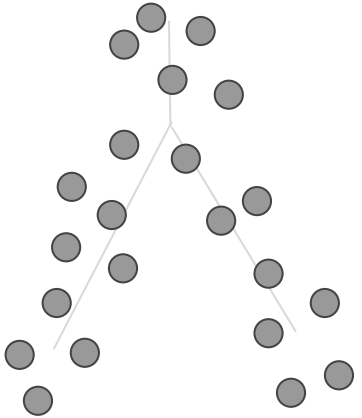
Line



Cluster



Loop



Flares

# Patterns/shapes are hidden

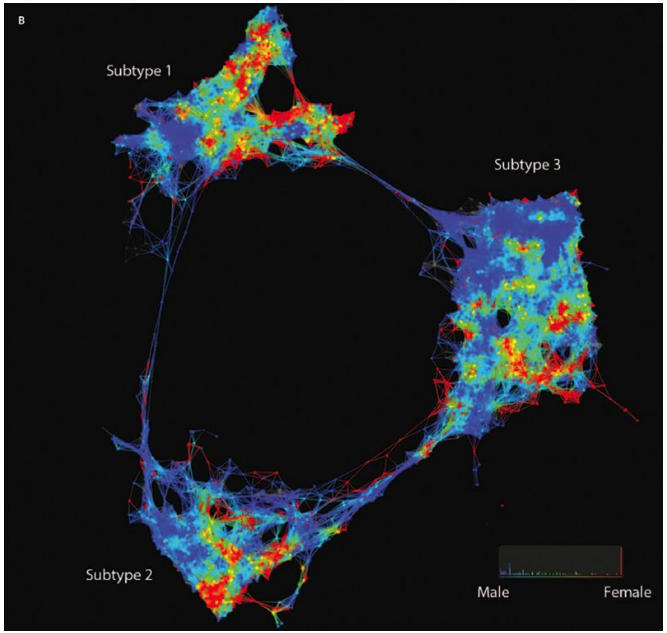
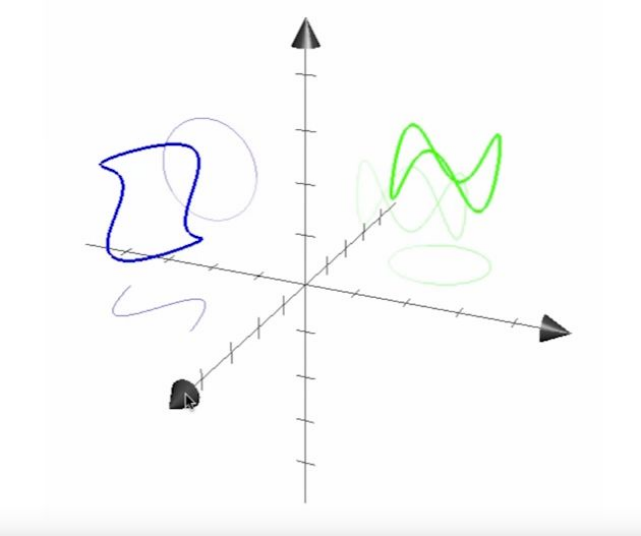



Image: https://www.ayasdi.com/blog/healthcare/identification-of-type-2-diabetes-subgroups-through-topological-data-analysis-of-patient-similarity/

# Dataset comes in multi-dimension

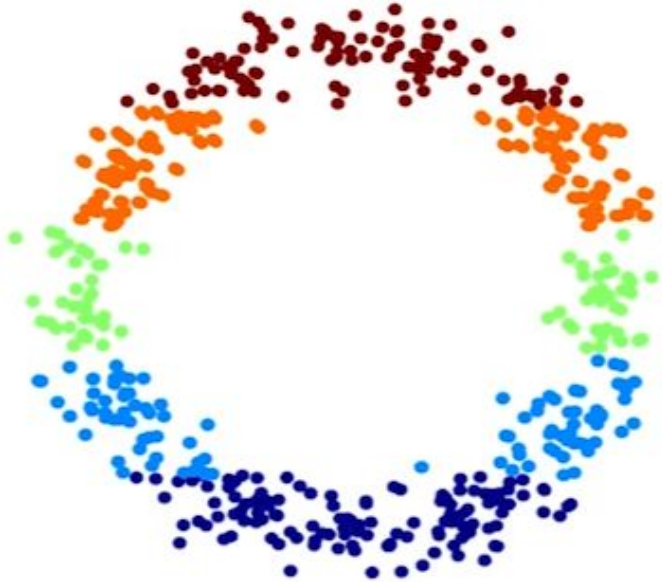
The topological feature (shape) could be obscure



