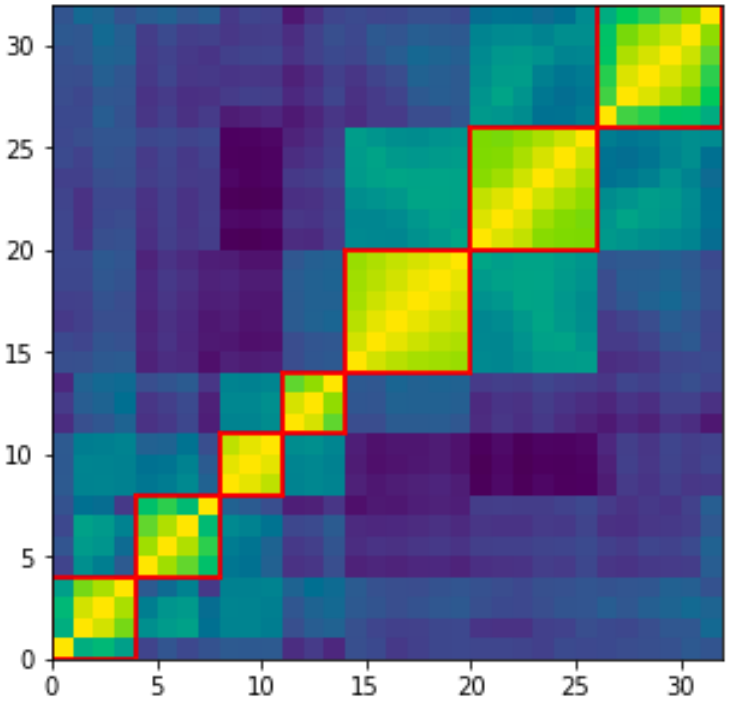
Creating Alpha with Machine Learning, some examples

Some background context and disclaimer:

- The following presentation is grounded on my personal experience…
- But, it will not discuss any specific alphas or strategies I have implemented for my current or past employers
- Illustrations for this talk were obtained using data sourced from the web, not belonging to current or past employers
- I will only present some of the ML tools I have been using in order to build new or improve existing alphas and strategies
- My experience relates to:
  - credit default swaps, HY bonds, distressed and special situations
  - systematic market neutral global equities

  => average holding period from a couple of days to several months;
  intraday or HFT may require a totally different set of (ML) tools

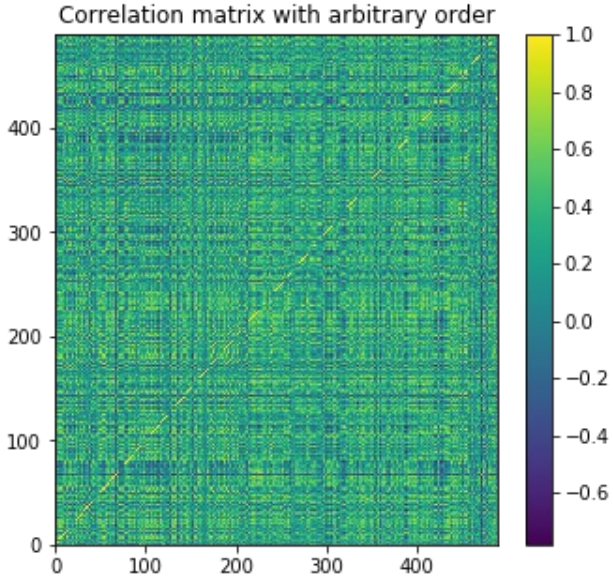
Combining alphas with **hierarchical clustering**

