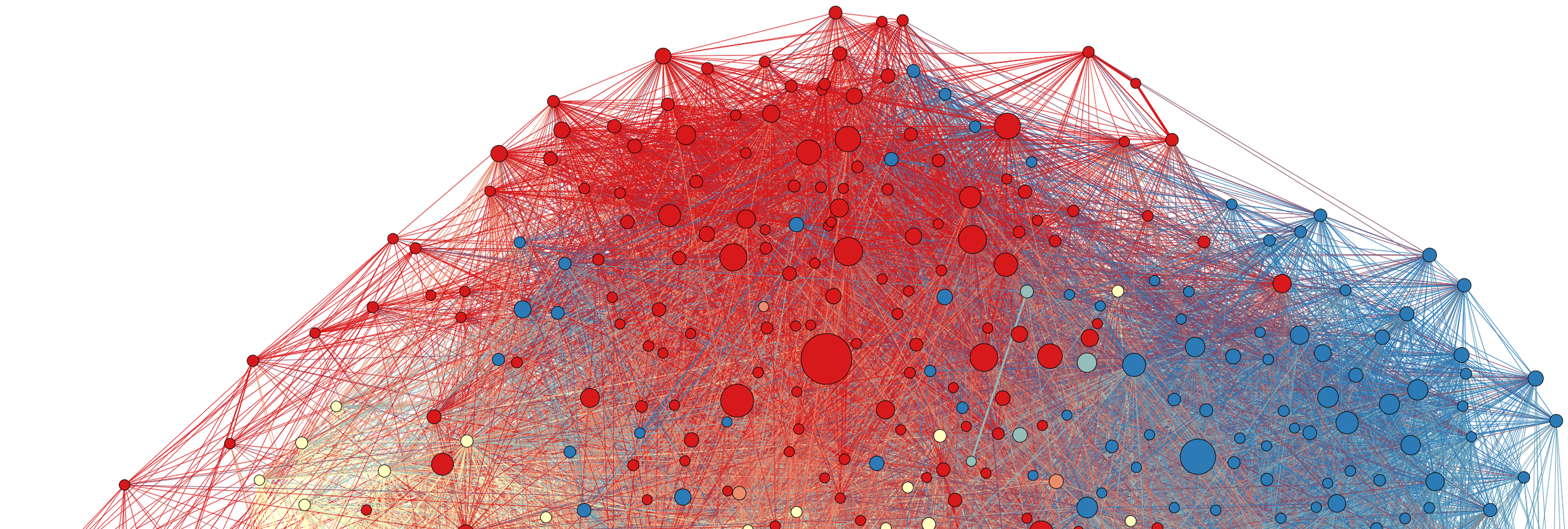
Writers, Designers, and Software Developers: Network-Based Measures of Worker Skills

Kate Anderson



Why Think about Skills?

The workplace is changing:

- Automation is affecting entirely new sectors of the workforce:
    - Accountants
    - Librarians
    - Office Workers
    - Travel Agents
    - Customer Service
    - Programmers
    - Health Care
    - Lawyers

Since 1990, library and administrative jobs have declined by 40%

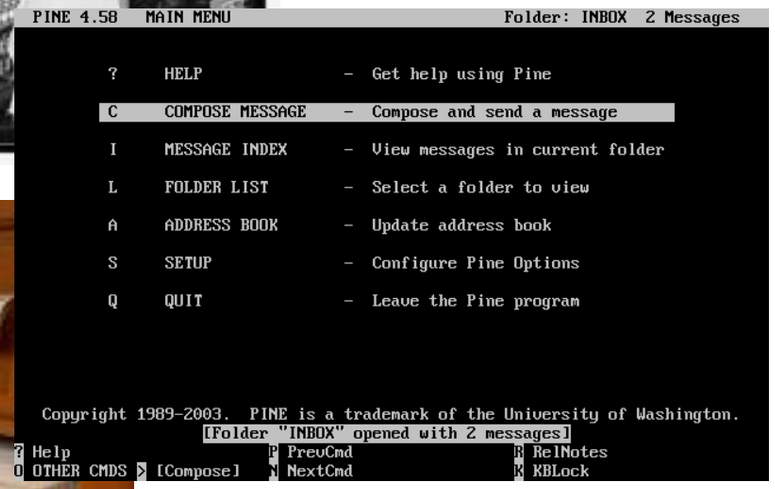
This is a loss of jobs on the scale of the steel industry, but much less visible!

