# Discerning Industrial Networks, Clusters and Competences – An Alternative View Using Web Mining Techniques

PRO-VE 2010

---




www.fabriam.com

# Outline

- Problem areas in industrial cluster research
- Use of the web to find industrial activity
- Comparison with conventional methods
- Shared competencies and overlaps between clusters
- Finding industrial networks using the web
- Analysis of linkages
- Concluding remarks

- Problem areas in industrial cluster research
  - The general problem of finding enough companies and other contributors to an economic milieu
  - The definition of an economic entity
  - Shortcomings of Standard Industry Classification (SIC) codes to describe company activity because:
    - SICs have difficulty capturing ‘new’ industries
    - They are often simply inaccurate
    - Many firms have multiple lines of business

*The net result is that many actors in a practical industrial cluster study are not found when using company lists with SICs as the sole indicator of activity*

Why not use the web to find out ‘what goes on’ in industry in a defined locale?

In this context the Research Question becomes:

*Can the use of the web as an information resource add anything useful to more conventional methods of researching industrial clusters and networks?*

- Use of the web to find industrial activity
  - Acquire a large number of URLs of appropriate firms and supporting actors in a defined locale
    - Credit reference database lists/Chamber of commerce/trade associations/other
  - Full text spidering of all the URLs with data then written to a linked database
- Metrics
  - Region – North East of England
  - 14000 URLs used
  - 177432 keywords

- Comparison with conventional methods
  - Used a subset of 8518 company records of known provenance (from DNB)
  - Compared the results of full text spidering with fields containing SIC based data
  - Not surprisingly whole text searching in a closed set of websites shown to be far superior for finding activity than by searching on SIC based activity alone (Table 1)

- Table 1 - Test of keyword frequency by number of firms

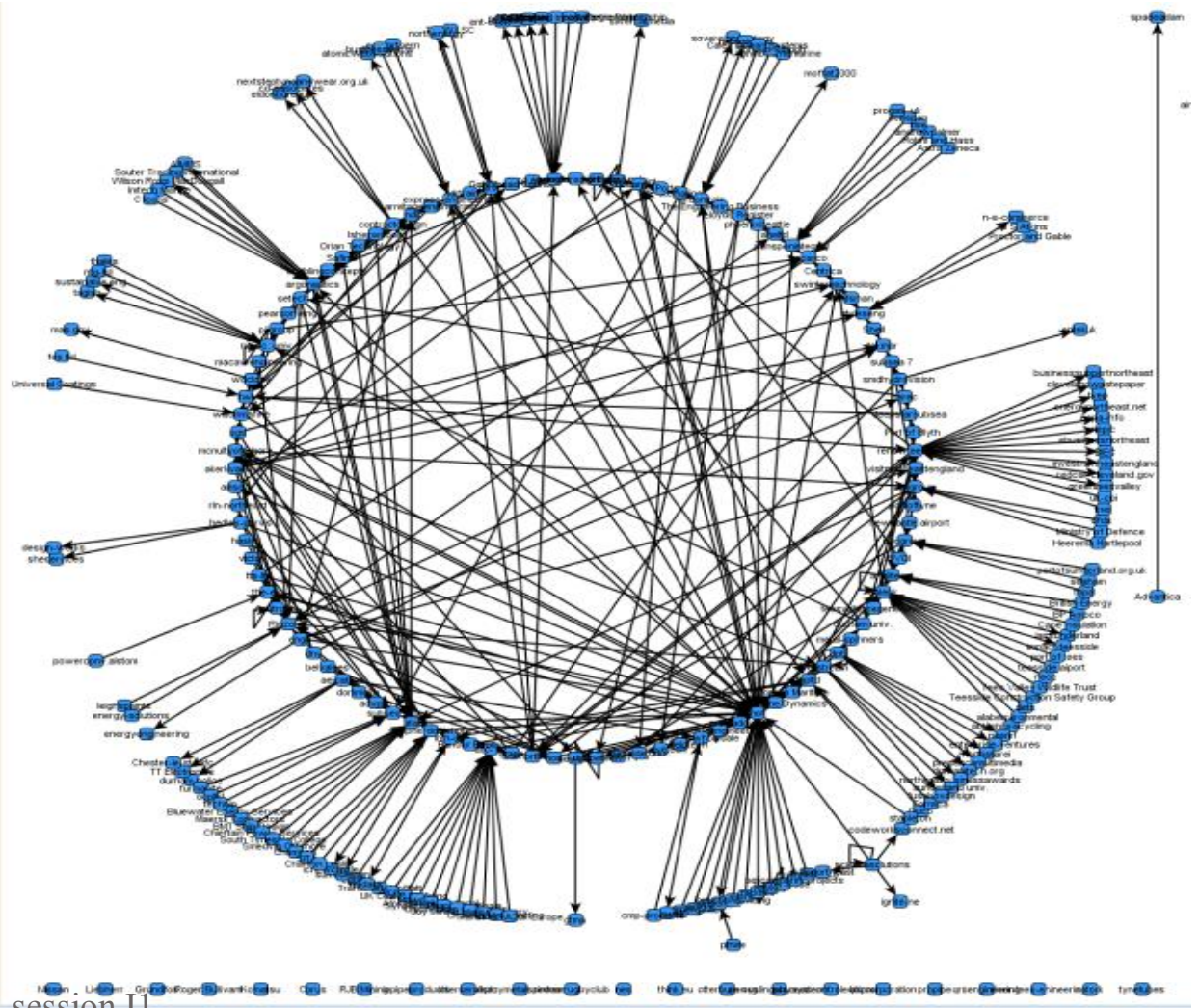
| Word | Count by DNB 'Main Activity' | Count by any DNB 'SIC text' | Count by Keyword |
|---|---|---|---|
| Motorsport | 0 | 0 | 11 |
| Motor racing | 0 | 0 | 9 |
| Motorcycle | 3 | 24 | 38 |
| Automotive | 10 | 8 | 104 |
| Defence | 0 | 0 | 48 |
| Military | 2 | 0 | 25 |
| Offshore | 5 | 0 | 105 |
| Subsea | 1 | 0 | 15 |
| Yacht | 2 | 0 | 13 |
| Sail | 1 | 0 | 9 |
| boat | 6 | 8 | 36 |
| TOTALS | 30 | 40 | 413 |