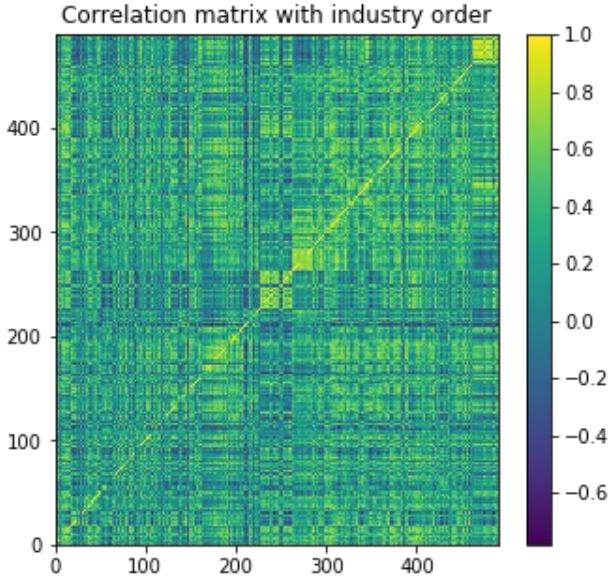


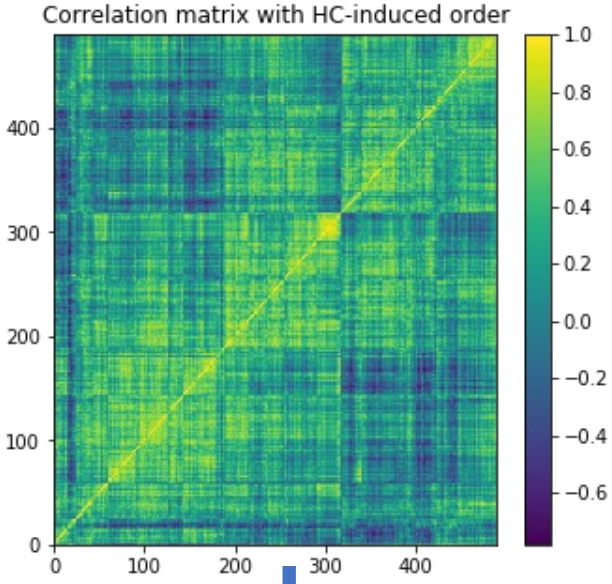
32 predictive signals -
but only 7 uncorrelated ideas in practice

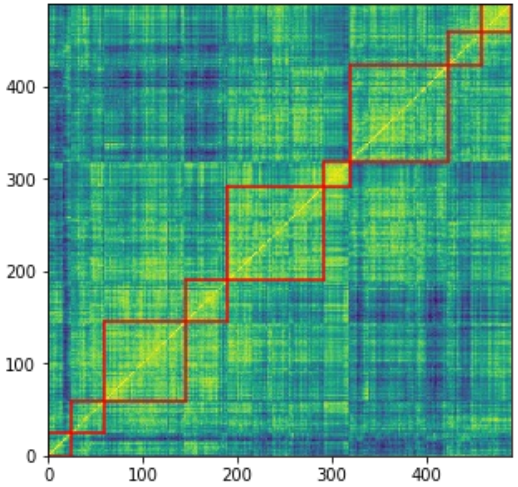
- Given a bunch of signals, can we ‘sort’ them?
- Using hierarchical clustering we can look at the correlation (or any kind of similarity) between the signals, or the ‘pnl’ of the signals
- Clusters can be used to ‘summarize’ many very similar signals into a single stronger one
- Unlike with PCA, clustering helps to keep a certain level of interpretability

Redefining sectors with **hierarchical clustering**









What are similar companies?
Sectors and industries are a good first approximation, but can we do “better”?

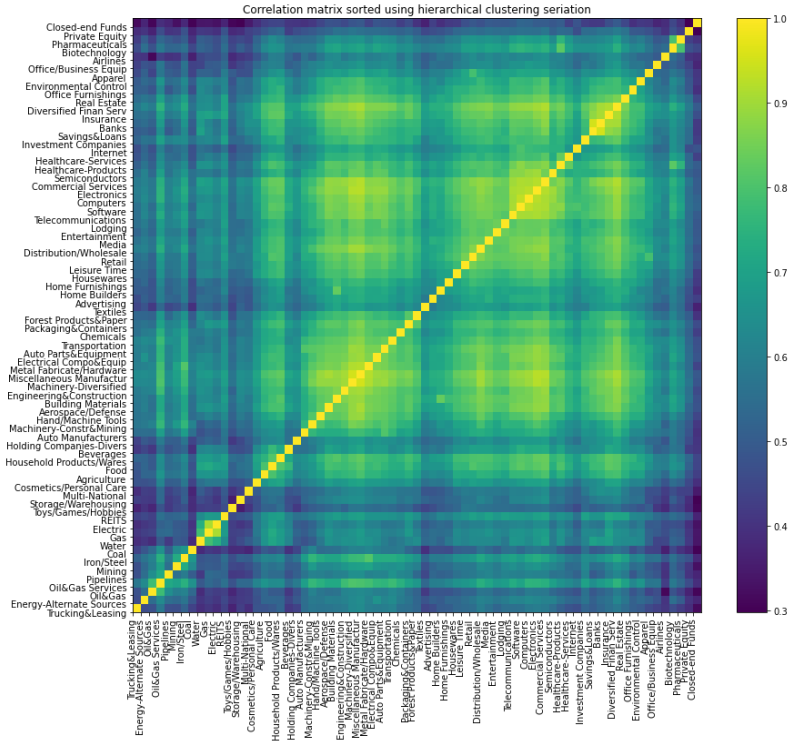
Hierarchical clustering is also useful to define custom proprietary classifications of stocks, bonds, strategies, portfolio managers, etc.