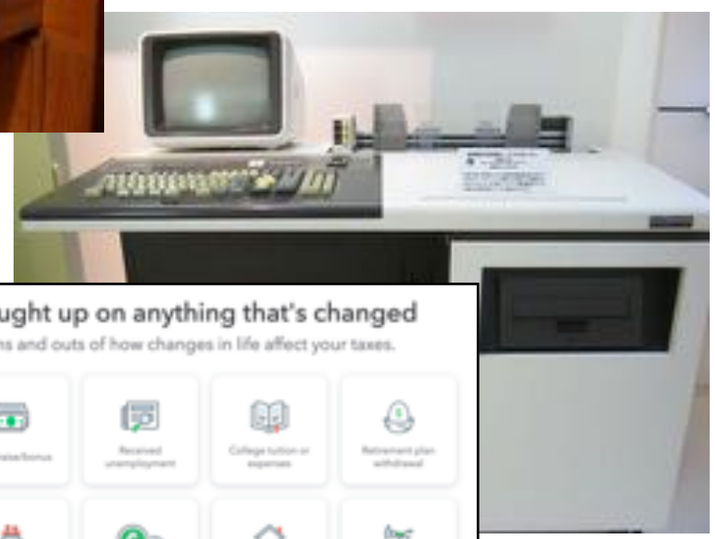
The Changing Labor Market

The labor market is constantly adjusting to new technology.

People retrain into jobs that use some or all of their skills

- Librarian → CS, Teaching, Communications, IT
- Admin → Marketing
- Travel Agent → Customer Service





As more jobs are changed or eliminated, this becomes more pressing

What will young lawyers retrain to do?

The Changing Labor Market

That raises important questions:

- How do skills translate into wages?
- Which combinations of skills are endangered?
    - Which of those skills are transferable to other occupations?
    - What skills should workers acquire to bolster their job prospects in the AI economy?
- How can the strengths of AI be used to *complement* human skills?

Economic Models of Skills and Production

Economists usually think about skills in terms of production:

Input
(skills + effort)

Output
(production)

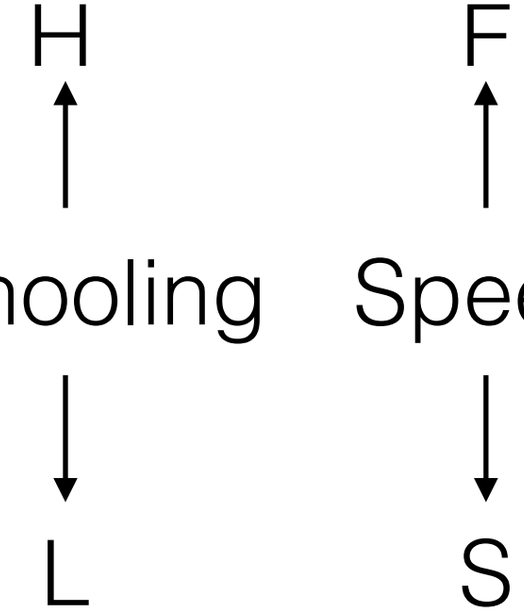
The value that people bring to production is called *human capital*

Measures of Human Capital

Traditional notions of human capital are motivated by manufacturing

One-dimensional skill level:

- Mankiw, Romer, and Weil (1992):
- Gibbons and Katz (1991)



Two types/occupations:

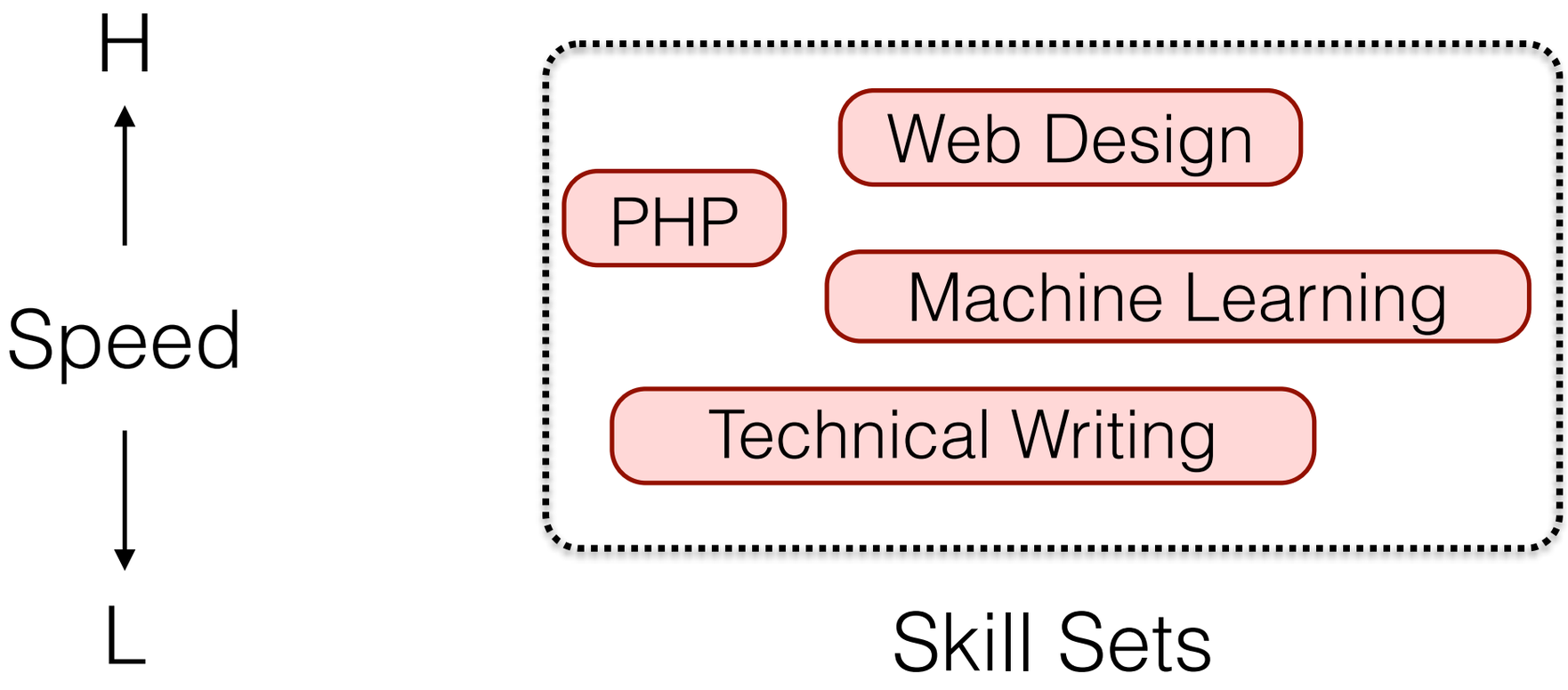
- Roy Model (1951)

