Economic Complexity and the emergence of ideas

Alex Bishop
Juan Mateos-Garcia
Overview

Study metrics of complexity that capture the composition of regional economies using labour-market statistics.
Overview

Study metrics of complexity that capture the composition of regional economies using labour-market statistics.

Move beyond official data which is laggy and aggregated to using web data to measure emergence and novelty at the business level.
Overview

Study metrics of complexity that capture the composition of regional economies using labour-market statistics.

Move beyond official data which is laggy and aggregated to using web data to measure emergence and novelty at the business level.

Analyse the links between regional economic complexity and levels of emergent/novel activity.
Overview

Study metrics of complexity that capture the composition of regional economies using labour-market statistics.

Move beyond official data which is laggy and aggregated to using web data to measure emergence and novelty at the business level.

Analyse the links between regional economic complexity and levels of emergent/novel activity.

Use web-data to generate a new measure of complexity.
Introduction to Economic Complexity
Economic complexity

**Economic complexity**: Diversity and sophistication of productive capabilities.
**Economic complexity**: Diversity and sophistication of productive capabilities.

Link between economic complexity of nations and their wealth in terms of GDP per capita.

Hidalgo et al. PNAS, 106(26):10570–10575, June 2009
Economic complexity: Diversity and sophistication of productive capabilities.

Link between economic complexity of nations and their wealth in terms of GDP per capita.
Hidalgo et al. PNAS, 106(26):10570–10575, June 2009

Economic complexity shapes a location’s future specialisation trajectory through the Principle of Relatedness.

nesta
Economic complexity: Diversity and sophistication of productive capabilities.

Link between economic complexity of nations and their wealth in terms of GDP per capita.
Hidalgo et al. PNAS, 106(26):10570–10575, June 2009

Economic complexity shapes a location’s future specialisation trajectory through the Principle of Relatedness.

Estimate complexity using Economic Complexity Index (ECI). 

nesta
Hidalgo and Hausmann, PNAS, (2009)
Originally defined through the *method of reflections*.
Originally defined through the method of reflections.

Calculates diversity and ubiquity, recursively using the information of one to correct the other.
Originally defined through the method of reflections.

Calculates diversity and ubiquity, recursively using the information of one to correct the other.

Equivalent to a spectral clustering algorithm which partitions a similarity graph into two.
Mealy and Farmer. arXiv:1711.08245
Economic Complexity in the UK
We use labour market statistics from the ONS: BRES & IDBR
https://www.nomisweb.co.uk/
We use labour market statistics from the ONS: BRES & IDBR
https://www.nomisweb.co.uk/

Gives employment count and business count by sector (4-digit SIC codes) and regional geography
Data

We use labour market statistics from the ONS: BRES & IDBR
https://www.nomisweb.co.uk/

Gives employment count and business count by sector (4-digit SIC codes) and regional geography

We cluster SIC codes into ‘Industrial segments’ using an algorithm which measures the relatedness based on employment, business co-location, and occupational composition of the workforce and B2B trade.
Data

We use labour market statistics from the ONS: BRES & IDBR
https://www.nomisweb.co.uk/

Gives employment count and business count by sector (4-digit SIC codes) and regional geography

We cluster SIC codes into ‘Industrial segments’ using an algorithm which measures the relatedness based on employment, business co-location, and occupational composition of the workforce and B2B trade.

Result in a [243x72] matrix for each dataset capturing either employment count or business count in each sector-TTWA combination.
What connects economic complexity with development? **Emergence**
COMPLEXITY AND EMERGENCE

What connects economic complexity with development? **Emergence**

But:

- Principle of relatedness can’t *currently* capture emergence of truly novel activities
COMPLEXITY AND EMERGENCE

What connects economic complexity with development? **Emergence**

But:

- Principle of relatedness can’t *currently* capture emergence of truly novel activities
- Structured data offers an aggregated, lagging view of the economy
Web data: moving beyond aggregates
Glass ([www.glass.ai](http://www.glass.ai)), an economic intelligence start-up has developed technology that can “understand” text at scale.
Glass ([www.glass.ai](http://www.glass.ai)), an economic intelligence start-up has developed technology that can “understand” text at scale.

Their web crawler has collected text from over a million UK business websites.
Glass (www.glass.ai), an economic intelligence start-up has developed technology that can “understand” text at scale.

Their web crawler has collected text from over a million UK business websites.

Addresses, descriptions of the business, and a predicted industrial sector (based on the LinkedIn taxonomy) are extracted using machine learning.
Web data

Glass (www.glass.ai), an economic intelligence start-up has developed technology that can “understand” text at scale.

Their web crawler has collected text from over a million UK business websites.

Addresses, descriptions of the business, and a predicted industrial sector (based on the LinkedIn taxonomy) are extracted using machine learning.

