Automatic Prioritizing of Abnormal Situations in the Context of Fraudulent Claims (with Graph-of-relations Analysis)

Prepared by

Marc-André Desrosiers
the co-operators
Expert, Claims Portfolio Analytics

Matthew Griffith
University of Bath
Ph.D. Student

Francis Duval
UQAM
Ph.D. Student

Bowei Zhang
University of Western Ontario
Master Student

2019-08-23

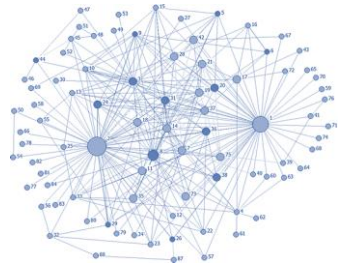
A win this morning ...

is that you

- understand the approach that the group took to make progress on the problem
- have fun listening to us

Une place de choix
the co-operators
A Better Place For You
UQÀM
McMaster University
Western
UNIVERSITY OF BATH
CENTRE DE RECHERCHES MATHÉMATIQUES
IVADO
Université Concordia University
MEMORIAL UNIVERSITY
UNIVERSITY OF OXFORD

Problem re-statement



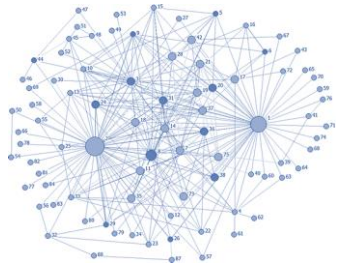
Produce an interpretable claim-level score that allows Special Investigation Unit analysts to decide which claims to investigate

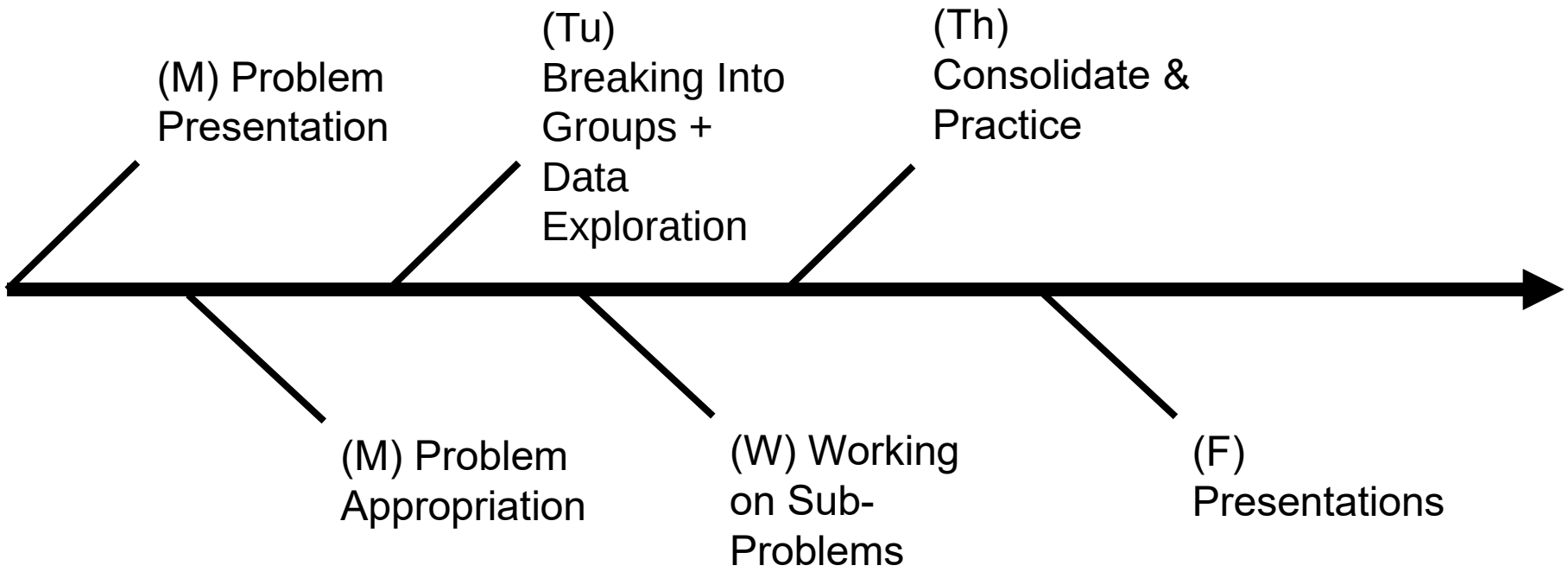
The team

Fabian Ying, Lindon Roberts, Matthew Griffith, David Mazurkiewicz, Anthony Forgetta
Marc-André Desrosiers, Steven Côté, Anas Abdallah, Helen Samara Dos Santos
Mathilde Bourget, Manuel Morales, Caio De Naday Hornhardt, Bowei Zhang, Francis Duval

Une place de choix
the co-operators
A Better Place For You
UQÀM
McMaster University
Western
UNIVERSITY OF BATH
CENTRE DE RECHERCHES MATHÉMATIQUES
IVADO
Université Concordia University
MEMORIAL UNIVERSITY
UNIVERSITY OF OXFORD

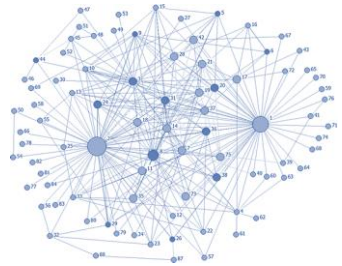
Timeline





Une place de choix
the co-operators
A Better Place For You
UQÀM
McMaster University
Western
UNIVERSITY OF BATH
CENTRE DE RECHERCHES MATHÉMATIQUES
IVADO
Université Concordia University
MEMORIAL UNIVERSITY
UNIVERSITY OF OXFORD

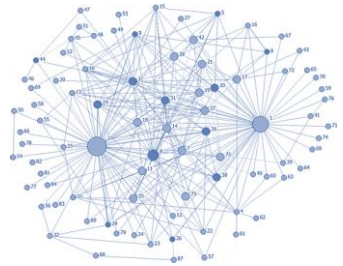
Academic point-of-view



- For supervised learning, more data!
- Better network data!
- More detailed annotations!
- High-dimensional anomaly detection is generally difficult!!!

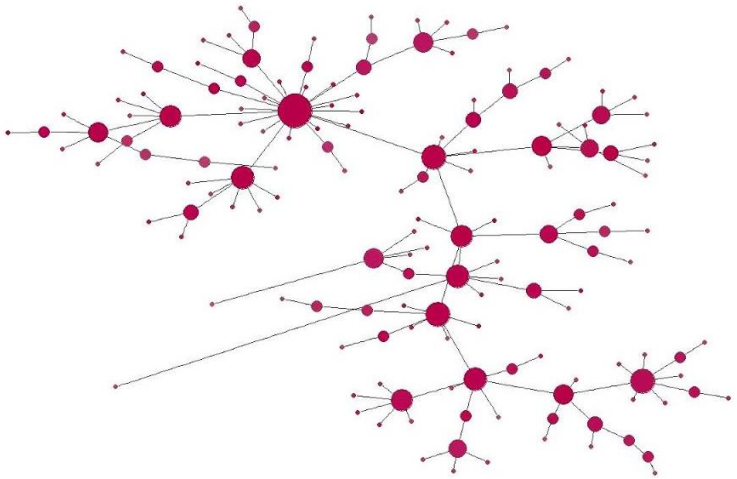
Une place de choix
the co-operators
A Better Place For You
UQÀM
McMaster University
Western
UNIVERSITY OF BATH
CENTRE DE RECHERCHES MATHÉMATIQUES
IVADO
UNIVERSITÉ Concordia UNIVERSITY
MEMORIAL UNIVERSITY
UNIVERSITY OF OXFORD

Taskforce creation



Data Exploration

Network Representation
Fabian, Matthew, Lindon



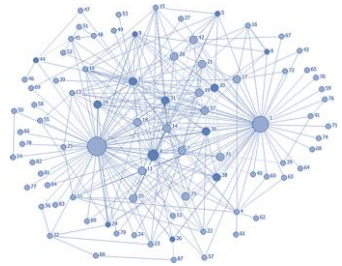
Anomaly Detection
Bowei, Anthony
Francis, Caio, Helen

