

# Co-supplying patterns in firm-to-firm transactions data are able to generate **vector representations (embeddings)** of firms that capture relevant aspects of competition.

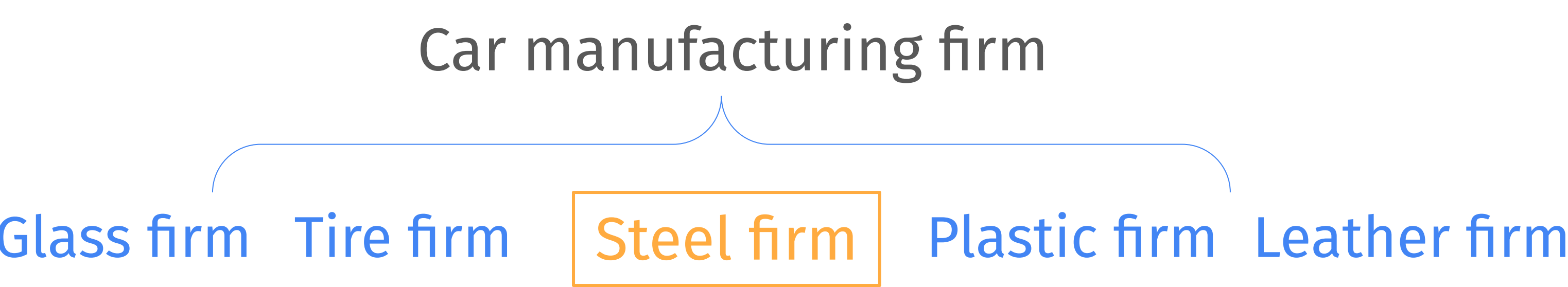
**Firm embeddings: A machine learning approach for characterizing firms with transactions data**  
Vasco Carvalho, Stephen Hansen, Glenn Magerman, Yabra Muvdi

## Motivation and methodology

Generating a measure of similarity between firms

The fundamental idea underlying firm embeddings is that firms can be characterized by exploiting the co-occurrence patterns in firm-to-firm transactions data.

“You should know a ~~word~~ firm by the company it keeps.”

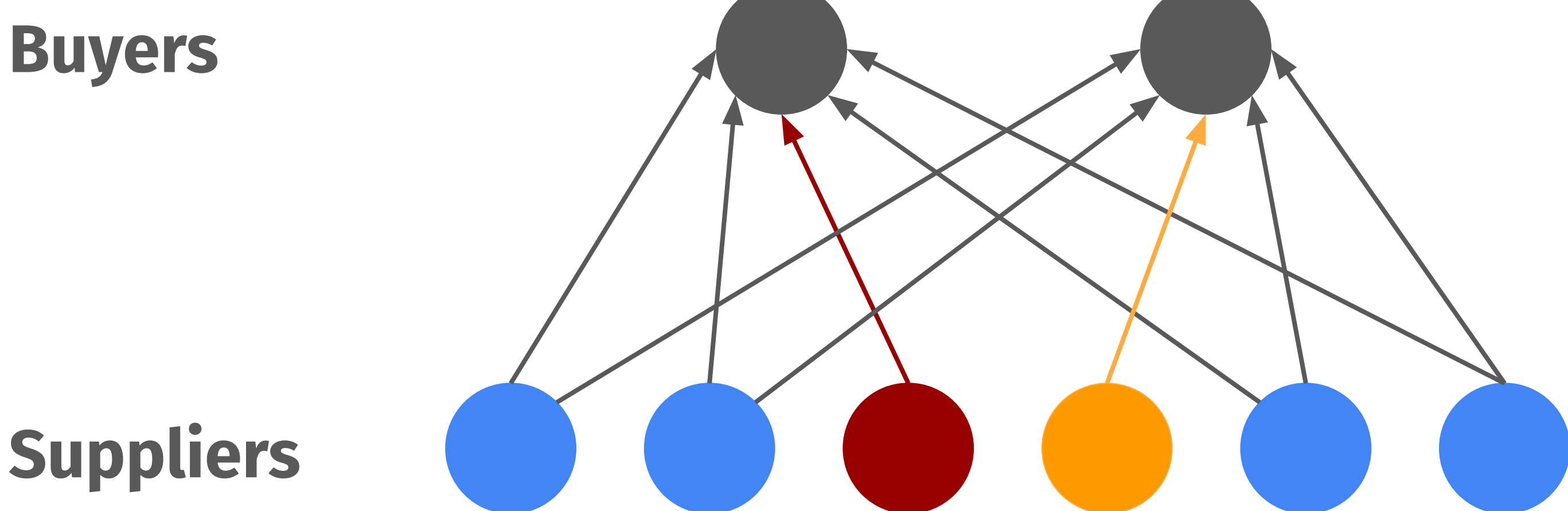


We use the **exponential family embeddings** framework developed by Rudolph et al. (2016) to formalize this idea:

$$Pr[f_t=s \mid C] = \frac{\exp(\bar{\alpha}_c \rho_s)}{\sum_{s'} \exp(\bar{\alpha}_c \rho_{s'})}$$

By modeling the co-supplying patterns in this way, we are able to find the  $\alpha$  and  $\rho$  that maximize the probability of the observed data. We call these vectors **firm embeddings**.

