Knowledge Construction for Science Jobs: Employing Network Analysis to Identify the Scaffold of Competencies

George Neofotistos

University of Crete, Heraklion, Greece

Nikitas Kastis

Lambrakis Foundation, Athens, Greece, and

MENON Network EEIG, Brussels, Belgium

Abstract

How can we optimize knowledge uptake by exploring knowledge networks (knowledge spaces)? Such an optimization can significantly affect the work-related skills of the labor force yielding a better match to the requirements of the jobs, since every occupation requires a different mix of knowledge, skills, and abilities.

In this paper, we identify critical competencies for a specific job family by analyzing the characteristics of the jobs-knowledge bipartite weighted network for the 48 jobs and the 33 knowledge domains comprising the "Life, Physical and Social Sciences" job family of the O*NET (Occupational Information Network) database, USA's primary source of occupational information.

We calculate centrality measures (degree centrality, betweenness centrality, eigenvector centrality) of the knowledge domains in this job-knowledge network, in order to identify the critical knowledge domains (according to the O*NET typology). Knowledge domains rankings are compared and explained. For example, using a degree centrality measure, English language, mathematics, computers and electronics, education and training, customer and personal service, administration and management, law and government, clerical, biology and communications and media, comprise the top-ten knowledge domains; only biology qualifies from science! This is due to the fact that knowledge domains such as communications and media, or clerical, appear consistently at the middle of each job's rankings (but not at the top).

On the other hand, using betweenness centrality (a measure considered to be important in characterizing "transport" in networks) and identifying nodes in the infinite incipient percolation cluster (a cluster of high betweenness centrality nodes, which can be interpreted as the "superhighways" of the network), English language, biology, customer and personal service, education and training, physics, chemistry and mathematics appear in top; a ranking, which closely conforms onto a typical school education or university's knowledge domain structure (revealing however competencies not explicitly covered, such as customer and personal service).

Our findings can contribute to a better understanding of knowledge construction paradigms attuned to a) specific job families, b) what types of knowledge competencies are important and should be taught at the secondary, tertiary and life-long learning education levels, and c) help identify knowledge needs in the workplace, which are not covered by the "knowledge superhighways" and can be offered by e-learning.

Keywords: Knowledge building paradigms, knowledge networks, key competencies, network analysis, centrality measures

1. Introduction

There is a widespread belief that workers' skills and education are not adequate for the demands of jobs in the current economy (Handel 2003). Journalistic reports, employer surveys, popular and policy debates on school quality and education reform, sociological writings on the economy, and economic accounts of the recent growth of wage inequality all suggest a mismatch between the skills workers possess and what jobs require, what economists call an imbalance between the supply of and demand for human capital. Many believe that the problems will become even more serious because the pace of change is accelerating and the workplace is becoming increasingly high tech, service-oriented, and reorganized to involve greater employee participation in the workplace (Bresnahan et al 2002; see also Smith 1997).

In this paper, we investigate the characteristics of a specific job-knowledge network (knowledge space) with the objective to improve the uptake of knowledge toward the requirements of science jobs. In particular, we identify critical knowledge competencies by calculating centrality measures [degree centrality, betweenness centrality, and eigenvector centrality of the knowledge nodes (domains)] of the bipartite weighted network for the 48 jobs and the 33 knowledge domains comprising the “Life, Physical and Social Sciences” job family of the O*NET (Occupational Information Network) database, USA's primary source of occupational information.

The paper is organized as follows. Section 2 presents the O*NET database and the particular dataset we have used. Section 3 presents the methodological approach, provides the definitions of the different centrality measures of the network nodes. In Section 4, results are presented. Finally, Section 5 ends the paper with discussion and conclusions.

2. The O*NET Database

The Occupational Information Network (O*NET) database contains information on standardized and occupation-specific descriptors, and is continually updated by surveying a broad range of workers from each occupation. Based on the Standard Occupational Classification (SOC), the O*NET-SOC taxonomy includes 812 occupations which currently have, or are scheduled to have, data collected from job incumbents or occupation experts. The most recent O*NET-SOC 2006 taxonomy includes 949 occupational titles, 812 of which represent data-level occupations. The O*NET Program is collecting and disseminating updated data for the 812 data-level occupations. Data are gathered on approximately 200 occupations each year, with the goal of replenishing the database every five years.

