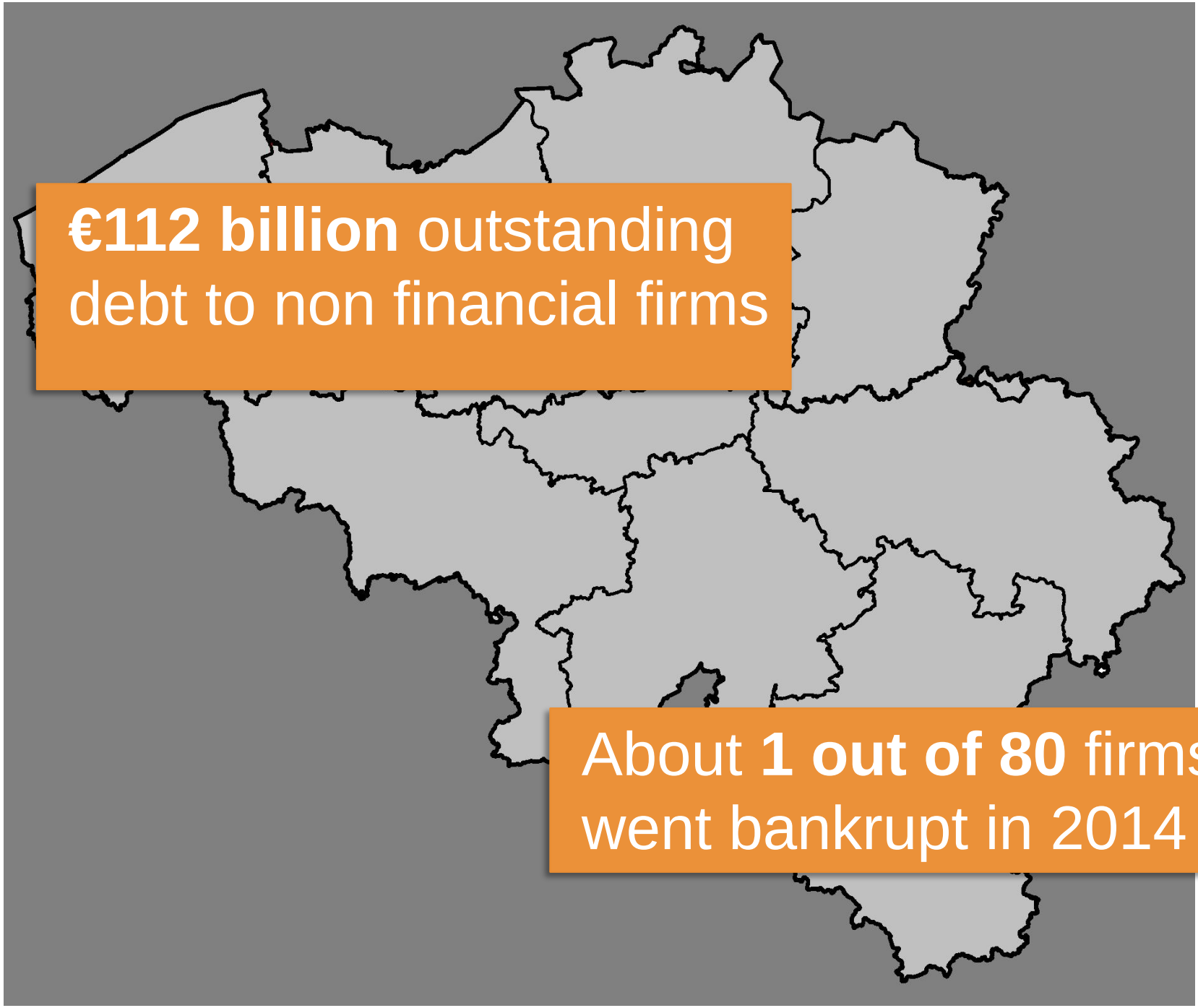
It's not all about the money

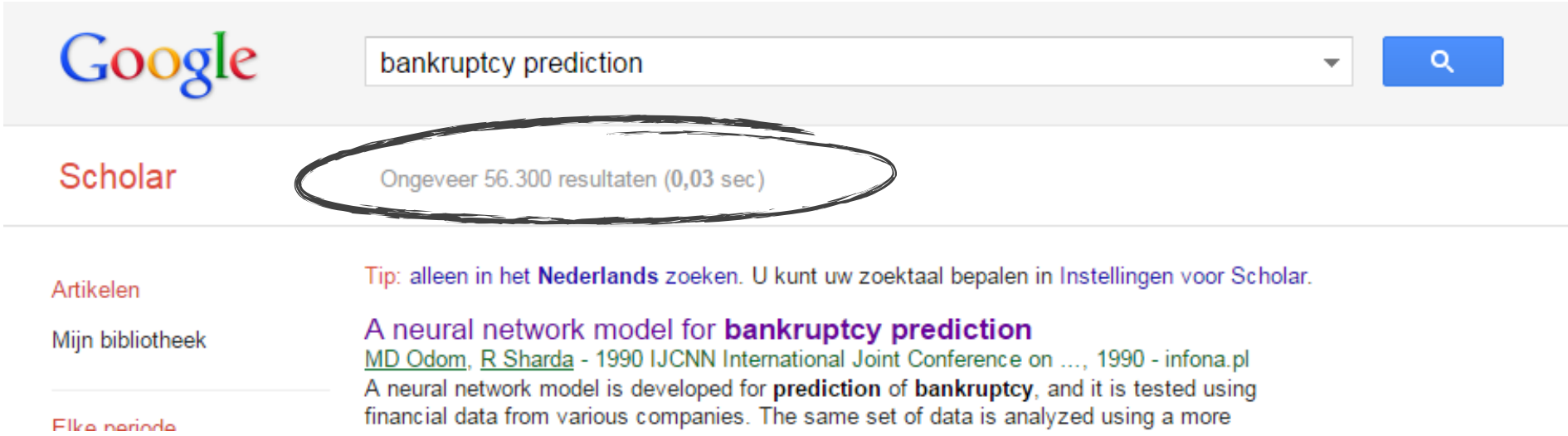
Bankruptcy prediction using relational data

Credit Scoring and Credit Control XIV conference 2015
University of Edinburgh
Ellen Tobback

APPLIED DATA MINING

Universiteit Antwerpen


€112 billion outstanding debt to non financial firms
About **1 out of 80** firms went bankrupt in 2014



A Google Scholar search on “bankruptcy prediction” lists **56,300** hits.

