Social Networks in Credit Risk Assessment

Credit Scoring & Credit Control Conference XVI
Edinburgh 28 – 30 Aug. 2019

Abd Rabuh, A. [1], Xu, M. [1] & Kizys, R. [2]

[1] Portsmouth Business School

[2] University of Southampton

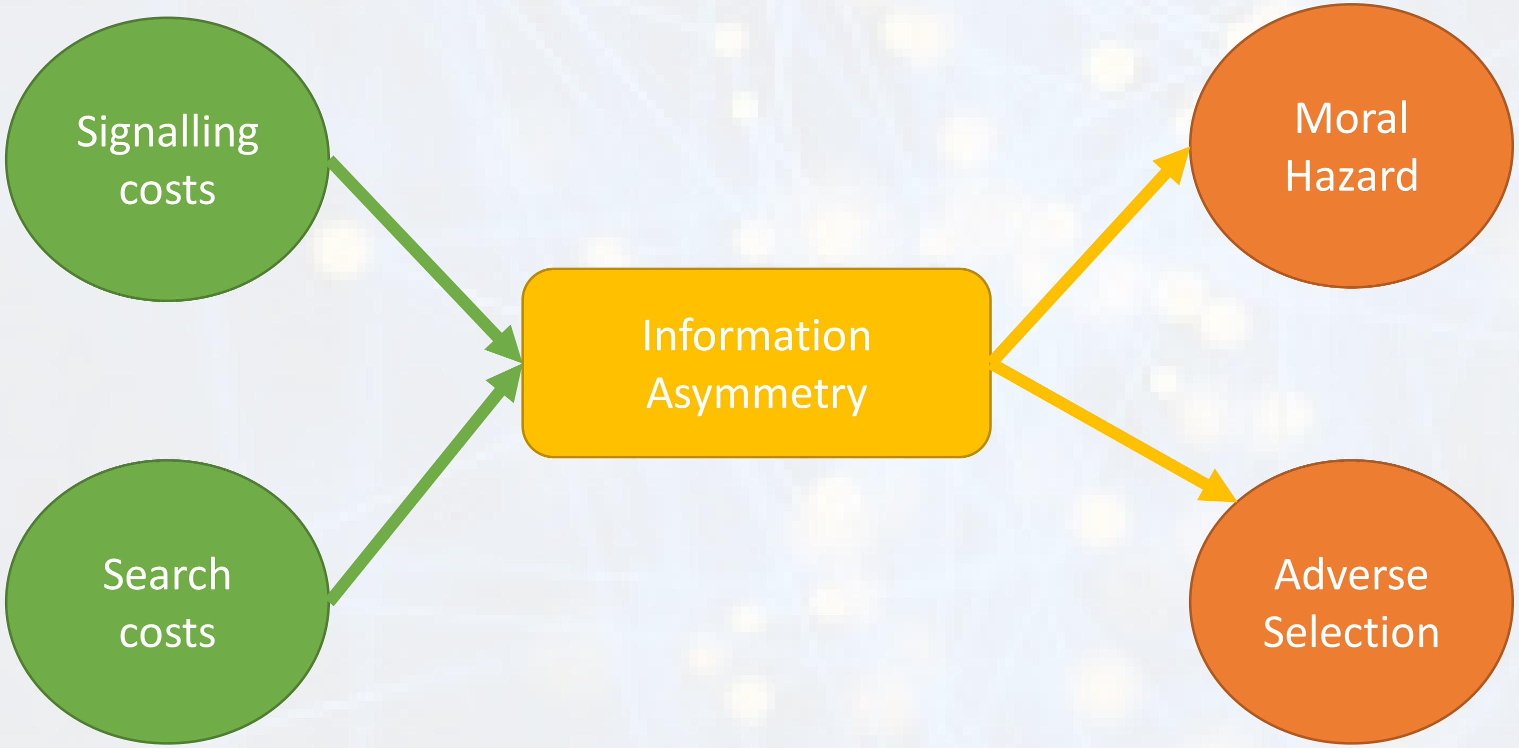