The O*NET jobs-knowledge network is a weighted bipartite network. A bipartite network has two kinds of nodes, say, J (denoted as such for jobs) and K (denoted as such for knowledge), in which there are only links between two nodes of different kinds. Table 1 presents the weights of the links between knowledge domains and selected jobs from the O*NET “Life, Physical, and Social Science” job family. For each job, the workers surveyed have graded the 33 knowledge domains with respect to the requirements of their particular job.

Table 1: Weights of the links between Knowledge and selected Jobs comprising the O*NET "Life, Physical, and Social Science" Job Family

KNOWLEDGE \ JOBS	Anthropologists	Archeologists	Astronomers	Atmospheric and Space Scientists	Biologists
Administration and Management	51	56	30	26	53
Biology	46	38	15	7	98
Building and Construction	9	20	4	3	30
Chemistry	14	27	48	20	63
Clerical	55	54	15	32	55
Communications and Media	60	49	31	60	48
Computers and Electronics	46	47	74	76	61
Customer and Personal Service	48	26	12	80	60
Design	14	35	26	10	36
Economics and Accounting	29	29	9	8	18
Education and Training	75	60	52	57	52
Engineering and Technology	10	16	60	22	55
English Language	90	86	80	81	62
Fine Arts	28	18	2	3	0
Food Production	18	5	0	2	1
Foreign Language	74	44	20	8	8
Geography	65	76	16	84	56
History and Archeology	84	99	8	13	26
Law and Government	55	49	11	24	77
Mathematics	57	57	95	78	57
Mechanical	9	24	23	4	36
Medicine and Dentistry	32	15	1	3	30
Personnel and Human Resources	44	40	19	22	24
Philosophy and Theology	65	45	10	4	6
Physics	6	18	99	82	47
Production and Processing	22	12	6	34	13
Psychology	69	23	9	17	28
Public Safety and Security	26	26	9	33	59
Sales and Marketing	22	17	11	18	12
Sociology and Anthropology	98	97	6	12	1
Telecommunications	21	14	24	41	23
Therapy and Counseling	25	6	3	6	6
Transportation	19	24	7	16	35

3. Methodological approach

3.1 Network Centrality Measures

In large complex networks, not all nodes are equivalent (Barabasi 2002; Strogatz 2001). Centrality measures address the question, "Which is the most important or central node in this network?" The simplest of centrality measures is the *degree centrality*, also called simply *degree*. The degree of a node in a network is the number of links attached to it. However, degree centrality is a local quantity, which does not inform about the overall importance of the node in the network. A more sophisticated centrality measure is the *eigenvector centrality*. Where degree centrality gives a simple count of the number of links a node has, the eigenvector centrality accords each node a centrality that depends both on the number as well as the quality of its links (that is, the centrality of the nodes with which it is connected).

However, in terms of transport (that is, paths in a network), a good measure of the centrality of a node has to incorporate more global information such as its role played in the existence of paths between any two given nodes in the network. The *betweenness centrality* (BC) is the number of times a node is used by the set of all shortest paths between all pairs of nodes (Barthelemy 2004). High values of the betweenness centrality indicate that a node can reach others on short paths. If one removes a node with large centrality it will lengthen the paths between many pairs of nodes. For simplicity we call the "betweenness centrality" here "centrality". and we use the notation "nodes" but similar results have been obtained for links. This centrality measure, BC, quantifies the "importance" of a node for transport in the network. Identifying the nodes with high BC enables to improve their transport capacity and thus improve the global transport in the network.

3.2 Identifying high centrality nodes (the knowledge "superhighways")

Transport in weighted networks is dominated by the minimum spanning tree (MST), the tree connecting all nodes with the minimum total weight. The MST can be partitioned into two distinct components, having significantly different transport properties, characterized by betweenness centrality. One component, the "superhighways", is the *infinite incipient percolation cluster* (IIC), for which nodes with high betwenness centrality dominate (Wu et al, 2006). For the other component, that is "roads", which includes the remaining nodes, low centrality nodes dominate. The distribution of the centrality for the infinite incipient percolation cluster satisfies a power law, with an exponent smaller than that for the entire MST; the global transport can be enhanced significantly by improving the small fraction of the network, the superhighways.