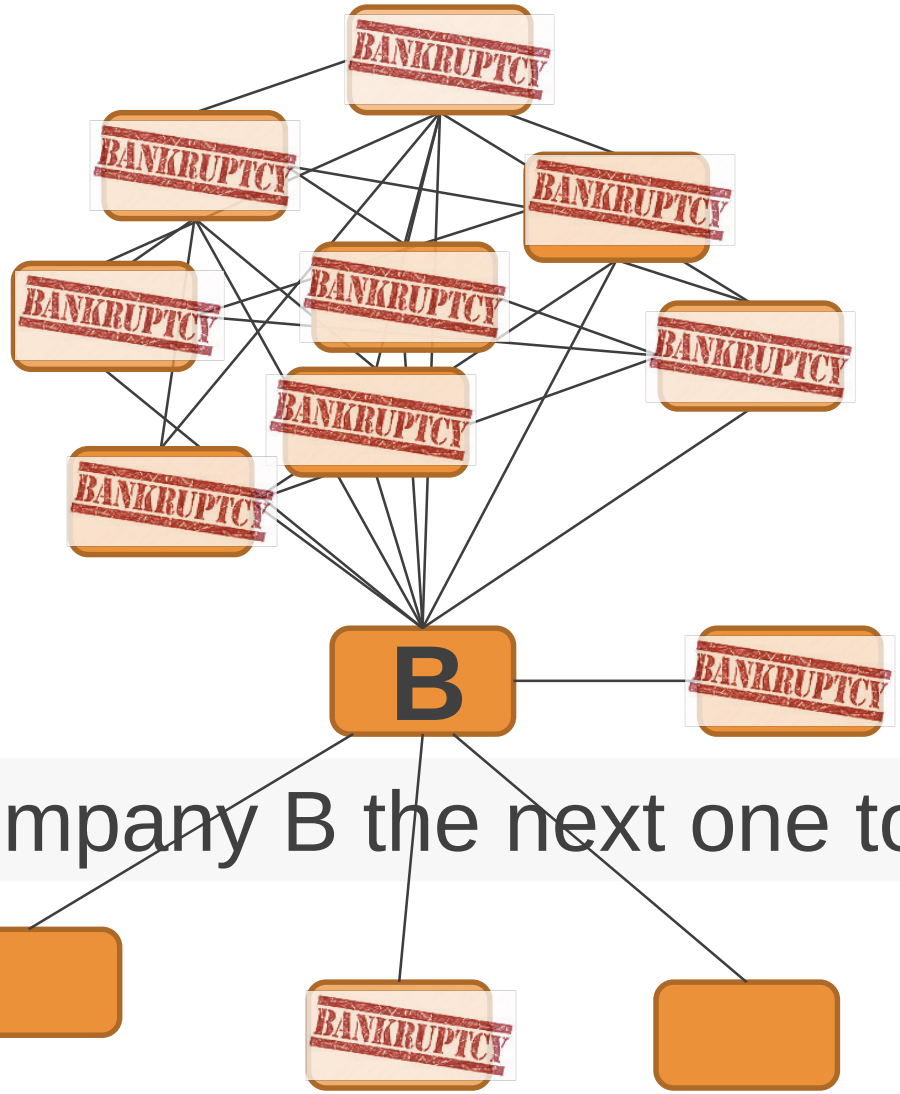
Most of them use financial and accounting variables.

80% of the bankruptcy cases are caused by **mismanagement** and **incompetence**.*

More than 5 % of the active Belgian SMEs have **at least one** manager or board member that was involved in a **previous bankruptcy**.

* According to a Belgian survey

If company A and B are linked through the same manager and company A fails, will company B have a **higher probability of failure**?



Is company B the next one to **fail**?

Data set in numbers

+ 400,000	Belgian SMEs
2012-2013	Training and validation period
2014	Test period
1%	Event rate in each set
+5%	“Bad link” rate