Provides an unparalleled depth of data on the UK’s digital economy with which to measure emergence!
Fuzzy matching methodology

How can we trust it?
Fuzzy matching methodology

How can we trust it?

Match Glass companies to Companies House using company names in order to verify the sectoral and geographic coverage.
Fuzzy matching methodology

How can we trust it?

Match Glass companies to Companies House using company names in order to verify the sectoral and geographic coverage.

Name similarity features constructed using a combination of Locality Sensitive Hashing and TF-IDF cosine similarity to generate name similarity scores.
Fuzzy matching methodology

How can we trust it?

Match Glass companies to Companies House using company names in order to verify the sectoral and geographic coverage.

Name similarity features constructed using a combination of Locality Sensitive Hashing and TF-IDF cosine similarity to generate name similarity scores.

Thresholds calibrated by extracting company numbers from websites (required by Companies act 2006) - less than 10% websites contained this information.

nesta
Data validation

- ~1.5m UK business websites
- ~800k with qualified depth
- ~400k after geocoding via NSPL & further filtering
Data validation

- ~1.5m UK business websites
- ~800k with qualified depth
- ~400k after geocoding via NSPL & further filtering

- ~215k websites matched to Companies House
- ~120k websites with same TTWA for Glass and Companies House

nesta
Data validation

- ~1.5m UK business websites
- ~800k with qualified depth
- ~400k after geocoding via NSPL & further filtering

- ~215k websites matched to Companies House
- ~120k websites with same TTWA for Glass and Companies House
Data validation

- ~1.5m UK business websites
- ~800k with qualified depth
- ~400k after geocoding via NSPL & further filtering

- ~215k websites matched to Companies House
- ~120k websites with same TTWA for Glass and Companies House

<table>
<thead>
<tr>
<th>Place</th>
<th>% over-represented by glass</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>8.707801</td>
</tr>
<tr>
<td>Manchester</td>
<td>1.202007</td>
</tr>
<tr>
<td>Bristol</td>
<td>0.399800</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>0.276403</td>
</tr>
<tr>
<td>Glasgow</td>
<td>0.231062</td>
</tr>
<tr>
<td>Oxford</td>
<td>0.214957</td>
</tr>
<tr>
<td>Norwich</td>
<td>0.182724</td>
</tr>
<tr>
<td>Nottingham</td>
<td>0.179391</td>
</tr>
<tr>
<td>Southampton</td>
<td>0.175164</td>
</tr>
<tr>
<td>Sheffield</td>
<td>0.174101</td>
</tr>
<tr>
<td>Guildford and Aldershot</td>
<td>0.156337</td>
</tr>
<tr>
<td>Brighton</td>
<td>0.147259</td>
</tr>
<tr>
<td>Leeds</td>
<td>0.112410</td>
</tr>
<tr>
<td>Bournemouth</td>
<td>0.096772</td>
</tr>
<tr>
<td>Bath</td>
<td>0.095922</td>
</tr>
<tr>
<td>Ipswich</td>
<td>0.090282</td>
</tr>
<tr>
<td>Redruth and Truro</td>
<td>0.084659</td>
</tr>
<tr>
<td>Newcastle</td>
<td>0.081182</td>
</tr>
<tr>
<td>Cheltenham</td>
<td>0.073572</td>
</tr>
<tr>
<td>Colchester</td>
<td>0.062851</td>
</tr>
</tbody>
</table>

Places over-represented by the Glass-House dataset compared to the IDBR
accelerator, accelerators, AI, algorithmic, algorithms, animation, animations, AR, Artificial Intelligence, Augmented Reality, autonomous, BI, big data, bioscience, biotech, biotechnology, bitcoin, blockchain, catapult, cleantech, cross platform, crowdfunding, crypto, cryptocurrencies, cryptocurrency, cyber, cybersecurity, 3d animation, 3d modelling, 3d printing, 3d scanning, data analytics, data science, data visualization, decentralised, decentralized, digital technologies, disruptive technologies, drone, drones, early adopters, emerging markets, fintech, fuel cell, gamification, highly scalable, IAAS, immersive, incubator, internet of things, IoT, IPTV, machine learning, medtech, motion graphics, oculus rift, PAAS, patent pending, patented, patented technology, predictive analytics, remote sensing, robotics, SAAS, SAAS software, sensor technology, sequencing, smart cities, smart grid, UAV, unique patented, virtual reality, VR, wearables

nesta
accelerator, accelerators, AI, algorithmic, algorithms, animation, animations, AR, Artificial Intelligence, Augmented Reality, autonomous, BI, big data, bioscience, biotech, biotechnology, bitcoin, blockchain, catapult, cleantech, cross platform, crowdfunding, crypto, cryptocurrencies, cryptocurrency, cyber, cybersecurity, 3d animation, 3d modelling, 3d printing, 3d scanning, data analytics, data science, data visualization, decentralised, decentralized, digital technologies, disruptive technologies, drone, drones, early adopters, emerging markets, fintech, fuel cell, gamification, highly scalable, IAAS, immersive, incubator, internet of things, IoT, IPTV, machine learning, medtech, motion graphics, oculus rift, PAAS, patent pending, patented, patented technology, predictive analytics, remote sensing, robotics, SAAS, SAAS software, sensor technology, sequencing, smart cities, smart grid, UAV, unique patented, virtual reality, VR, wearables

nesta
Top 20 TTWA’s by share of emergent companies in Glass data
Complexity and emergence
Emergence regression