To identify the IIC of the network, we start with the fully connected network and remove links in ascending order of their weights. After each removal of a link, we calculate

$$\kappa \equiv \langle k^2 \rangle / \langle k \rangle,$$

(where $\langle k \rangle$ is the average degree –that is, the average number of links- and $\langle k^2 \rangle$ is the average squared degree), which decreases with link removals. When $\kappa < 2$, we stop the process because at this point, the largest remaining component is the IIC (Wu et al, 2006), the knowledge nodes, which comprise the "superhighways" in the network.

4. Results

The second column of Table 2 presents the rank of the knowledge domains with respect to the total sum of weights, for the entire O*NET "Life, Physical, and Social Science Job Family". This is the "view from the workplace". As can be seen, the top 10 ranked knowledge domains are: English Language; Mathematics; Computers and Electronics; Education and Training; Customer and Personal Service; Administration and Management; Law and Government; Clerical; Biology; Communications and Media. Only Biology represents the "sciences" in this top-10 rank.

Such a rank seems to provide support to the widespread belief that workers' education is not adequate for the demands of jobs in the current economy (journalistic reports, employer surveys, popular and policy debates on school quality and education reform, all seem to suggest a mismatch between the skills workers possess and what jobs require, what economists call an imbalance between the supply of and demand for human capital). For example, it seems that universities do not adequately supply the "right" knowledge to the scientists at the workplace.

However, this is not actually true. The higher rank of "non-scientific" knowledge domains such as communications and media, or clerical, is due to the fact that they appear consistently at the middle of each job's rankings (*but not at the top*). The "pure scientific" point of view emerges when we calculate the centrality of the knowledge domains (nodes) in the network and identify the IIC, that is, the "knowledge superhighways" in the network.

The third, fourth, fifth, sixth and seventh columns of Table 2 present the total sum of weights for each knowledge domain but in ascending order of the weights. For example, the third column presents the total sum of weights above 50 (W>49) [that is, after removing links with weights with values less than 49]. The subsequent columns present the remaining sum of weights after removing, successively, weights with values less than 59, 69, 79, and 87, respectively.

As can be seen in Table 2, as weights are removed in ascending order, the rank of the knowledge domains changes. "Pure scientific" knowledge domains appear more prominent; their centrality changes. To manifest this effect more clearly, we calculate centrality measures (degree centrality, betweenness centrality, eigenvector centrality) of the knowledge nodes in this job-knowledge network, in order to identify the critical knowledge domains (according to the O*NET typology).

Table 3 presents the values of the degree centrality, the betweeness centrality and eigenvector centrality, of the knowledge nodes, for the W>79 case. As can be seen, the rank of "pure scientific" knowledge domains become prominent. Biology, physics and psychology enter the top-10 rank (which consists of only 8 non-zero values in the betweenness centrality measure).

The important question now becomes which ascending order of the weights to choose in order to calculate the "true" centralities. The answer to this question is provided by identifying the IIC of the network, a process which starts with the fully connected network, removing links in ascending order of their weights, and, after each removal of a link, calculating the quantity $\kappa \equiv \langle k^2 \rangle / \langle k \rangle$, which decreases with link removals. When $\kappa < 2$, we stop the process because at this point, the largest remaining component is the IIC, the knowledge nodes, which comprise the "superhighways" in the network.

Research findings from a companion paper (Neofotistos, 2007) on the identification of the IIC for the same O*NET jobs-knowledge network has demonstrated that W>87 provides the tipping point and, for this case, knowledge competencies comprising the "superhighways" of the jobs-knowledge network are: a) English language, b) "pure sciences" such as biology, chemistry, physics, geography, history and archeology, psychology, sociology and anthropology, and c) "general competencies" such as administration and management, customer and personal service, education and training, mathematics, law and government, personnel and human resources, and therapy and counseling.