Advanced Analytics
Bowei, Anthony
Francis, Caio, Helen

LASSO

Scoring Methodology

Advanced Analytics



Data Exploration

Network Representation
Fabian, Matthew, Lindon

Anomaly Detection
Bowei, Anthony
Francis, Caio, Helen

Advanced Analytics
Bowei, Anthony
Francis, Caio, Helen

LASSO

Scoring Methodology

Une place de choix
the co-operators
A Better Place For You

UQÀM

McMaster University

Western

UNIVERSITY OF BATH

CENTRE DE RECHERCHES MATHÉMATIQUES

IVADO

Université Concordia University

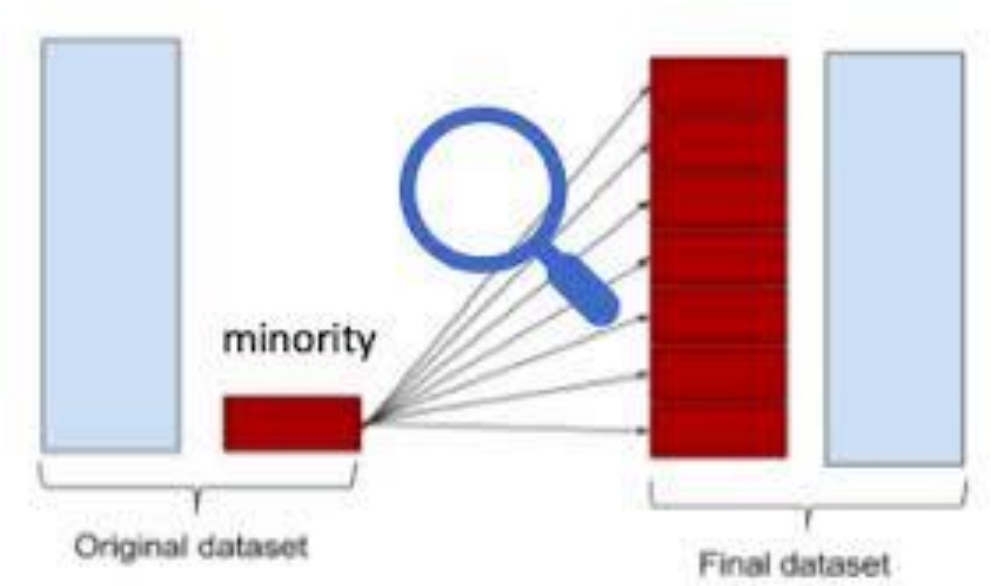
MEMORIAL UNIVERSITY

UNIVERSITY OF OXFORD

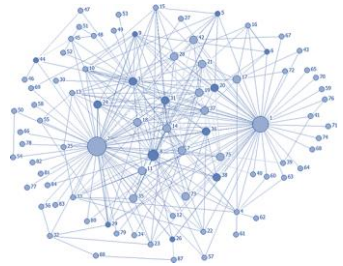
Advanced Analytics

Uncommon events

Oversampling (on steroids)