UNIVERSITY OF PORTSMOUTH

Outline

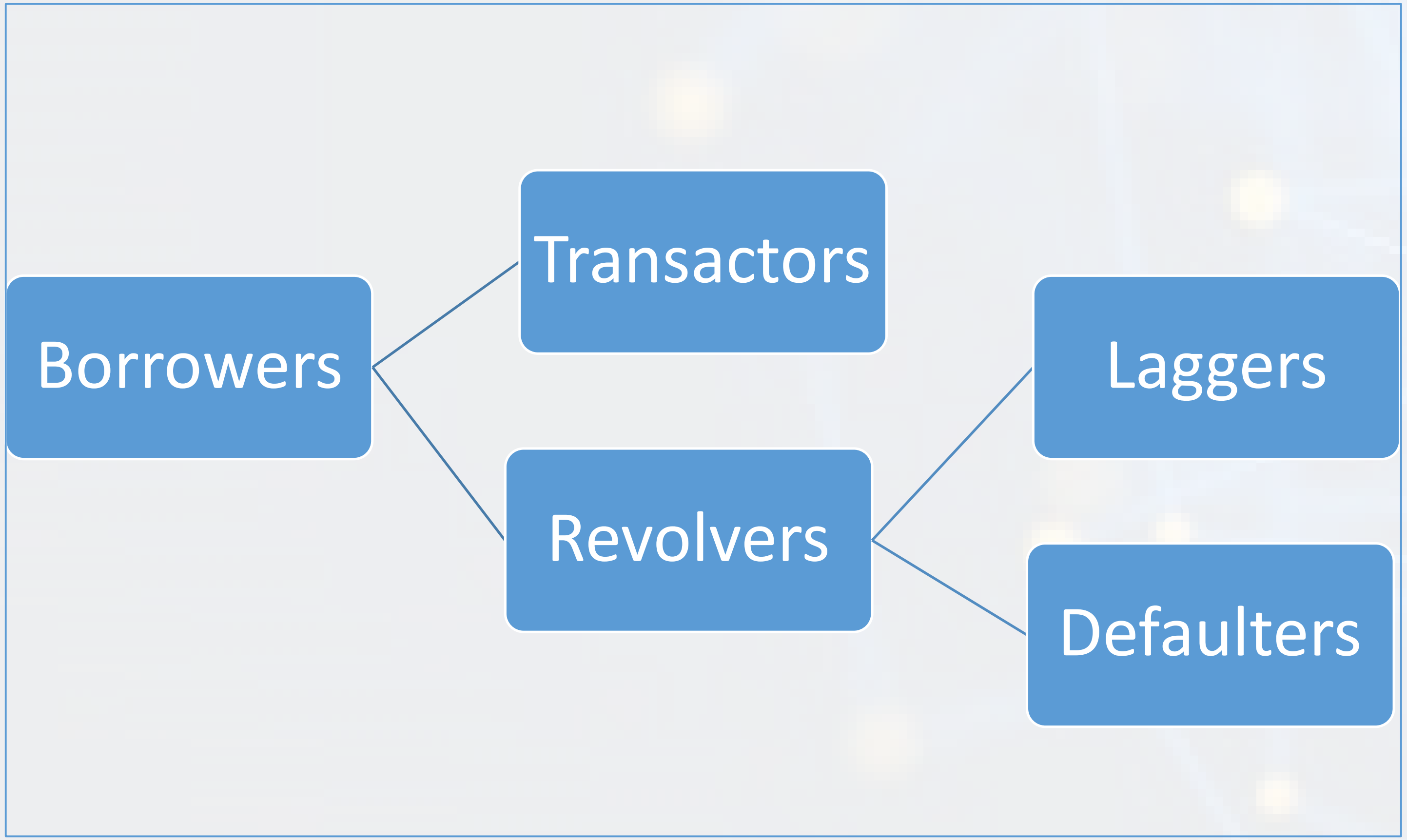
- Financial Inclusion
  - Thin profiles & unbanked
  - Existing groups
- Social Networks
  - Features
  - Types
- Related Work
- Model
- Results & Conclusion
- Future Agenda & Challenges