Some code available at:
https://marti.ai/qfin/2020/07/05/hierarchical-pca-avellaneda-paper.html

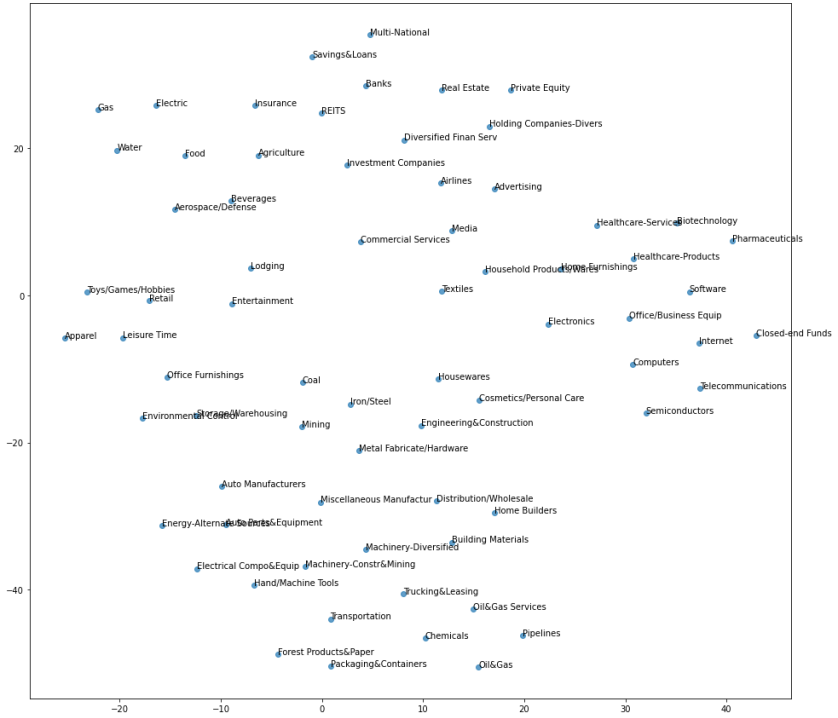
Embedding sectors (with **t-SNE** or **GNN**)

	Utilities	Technology	Transportation	Airlines	Internet	Electric	Gas
GOOG	0	1	0	0	1	0	0
DYN	1	0	0	0	0	1	0
AMZN	0	1	0	0	1	0	0
NJR	1	0	0	0	0	0	1
DAL	0	0	1	1	0	0	0

Dummy variables for (sector, industry)



Based on some proprietary features
build a similarity matrix or a network

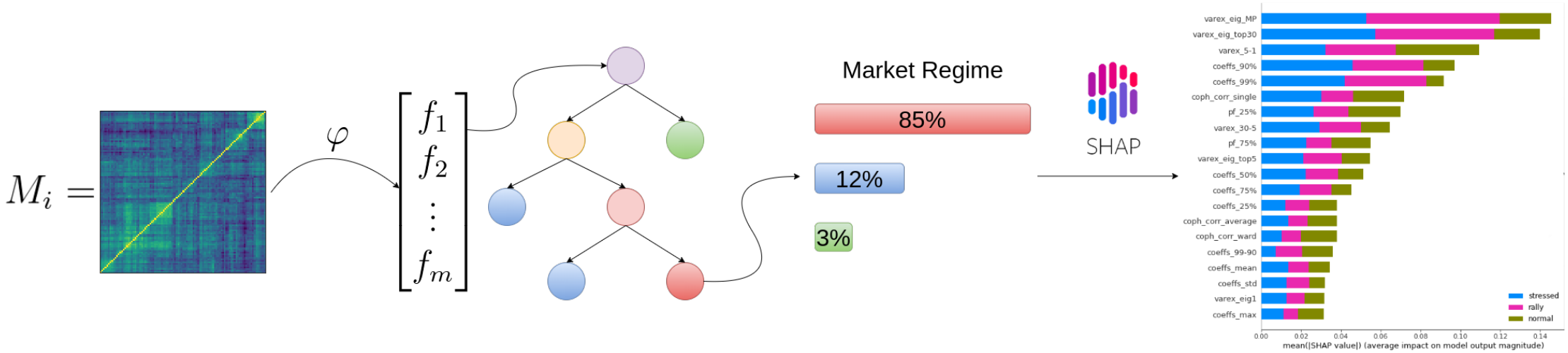


https://www.stellargraph.io
https://lvdmaaten.github.io/tsne/

More details available there: https://marti.ai/quant/2021/08/13/gnn-sector-embedding.html