Using these vectors we are able to generate precise measures of similarity between firms (e.g. cosine similarity). Intuitively, two firms will have similar embeddings if they have similar co-suppliers.



## Data

- National Bank of Belgium firm-to-firm transactions database (**7,343,476 transactions in 2014**)
- Standard firm characteristics from the annual accounts and VAT declarations

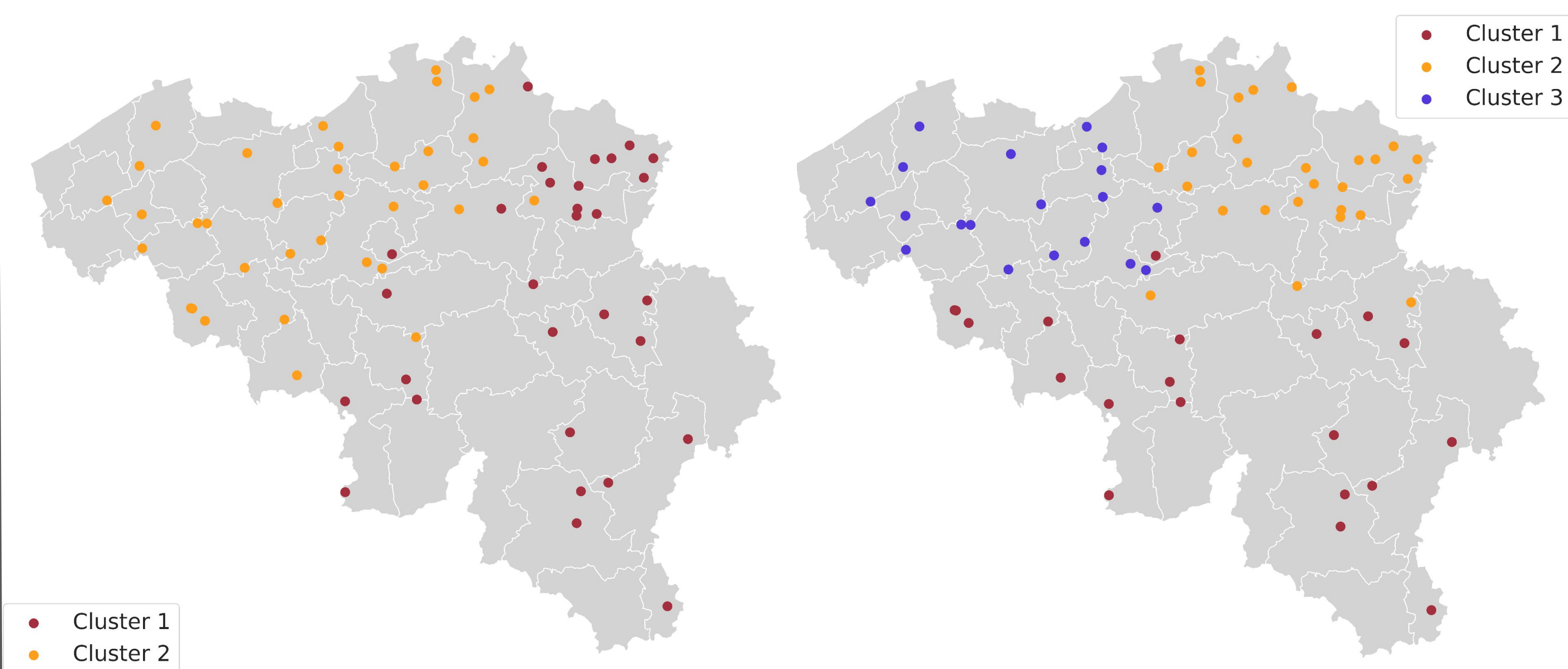
## Preliminary results

Our embeddings are able to capture the role of geography in four-digit NACE sectors where this is a fundamental aspect of competition.

$$\cos sim_{i,j} = \beta_0 + \beta_1 \log distance_{i,j} + \varepsilon_{i,j} \quad \forall i,j \in NACE_k$$

NACE-4 Sector	$\beta_1$
Landscape service activities	-0.134
<b>Manufacture of ready-mixed concrete</b>	<b>-0.131</b>
Support activities for crop production	-0.128
Cutting, shaping and finishing of stone	-0.119
Floor and wall covering	-0.116

## Clustering of ready-mixed concrete firms embeddings



## Further directions and validation exercises

- Test embeddings on a different dimension of competition (e.g. quality)
- Incorporate price data
- Assess the capacity of the embeddings to capture similarity between firms involved in merger cases
- Market definitions and monopoly tests