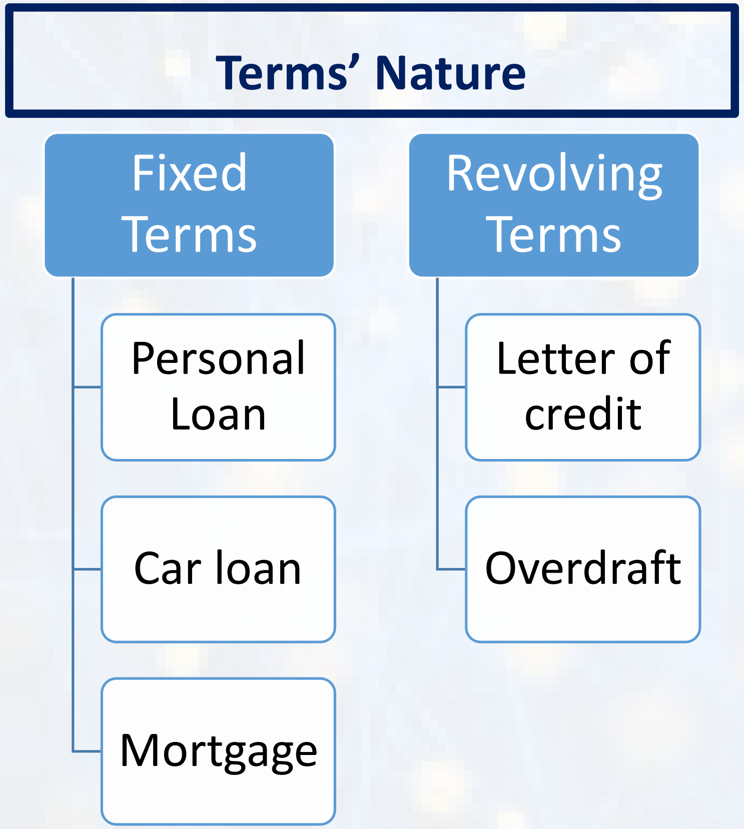
Financial Inclusion

- Economic Driver
- Unemployment
- Consumers and Suppliers
- Economic Growth



Classifications associated with credit offerings

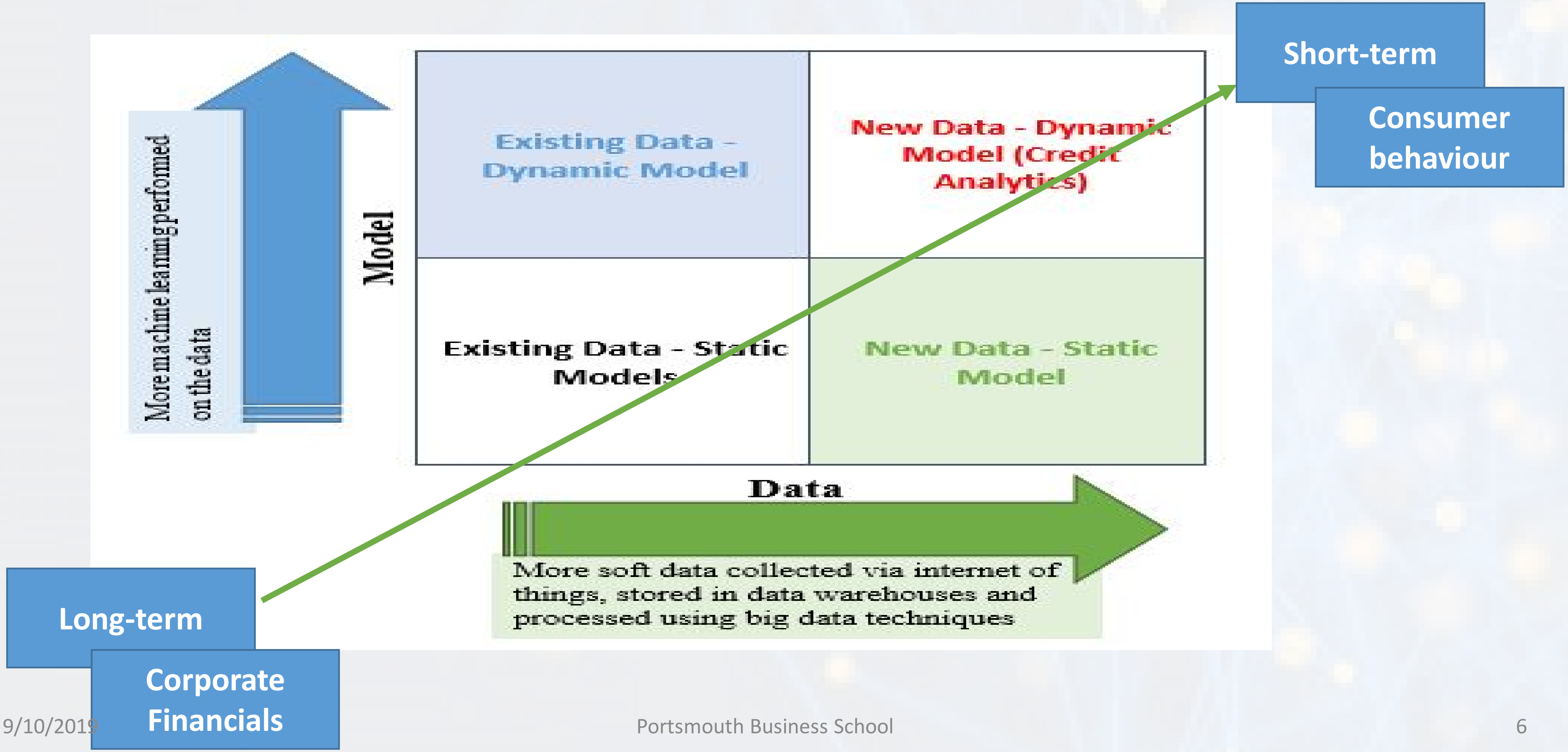




Example Criteria for Credit Scoring

- *Average transaction value*
- *Number of cash withdrawals*
- *Credit limit changes*
- *Rate of total jumps*
- *Proportion of months in arrears*
- *Repayment amount and outstanding balance, etc.*