hunter

fisherman

Skills and Production

But human capital is much different in knowledge-based production: skill *combinations* become more important



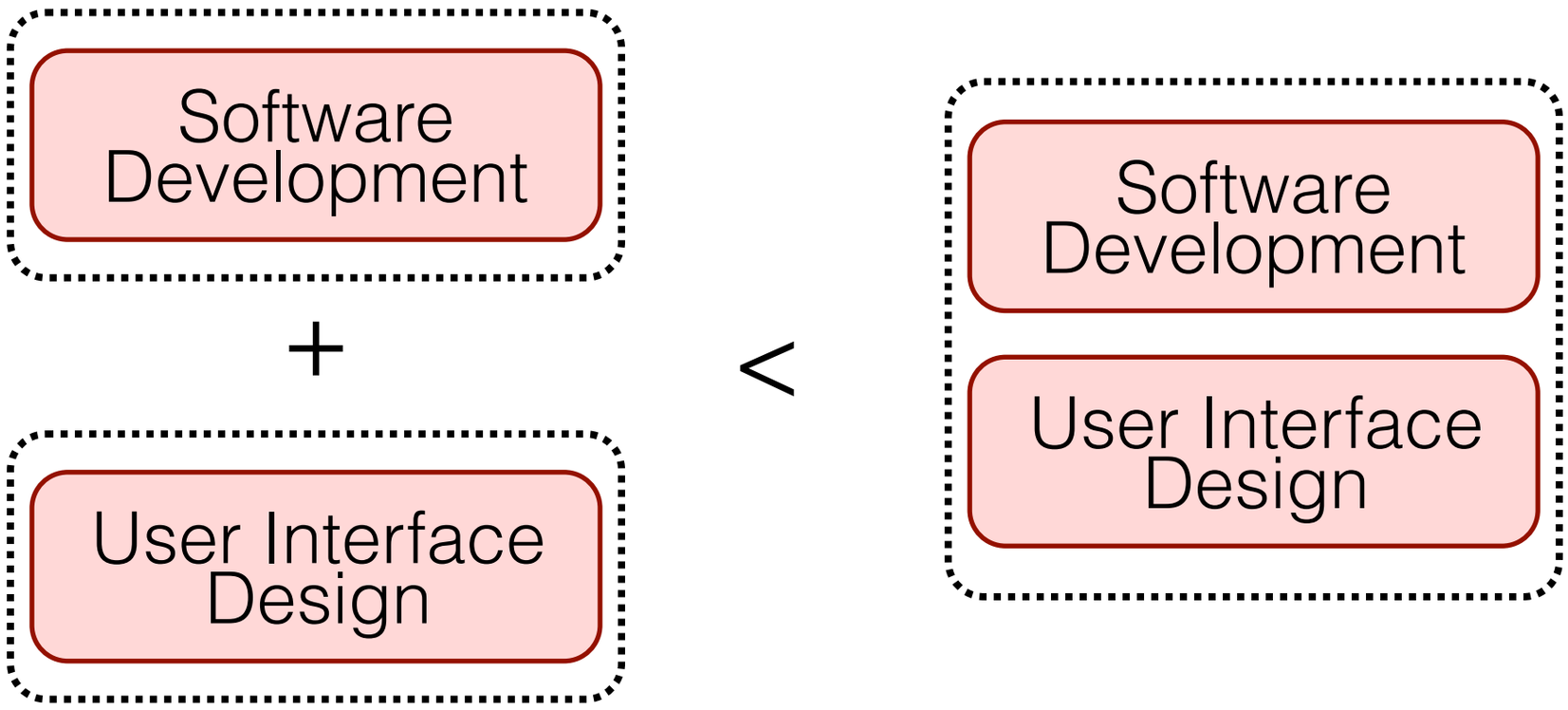
Multi-dimensional Human Capital

Interest in a more nuanced view of human capital:

- Leung and Ng (2014), Leung (2014): boundary spanning between job categories
- Heckman et al (2006): Cognitive and Non-cognitive skills
- Lazear (2005): breadth of experience by number of roles in employment history
- Adamic et al (2010): focus, with skills = topics
- Autor and Handel (2009): O*NET categorizing occupations by required skills
- Psychology and OBT literatures: team-level diversity related to performance/ productivity with skills = assessment-based measures

Complex Human Capital

A crucial factor in multi-dimensional human capital is that some skills work together: they are more valuable together than they are apart

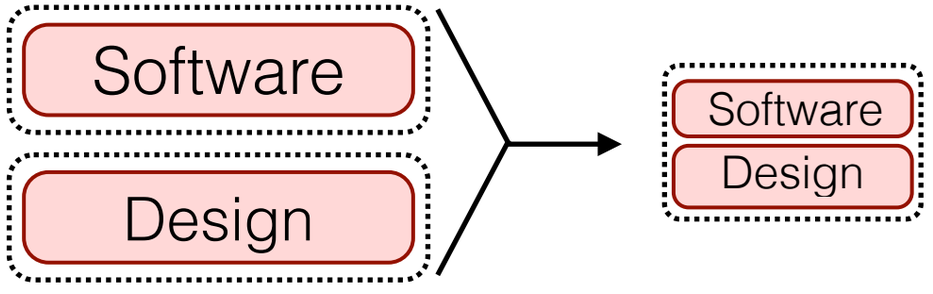


These synergies between skills often distinguish workers in the labor market

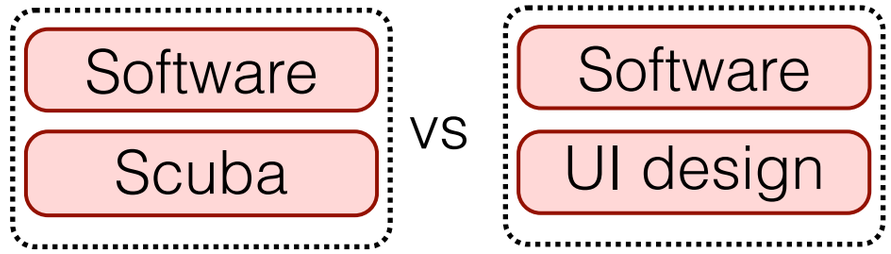
Complex Human Capital

But this makes determining the potential value of a worker *much* more complicated

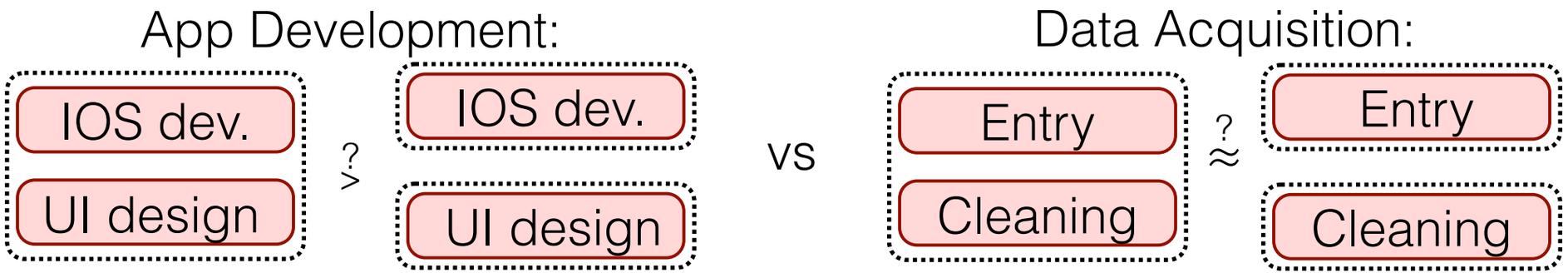
- Skills that are common individually may be rare in combination