The above-mentioned knowledge competencies comprise a cluster of high betweenness centrality nodes, which can be interpreted as the "superhighways" of the jobs-knowledge network, which closely conform onto a school-education and university's knowledge-domain structure (revealing however competencies not explicitly covered, such as customer and personal service).

5. Discussion

Our findings can contribute to better understanding of knowledge construction paradigms attuned to specific job families, b) key knowledge competencies (knowledge "superhighways"), which should be focused upon at the secondary, tertiary and life-long learning education levels (English language, "pure sciences", general competencies) and c) knowledge competencies, which can be interpreted as "roads" leading to the specific (life, physical and social science) jobs. Our methodological approach can systematically monitor the "coupling" between education systems and the evolution in the workplace (whether -and how- workers' skills and education are, or are not, adequate for the demands of jobs in the current economy, a problem which many believe will become even more serious because the pace of change is accelerating and the workplace is becoming increasingly high tech, service-oriented, and reorganized to involve greater employee participation).

Acknowledgements

We would like to thank Nikos Neofotistos for compiling the data and Chrisostomos Maraggoulis for the computational aspect of the work and the calculation of centralities.

References

Barabasi A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*. (74):47-97.

Barthelemy M. (2004). Betweeness centrality in large complex networks. *The European Physical Journal B*. 38: 163-168.

Bresnahan T. F., Brynjolfsson E, and Hitt L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. The *Quarterly Journal of Economics*. 117:339-76.

Cunha F., Heckman J. J., Lochner L., and Masterov D. V. (2005). Interpreting the Evidence on Life Cycle Skill Formation. *IZA Discussion Paper* No. 1675. Available at SSRN: http://ssrn.com/abstract=766744.

Handel M. J. (2003). Skills mismatch in the labor market. *Annual Review of Sociology*. 29:135-165.

Neofotistos G. (2007). Which is the most important knowledge domain for science jobs? A network analysis of the O*NET database. Technical Report. TR-01-02-2007. University of Crete.

O*NET (Occupational Information Network). Available at http://www.onetcenter.org/overview.html.

Smith V. (1997). New forms of work organization. *Annual Review of Sociology*. 23:315-39.

Strogatz S.H. (2001). Exploring complex networks. *Nature.* 410(6825):268–276.

Wu Z., Braunstein L. A., Havlin S., and Stanley H. E. (2006). Transport in weighted networks: partition into superhighways and roads. *Physical Review Letters.* 96(14):148702-1–148702-4.

Table 2: Knowledge categories' total sum of weights and ascending order of weights, for the Jobs-Knowledge network of the O*NET "Life, Physical, and Social Science" Job Family

KNOWLEDGE	TOTAL SUM OF WEIGHTS	W>49	W>59	W>69	W>79	W>87
Administration and Management	2372	1239	697	246	176	94
Biology	1931	1262	1103	830	610	280
Building and Construction	822	247	137	0	0	0
Chemistry	1862	1304	1027	636	190	190
Clerical	2068	922	331	70	0	0
Communications and Media	1912	719	395	74	0	0
Computers and Electronics	2471	1635	808	295	0	0
Customer and Personal Service	2383	1686	1264	627	262	182
Design	1278	323	156	156	80	0
Economics and Accounting	1134	246	86	86	86	0
Education and Training	2436	1502	1057	540	173	88
Engineering and Technology	1692	813	487	301	0	0
English Language	3424	3291	3183	2537	1429	268
Fine Arts	254	0	0	0	0	0
Food Production	636	377	216	153	81	0
Foreign Language	763	74	74	74	0	0
Geography	1798	1250	928	549	182	98
History and Archeology	1186	514	457	327	183	99
Law and Government	2146	1302	872	615	174	94
Mathematics	2909	2518	1691	1229	341	95
Mechanical	1192	277	62	0	0	0
Medicine and Dentistry	874	274	221	221	0	0
Personnel and Human Resources	1532	206	155	95	95	95
Philosophy and Theology	924	234	126	0	0	0
Physics	1587	894	586	586	363	199
Production and Processing	1254	310	202	79	0	0
Psychology	1715	786	627	438	291	291
Public Safety and Security	1662	544	275	77	0	0
Sales and Marketing	1147	249	137	73	0	0
Sociology and Anthropology	1457	842	576	514	295	295
Telecommunications	991	0	0	0	0	0
Therapy and Counseling	732	297	297	297	297	297
Transportation	957	52	0	0	0	0