Prediction models

1. Base model
2. Network model
3. Ensemble model

Base model

Financial Variables

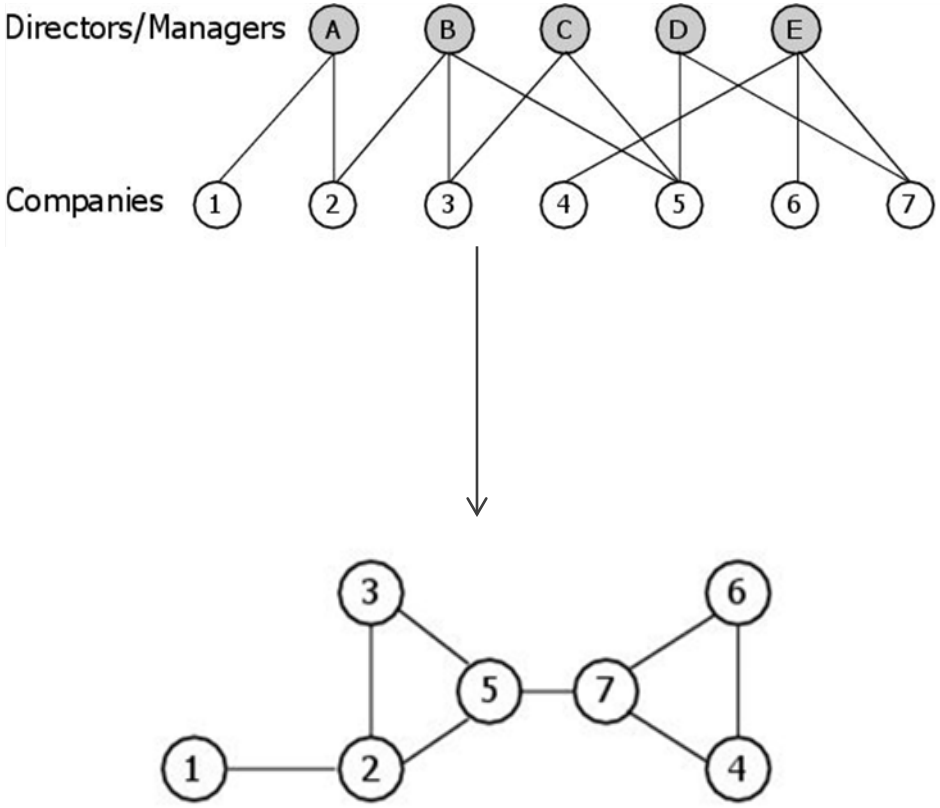
- Return on Assets
- Return on Equity
- Current Ratio
- Debt to Total Assets
- Cash Flow to Equity
- Profit/Loss

Dummy Variables

- Missing Values
- Small or Large?
- Newly Founded?

Linear SVM model

Network model



From a bigraph

To a unigraph

How do we **quantify** the links?

Bottom Node Projection

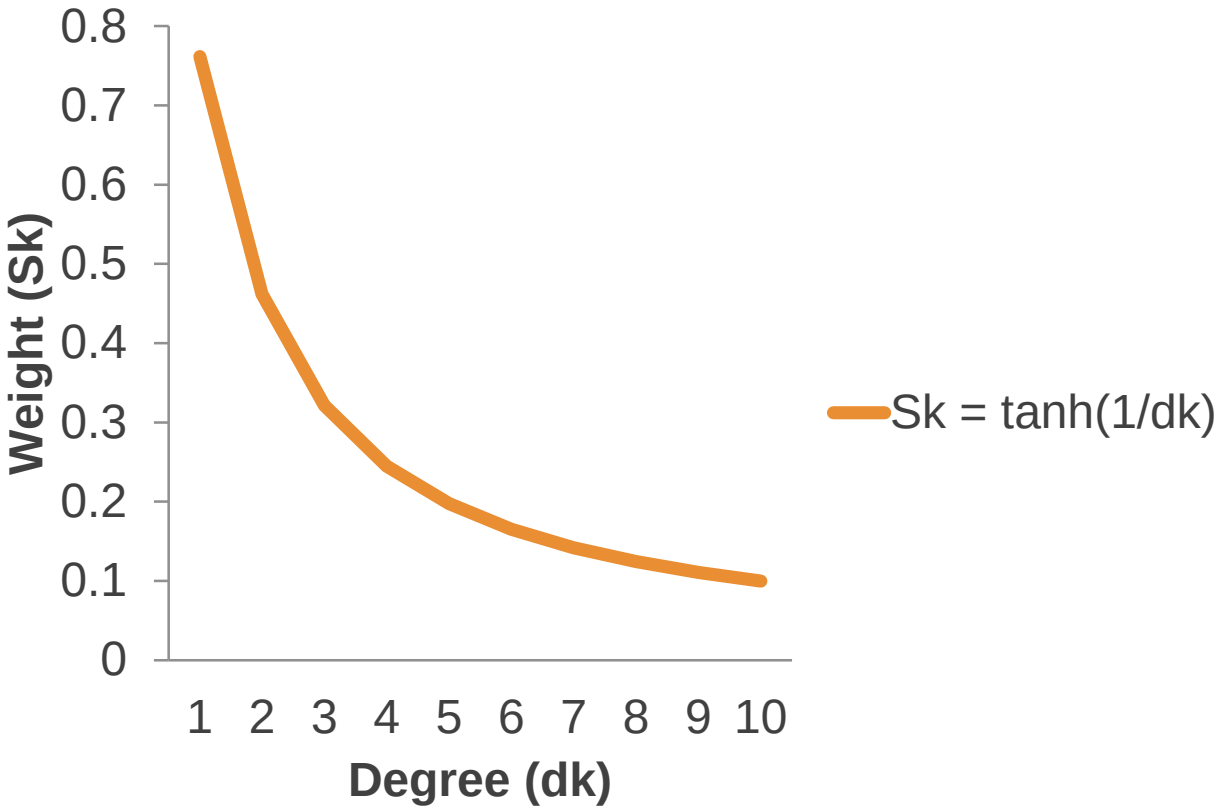
How do we get a **prediction**?