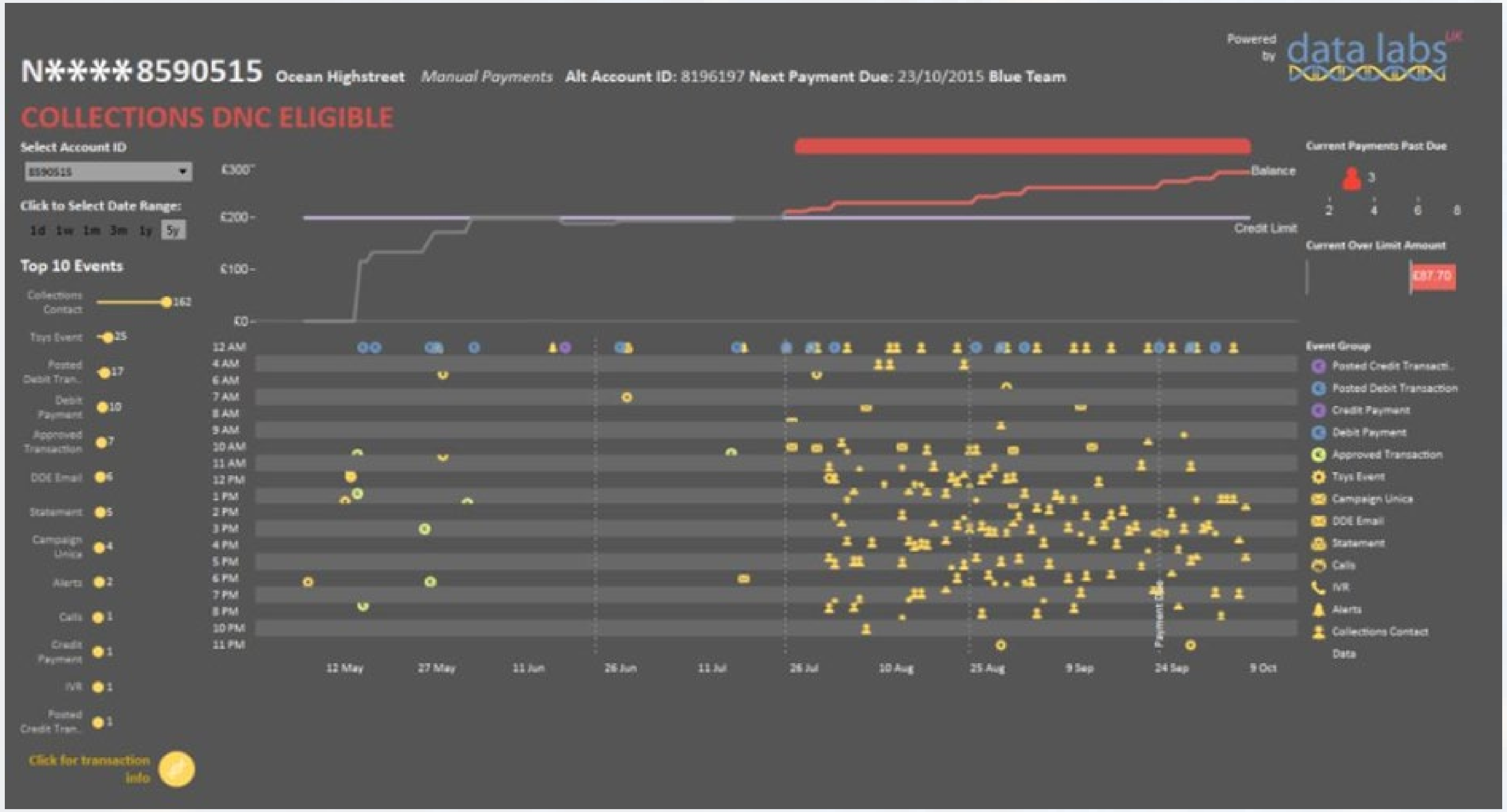
Credit Risk Assessment Models



Transactional and Behavioural

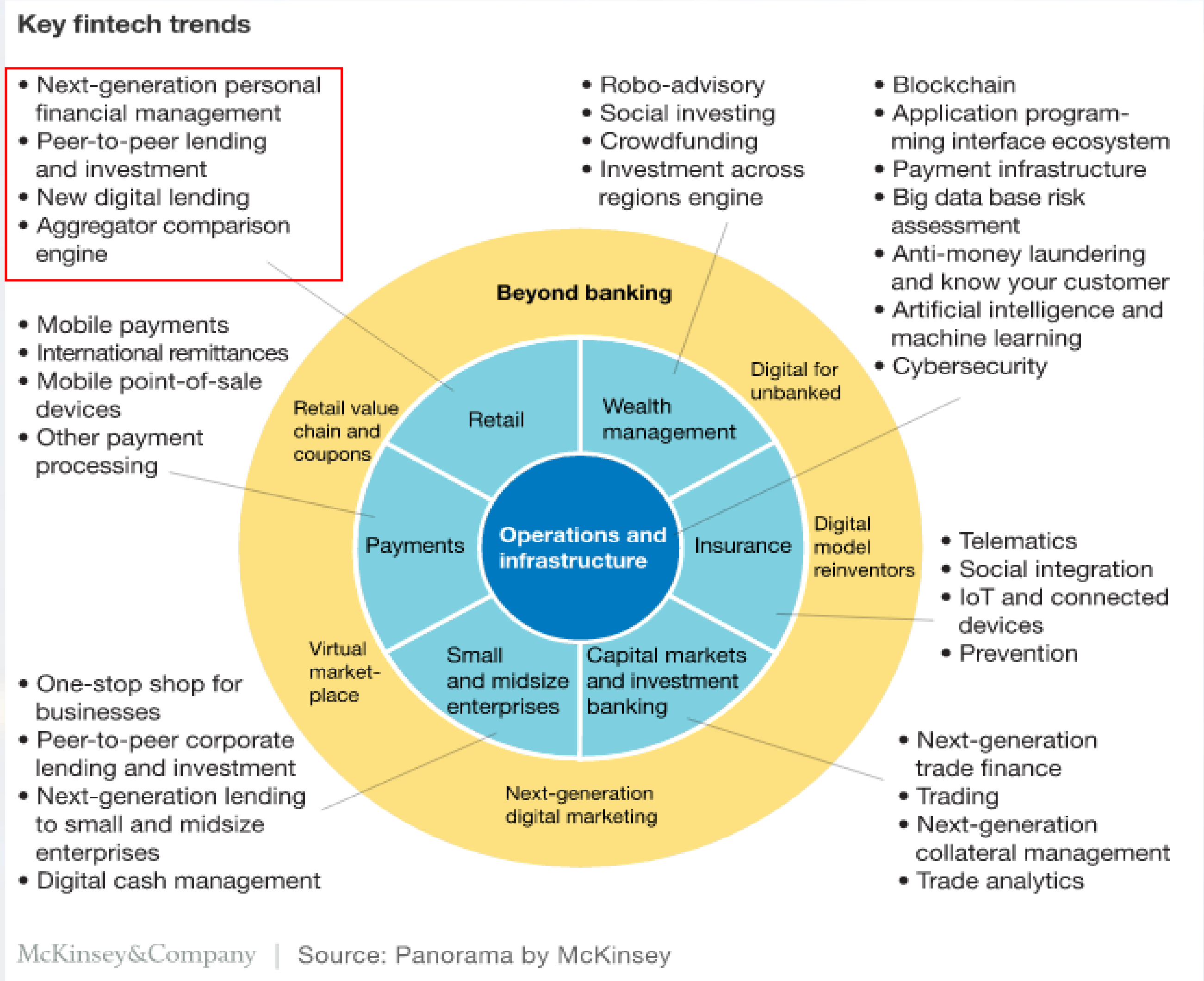
- Techniques Used:

1. Neural Network
2. Classification Trees
3. XGBoost

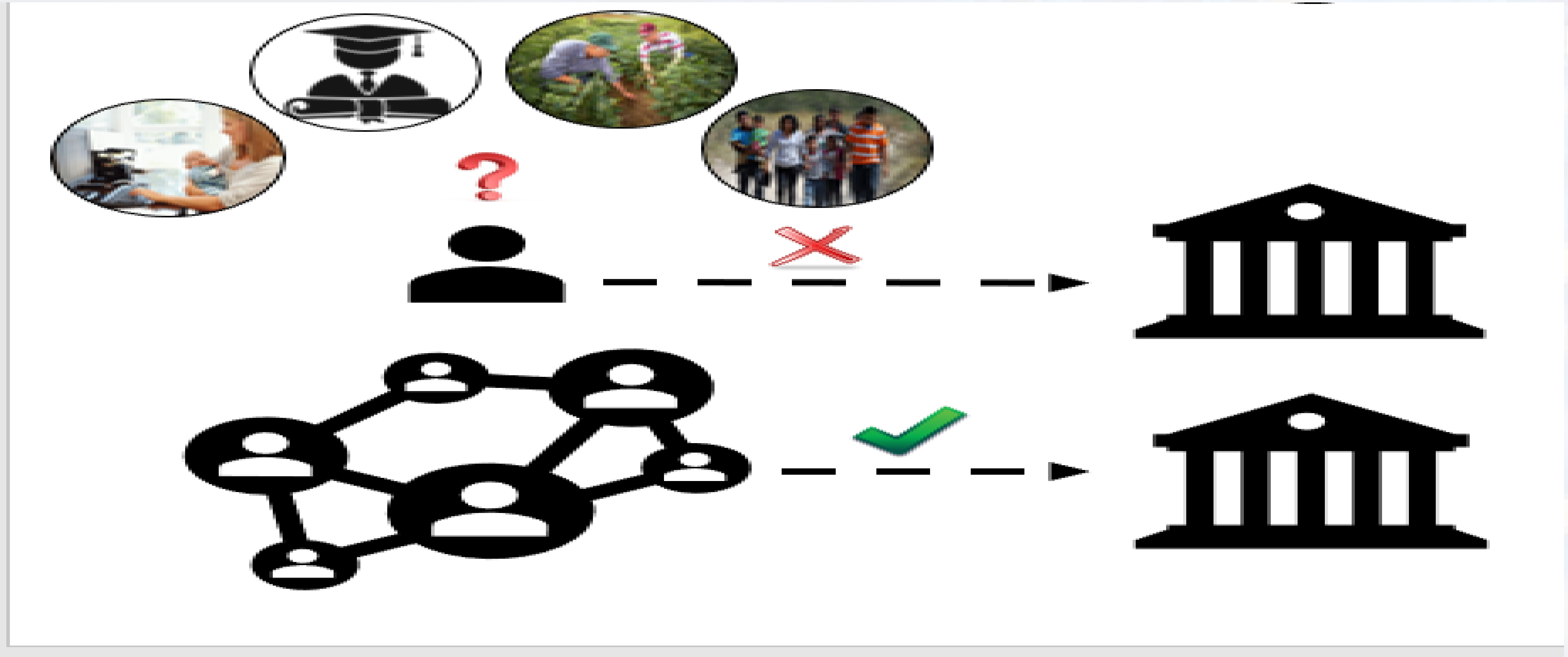


Beyond Banking (Fintech)

Digitisation of wealth and asset management…



Limited Profile?



What are Social Networks Statistics?

- Centrality
- Successors
- In/out degree (directed graph)
- Size
- Strength
- Density

Why Social Networks?