- Some skills are more valuable in combination and others are not



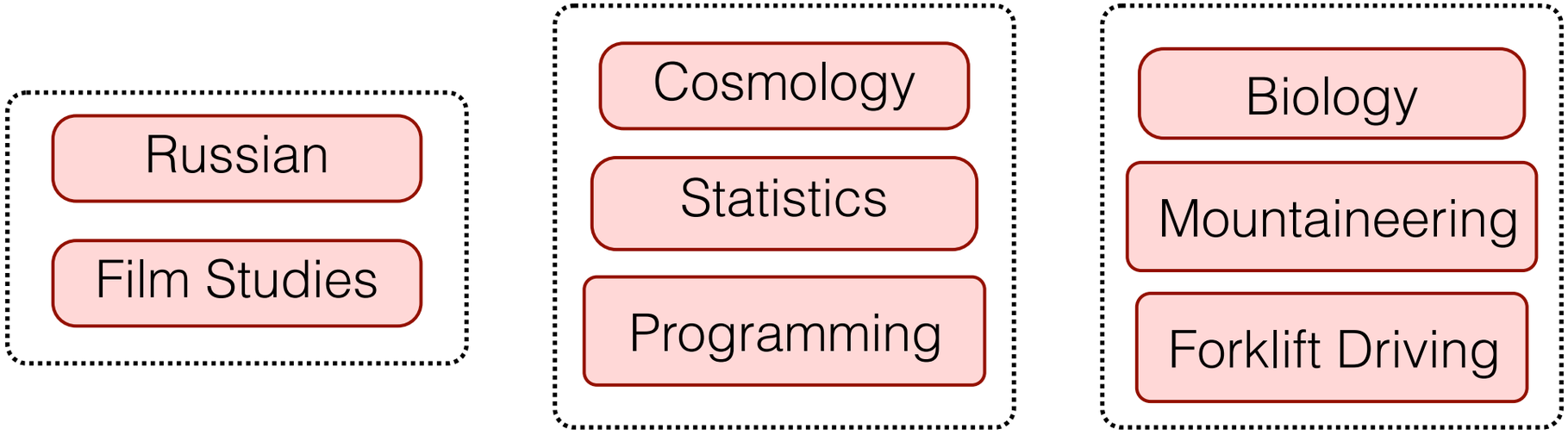
- Some tasks are more “separable” than others



Skill Sets and Productivity

So we cant assign prices to individual skills

- The value of adding a skill to a workers skill set depends on the skills she already has
- The value of a combination may be more than the sum of its parts



Complex Human Capital

So does this make a difference in the labor market?

Yes!

Both in theory and in practice!

Wages with Skill Combinations

A labor market:

Workers $I = \{1,2,3...N\}$

Each endowed with a skill set, A_i drawn from $S = \{s_1 \ldots s_K\}$

Skill sets distributed according to $\delta : \mathscr{P}(S) \rightarrow [0,1]$

Workers compete for jobs: $\mu = \{\mu_1 \ldots \mu_M\}$

A job requires skills $\mu_j \subseteq S$

Job requirements distributed according to $\sigma : \mathscr{P}(S) \rightarrow [0,1]$

With some additional assumptions:

$$E\left[w(A)\right] = \sum_{C \subseteq A} \frac{\sigma(C)}{\sum_{D \subseteq S} \delta(C \cup D)}$$

Wages with Skill Combinations

$$E\left[w(\{a\})\right] = \frac{\sigma(\{a\})}{\delta(\{a\})}$$

$$E\left[w(\{a,b\})\right] = \frac{\sigma(\{a\})}{\delta(\{a\})} + \frac{\sigma(\{b\})}{\delta(\{b\})} + \frac{\sigma(\{a,b\})}{\delta(\{a,b\})}$$

When there is no demand for skills in combination, wages are additive $E\left[w(A \cup B)\right] = E\left[w(A)\right] + E\left[w(B)\right]$

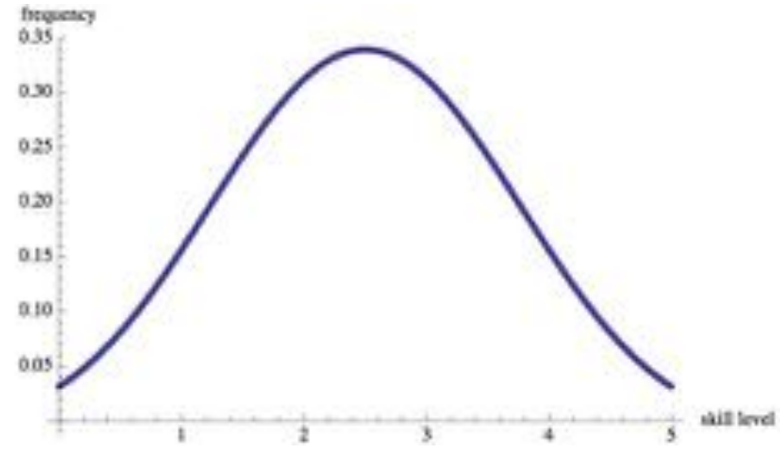
When there is, wages are super additive:

$$E\left[w(A \cup B)\right] > E\left[w(A)\right] + E\left[w(B)\right]$$

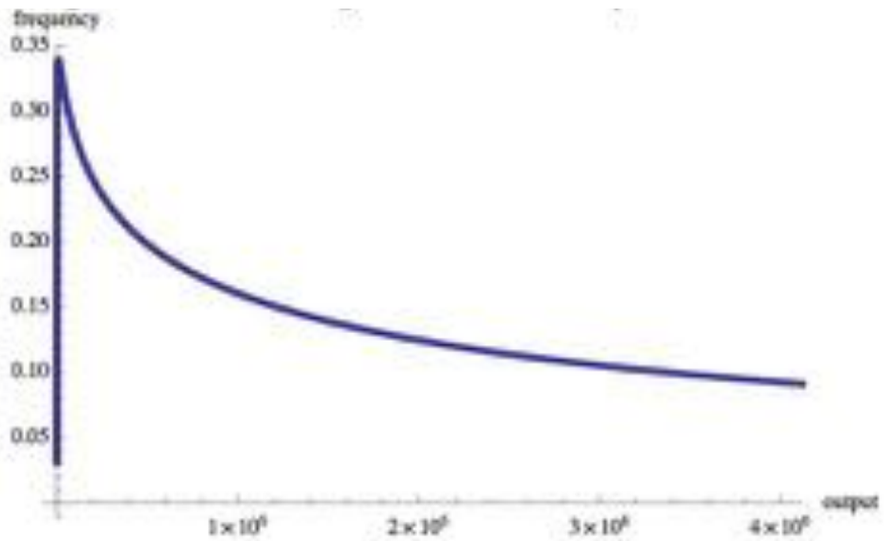
Wages with Skill Combinations

Suppose skills are distributed to workers independently with equal probability

When skills are only used individually, wages mirror the distribution of skills



When jobs require combinations of skills, skill differences are exaggerated



Considering skill combinations should help us understand more variation in wages

Wages with Skill Combinations

But we also cant think about all possible interactions—
the problem is too big!

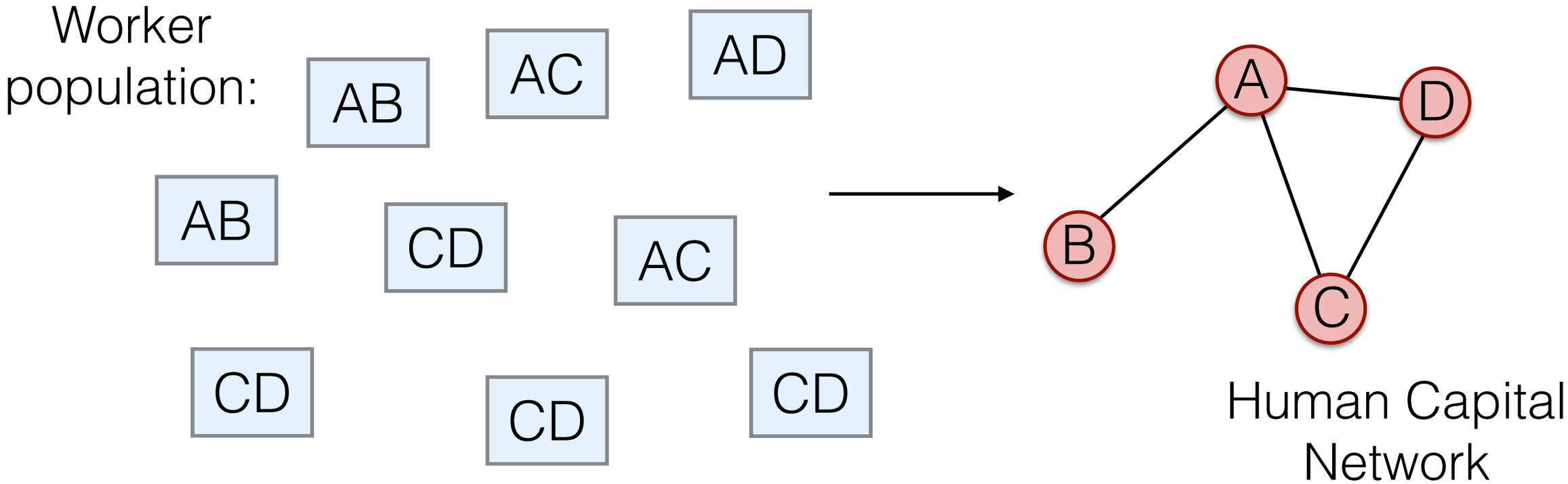
So we need to find different ways to measure and assign
value to human capital

Human Capital Networks

Proposal: represent the interactions between worker skills using a network

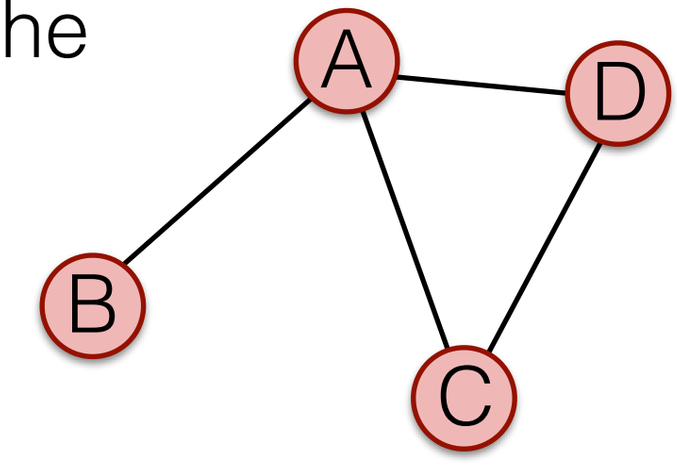
Nodes = skills

Two nodes are linked if at least one worker has both skills



Human Capital Networks

We want to weight links to reflect how often the skills co-occur



Some possible weighting schemes:

Counts: $w_{AB} = n_{AB}$

Jaccard Similarity Index: $w_{AB} = \frac{A \cap B}{A \cup B} = \frac{n_{AB}}{n_A + n_B - n_A B}$

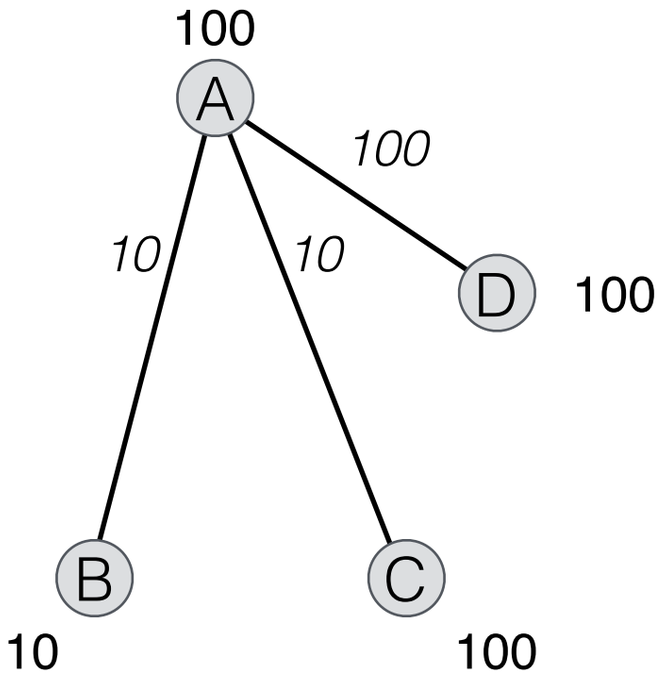
Conditional Probability: $w_{AB} = \frac{n_{AB}}{n_A}$ $w_{BA} = \frac{n_{AB}}{n_B}$

Modified Conditional Probability: $w_{BA} = \frac{n_{AB}}{min\left(n_A, n_B\right)}$

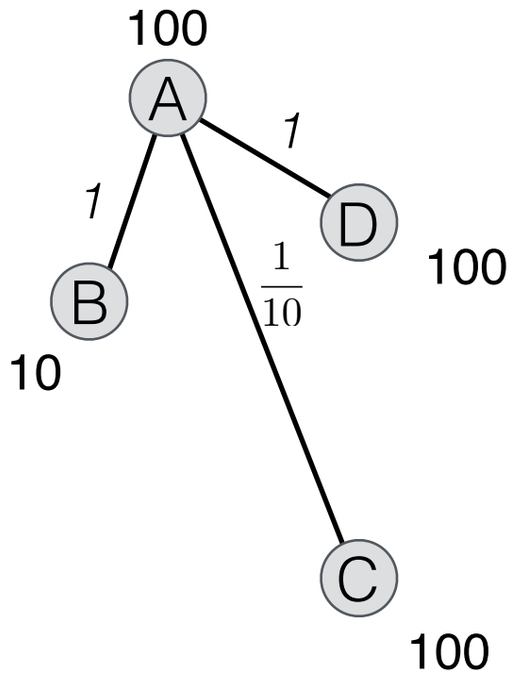
Human Capital Networks

These different weightings emphasize different aspects of the relationship between skills