Weighted vote Relational Neighbour

Bottom node projection

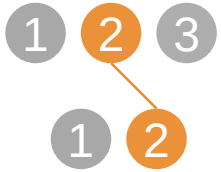
1. Hyperbolic Tangent Function
2. Sum of Shared Nodes

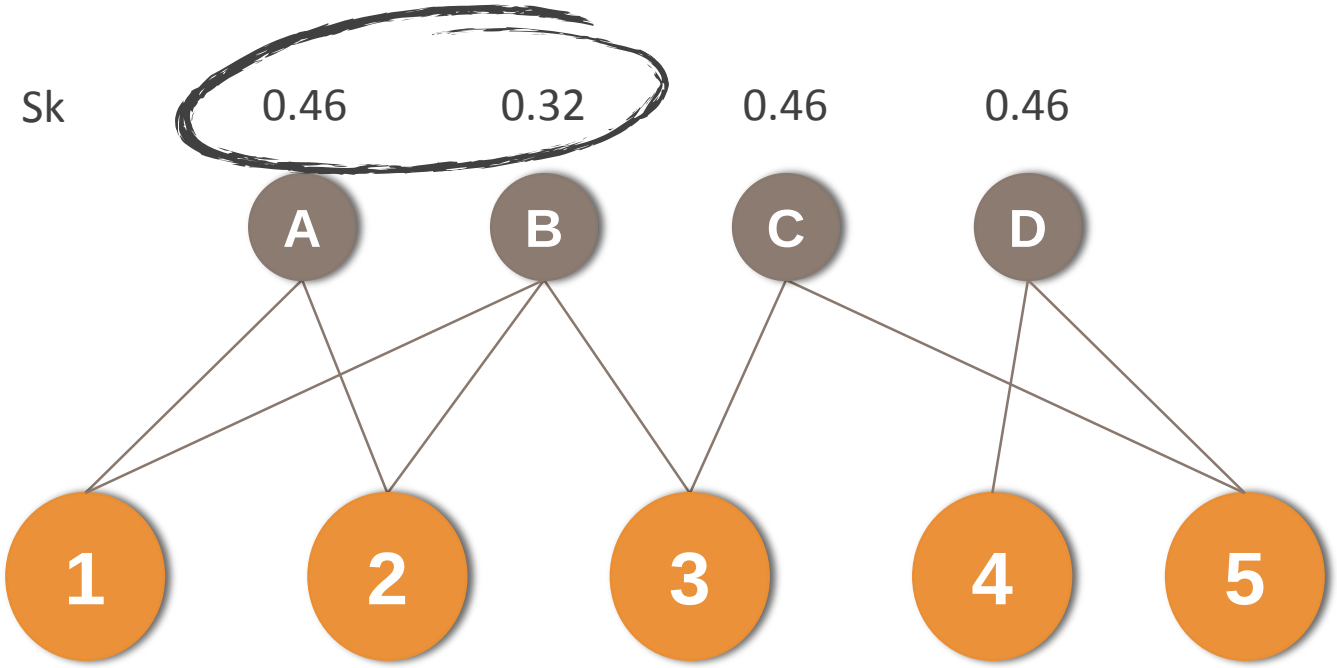
Hyperbolic Tangent Function



Downweight managers that are part of many firms

Sum of Shared Nodes





Company 1 and 2 have A and B in common.
The weight of the link between 1 and 2 is
0.46 + 0.32 = 0.78.