Analysis of such multi-dimensional data could be difficult

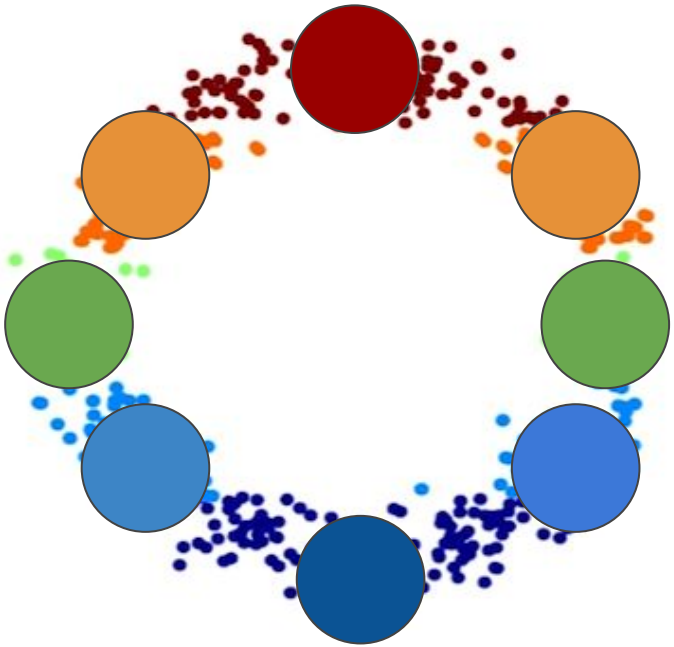
Source: https://www.youtube.com/watch?v=F2t_ytTLrQ4

# How to recognize shape using algorithm



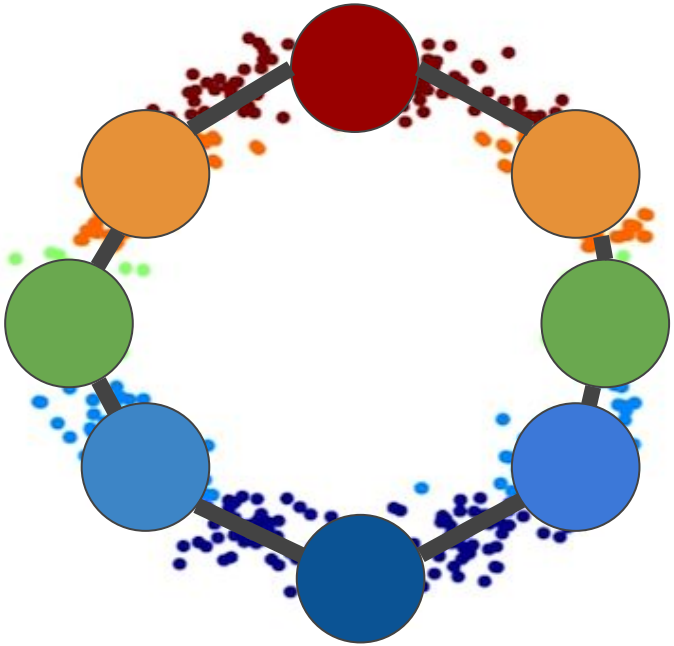
Source: https://www.youtube.com/watch?v=F2t_ytTLrQ4

# How to recognize shape using algorithm



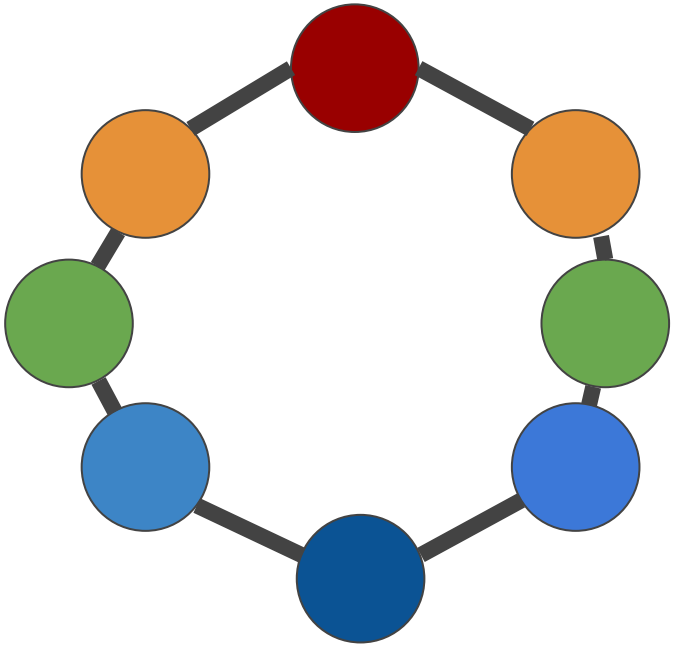
Source: https://www.youtube.com/watch?v=F2t_ytTLrQ4