- Overlaps between clusters and shared competencies
  - Looked at specific groups of engineering firms
    - 1502 in Marine, Offshore, Subsea, Defence and Environmental tech.
  - Firms had ‘membership’ of more than one group
  - Commonalities of capability across sectors
    - manufacture of medium and large sized artefacts such as for example ships, armaments, rigs and subsea ploughs.
    - Particular capability in engineering for more hostile environments

- Finding industrial networks using the web (1)
  - Assumption made that a link exists if, on a company website:
    - a firm puts a reference to another organisation on its own website as a clickable link (found by a short search program)
    - there are visual clues to other organisations on a company website e.g. a client list (carried out by inspection)
    - an external firm references the URL being looked at, so called in-links (using an in-link finder such as www.linkpopularity.com)

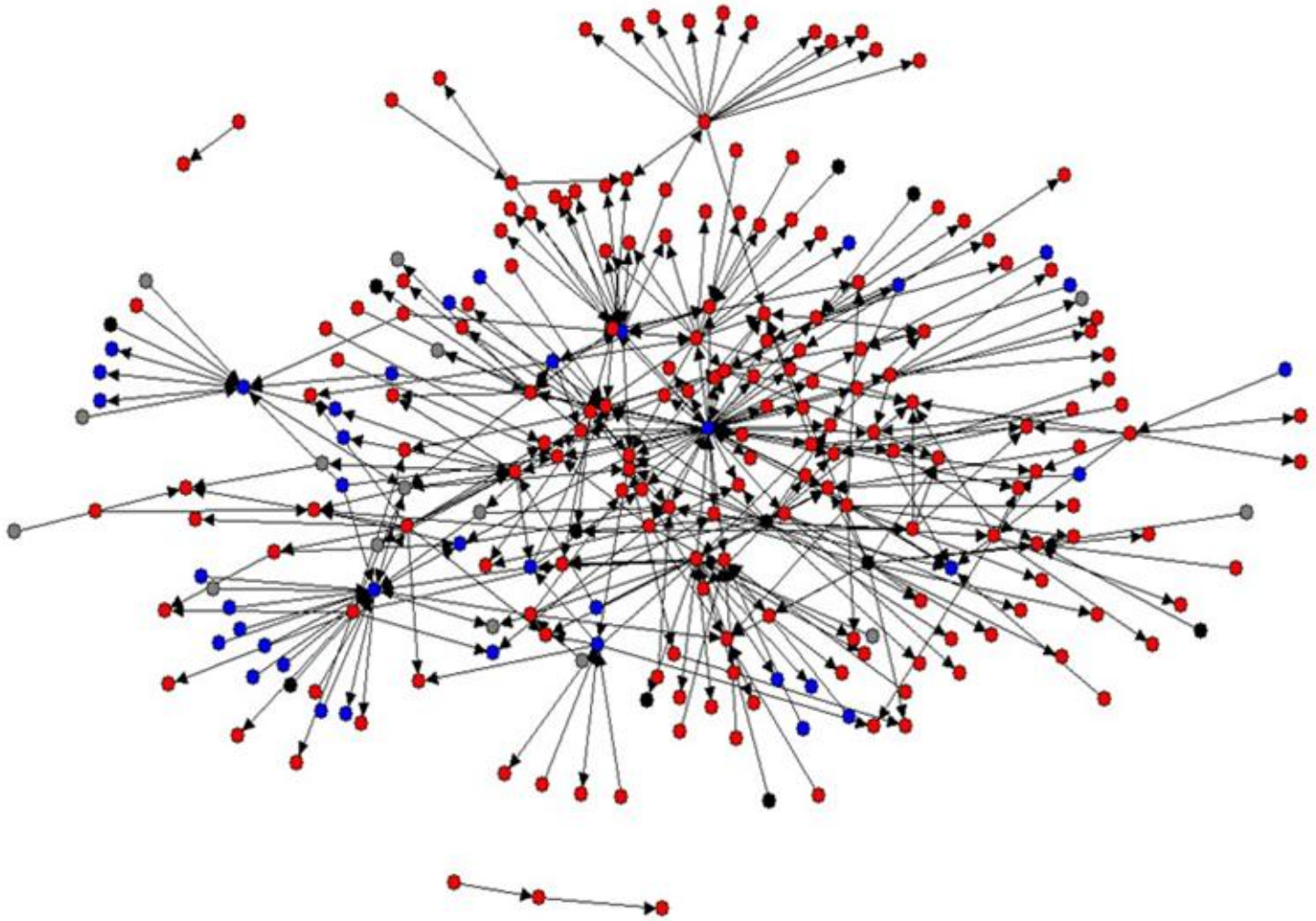
- Finding industrial networks using the web (2)
  - Base URLs taken from a subsea engineering grouping identified from earlier work i.e. organisations with the word ‘subsea’ anywhere on their website
  - Connectivity only within the N.E. region
  - Only looked at linkages and direction of knowledge or trades, no strength inferred
  - Some companies had large numbers of in-links therefore some selective filtering was applied
  - Other quality checks applied manually
  - 3 types of links merged into single matrix