- No prior knowledge is required
  - Clustering – node similarity
  - Communities detection algorithms – *finding cliques, maximizing modularity, random walk and stochastic block models*

- Sources:
  - Communication
  - Citations
  - Societal memberships and social relations
  - Web analytics
  - Social media
  - Remittances / transfers

Related Work

- Manski 1993 – Individual factors vs. socio-economic factors

$$Y_i = \beta\, x_i + \lambda\, w_i + \varepsilon$$

- Freedman and Jin 2014 – Endorsing a friend (with/w'out a bid) vs. no endorsement.
- Wei, Yildirim et al. 2015 – Homophily & social discrimination

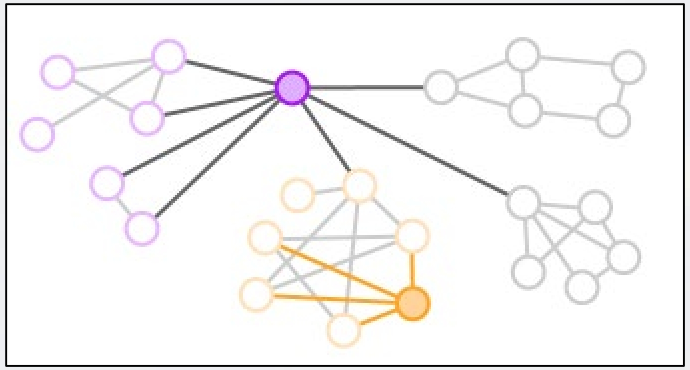
$$U_i = \sum (m_{ij} - |x_i - x_j|) + \alpha\mu_i$$

- Wan, Peng et al. 2016 – Stochastic Block Model

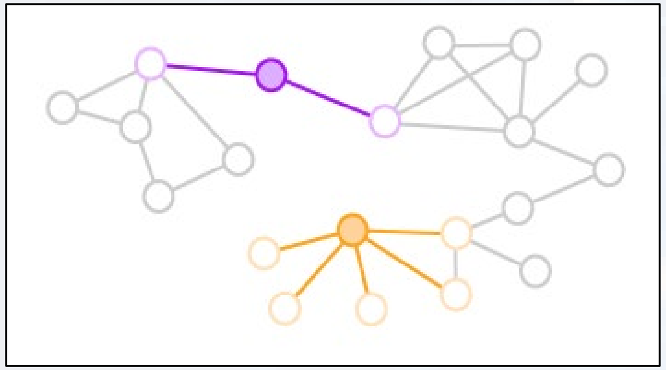
Probability of belonging to a community r: $$\alpha_{ir} = \frac{\sum_{s} \omega_{rs}\theta_{ri}}{\sum_{rs} \omega_{rs}\theta_{ri}}$$

What can we infer from Social Networks?

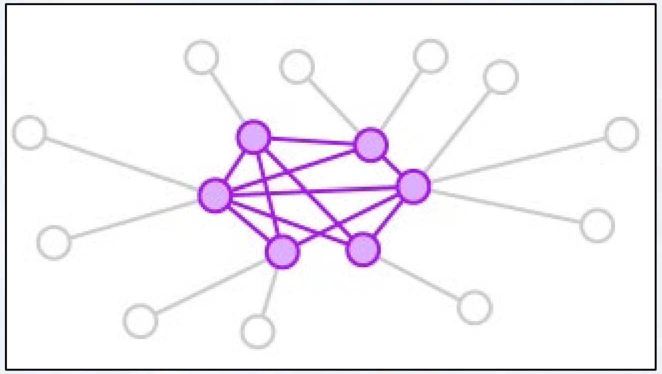
Ideation



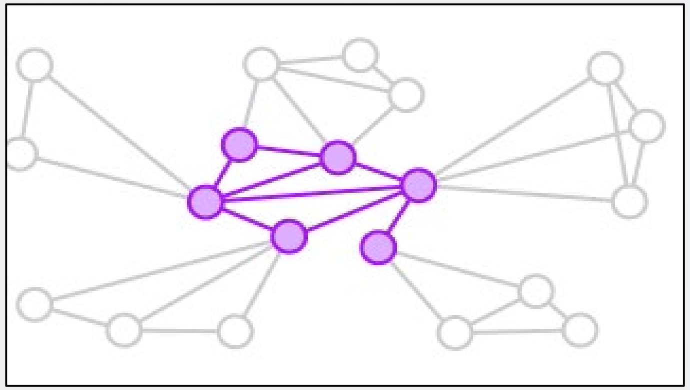
Influence



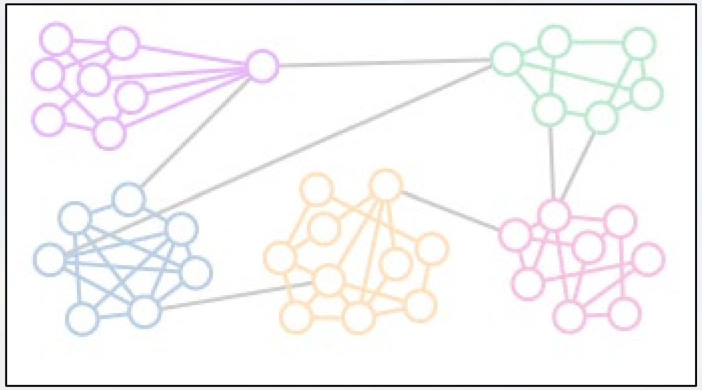
Efficiency



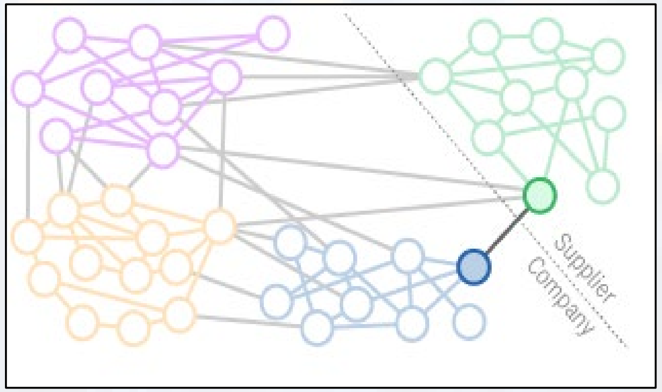
Innovation



Siloed



Vulnerability



Gap & Research Question

GAP:

- No empirical study on credit scoring using social network data.