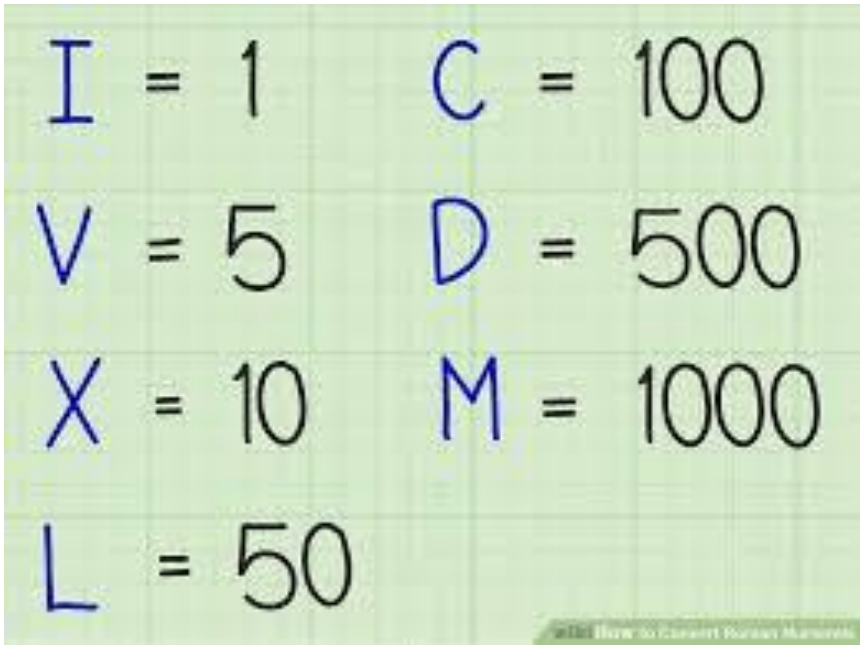


Advanced Analytics

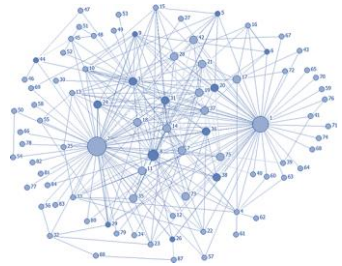


Categorical variables

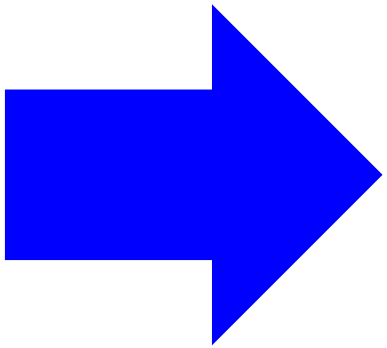
Target encoding



Advanced Analytics

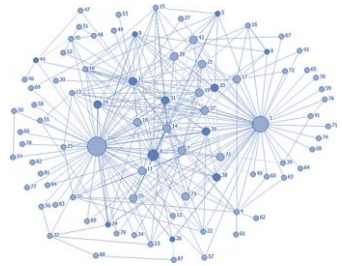


Missing values

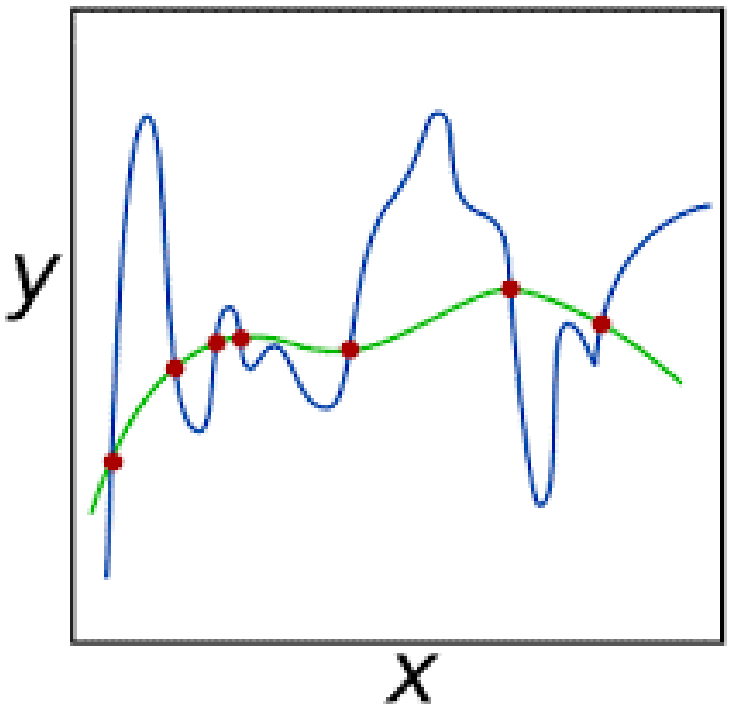


Appropriate treatment

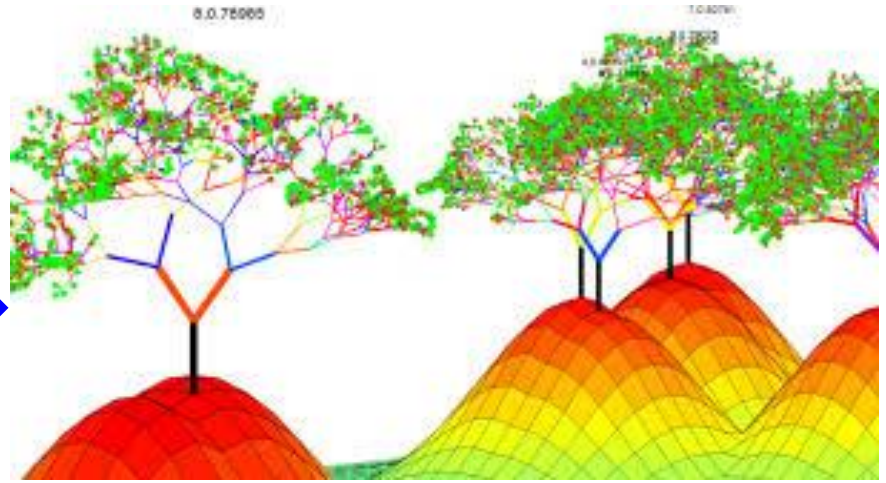
Advanced Analytics



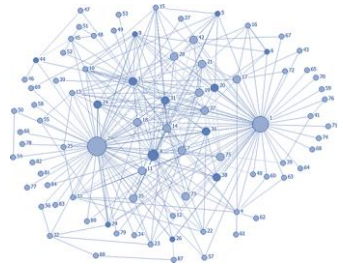
Selection of variables

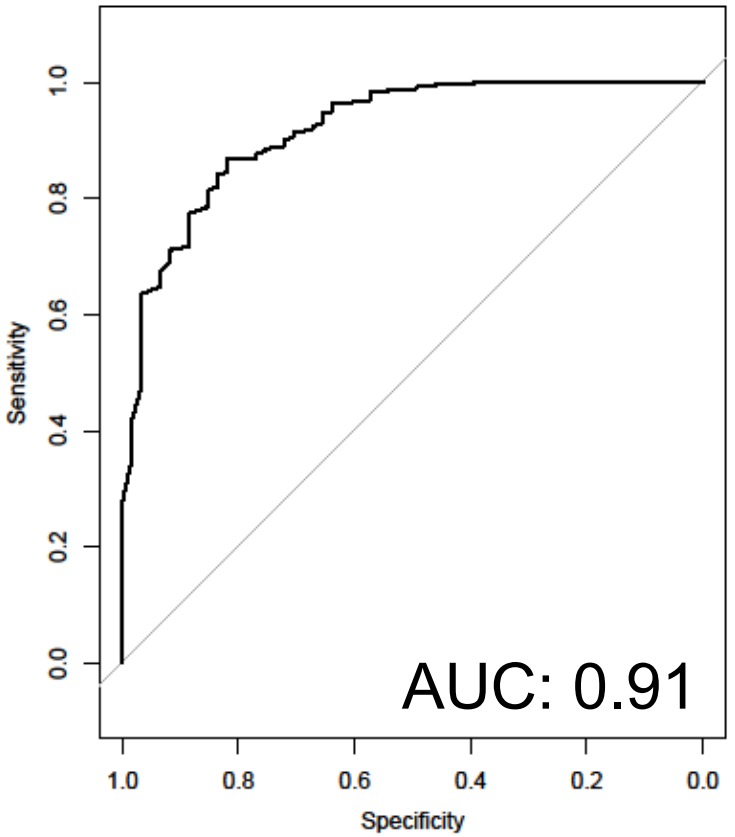


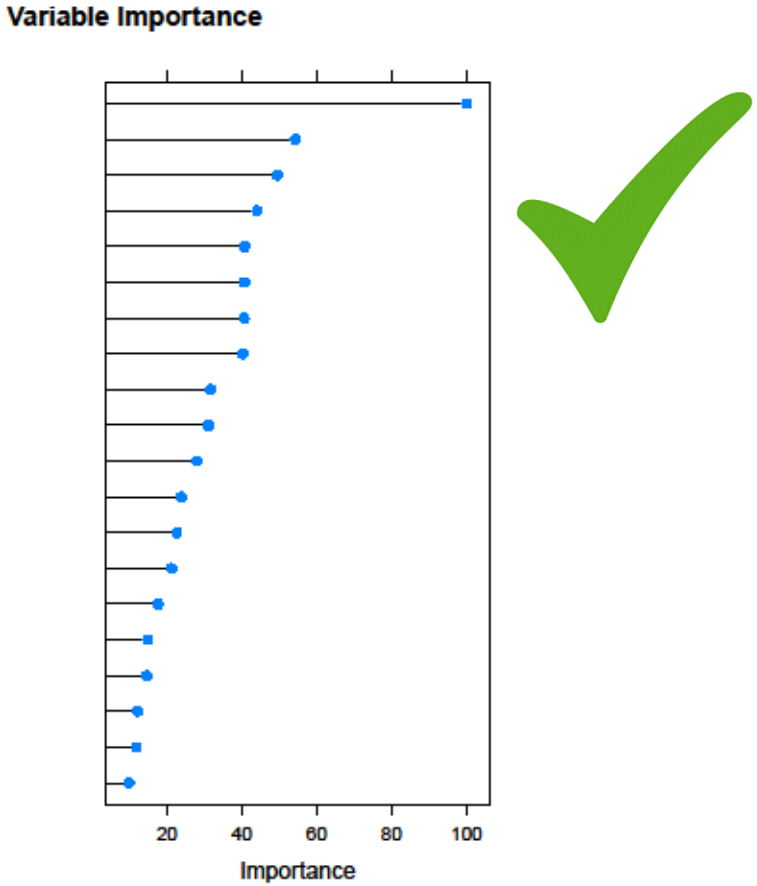
Importance of variables



Advanced Analytics







Une place de choix
the co-operators
A Better Place For You

UQÀM

McMaster University

Western

UNIVERSITY OF BATH

CENTRE DE RECHERCHES MATHÉMATIQUES