# Figure 1 – Subsea network, circular representation

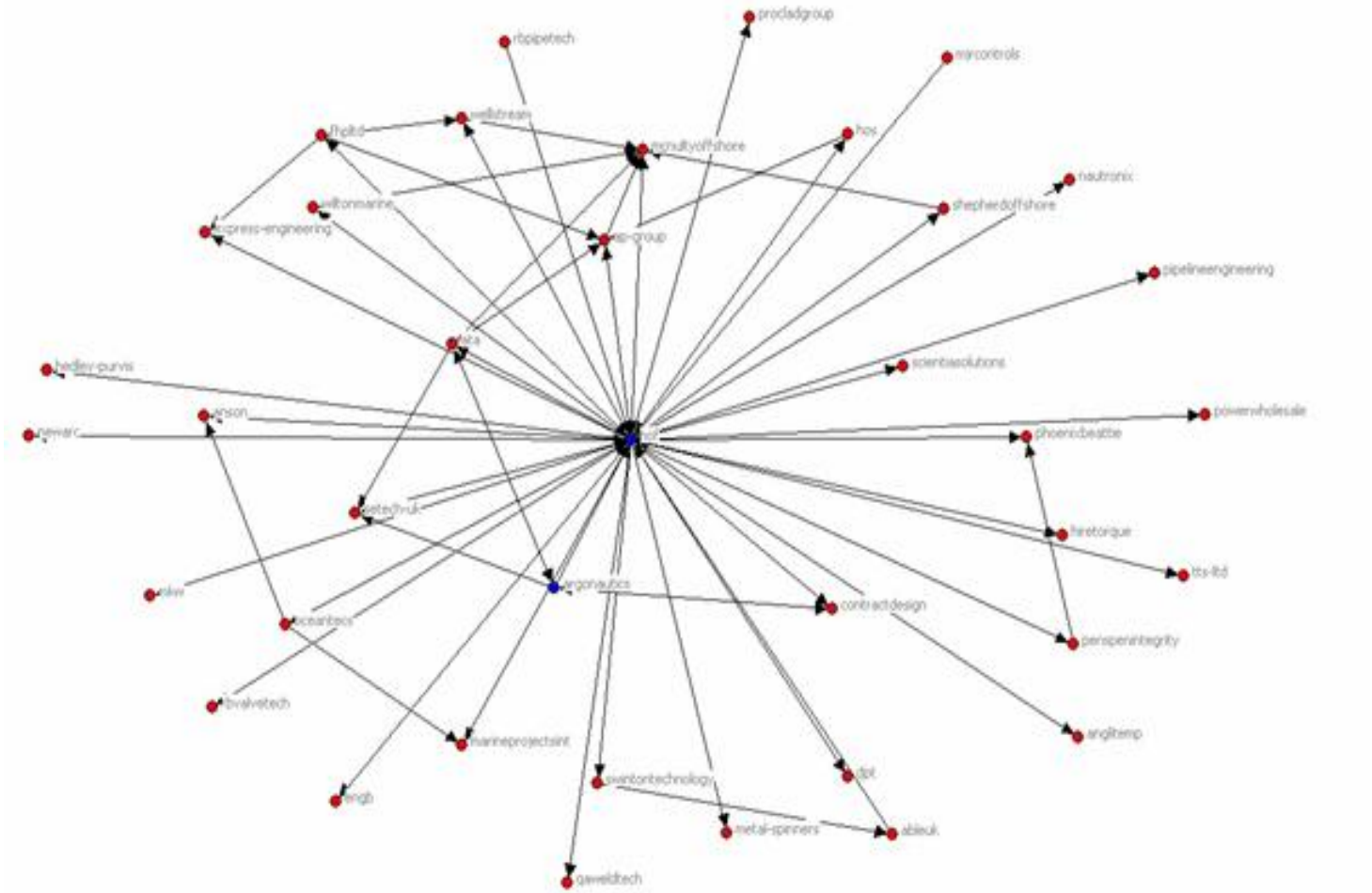


- Analysis of linkages
  - Simple matrix output analysed using conventional graphing methods – circular representation (Figure 1)
  - Linear network (Figure 2)
  - Sample egonet (Figure 3)
  - Metrics
    - 282 nodes
    - 394 intraregional links
    - Average geodesic distance 4.306
    - Overall clustering co-efficient 0.059

Figure 2. Subsea Network (isolates removed)



# Figure 3. Egonet for a trade association in Oil & Gas sector



- Concluding remarks (1)
  - Comparison made with nets derived from the web and one by snowball sampling using the same firm population
    - About 50% match although more links found by web based means
  - Method works well for finding:
    - Specialised groups of firms by keyword searches
    - Linkages in the absence of co-operation from individual companies
    - Key players and gatekeepers of knowledge

- Concluding remarks (2)
  - Problem areas/more research required
    - Acquisition of large numbers of seed URLs
      - The presence of directories can confuse and frustrate attempts to automate this process
    - Dependency on the quality of individual websites as regards basic firm information and the number of linkages to or from the site
    - Very large sites take a long time to spider and may not be appropriate
    - Problem of definition  of the firm or other contributing or supporting economic entity
    - The ‘nature’ of links

# Thanks for your attention

jrw@fabriam.com
J.Williams@ncl.ac.uk