R.Q:

- Can credit score be improved when using identifying borrowers’ network types?

Data

- 7 data sets from a lender.
- 356k loan applicants (120 attr.) in Southeast Asia
- 1.72 m. historical loans (credit bureau).
- Extraction SN data

ID	TARGET	OBS 30 CNT SOCIAL	DEF 30 CNT SOCIAL CIRCLE	OBS 60 CNT SOCIAL	DEF 60 CNT SOCIAL CIRCLE
453139	1	0	0	0	0
115000	0	0	0	0	0
125261	0	1	0	1	0
119758	0	2	0	2	0
107986	0	0	0	0	0
126821	0	1	0	1	0
388734	1	2	0	2	0
107473	0	3	0	3	0
122833	0	5	3	5	1
315722	1	0	0	0	0

	Cross-sectional	Time-series	Financial	Macro-economic	Behavioural	Static	Dynamic
Definition	Multi-dimensional at a single point of time	One attribute's trend over time	Can be sourced from financial records / banks	Systematic and tied with the economy as a whole	Varies from one person to another based on other (non-financial) interests	Do NOT change throughout the life of the loan	High variability
Support	95%	5%	53%	7%	40%	42%	58%
Example	Purpose of the loan	Instalment payments	Average account balances in the last 6 months	Interest rate at the time of application	Car's ownership	Amount applied for	Address of the borrower

Exploratory Data Analysis

<u>Pre-processing:</u>

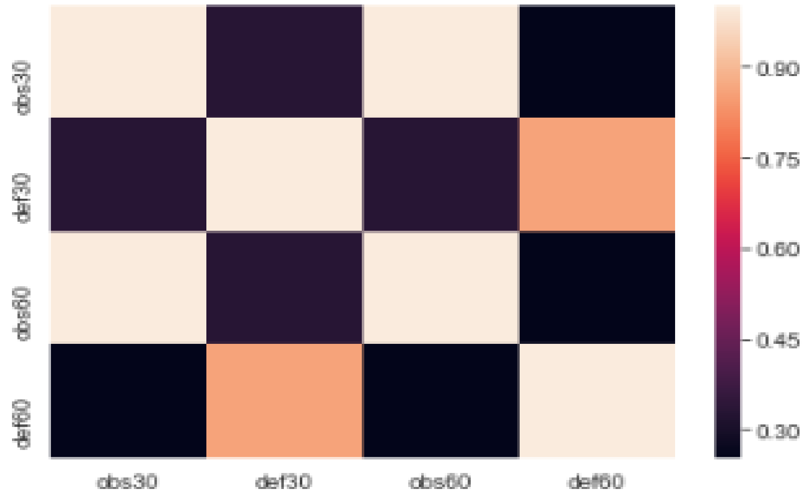
- Removing NA's
- Run a correlation test
- Reduce dimensions
- Balancing a subset to eliminate bias

```
cln_isol.iloc[:,1:5].corr()
```

	obs30	def30	obs60	def60
obs30	1.000000	0.329338	0.998490	0.253499
def30	0.329338	1.000000	0.331571	0.860517
obs60	0.998490	0.331571	1.000000	0.255570
def60	0.253499	0.860517	0.255570	1.000000

```
sb.heatmap(cln_isol.iloc[:,1:5].corr())
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x21117f54198>
```



Descriptive Analysis on Network Size

Group Statistics

	TARGET	N	Mean	Std. Deviation	Std. Error Mean
OBS 30 CNT SOCIAL CIRCLE	0	281701	1.42	2.403	.005
	1	24789	1.50	2.378	.015
DEF 30 CNT SOCIAL CIRCLE	0	281701	.14	.440	.001
	1	24789	.19	.517	.003

		Levene's Test for Equality of Variances			
		F	Sig.	t	df
OBS 30 CNT SOCIAL CIRCLE	Equal variances assumed	36.157	.000	-5.055	306488
	Equal variances not assumed			-5.099	29419.683
DEF 30 CNT SOCIAL CIRCLE	Equal variances assumed	1072.819	.000	-17.862	306488
	Equal variances not assumed			-15.614	28041.162

Log Odds and Information Gain

- $P(G) = \dfrac{281{,}701}{306{,}490} = 0.92$
- $P(B) = \dfrac{24{,}789}{306{,}490} = 0.08$
- $O_{pop} = \dfrac{P(G)}{P(B)} = \dfrac{281{,}701}{24{,}789} = 11.36$

obs30	G	B	Odds	Px	P(G/x)	P(B/x)	I(x)	WoE
0.0	150960	12950	11.657143	0.534797	0.920993	0.079007	1.026157	0.025821
1.0	44888	3895	11.524519	0.159167	0.920157	0.079843	1.014482	0.014378
2.0	27350	2458	11.126932	0.097256	0.917539	0.082461	0.979483	-0.020730
3.0	18647	1675	11.132537	0.066306	0.917577	0.082423	0.979977	-0.020226
4.0	12895	1248	10.332532	0.046145	0.911758	0.088242	0.909554	-0.094801
5.0	8756	797	10.986198	0.031169	0.916571	0.083429	0.967095	-0.033459
6.0	5888	565	10.421239	0.021055	0.912444	0.087556	0.917363	-0.086252
7.0	3979	411	9.681265	0.014323	0.906378	0.093622	0.852224	-0.159906