# How to recognize shape using algorithm



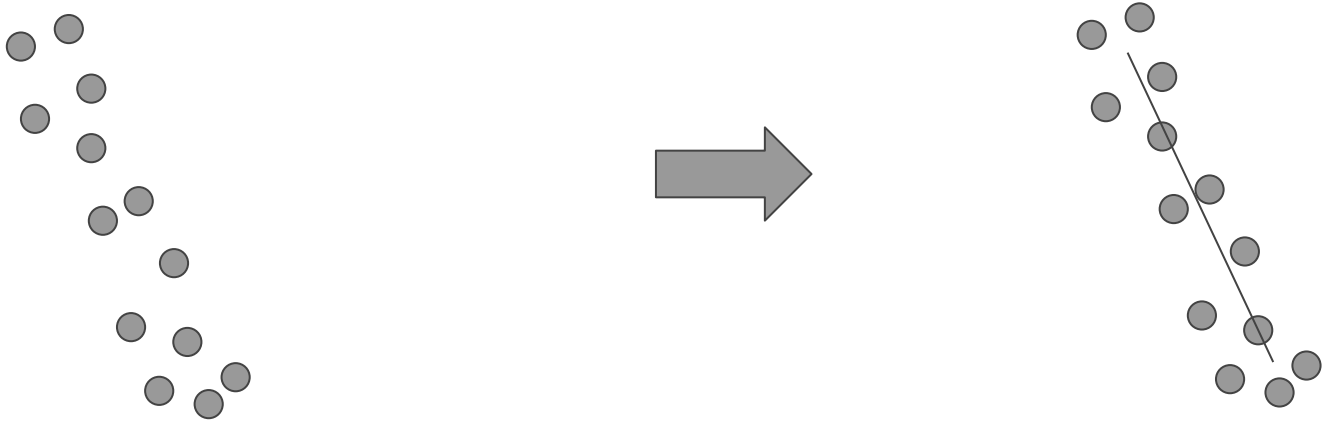
Source: https://www.youtube.com/watch?v=F2t_ytTLrQ4

# How to recognize shape using algorithm

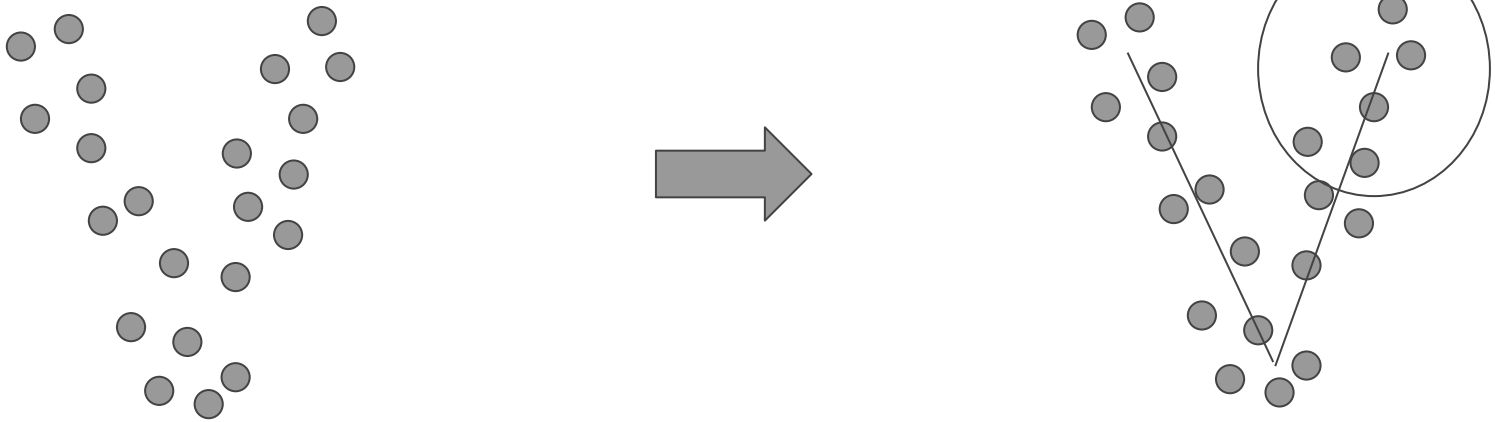


Source: https://www.youtube.com/watch?v=F2t_ytTLrQ4

# From shapes to Regions



# From shapes to Regions

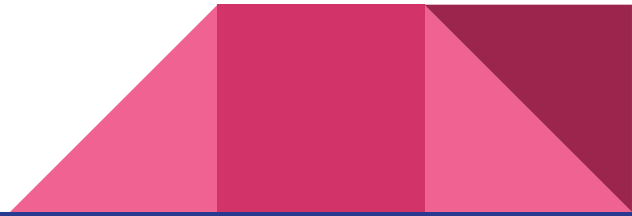


# Modelling is much easier now

Modelling is much easier when data share similarity

Algorithm could now be apply to data regions to summarize data pattern, and generate topological summary.

- Statistic
- Geometry
- Machine learning
- etc.



