

Economic Growth Goes ‘Fractal’: The Changing Structure of the UK’s High-Growth Economy

Anna Valero, CEP and POID, LSE

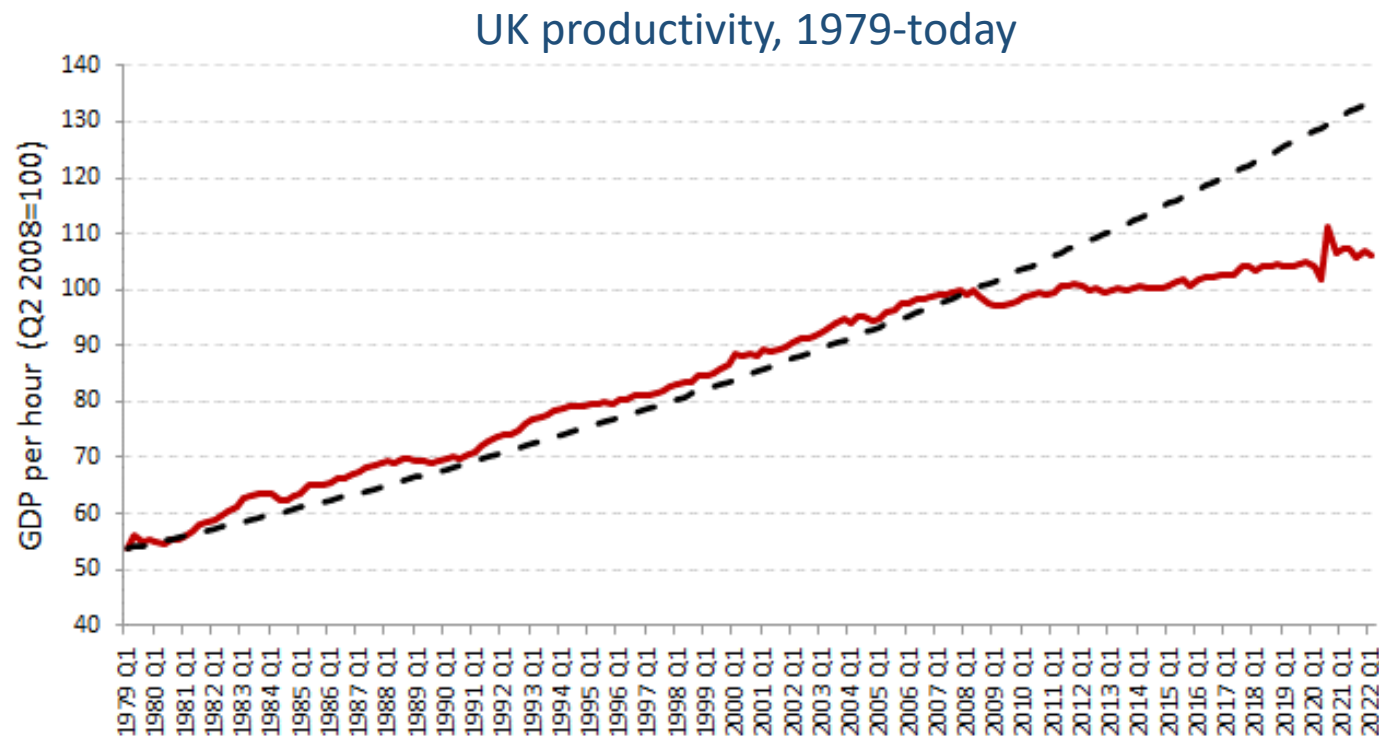
(with Mirko Draca, Max Nathan, Viet Nguyen, Juliana Oliveira-Cunha, Anna Rosso and Shengxing Zhang)

Modelling an Evolving Economy, 7th October 2022

Overview

- We study the structure of the UK's 'high-growth' economy using comprehensive data on firm activities and performance from Beahurst
- We create measures of textual similarity between firms which enable us to locate them in clusters and networks
- We then relate such features to firm outcomes
- **Key findings:**
 - We document a 'fractal' structure amongst high-growth firms – which are split into meaningful clusters
 - High-growth firms appear to be getting more differentiated from each other
 - Originality pays – up to a point – there are better outcomes for firms doing something new, but doing so amongst peers

Motivation 1: High-growth firms are an important source of much needed growth in the UK



Source: ONS Output per hour worked, release date 7 July 2022. Table 32.



Policy paper

Patient Capital Review

The Treasury has now concluded its Patient Capital Review, which considered how to support innovative firms to access the finance that they need to scale up.

From: [HM Treasury](#) and [Department for Business, Energy & Industrial Strategy](#)

Published 23 January 2017

Motivation 2: Traditional datasets are limited in their ability to shed light on high-growth firms

- Typically we see revenue or employment growth, but not much about growth potential or intention
- And SIC codes are not very informative on emerging technologies