Table 3: Centrality measures for W>79

Degree Centrality Measure (W>79)		
RANK	KNOWLEDGE	Degree
1	English Language	0,43750
2	Customer and Personal Service	0,25000
3	Education and Training	0,18750
4	Psychology	0,18750
5	Administration and Management	0,15625
6	Personnel and Human Resources	0,15625
7	Biology	0,12500
8	Physics	0,12500
9	Geography	0,09375
10	Mathematics	0,09375
11	Therapy and Counseling	0,09375
12	Economics and Accounting	0,06250
13	History and Archeology	0,06250
14	Sociology and Anthropology	0,06250
15	Design	0,03125
16	Food Production	0,03125
17	Law and Government	0,03125
18	Building and Construction	0,00000
19	Chemistry	0,00000
20	Clerical	0,00000
21	Communications and Media	0,00000
22	Computers and Electronics	0,00000
23	Engineering and Technology	0,00000
24	Fine Arts	0,00000
25	Foreign Language	0,00000
26	Mechanical	0,00000
27	Medicine and Dentistry	0,00000
28	Philosophy and Theology	0,00000
29	Production and Processing	0,00000
30	28. Public Safety and Security	0,00000
31	29. Sales and Marketing	0,00000
32	31. Telecommunications	0,00000
33	33. Transportation	0,00000

Betweenness Centrality Measure (W>79)		
RANK	KNOWLEDGE	Betweenness
1	English Language	0,16028
2	Biology	0,05847
3	Customer and Personal Service	0,02621
4	Education and Training	0,00202
5	Physics	0,00202
6	Psychology	0,00202
7	Mathematics	0,00101
8	Administration and Management	0
9	Building and Construction	0
10	Chemistry	0
11	Clerical	0
12	Communications and Media	0
13	Computers and Electronics	0
14	Design	0
15	Economics and Accounting	0
16	Engineering and Technology	0
17	Fine Arts	0
18	Food Production	0
19	Foreign Language	0
20	Geography	0
21	History and Archeology	0
22	Law and Government	0
23	Mechanical	0
24	Medicine and Dentistry	0
25	Personnel and Human Resources	0
26	Philosophy and Theology	0
27	Production and Processing	0
28	Public Safety and Security	0
29	Sales and Marketing	0
30	Sociology and Anthropology	0
31	Telecommunications	0
32	Therapy and Counseling	0
33	Transportation	0

Eigenvector Centrality Measure (W>79)		
RANK	KNOWLEDGE	Eigenvector
1	English Language	0,4883
2	Customer and Personal Service	0,3872
3	Education and Training	0,3390
4	Psychology	0,3390
5	Administration and Management	0,3111
6	Personnel and Human Resources	0,3111
7	Physics	0,1980
8	Therapy and Counseling	0,1946
9	Geography	0,1791
10	Biology	0,1547
11	Mathematics	0,1318
12	Economics and Accounting	0,1035
13	History and Archeology	0,0978
14	Sociology and Anthropology	0,0978
15	Law and Government	0,0815
16	Design	0,0258
17	Food Production	0,0258
18	Building and Construction	0,0000
19	Chemistry	0,0000
20	Clerical	0,0000
21	Communications and Media	0,0000
22	Computers and Electronics	0,0000
23	Engineering and Technology	0,0000
24	Fine Arts	0,0000
25	Foreign Language	0,0000
26	Mechanical	0,0000
27	Medicine and Dentistry	0,0000
28	Philosophy and Theology	0,0000
29	Production and Processing	0,0000
30	Public Safety and Security	0,0000
31	Sales and Marketing	0,0000
32	Telecommunications	0,0000
33	Transportation	0,0000