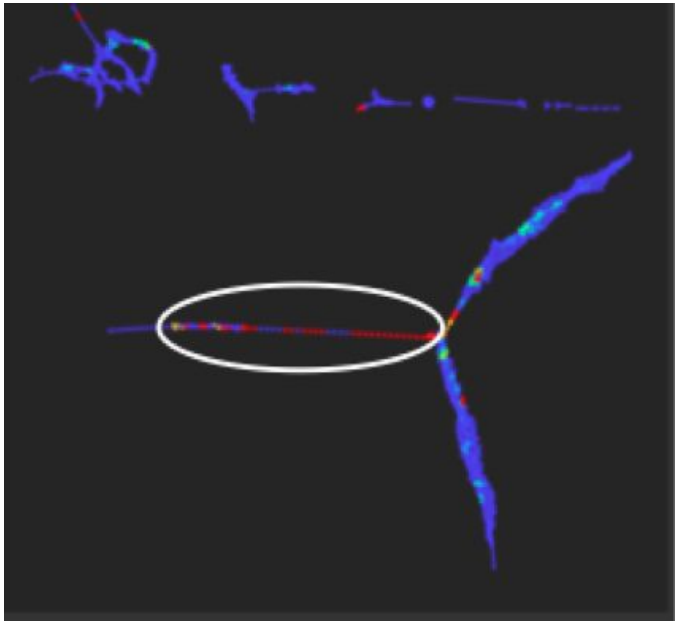
# Overview of TDA

1. Using topological algorithms to recognize shapes in multi-dimensional data.
2. Recognizing shapes and aggregate them into regions.
3. Applying various algorithms to summarize the characteristics of each regions.

Benefits:

- Fast response
- **Unsupervised system. Less bias. -** *The TDA process could recognize abnormality in input data without human assistance.*
- Multi-dimension analysis

# TDA Example: Healthcare Providers Network



**Data includes:**

- Utilization of Services
- Payments
- Units Per Claim
- Many Other Arbitrary Variables

**Did you notice?**

The circular white region is contour to the y-coordinate? That is because clusters of information was found along that region. (The blue dotted region)

**Why does this matter?!**

One can change the color of different nodes of the network based on past committed fraud or data used to analyze fraud (such data may include payment thresholds, multiple services provided in a short time, etc..)

Image: https://www.ayasdi.com/resources/whitepaper/topology-and-topological-data-analysis/

# Roadmap

- ✓ **Introduction**
  - ✓ What is fraud?
  - ✓ Why does it matter?
  - ✓ What is fraud detection?
- ❑ Fraud detection techniques
  - ✓ **Topological Data Analysis**
  - ❑ Case Based Reasoning
  - ❑ P-RCE Neural Network
  - ❑ Self-Organizing Maps

# Case Based Reasoning (CBR)

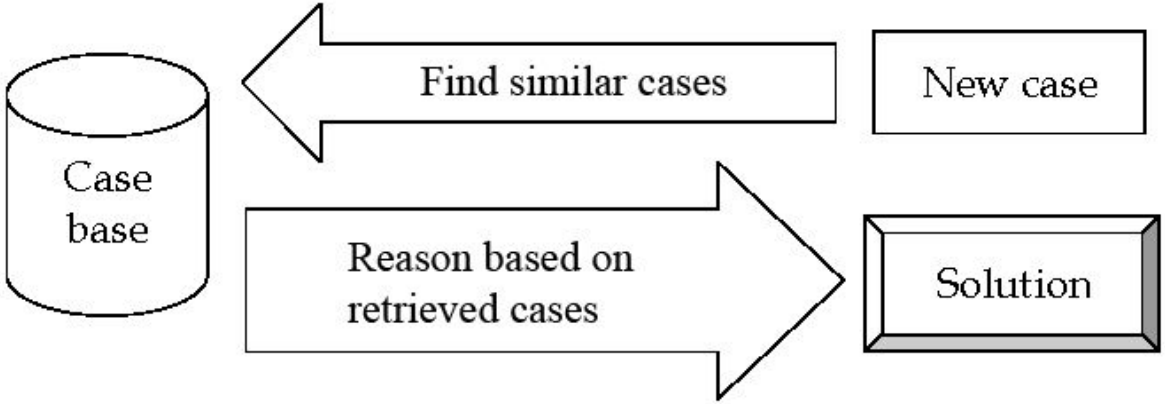



Image: http://www.intechopen.com/books/advances-in-industrial-design-engineering/design-for-automotive-panels-supported-by-an-expert-system