Shapley values and features importance

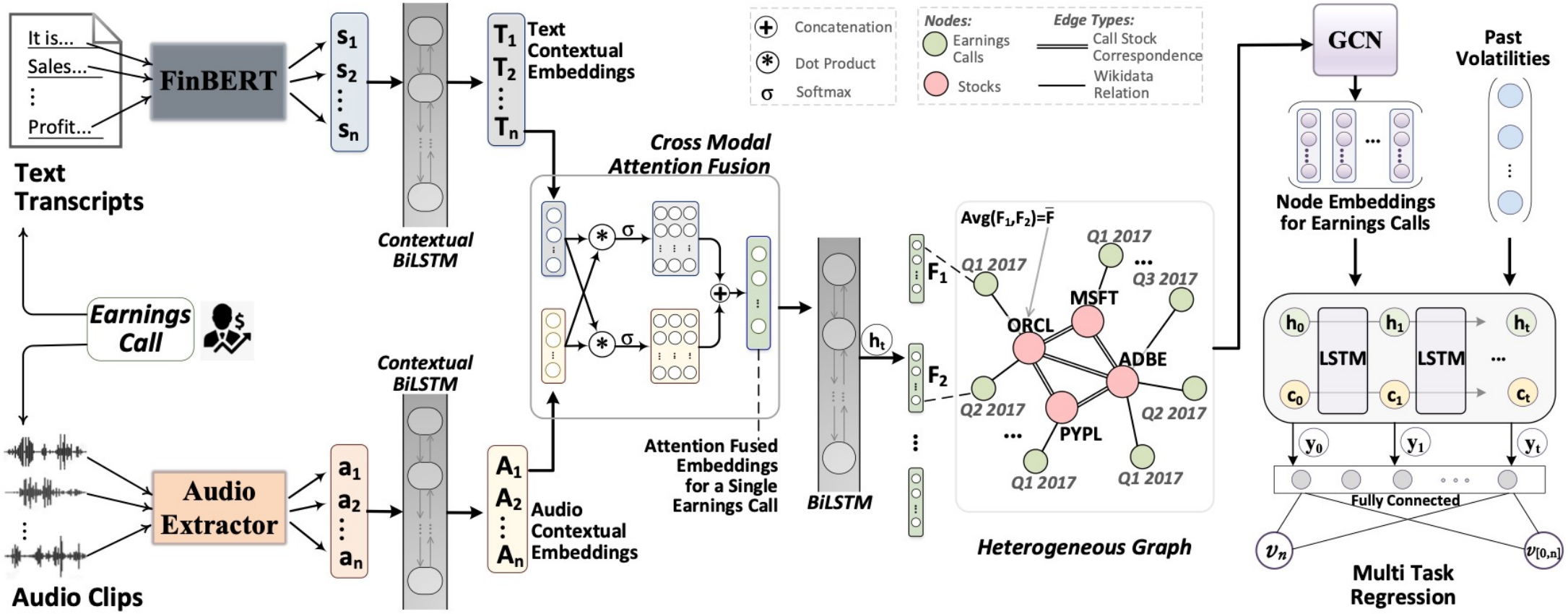
- For long horizon predictions, and relatively low Sharpe signals, black-box models may not be suitable.
- Features importance, and Shapley values in particular, can provide confidence in their predictions.
- With good understanding of the features and their interactions, one can even remove the ML model.