Complexity helps predict emergence
ECI > 0 highly predictive of GVA, no predictive power for ECI < 0

Share of ‘emergent’ companies not very predictive

CAVEAT: Experimental statistics not designed to be measure of regional productivity
Complexity of language
Each document is a mixture of topics
Each topic is a distribution over words
Each word is drawn from one of those topics
'101 Computer Services was founded in 1993 by Les Gutteridge. Les has been involved with computers in business and in education for more than twenty years. We are based in Birmingham, so our clients are mainly in the West Midlands, but we are happy to work with organisations much further afield. '101 Computer Services was founded in 1993 by Les Gutteridge. Les has been involved with computers in business and in education for more than twenty years. We are based in Birmingham, so our clients are mainly in the West Midlands, but we are happy to work with organisations much further afield.'
TOPIC MODELLING

Tokenise

(Stem)

Stopword removal

Bi-gram detection

Fit topic model

Clusters for each organisation

- 3 levels in both the topic distribution and the document cluster hierarchy. 
  Number of topics:  [359, 50, 10, 2] 
  Number of clusters:  [212, 44, 8, 3]
Clusters for each organisation

- 3 levels in both the topic distribution and the document cluster hierarchy.
  Number of topics: [359, 50, 10, 2]
  Number of clusters: [212, 44, 8, 3]

- Obtain a topic distribution and cluster membership for each organisation
Clusters for each organisation

- 3 levels in both the topic distribution and the document cluster hierarchy.
  Number of topics: [359, 50, 10, 2]
  Number of clusters: [212, 44, 8, 3]

- Obtain a topic distribution and cluster membership for each organisation

- Can aggregate these into Travel To Work Areas (TTWAs)
# Topics over-represented in the areas with highest ECI

<table>
<thead>
<tr>
<th>breseci</th>
<th>idbroeci</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>software, data, platform, communications, infrastructure</td>
</tr>
<tr>
<td>1</td>
<td>approach, understand, process, understanding, relationships</td>
</tr>
<tr>
<td>2</td>
<td>systems, innovative, solution, technical, industries</td>
</tr>
<tr>
<td>3</td>
<td>deliver, leading, delivering, commitment, operate</td>
</tr>
<tr>
<td>4</td>
<td>recruitment, talent, candidates, career, roles</td>
</tr>
<tr>
<td>5</td>
<td>focus, success, successful, existing, manage</td>
</tr>
<tr>
<td>6</td>
<td>value, professionals, long-term, provider, successfully</td>
</tr>
<tr>
<td>7</td>
<td>aim, helping, future, partnership, committed</td>
</tr>
<tr>
<td>8</td>
<td>investment, growth, portfolio, finance, capital</td>
</tr>
</tbody>
</table>

*nesta*
ECI revisited

Hidalgo and Hausmann, PNAS, (2009)
Topic model approach
Pros

- No explicit taxonomy needed
Pros
- No explicit taxonomy needed

Cons
- No explicit taxonomy needed
Pros

- No explicit taxonomy needed
- Interpretable

Cons

- No explicit taxonomy needed
Pros

- No explicit taxonomy needed
- Interpretable

Cons

- No explicit taxonomy needed
- Not easily interpretable
<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>- No explicit taxonomy needed</td>
<td>- No explicit taxonomy needed</td>
</tr>
<tr>
<td>- Interpretable</td>
<td>- Not easily interpretable</td>
</tr>
<tr>
<td>- Timely</td>
<td></td>
</tr>
</tbody>
</table>
Pros

- No explicit taxonomy needed
- Interpretable
- Timely

Cons

- No explicit taxonomy needed
- Not easily interpretable
- Not yet scalable
Conclusions
Conclusions

- Economic Complexity is a useful economic indicator for the UK
Conclusions

- Economic Complexity is a useful economic indicator for the UK

- Web data can be used to detect where emergent activity is happening at the business level
Conclusions

- Economic Complexity is a useful economic indicator for the UK.
- Web data can be used to detect where emergent activity is happening at the business level.
- Web data could be used to “nowcast” economic complexity using a constantly evolving taxonomy.