Independent Review of
UK Economic Statistics

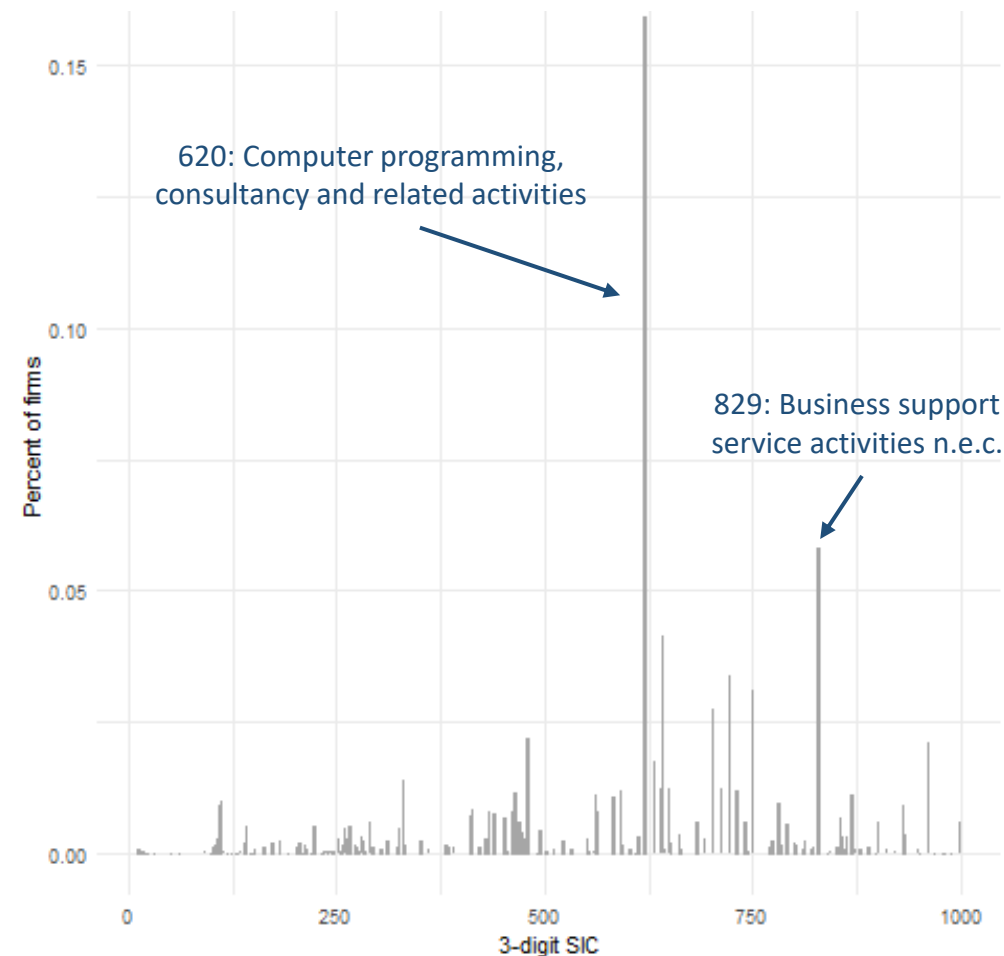
Professor Sir Charles Bean



March 2016

“... the changing structure of the economy means that SIC will constantly lag reality, under-representing newer industries and over-representing ones that are declining in importance.”

Distribution of Beahurst firms across SIC



Related literature

- **High-growth firms and business dynamism:**

- Haltiwanger et al. (2017), Decker et al. (2016), Calvino et al. (2018), De Loecker et al. (2022) [Deaton]; Oliveira-Cunha et al. (2021) [Economy 2030]

- **Entrepreneurship and venture capital:**

- Guzman & Stern (2020), Ewens et al. (2018), Dalle et al. (2017), Kerr et al. (2014)

- **Classifying firms based on textual information:**

- Competitive dynamics in product markets, firm similarity: Hoberg & Phillips (2010, 2014, 2016), Menon et al. (2018) [public firms, 10-K forms]
- Start up strategies: Guzman & Li (2022) [start-up websites vs incumbent 10-K forms]
- Mapping emerging sectors using website text: Nathan & Rosso (2015) [digital economy], Mateos-Garcia et al. (2014, 2018) [video games/VR/AR], Bishop et al. (2022) [within SIC]
- Technological innovation: Kelly et al. (2021) [patents], Kogan et al. (2022) [patents-occupations]

Data

- Beauhurst tracks UK firms that have hit any of 8 triggers since 2011
- Tracking ends upon exit (successful or unsuccessful)
- Comprehensive and curated profile including company descriptions, financials, fundraising activity, outcomes...
- We use the web-scraped company description



- ✓ Received equity investment
- ✓ Underwent an MBO/MBI
- ✓ Received venture debt
- ✓ Reached scaleup status
- ✓ Received a large innovation grant
- ✓ Spun out from an academic institution
- ✓ Graduated from a selected accelerator
- ✓ Featured on a selected high-growth list

	Web-scraped description	Analysts' description
Deliveroo	Deliveroo is on a mission to transform the way the world thinks about food delivery. It's not a chicken chow mein and a night on the sofa anymore, it's your favourite local restaurant, it's a dinner party, a date. We're five years in, and along the way our team have taken hundreds of ideas from brainstorm to global roll-outs, like Deliveroo Editions & bespoke kitchens designed to host a locally-curated selection of restaurants. Editions are our solution to ensuring that our customers have access to the best of the food-scene, no matter where they live. And that's just what we're like at Deliveroo, no compromise allowed and lots of food-inspired challenges to get your teeth into. Out-of-the-box thinking is actively encouraged and we move quickly to make great ideas happen. We're energetic, fast-paced and blow off steam with free-for-all Friday lunches. It's a formula that's working too & we're bringing great food to customers in 13 countries and over 200 cities.	Deliveroo provides delivery services for restaurants, using technology to predict the time taken to prepare meals and efficient ways of delivering orders using the location of restaurants, customers and riders.

Cleaning up the text

- We apply standard pre-processing steps to construct a vocabulary of unique words

	Mean	Median	Max	Min	Sample total
(A) Beauhurst web-scraped (N = 33,973)					
Total words	76.38	65	1,655	13	
Unique words	62.95	56	826	13	94,298
(B) Beauhurst analysts (N = 33,973)					
Total words	11.58	11	49	1	
Unique words	11.34	11	44	1	41,385

Notes: This table provides information on the amount of text available in different UK data sources. The sample frame is the set of 33,973 BH high-growth firms, obtained after the pre-processing steps



Notes: The top 100 unique words in our BH web-scraped data, obtained after the pre-processing steps. The size of the word reflects the number of times it has been used across firm descriptions.