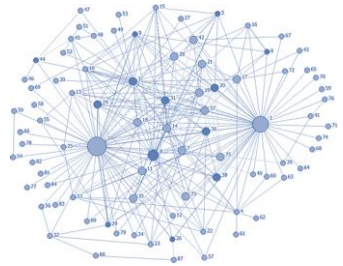
IVADO

UNIVERSITÉ Concordia UNIVERSITY

MEMORIAL UNIVERSITY

UNIVERSITY OF OXFORD

Anomaly Detection



Data Exploration

Network Representation
Fabian, Matthew, Lindon

Anomaly Detection
Bowei, Anthony
Francis, Caio, Helen

Advanced Analytics
Bowei, Anthony
Francis, Caio, Helen

LASSO

Scoring Methodology

Une place de choix
the co-operators
A Better Place For You

UQÀM

McMaster University

Western

UNIVERSITY OF BATH

CENTRE DE RECHERCHES MATHÉMATIQUES

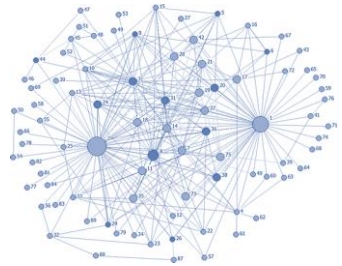
IVADO

Université Concordia University

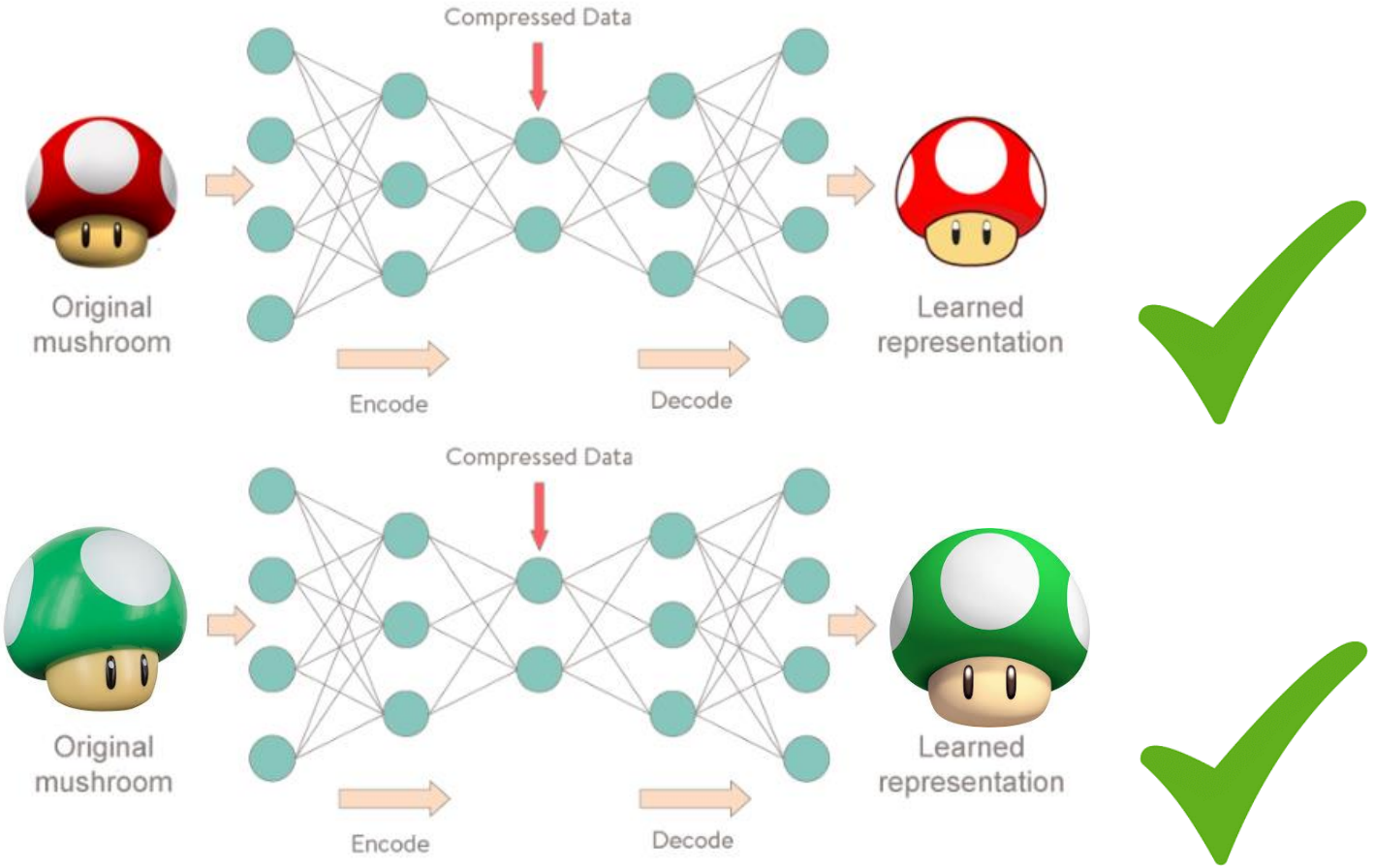
MEMORIAL UNIVERSITY

UNIVERSITY OF OXFORD

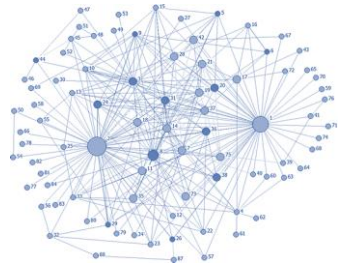
Anomaly Detection

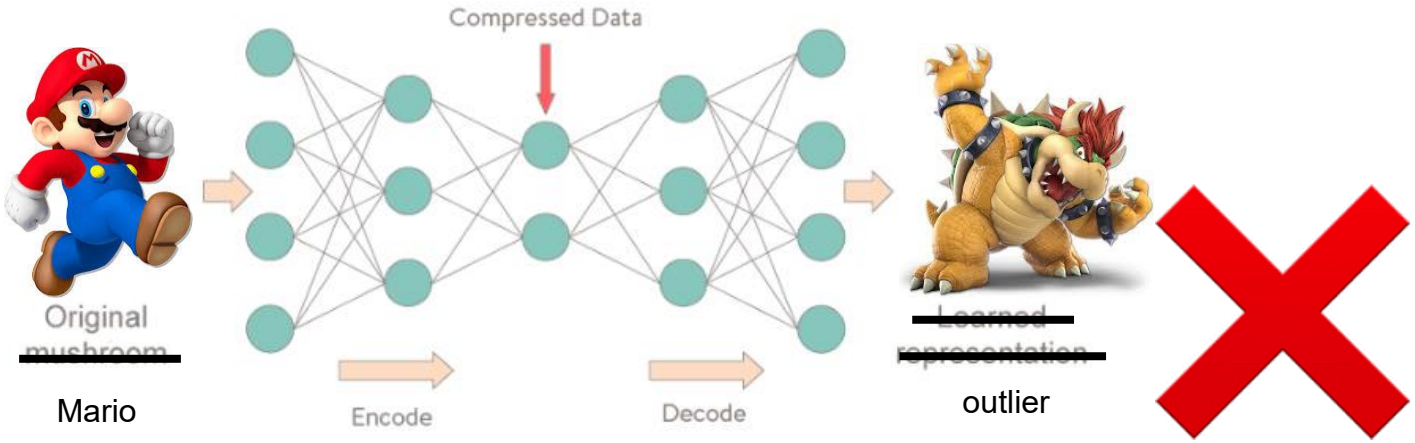


Autoencoder approach