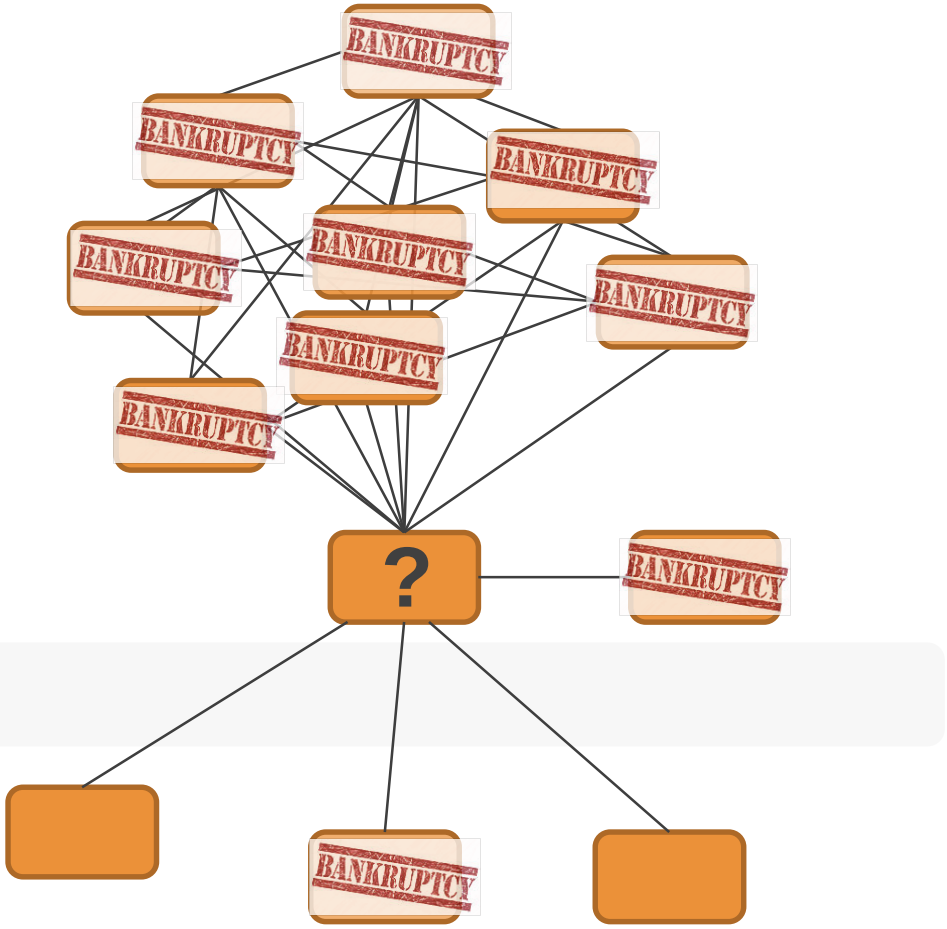
How do we **quantify** the links?

Bottom Node Projection

How do we get a **prediction**?

Weighted vote Relational Neighbour

Weighted vote relational neighbour



W

$$P(B = \text{bankrupt}|N_B) = \frac{1}{Z} \sum_{Vertex\ j \in N_i} w_{i,j} * P(J = "Bankrupt"|N_j)$$

Ensemble model

Financial Variables

- Return on Assets
- Return on Equity
- Current Ratio
- Debt to Total Assets
- Cash Flow to Equity
- Profit/Loss

Dummy Variables

- Missing Values
- Small or Large?
- Newly Founded?