Methodology overview

Cosine similarities across firm pairs (H&P, 2016):

- Start with vocabulary of words
- Each firm represented by a vector where each element is populated with a 1 if the firm uses a word, 0 if not
- These are converted into frequencies and then stacked into a matrix
- Cosine similarity is calculated between each two firms
- $[0,1]$, higher when two firms use more of the same words
- Similarities stacked into $N \times N$ firm matrix



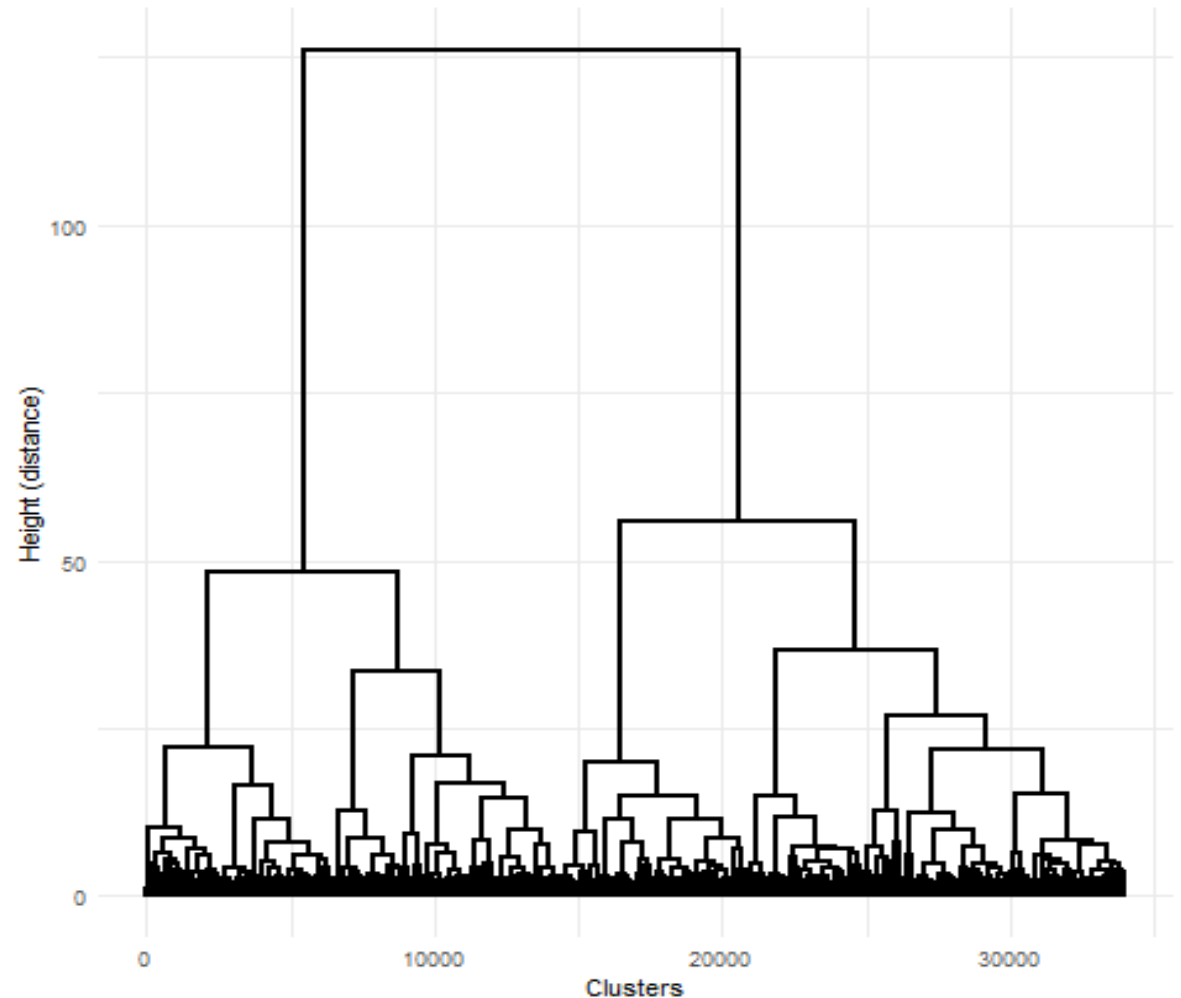
...form the basis of 3 analyses:

- Hierarchical clustering
- Network connections
- Differentiation across and within cohorts