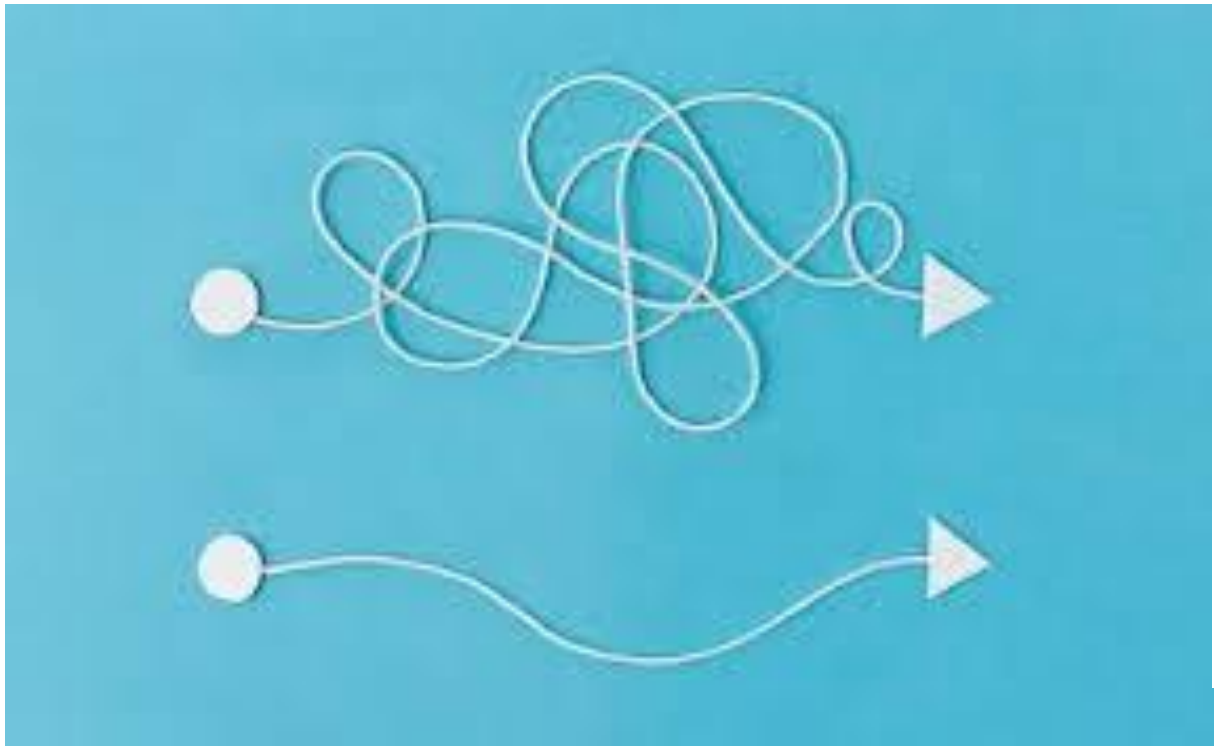


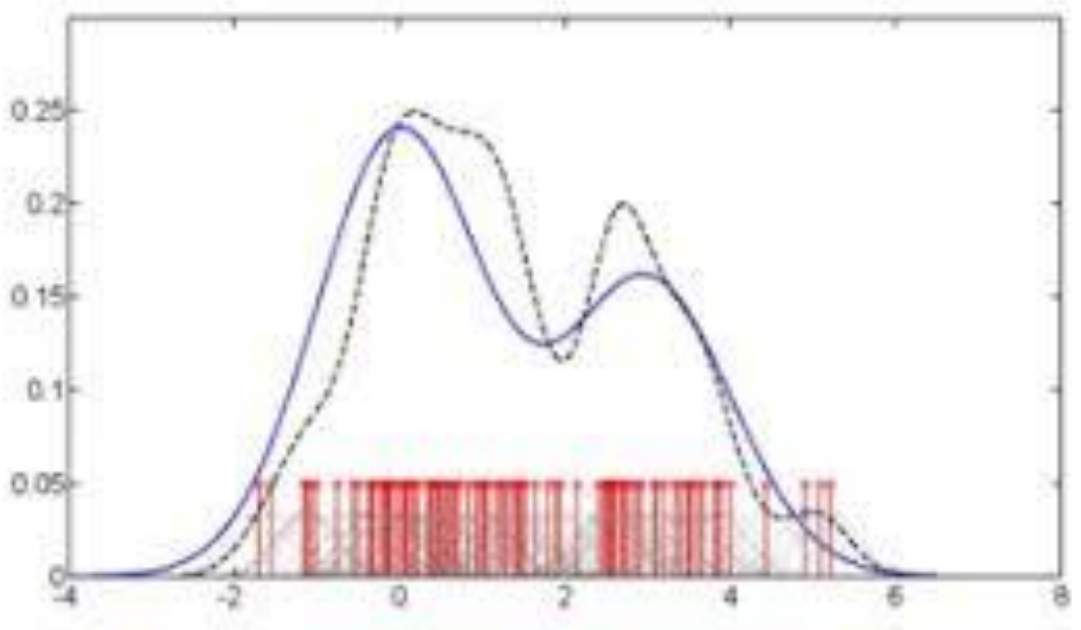
Anomaly Detection





Anomaly Detection



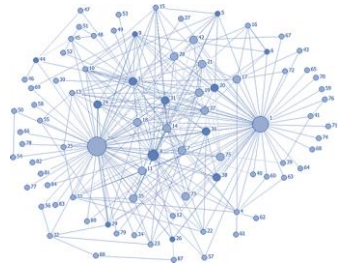


A couple of variables at a time

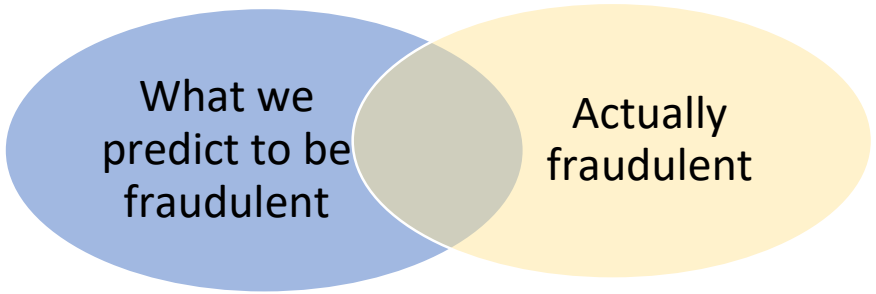
Worked with ~93 variables but not full dataset: convergence issues

Une place de choix
the co-operators
A Better Place For You
UQÀM
McMaster University
Western
UNIVERSITY OF BATH
CENTRE DE RECHERCHES MATHÉMATIQUES
IVADO
Université Concordia University
MEMORIAL UNIVERSITY
UNIVERSITY OF OXFORD

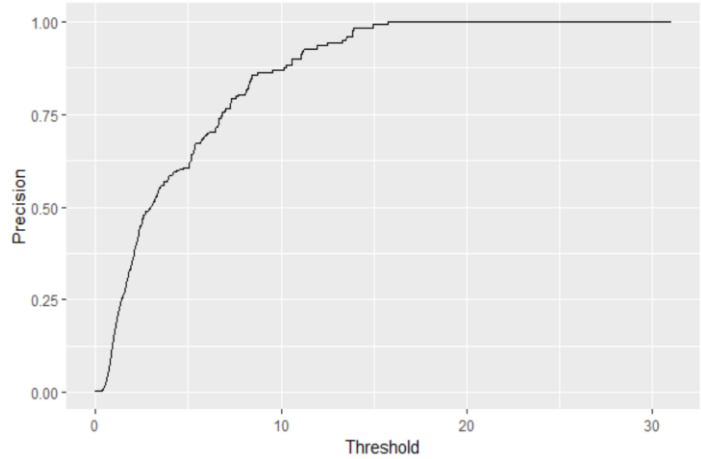
Anomaly Detection

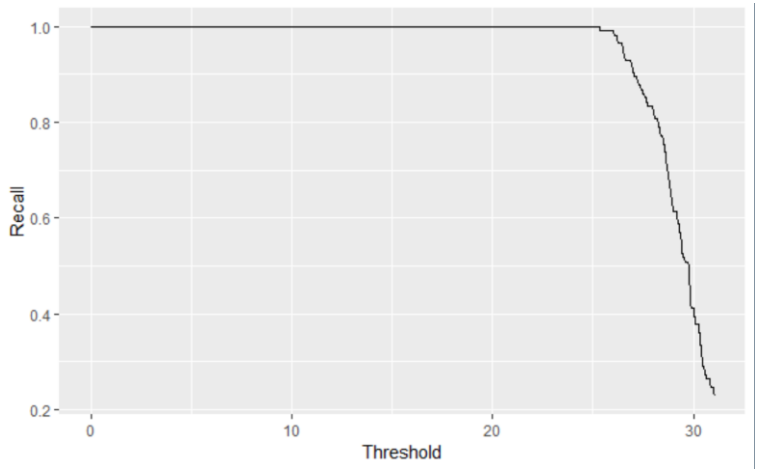


Precision = overlapping/BLUE



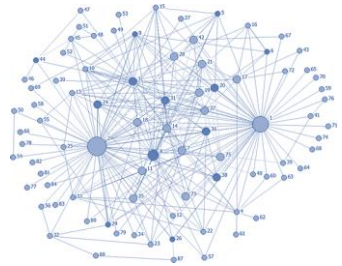
Recall = overlapping/YELLOW





(0.96; 1.00)