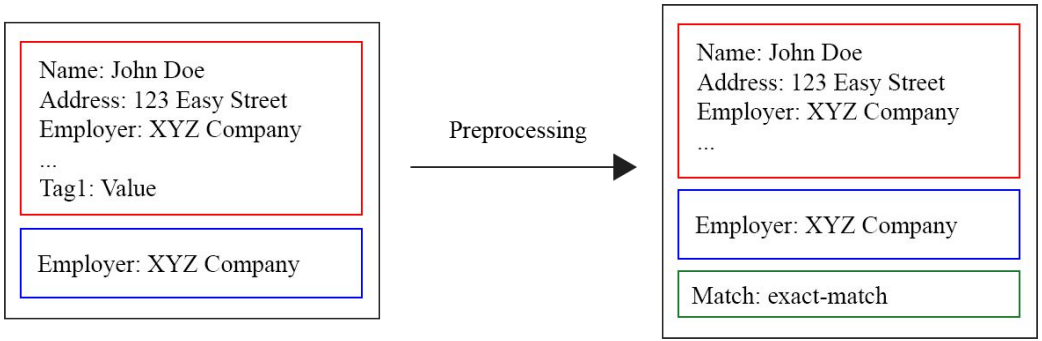
# Why use CBR?

- Provide meaningful confidence scores
- Little or no expert knowledge acquisition
- Easily updated and maintained
- Easy to see reasoning behind decision making
- Flexible to noisy or missing data
- Easily integrated with various databases

Source: Wheeler & Aitken (2000)

# CBR System: Data

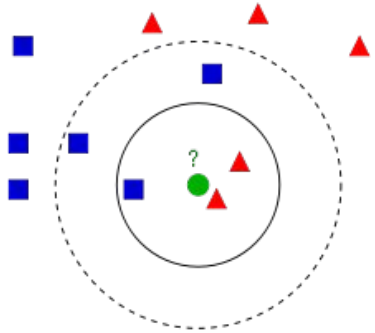
Source data consisted of cases containing application and evidence records. An additional fuzzy match descriptor was added to each pair during preprocessing…




Final data set contained **584 non-fraud cases** and **96 fraud**. In total, each case was described by 128 attributes.

# CBR System: Case Retrieval

- Wheeler & Aitken use a *thresholded retrieval* to retrieve relevant cases
- How does this work?
  1. Compute a similarity score for each existing case (all features weighted equally)
  2. Retrieve all cases with a similarity score above a certain threshold (can be adjusted dynamically)

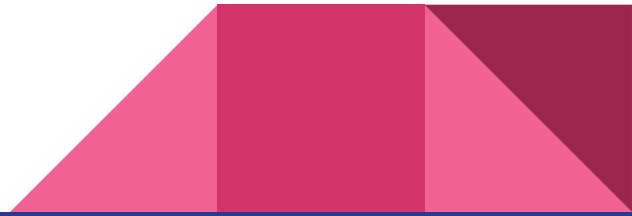



k-NN
Test case is green

Image: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

# CBR System: Diagnosis

- Wheeler & Aitken employ multiple algorithms to analyze the cases retrieved and make a fraud or non-fraud decision…
  - Density Selection
  - Negative Selection
  - Probabilistic Curve
  - Best Match
  - Default
- To resolve conflicts…
  - Run the algorithms in sequence
  - Use prediction of first one that fires



# CBR System: Results

- The proposed system was able to accurately identify **80%** of fraud cases and **52%** of non-fraud cases
- Typical systems (at the time of publishing) could identify only **60%** of fraud cases and **30%** of non-fraud cases
- Wheeler & Aiken conclude that “... the initial testing results… strongly suggest that multi-algorithmic CBR will be capable of high accuracy rates.”

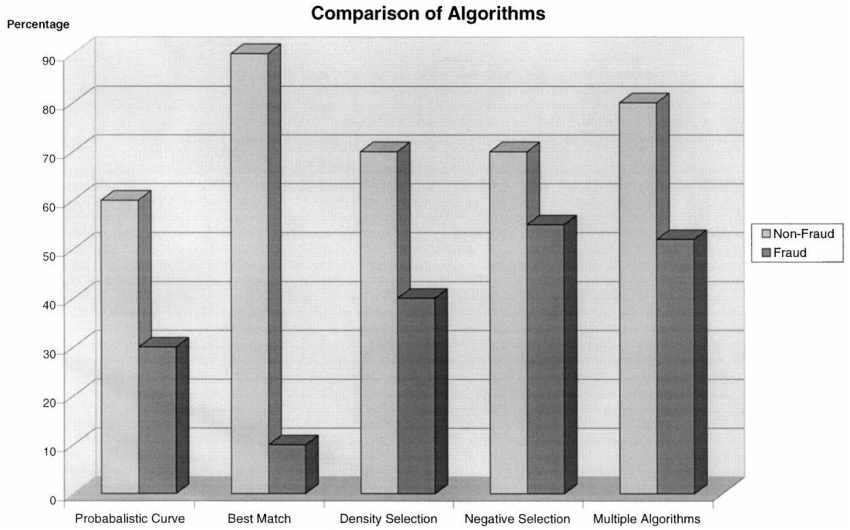



Fig. 1. Comparison of Algorithms

Image: Wheeler & Aitken, 2000

# Roadmap

- ✓ **Introduction**
  - ✓ What is fraud?
  - ✓ Why does it matter?
  - ✓ What is fraud detection?
- ❑ Fraud detection techniques
  - ✓ **Topological Data Analysis**
  - ✓ **Case Based Reasoning**
  - ❑ P-RCE Neural Network
  - ❑ Self-Organizing Maps