Clustering

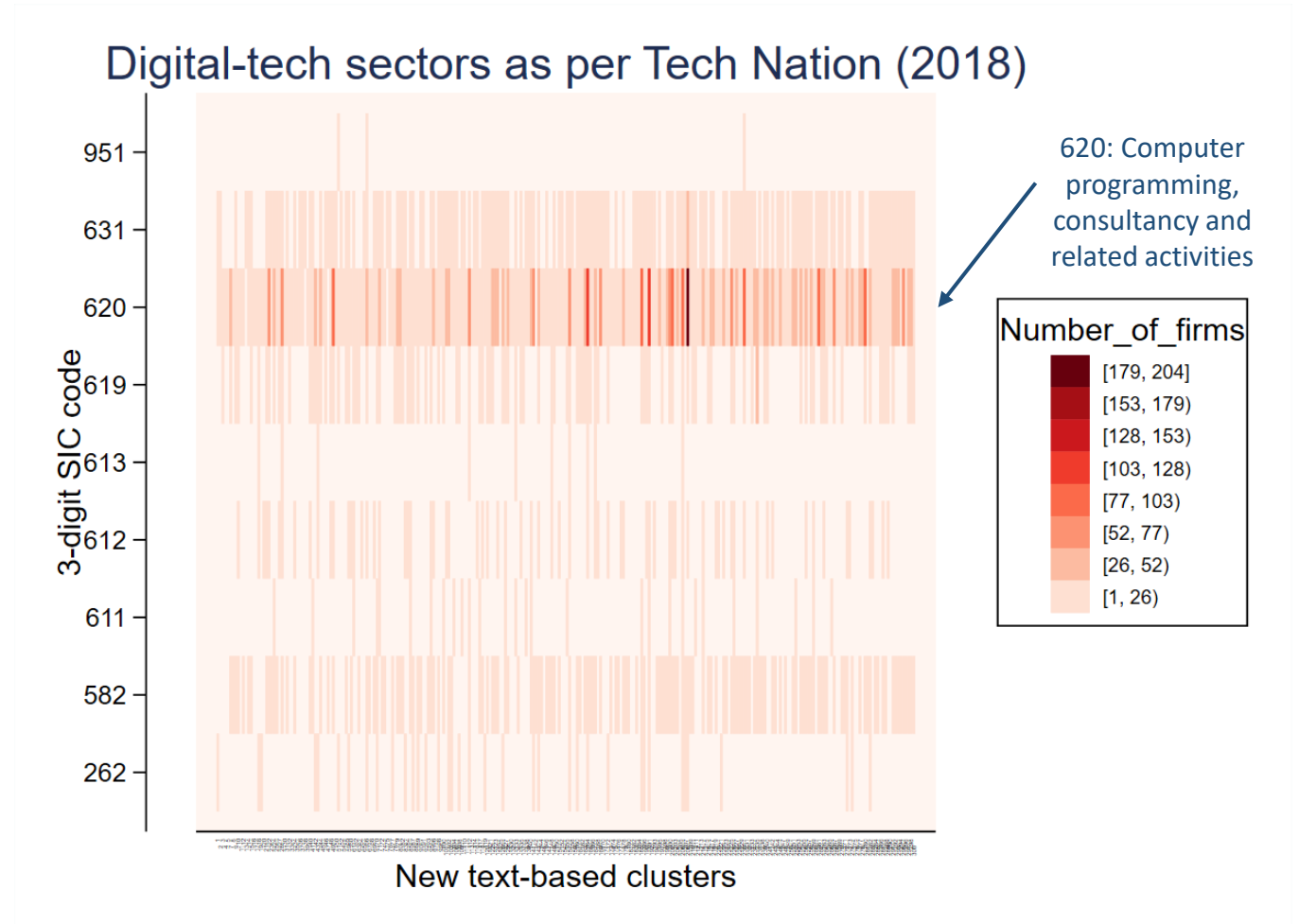
Hierarchical clustering of high-growth firms

- We run a clustering algorithm on the firm matrix
- This creates a branching structure ('fractal' groupings of firms)
- We place similar firms into 300 discrete bins (vs 286 3 digit SIC codes in our data)