Network Representation

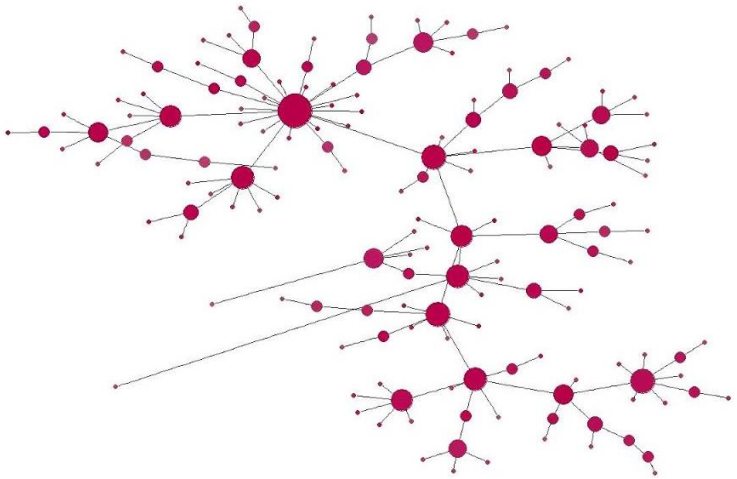


Data Exploration

Network Representation
Fabian, Matthew, Lindon

Anomaly Detection
Bowei, Anthony
Francis, Caio, Helen

Advanced Analytics
Bowei, Anthony
Francis, Caio, Helen



LASSO

Scoring Methodology

Une place de choix
the co-operators
A Better Place For You

UQÀM

McMaster University

Western

UNIVERSITY OF BATH

CENTRE DE RECHERCHES MATHÉMATIQUES

IVADO

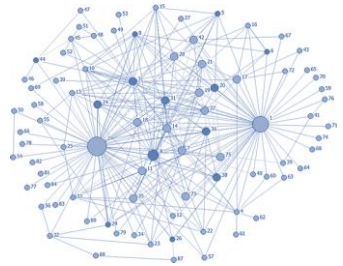
Université Concordia University

MEMORIAL UNIVERSITY

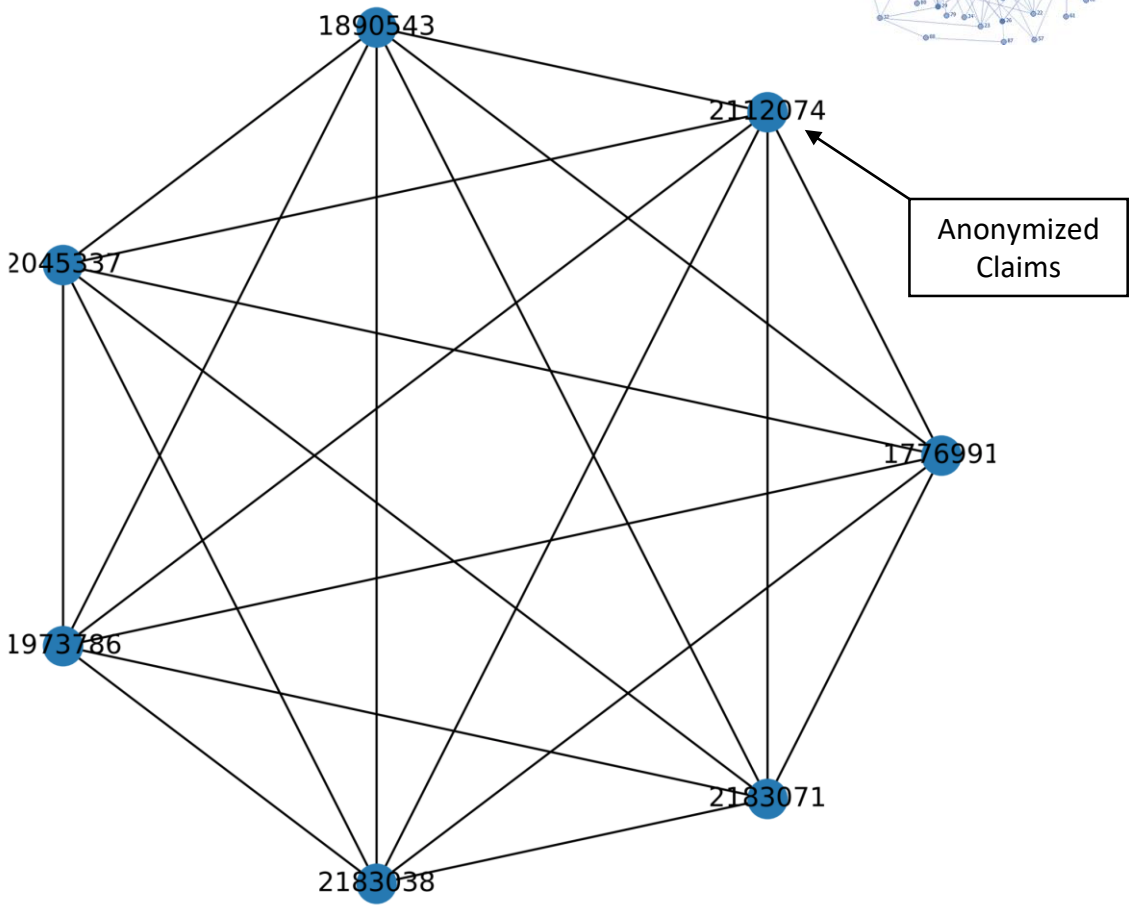
UNIVERSITY OF OXFORD

Network Representation

→ Motivation

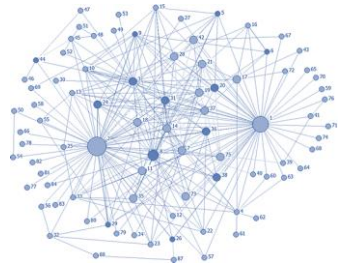


- Exploration of underlying network structure in data.
- Initial investigation using address.
- Important fraudulent case flagged. ⚠
- Suggestive of network structure being useful to detect fraudulent cases.
- Validated continuing with this method.