Logistic Regression

Precision	Recall	F1-score	Support
0.57	0.89	0.65	49,870

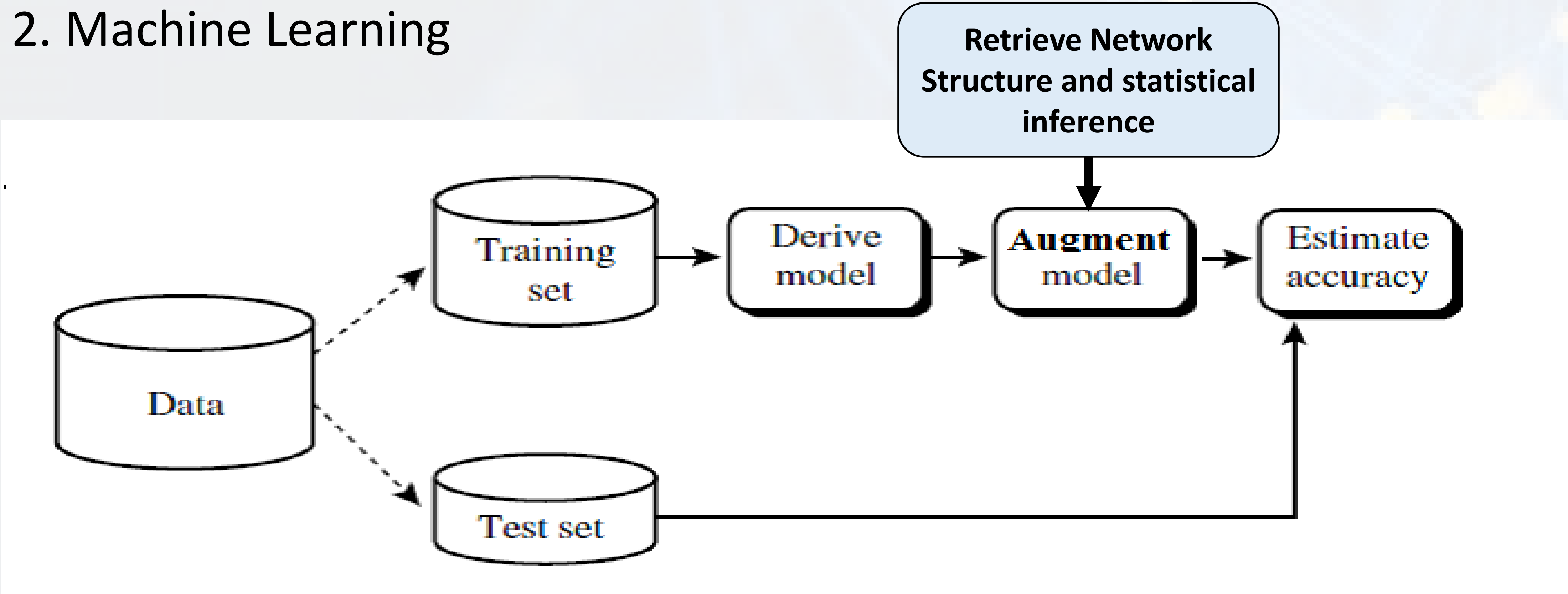
Conclusion

A. Providing a legal framework, networks carry valuable information when less traditional data is available.

B. The collective performance of a well-connected social group can be estimated, but there are individualised margins.

Future Work

1. Ensemble SN with behavioural or transactional criteria
2. Machine Learning



3. Time-series Analysis

Limitations

- How to link SN with credit default data in banking sector.
- Privacy concerns and ethical challenges.
- Social networking does not always convey the right information (*prosper.com*).

Thank You!

Ahmad Abd Rabuh
ahmad.abdrabuh@myport.ac.uk

References

- Coscia M., Giannotti F., & Pedreschi D. (2012). A classification for community discovery methods in complex networks.
- Fintechs can help incumbents, not just disrupt them
- Freedman, S., & Jin, G. Z. (2017). The information value of online social networks: lessons from peer-to-peer lending.
- Leow, M. & Crook J. (2016). A new Mixture model for the estimation of credit card Exposure at Default
- Leskovec, J. Lang, K. J. & Michael Mahoney (2010). Empirical comparison of algorithms for network community detection.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem.
- Newman, M. (2010). Networks: an introduction: Oxford University Press, Oxford.
- Thomas, Lyn C. (2009). Consumer Credit Models: Pricing, Profit and Portfolios: Pricing, Profit and Portfolios.
- Tong, E. N., Mues, C., Brown, I., & Thomas, L. C. (2016). Exposure at default models with and without the credit conversion factor.
- Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2015). Credit scoring with social network data.
- Yang, Y. (2007). Adaptive credit scoring with kernel learning methods