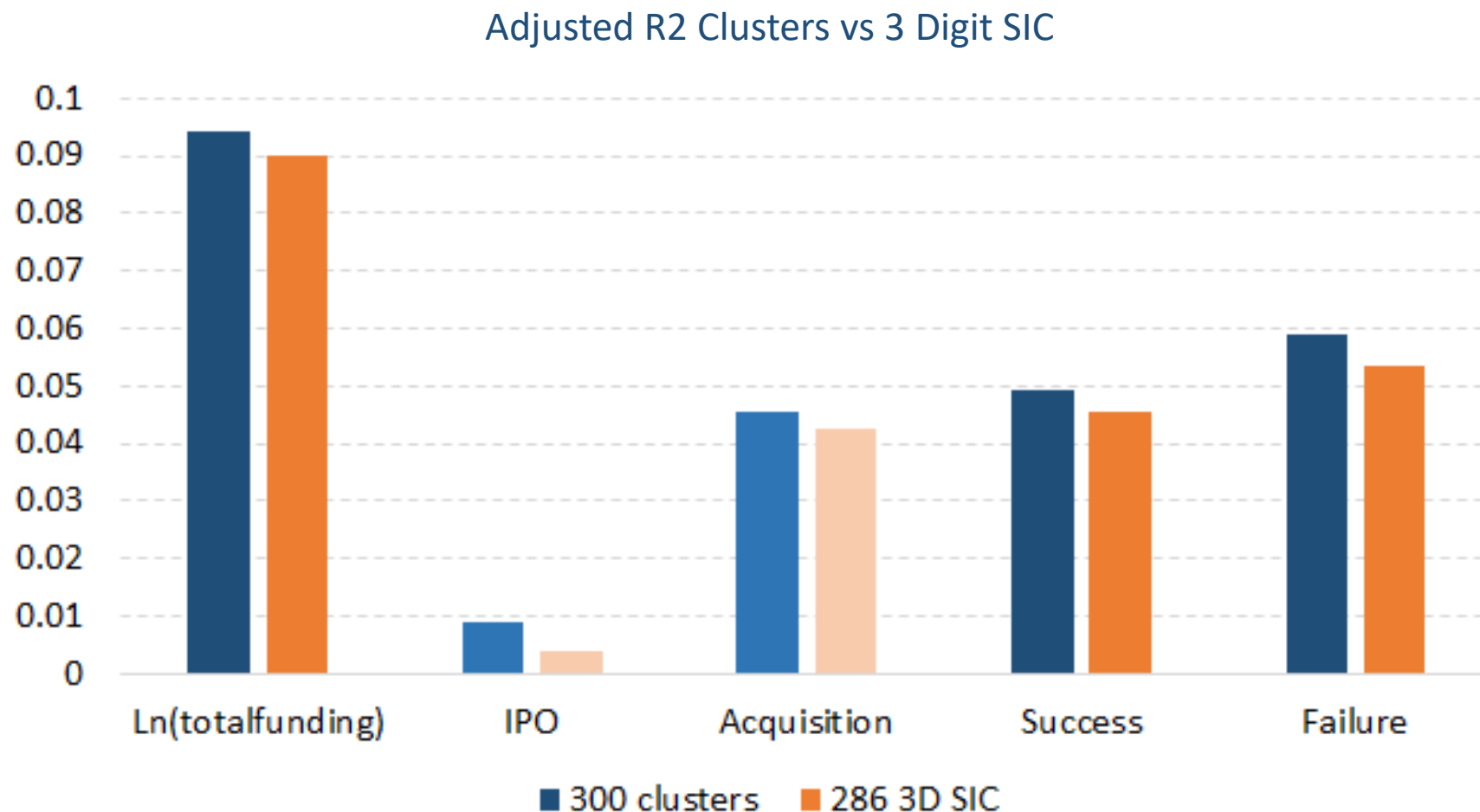


Some of the key sectors are spread across clusters

- Around 20% of sample are in 'digital sectors' – mainly 620
- Such codes are split across many clusters (272 out of the 300)



Clusters do a slightly better job of explaining key outcomes

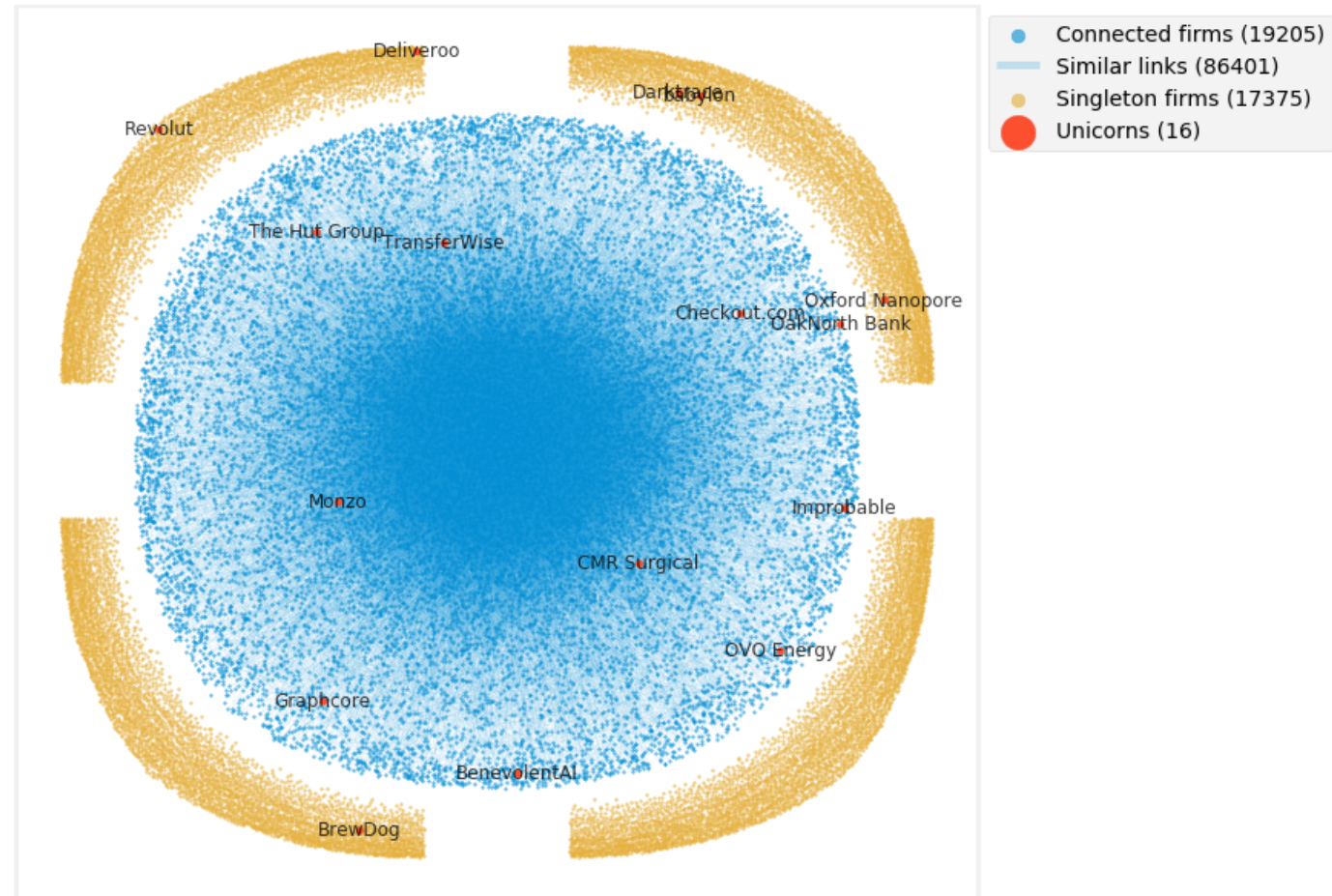


Notes: Regression sample includes Beauhurst tracked companies. Adjusted R2 from OLS regressions of dependent variable on clusters or 3 digit SIC dummies, respectively. Birth cohort dummies included in all regressions. Failure = business death or zombie status

Network connections

Network based structures

- We adopt a threshold of 0.2132 cosine-similarity to define a connection between firms (H&P, 2016)
- “Singleton” firms vs “connected” firms
 - Those closer to the centre are more connected



Notes: Fruchterman-Reingold force directed algorithm to position connected nodes.