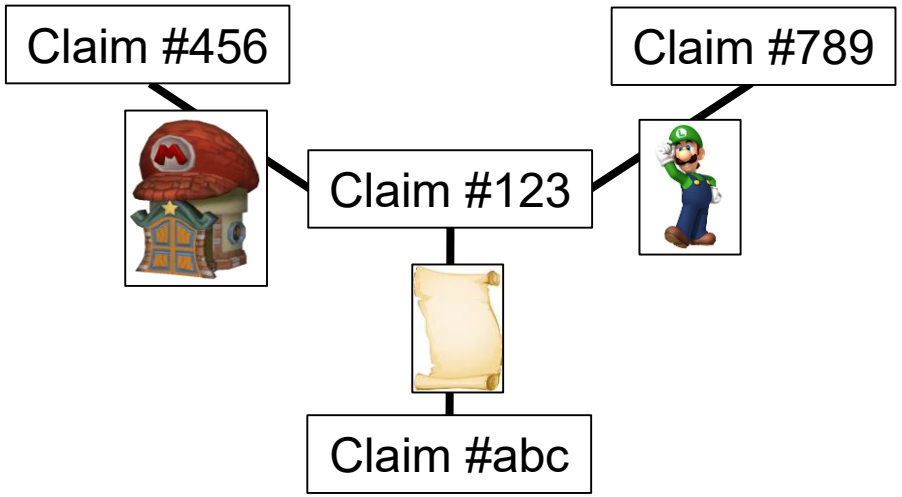
Network Representation

→ Construction



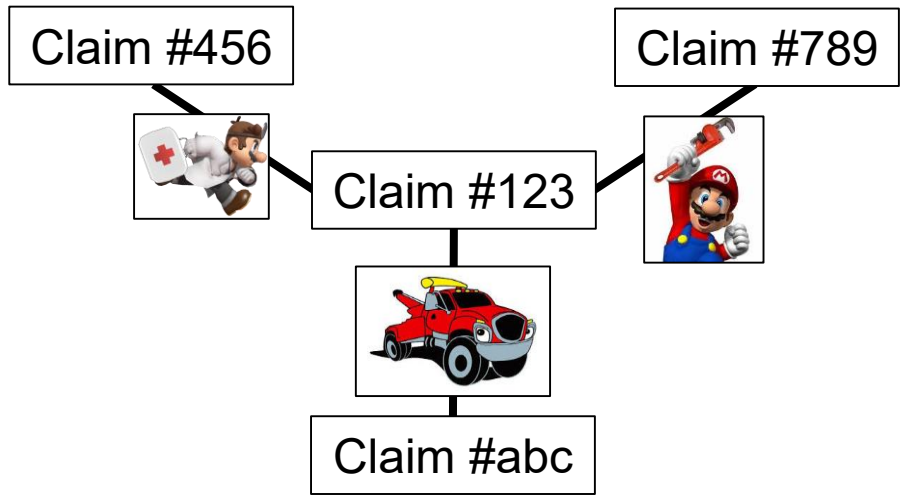
Personal Network

Connected by sharing personal properties



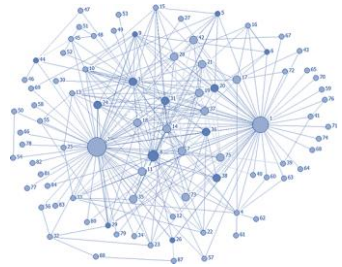
Vendor Network

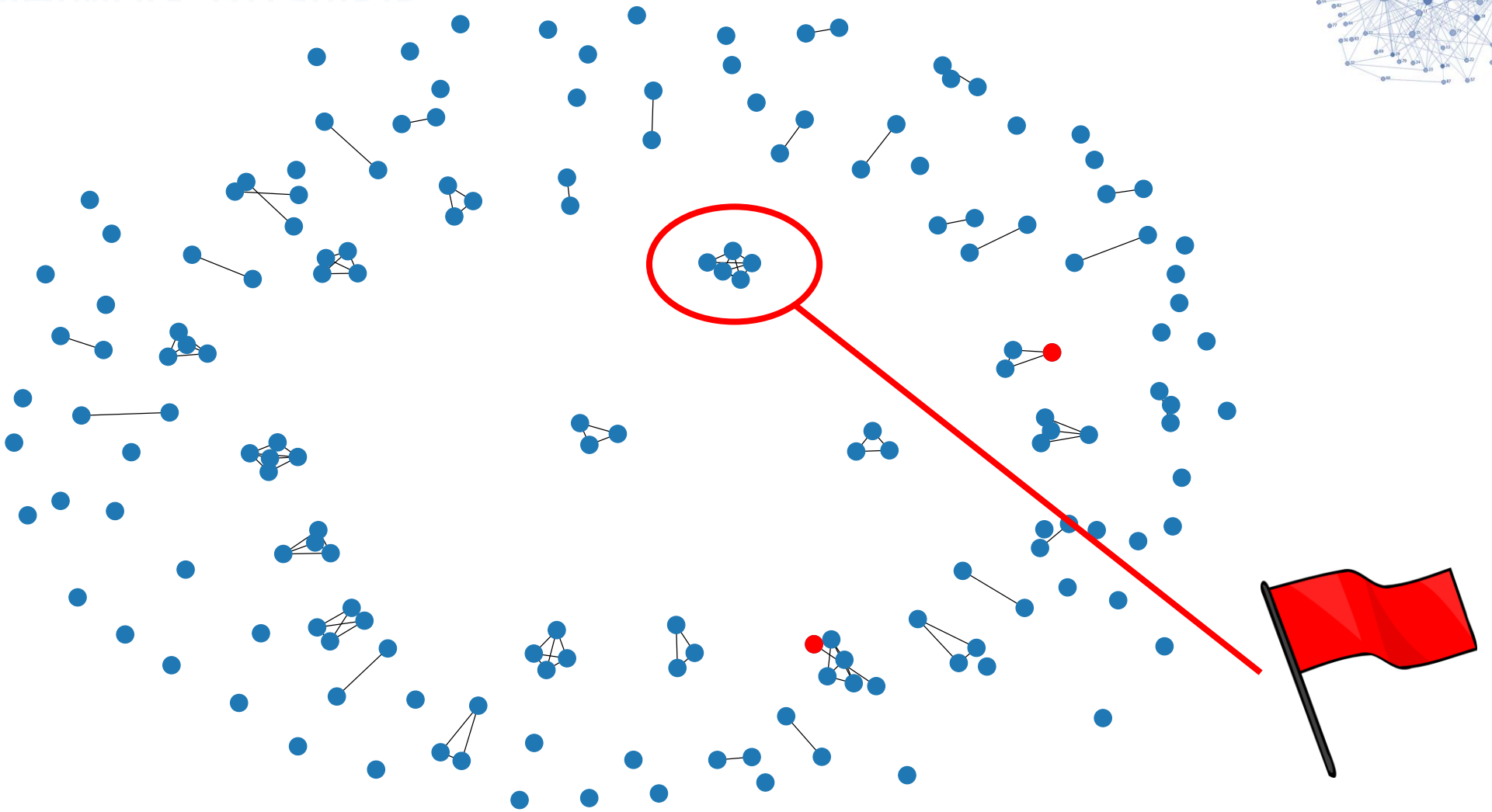
Connected by sharing other types of parties involved



Network Representation

→ Personal Network Analysis

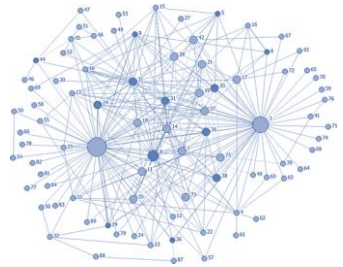


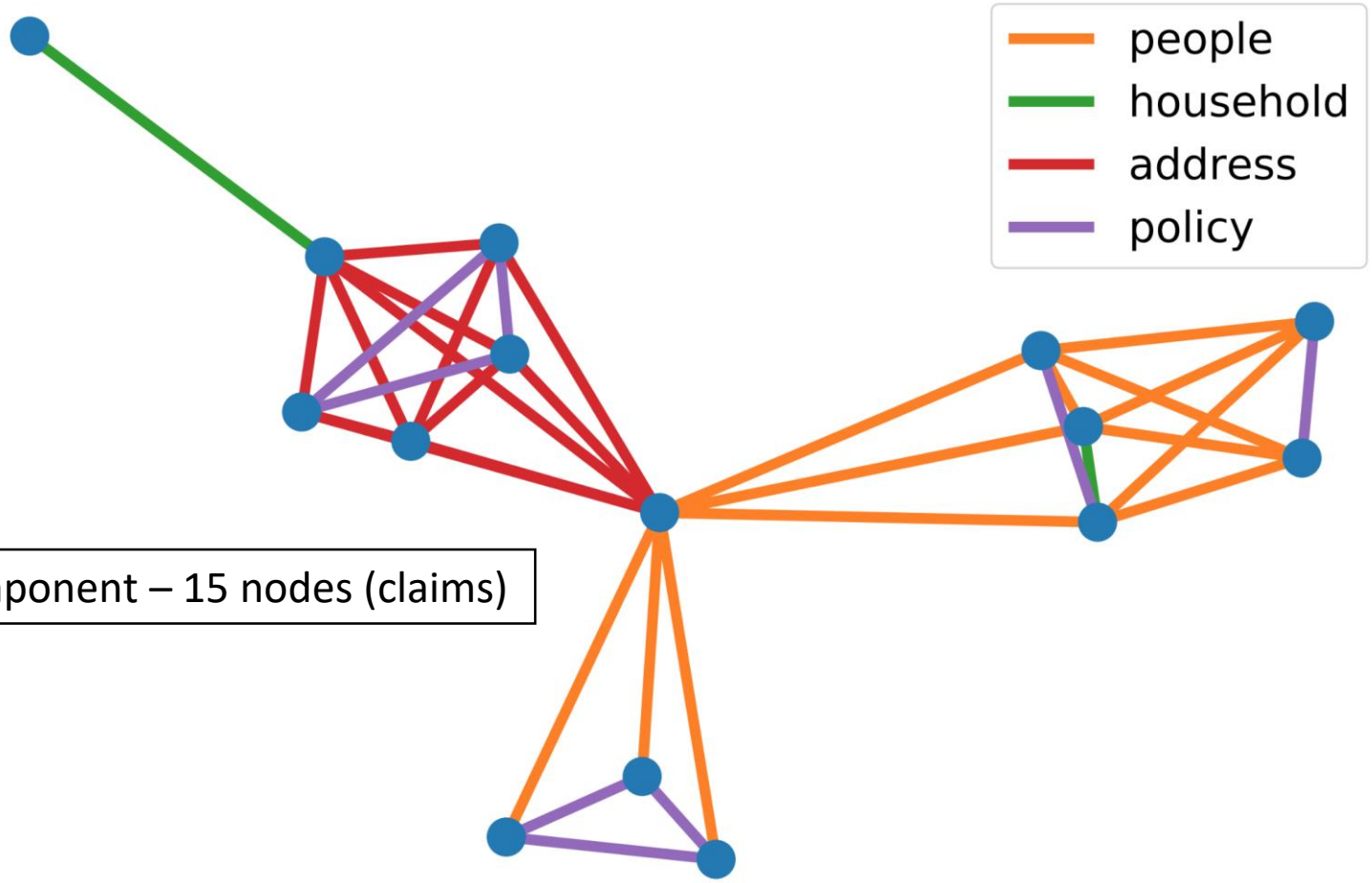


⚠ Component size provides good metric of cases to flag ⚠

Network Representation

→ Personal Network Analysis



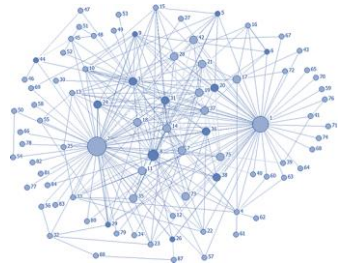


Largest Connected Component – 15 nodes (claims)