# P-RCE Neural Network

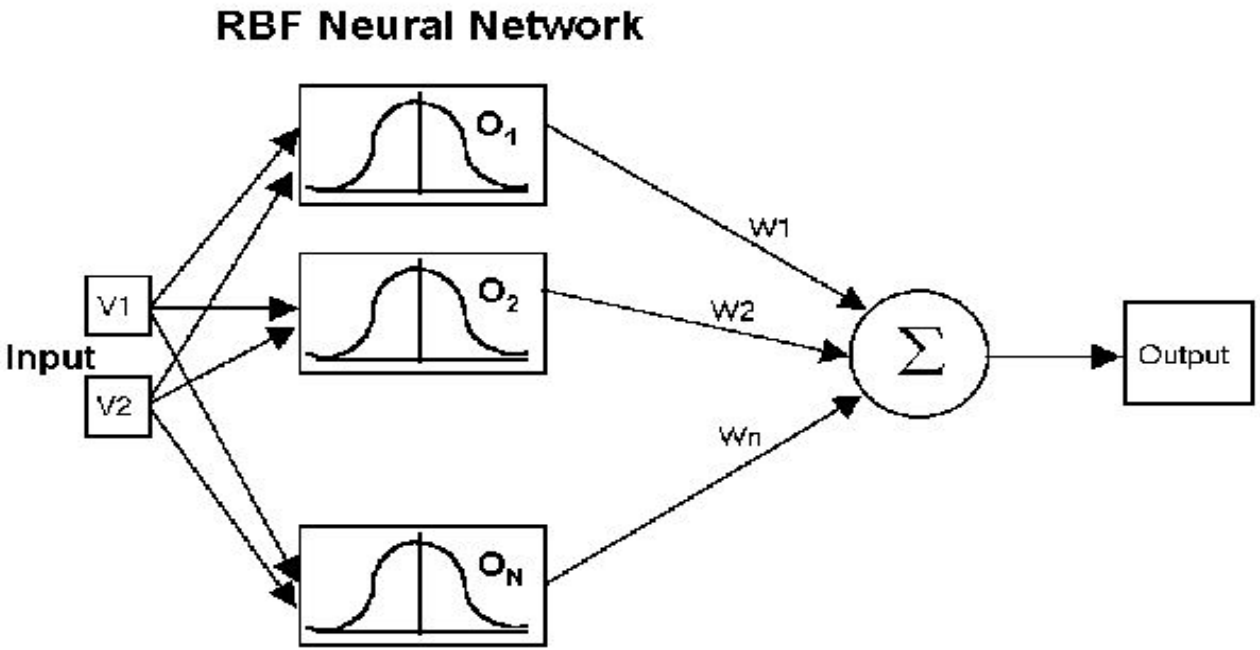



Image: https://www.dtreg.com/uploaded/pageimg/RBFarchitecture.gif

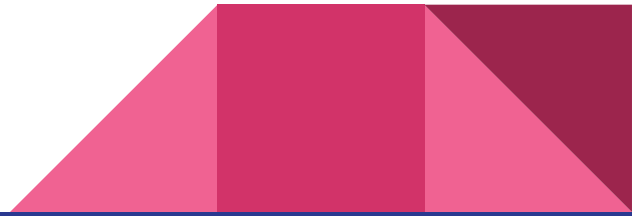
# P-RCE Neural Network Motivation

- Fraud is growing more quickly than the credit card business itself
- There is a lot to improve on the previous method
- There are many different types of credit card frauds
- Fraud needs to detected more quickly to recover more money

Image: http://il8.picdn.net/shutterstock/videos/245662/thumb/1.jpg

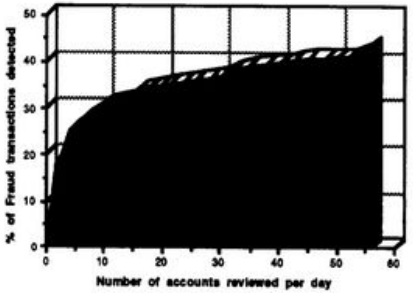
# How P-RCE Neural Networks Work

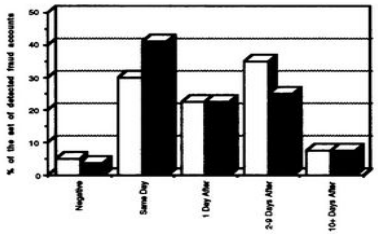
- Three layer feed-forward system
- Trained on a large sample of labeled credit card transactions that includes good transactions and different types of fraudulent transactions
- P-RCE Networks perform two passes during the training
- It outputs a fraud score for each transaction