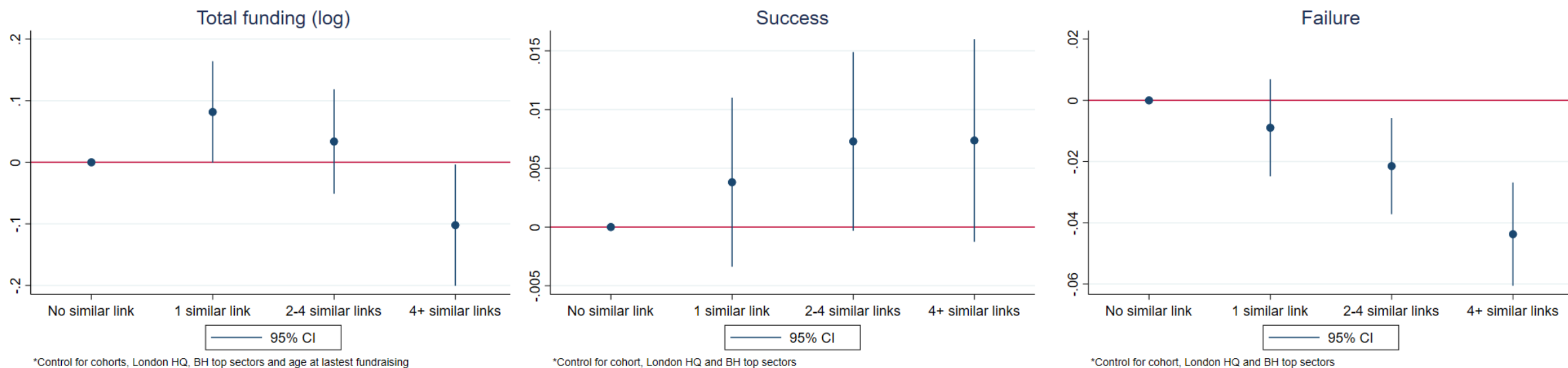
Some connections appear to be a good thing (1)

	(1)	(2)	(3)
	Ltotalfund	success	failure
1 similar link	0.0817* (0.042)	0.00382 (0.004)	-0.00894 (0.008)
2-4 similar links	0.0338 (0.043)	0.00729* (0.004)	-0.0215*** (0.008)
4+ similar links	-0.102** (0.050)	0.00737* (0.004)	-0.0437*** (0.009)
HQ region is London	0.538*** (0.031)	0.00835*** (0.003)	-0.0351*** (0.006)
Age at latest funding	0.437*** (0.009)		
Cohort	Yes	Yes	Yes
Top sectors	Yes	Yes	Yes
N	12,663	18,231	18,231
r ²	0.243	0.0195	0.0375
y _{mean}	12.94	0.0327	0.192

Notes: Sample includes Beauhurst startups founded from 2010-2019. Robust standard errors in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

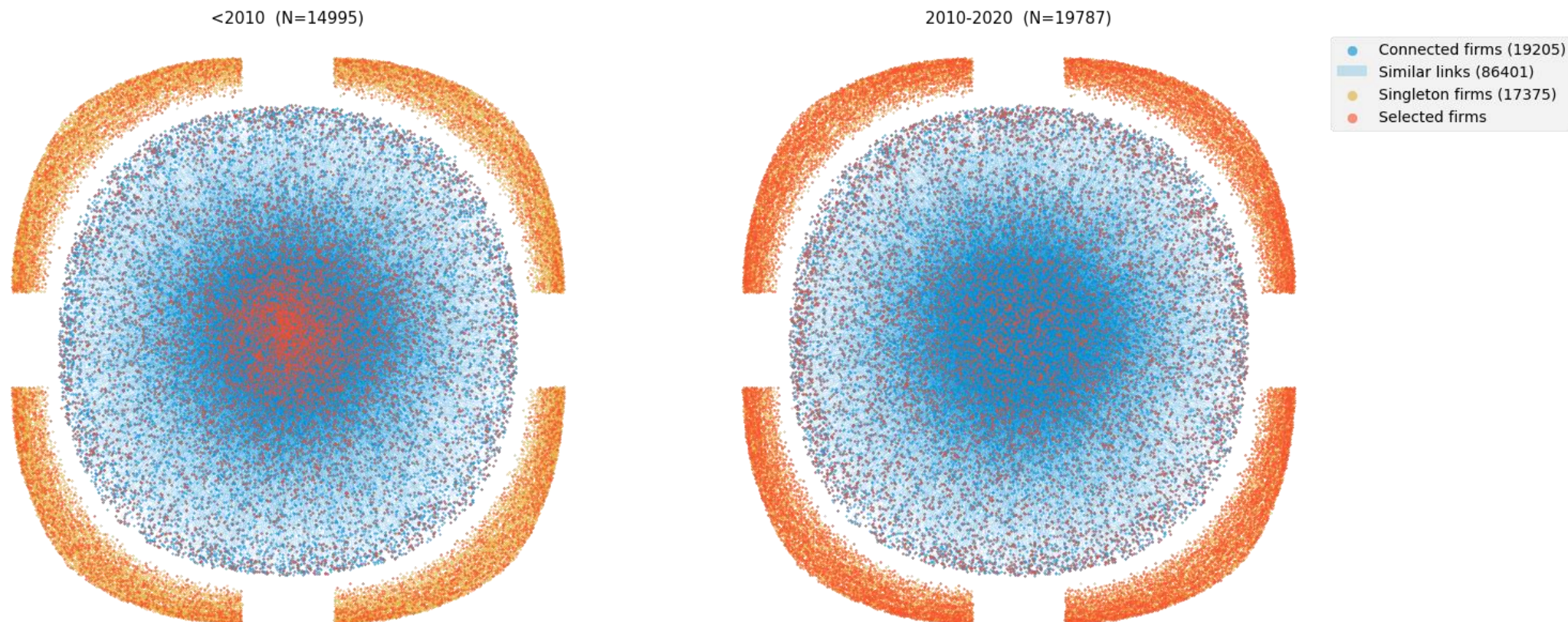
Some connections appear to be a good thing (2)

High growth firm outcomes and network connections



Notes: Sample includes Beauhurst startups founded from 2010-2019.