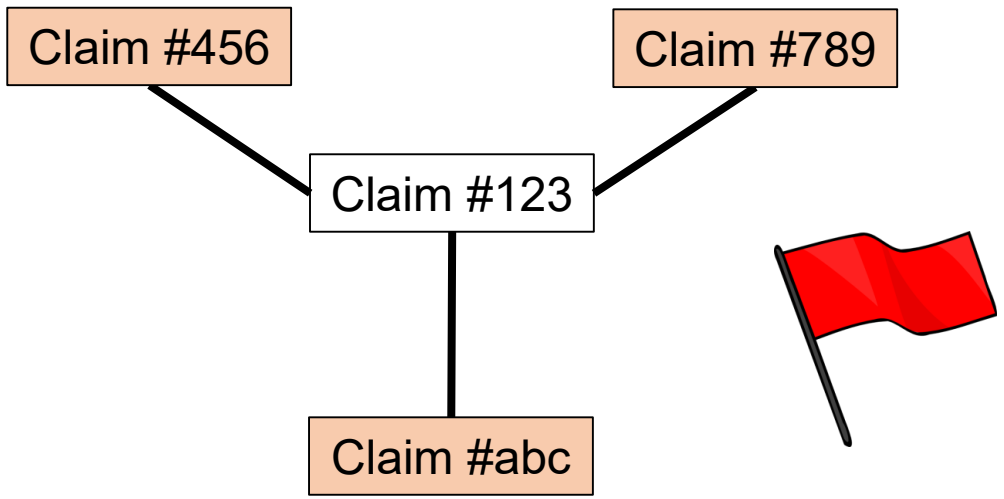
⚠ Typical cases worthy of escalation ⚠

Network Representation

→ Vendor Network Analysis

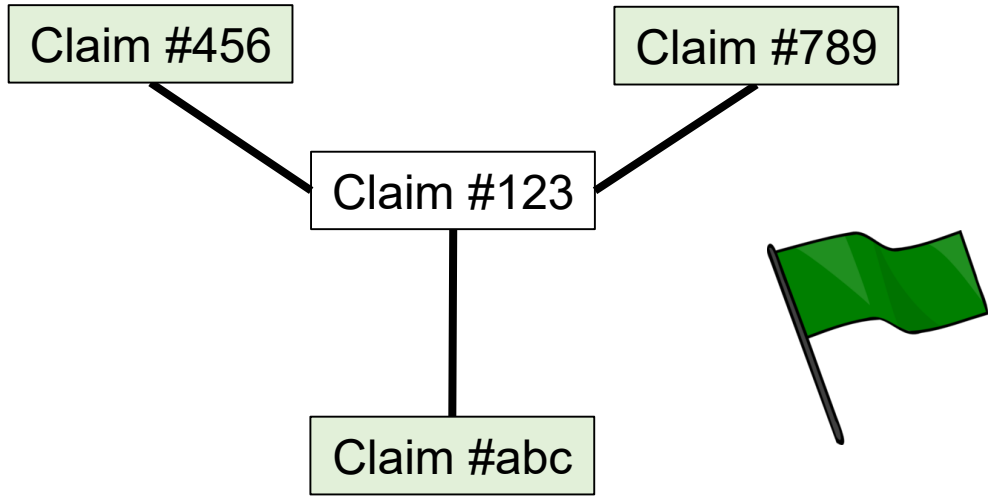


Claim surrounded by claims with 'bad' features



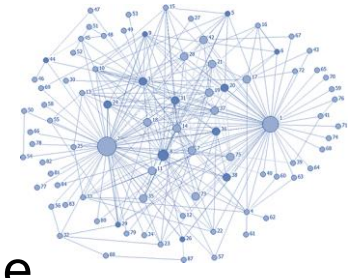
vs.

Claim surrounded by claims with**out** 'bad' features



⚠ Distance to labelled fraudulent cases provides good metric of cases to flag ⚠

Network Representation



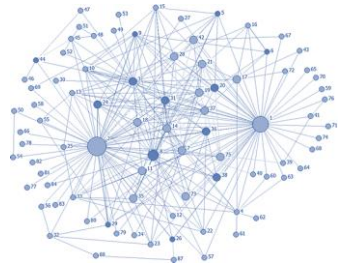
We combine information from all sources to provide a score to help SIU analysts decide which claims to investigate:

⚠ Average risk score of connected claims ⚠

Risk of a claim is then defined by

- A claim's well connectedness in the personal network
- More links to fraudulent cases in the vendor network
- Any other score of risk, e.g. results of regression model and auto-encoder.

Taskforce creation



Data Exploration

Network Representation
Fabian, Matthew, Lindon

Anomaly Detection
Bowei, Anthony
Francis, Caio, Helen

Advanced Analytics
Bowei, Anthony
Francis, Caio, Helen

LASSO

Scoring Methodology

Une place de choix
the co-operators
A Better Place For You

UQÀM

McMaster University

Western

UNIVERSITY OF BATH

CENTRE DE RECHERCHES MATHÉMATIQUES

IVADO

UNIVERSITÉ Concordia UNIVERSITY

MEMORIAL UNIVERSITY

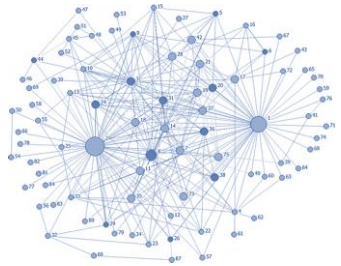
UNIVERSITY OF OXFORD

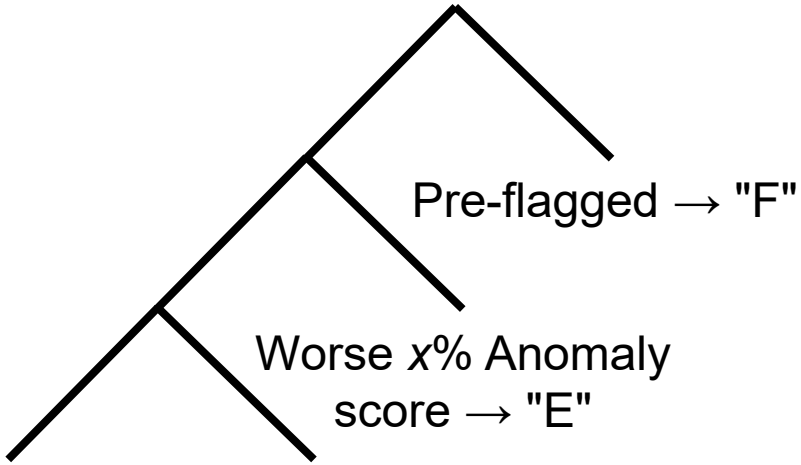
Prioritization (score)

Only reproduces what human flagged already: potentially missing new trends

While it is difficult how much fraud it will detect, we are willing to accept false positives

Prioritization (score)





$$\hat{p} = p_n \cdot \hat{n}_n + p_a \cdot \hat{a}_n + p_s \cdot \hat{s}_n$$

where $p_\pi + p_c + p_s = 1$ and subject to business decision, and

- $\blacksquare_n$ denotes normalized variables, that is, percentiles