Can we predict a market regime from correlation matrix features?

https://marti.ai/qfin/2020/09/04/correlation-matrix-features-market-regimes.html

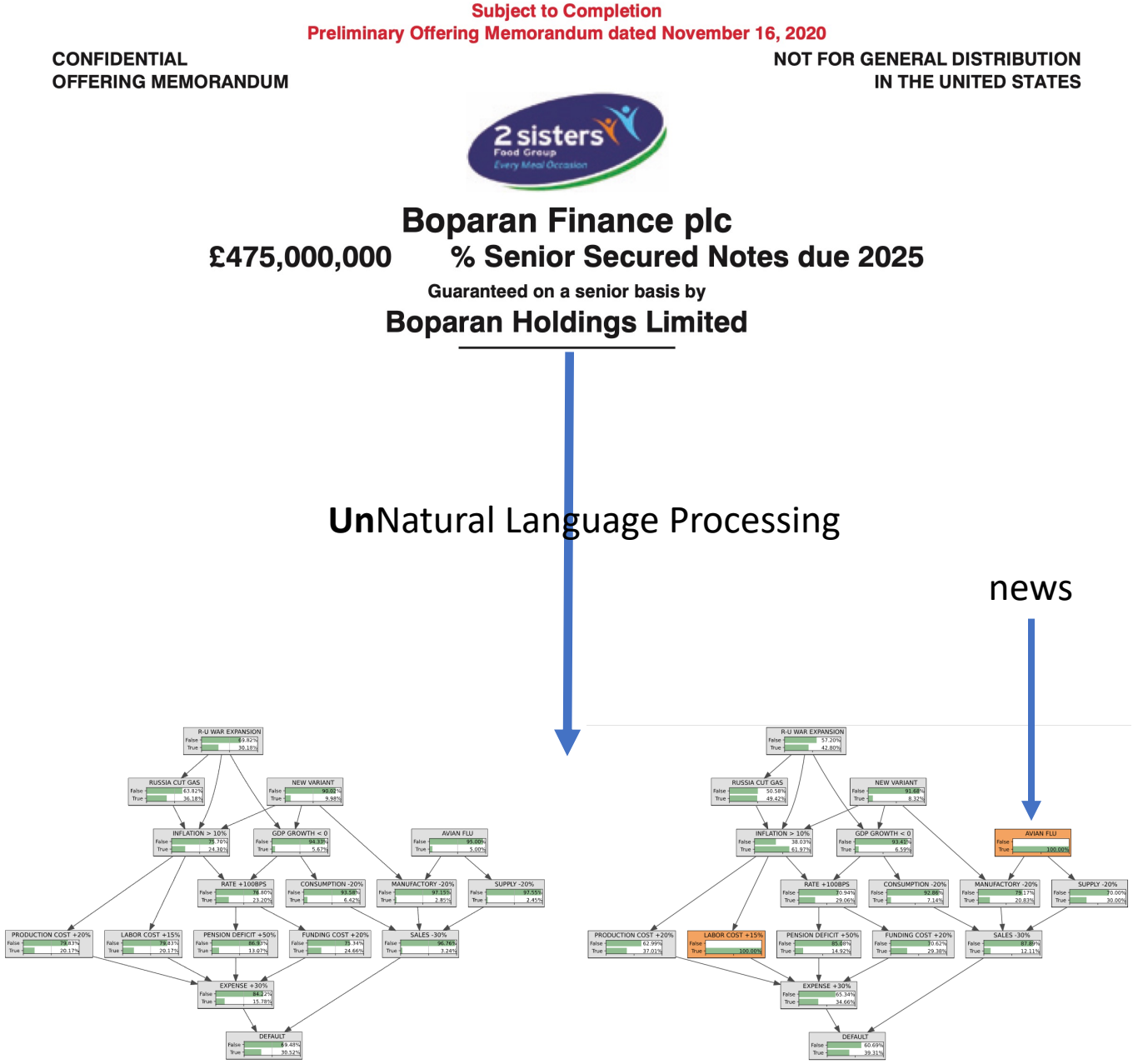
Propagating information with **GNNs**…



VolTAGE: Volatility Forecasting via Text Audio Fusion with Graph Convolution Networks for Earnings Calls (Sawhney, 2020)

… or with Bayesian networks and (Un)Natural Language Processing of news and docs

- A trove of information can be found in legal docs and various prospectuses
- NLP models advertised by HuggingFace, FAIR, OpenAI are of little use for these jargon heavy documents; => What does a ‘sentiment’ even mean?
- Fully leveraging this information and combining it with other data sources (e.g., news or fundamentals nowcasts) seem to be still out of reach for most firms in the industry



pyAgrum: https://agrum.gitlab.io

Conclusion

- ML tools are useful for improving and revisiting existing strategies
- ML tools can be used to automate strategies that were previously ran by discretionary managers but too complex for systematic quants
- ML is only a piece of the puzzle when designing strategies with a relatively low turnover; Understanding why the strategy makes sense, and whether the opportunity should continue to exist is important.