# Impact

- The previous system ran 750 transactions per day to detect one fraud per week whereas this system runs 50 a day and catches 40 percent
- Up to 60% of detected fraud is caught on first or second day
- It can detect all types of fraud



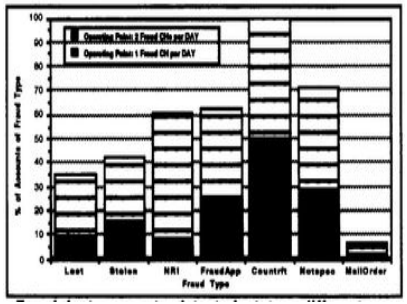



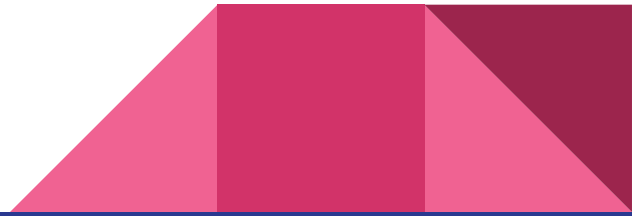

Image: http://bit.csc.lsu.edu/~jianhua/quang.pdf

# Roadmap

- ✓ **Introduction**
  - ✓ What is fraud?
  - ✓ Why does it matter?
  - ✓ What is fraud detection?
- ❑ Fraud detection techniques
  - ✓ **Topological Data Analysis**
  - ✓ **Case Based Reasoning**
  - ✓ **P-RCE Neural Network**
  - ❑ Self-Organizing Maps

# Self-Organizing Maps

- Used in credit card fraud detection.
- Provide an easy way to visualize data, as well as visualize connections in data.
- Relatively simple to implement.
- Once constructed, classifying the data can be transformed into classical graph problems, many of which have solutions in O(m+n) time, where m is number of edges and n is number of vertices.



# Why use a SOM?

- SOMs excel in clustering data. Once these clusters are created then patterns can be concluded.
- Since fraudulent transactions are continuously evolving and changing, self organizing maps can identify fraudulent transaction methods with no prior knowledge about them.
- Most of the time companies do not have any prior knowledge to fraudulent transactions, hence why SOMs are so useful.
- SOMs also provide insight into transaction patterns, and can help with association of transactions.

# More on Self Organizing Maps

- They are very good at finding and visualizing clusters.
- Since they do not use error correction learning (i.e. back propagation) instead they use what is known as competitive learning and use a neighborhood function to preserve topological properties of the input data, represented in a n-dimensional space.
- Self organizing maps work by mapping neurons onto data, and then adjusting nearby neurons, for each epoch until an acceptable configuration state is reached.
- As more data is processed efficiency is increased as nearby neurons will not move around as much.