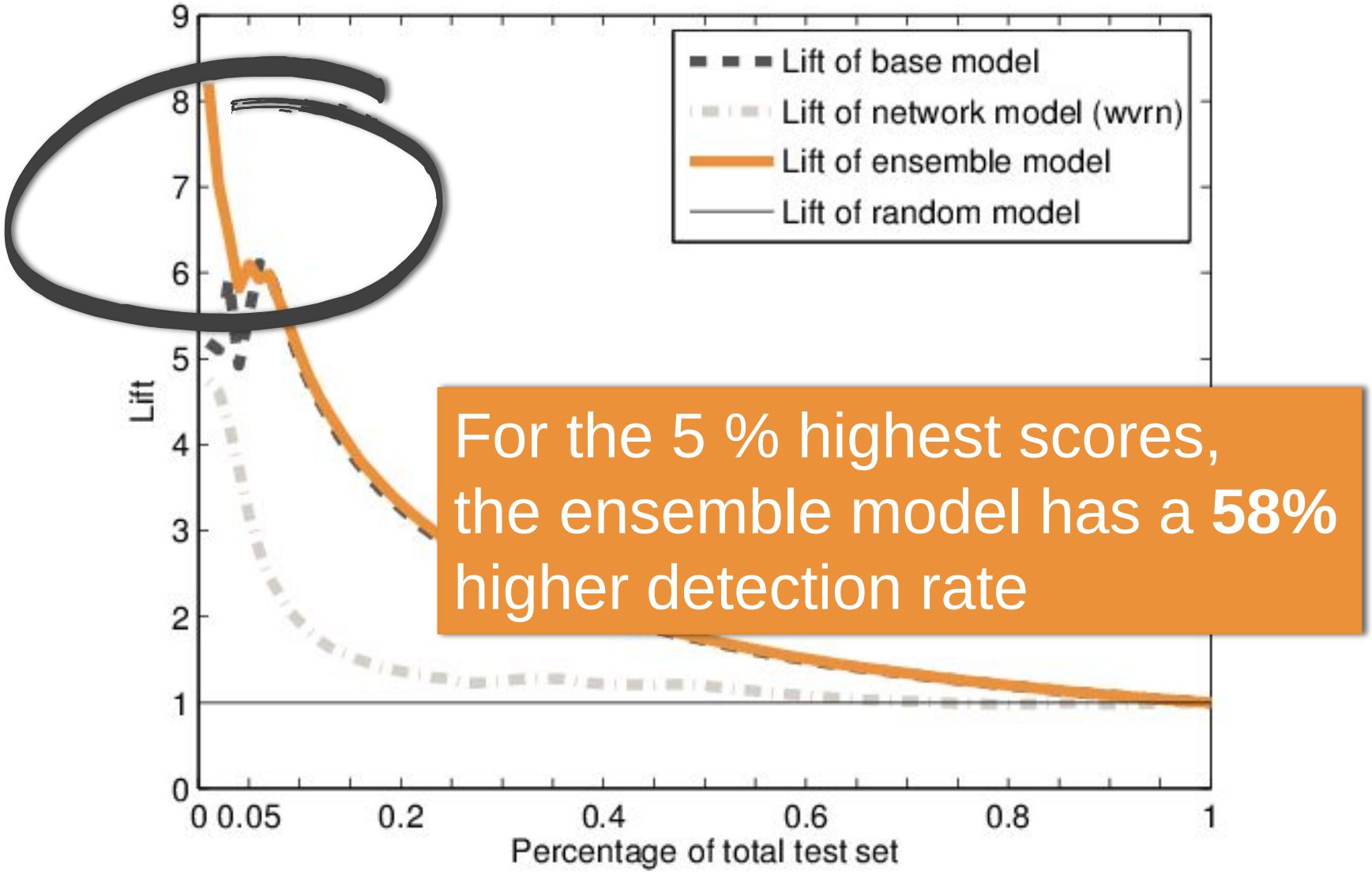
Network score

- wvRN

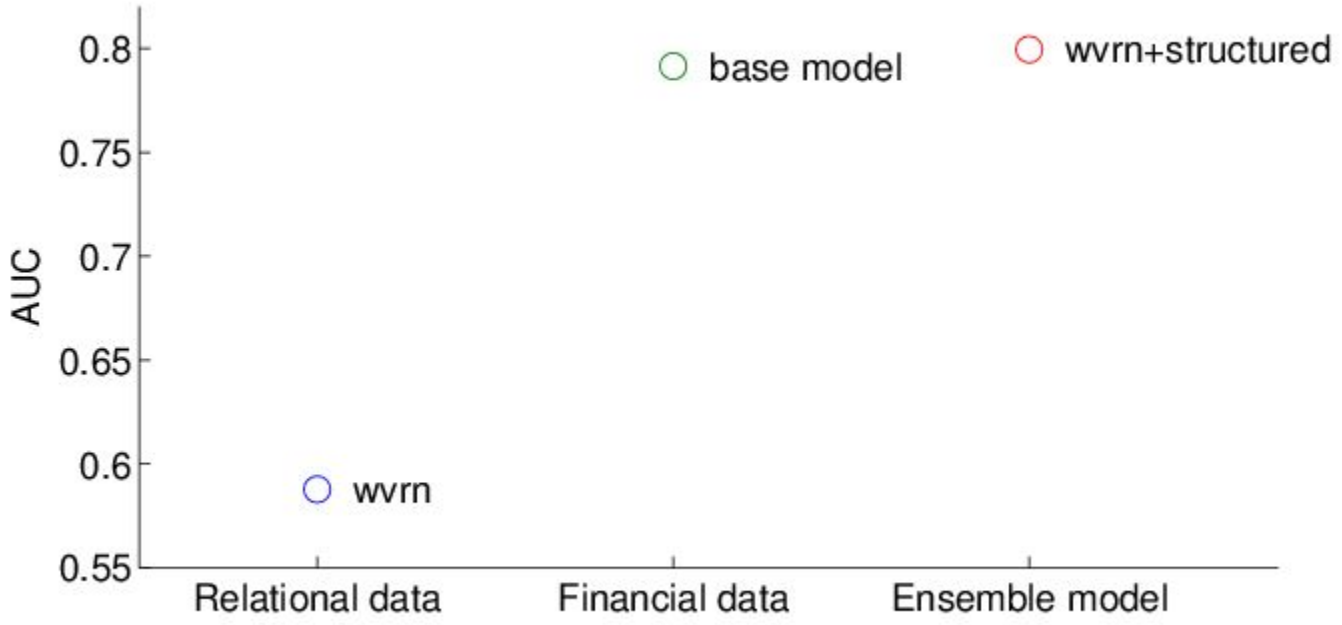
Linear SVM model

Performance assessment

1. Lift Curve
2. AUC Performance
3. Predictors analysis



For the 5 % highest scores, the ensemble model has a **58%** higher detection rate



The AUC-value is **not** significantly higher

	Base Model	Ensemble model
+	Missing ROE Missing Equity Ratio Missing ROA	Network scores Missing ROE Missing Equity Ratio
−	Newly Founded Equity Ratio Profit(+)/Loss(-)	Newly Founded Equity Ratio Profit(+)/Loss(-)

Three conclusions to take home

1

Companies linked to many (or only) bankrupt firms have a **higher probability of bankruptcy**, especially when they find themselves ***in a bad financial position***.

2

Combining the network scores of the relational model with the financial data results in a **better detection** of the companies that are more likely to fail.

3

Being related to healthy firms only **does not compensate for a bad financial situation** and does not lower your probability of failure.


FINISH