Across cohorts, firms appear to be less connected



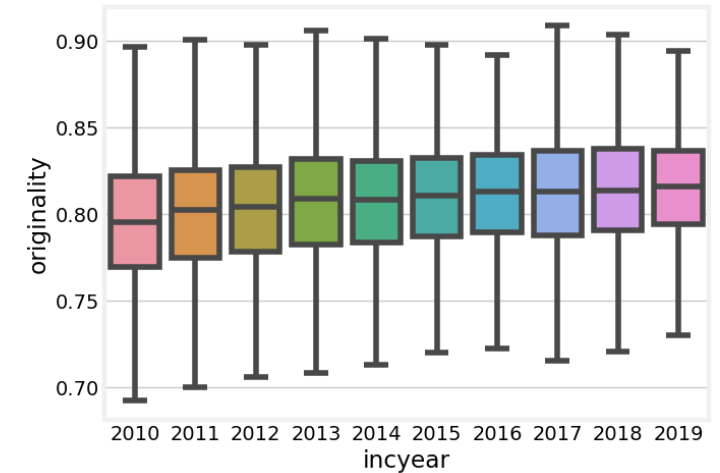
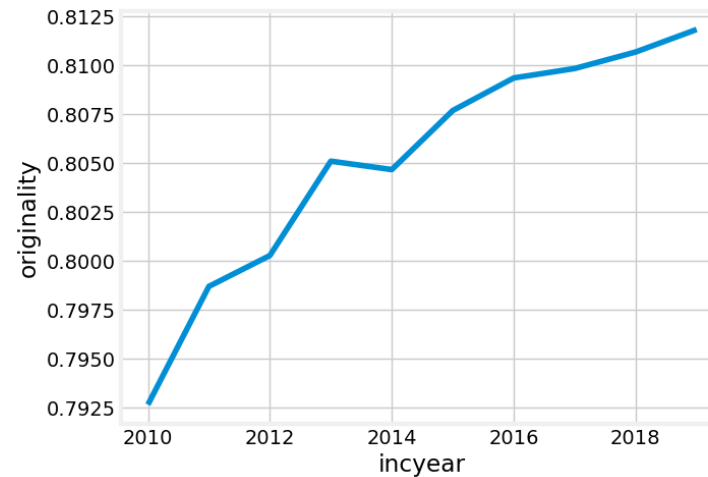
Differentiation across and within cohorts

Originality and trendiness measures

- **Originality:**

1-similarity with the most similar of high-growth start-ups in pre-2010 cohort

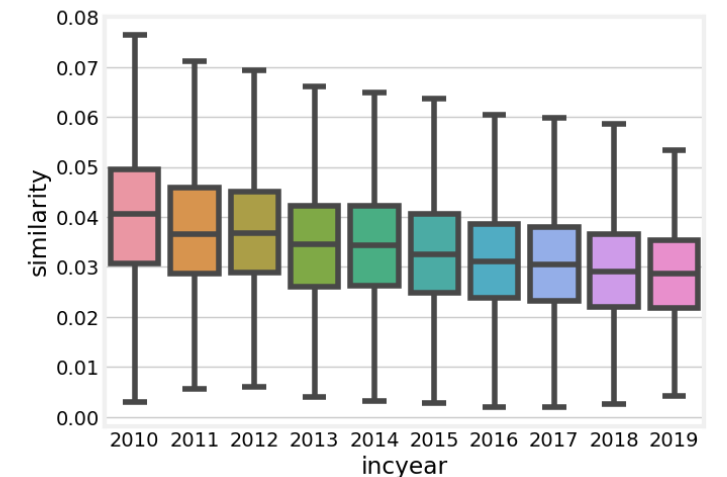
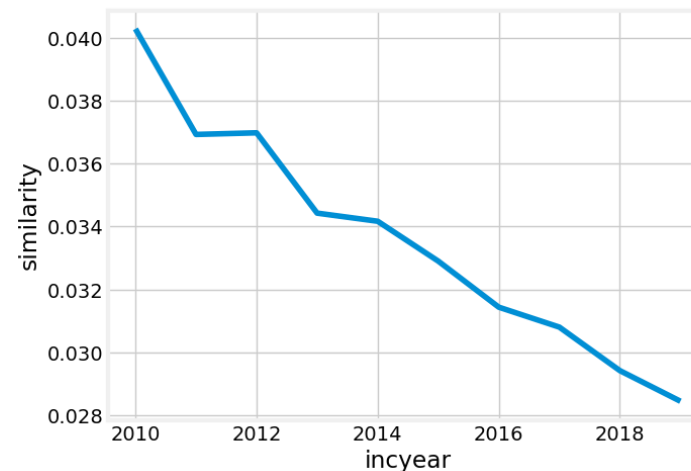
Originality against pre-2010 firms by cohorts