# A Visual Overview

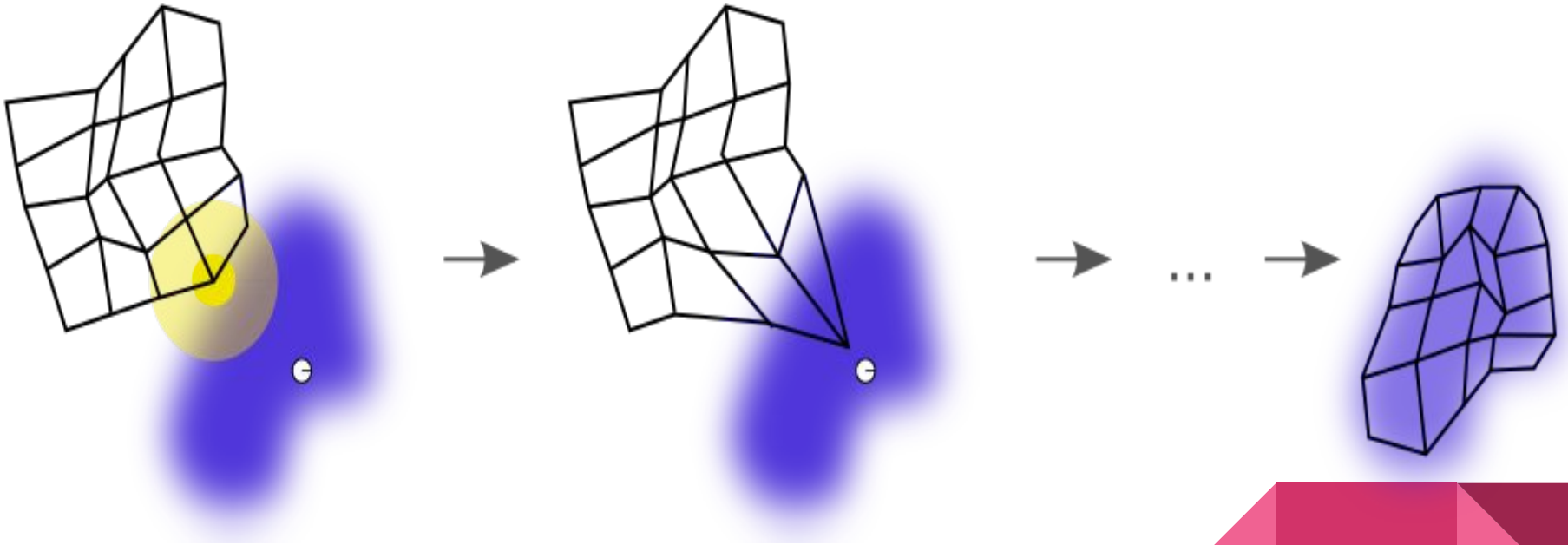


Image: https://en.wikipedia.org/wiki/Self-organizing_map

# Pseudo Code

Initialize weights

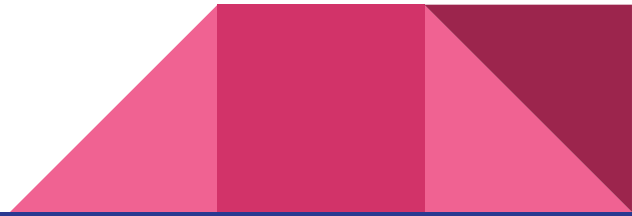
For 0 to X number of training epochs

    Select a sample from the input data set

    Find the "winning" neuron for the sample input

    Adjust the weights of nearby neurons

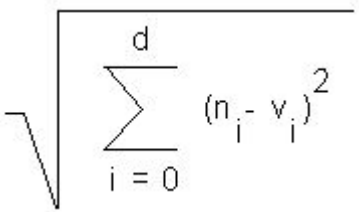
End for loop



# Subroutines (Winning Neuron)

Winning Neuron:

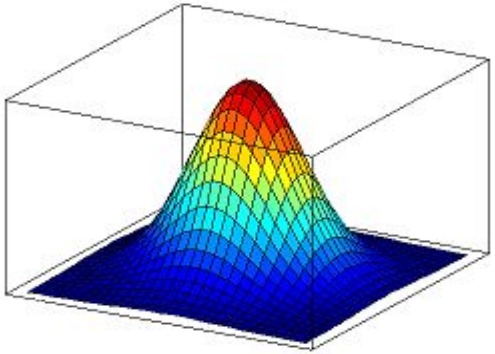
The winning neuron, is selected by a measure of distance from the point in, d, dimensional space we are examining to the neural map. In most cases this is the euclidean distance, that is:

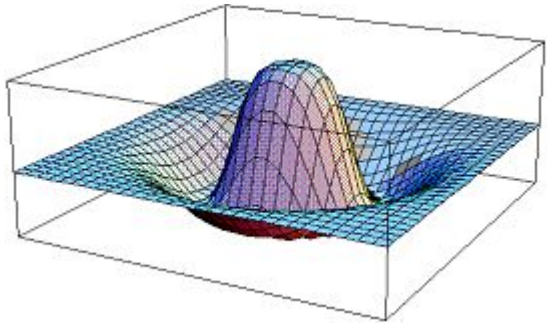



Where d is number of dimensions of data, v_i is the ith dimension value of the vertex we are examining, and n_i is the ith dimension value of the neuron

# Subroutines
# (Adjusting Weights)

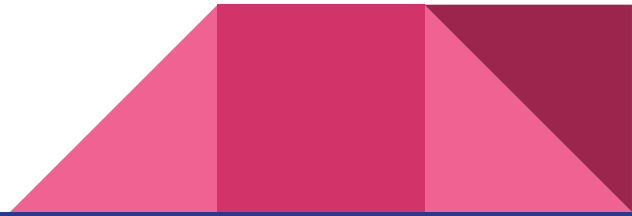
To adjust the weights, the procedure in general is to move all neurons, in a general subspace $S_d$, (of same dimension as vertices in graph: i.e. d), by a scaleable amount where the scale decreases as a function of the distance from the examined vertex. In most cases this is a gaussian function. A sample gaussian function on 3 dimensions is shown below. Also a normalized second derivative gaussian function can be used, (mexican hat function).





# Summary of SOM

- Thus because of the clustering properties of SOMs they are good filtering mechanisms.
- Therefore often times they are used in conjunction to a feed forward neural network, and used to decrease the complexity of parsing through a large selection of input.
- Also other types of classifiers, can be used in conjunction to SOMs, it is not limited to only using neural networks and this then adds both efficiency and accuracy if implemented correctly.



# Roadmap

- ✓ **Introduction**
  - ✓ What is fraud?
  - ✓ Why does it matter?
  - ✓ What is fraud detection?
- ✓ **Fraud detection techniques**
  - ✓ Topological Data Analysis
  - ✓ Case Based Reasoning
  - ✓ P-RCE Neural Network
  - ✓ Self-Organizing Maps