- **Trendiness:**

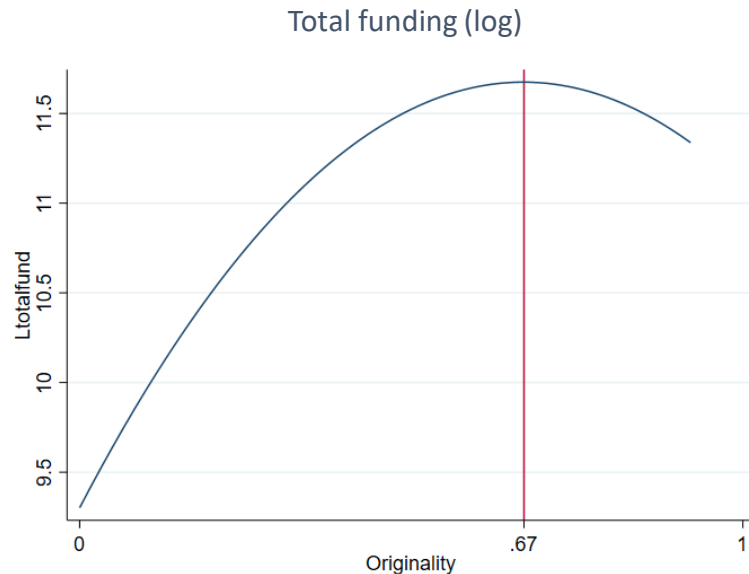
Average similarity with all high-growth start ups in the same birth cohort

Similarity within cohorts (Trendiness)

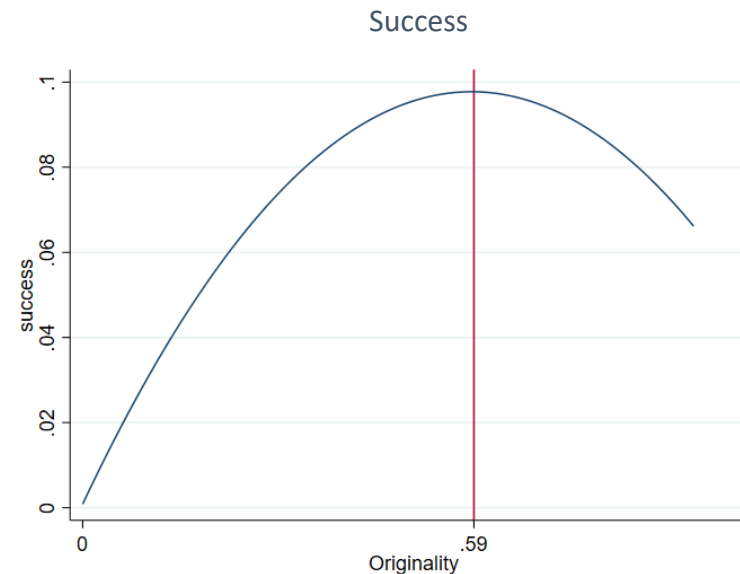


It's good to be different to the past, to some extent

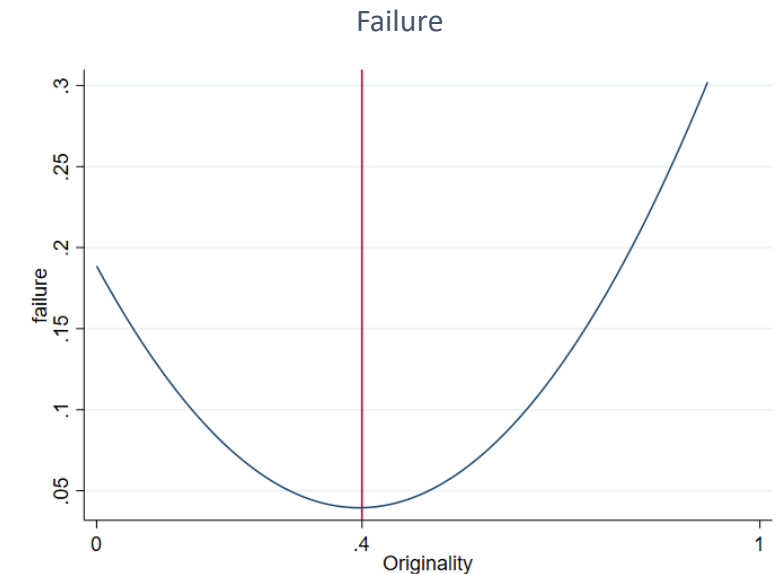
High growth firm outcomes and originality



*Fitted curve for Technology/IP-based businesses, HQ London, incorporated in 2010



*Fitted curve for Technology/IP-based businesses, HQ London, incorporated in 2010



*Fitted curve for Technology/IP-based businesses, HQ London, incorporated in 2010

Notes: Functions implied by coefficients estimated in regressions of specified dependent variable on originality, originality², London dummy, cohort fixed effects (and age at last funding). Coefficients on originality and originality² statistically significant. Sample includes Beahurst startups founded from 2010-2019. Robust standard errors in parentheses.

But also it's good to be in a currently 'trendy' area

	(1)	(2)	(3)
	Ltotalfund	success	failure
trendiness	5.791*** (1.39)	0.582*** (0.12)	-1.537*** (0.27)
HQ region is London	0.548*** (0.03)	0.00863*** (0.00)	-0.0348*** (0.01)
Age at latest funding	0.437*** (0.01)		
Cohort	Yes	Yes	Yes
Top sectors	Yes	Yes	Yes
N	12,663	18,231	18,231
r2	0.24	0.02	0.04

Notes: Sample includes Beahurst startups founded from 2010-2019. Robust standard errors in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Conclusions and future work

- Text based measures of similarity can form the basis of a number of informative measures about high-growth firms
 - Including new measures of originality ~ innovation
- We find that our clusters are informative
- And high-growth firms appear to be getting more differentiated
 - Originality pays – up to a point – there are better outcomes for firms doing something new, but doing so amongst peers
- Future work
 - Other measures of similarity / networks – looking deeper within SIC codes
 - Other outcomes (e.g. innovation explicitly, stages of fundraising – seed/growth)
 - Dynamics - changes in firm descriptions over time
 - Networks of individuals in the high-growth economy

Thank you!