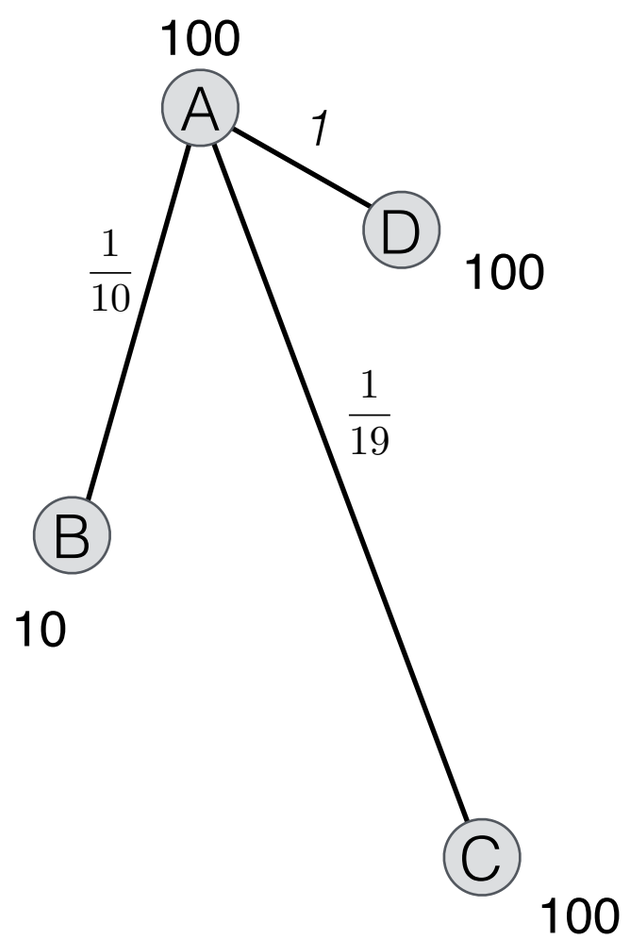
a) frequency weighting



b) modified conditional probability



c) Jaccard Similarity Index



An illustration: a online freelance labor market

In 2014 there were 53 million contract workers, generating $700 billion

Upwork is the largest online labor market

- 9 million workers
- 4 million clients
- $1 billion/year

Worker Profiles (skills, biography, test scores, employment history, ratings…)

Job Listings: (skills required, description, wage rate…)

Why Online Labor Markets?

1) Freelance and contract labor is an increasing fraction of the US labor force

53 million workers in 2014
Generating $715 billion

- Worker flexibility
- Geography-free = larger and more diverse labor pool for employers
- Less costly search process

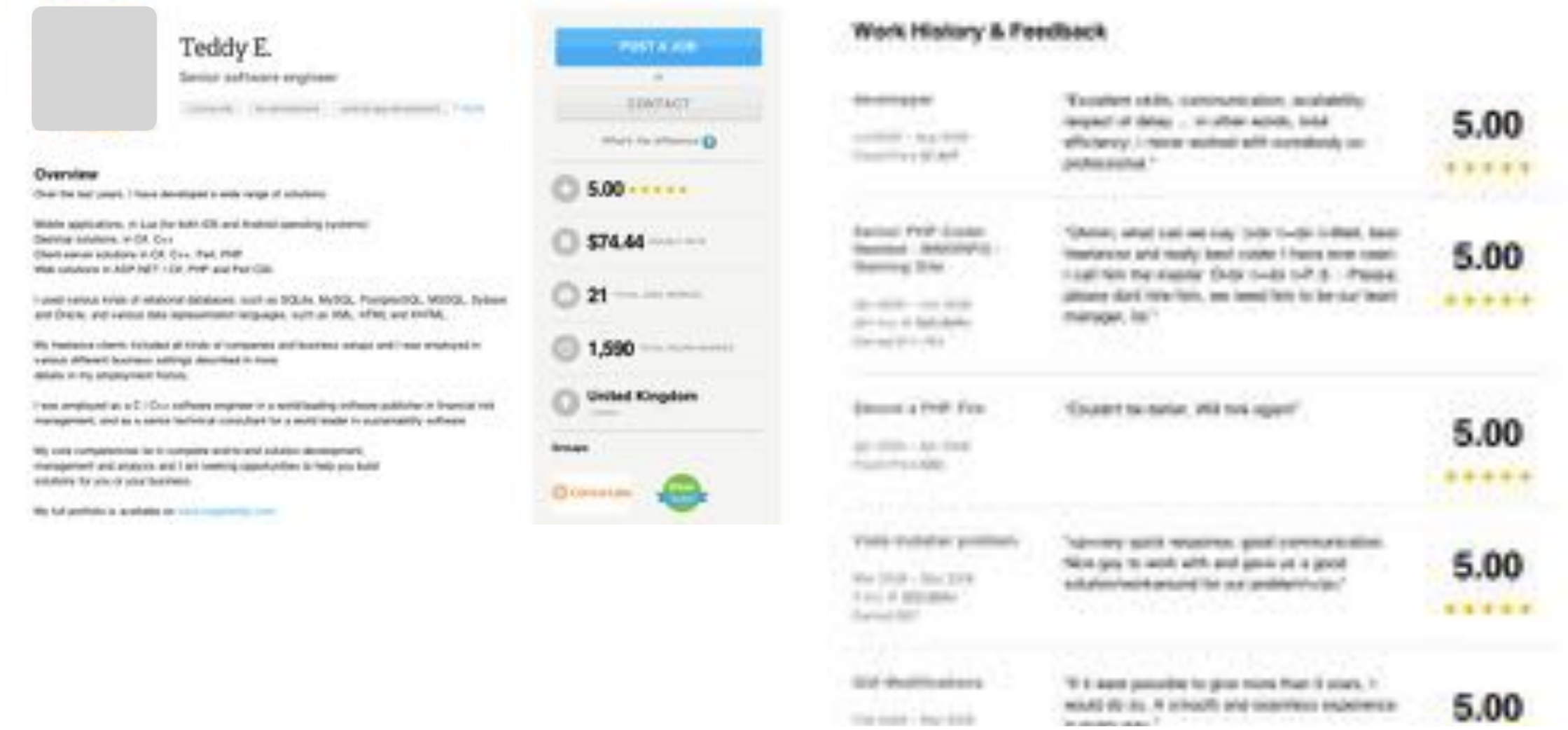
Why Online Labor Markets?

2) Employers increasingly search for employees using online tools

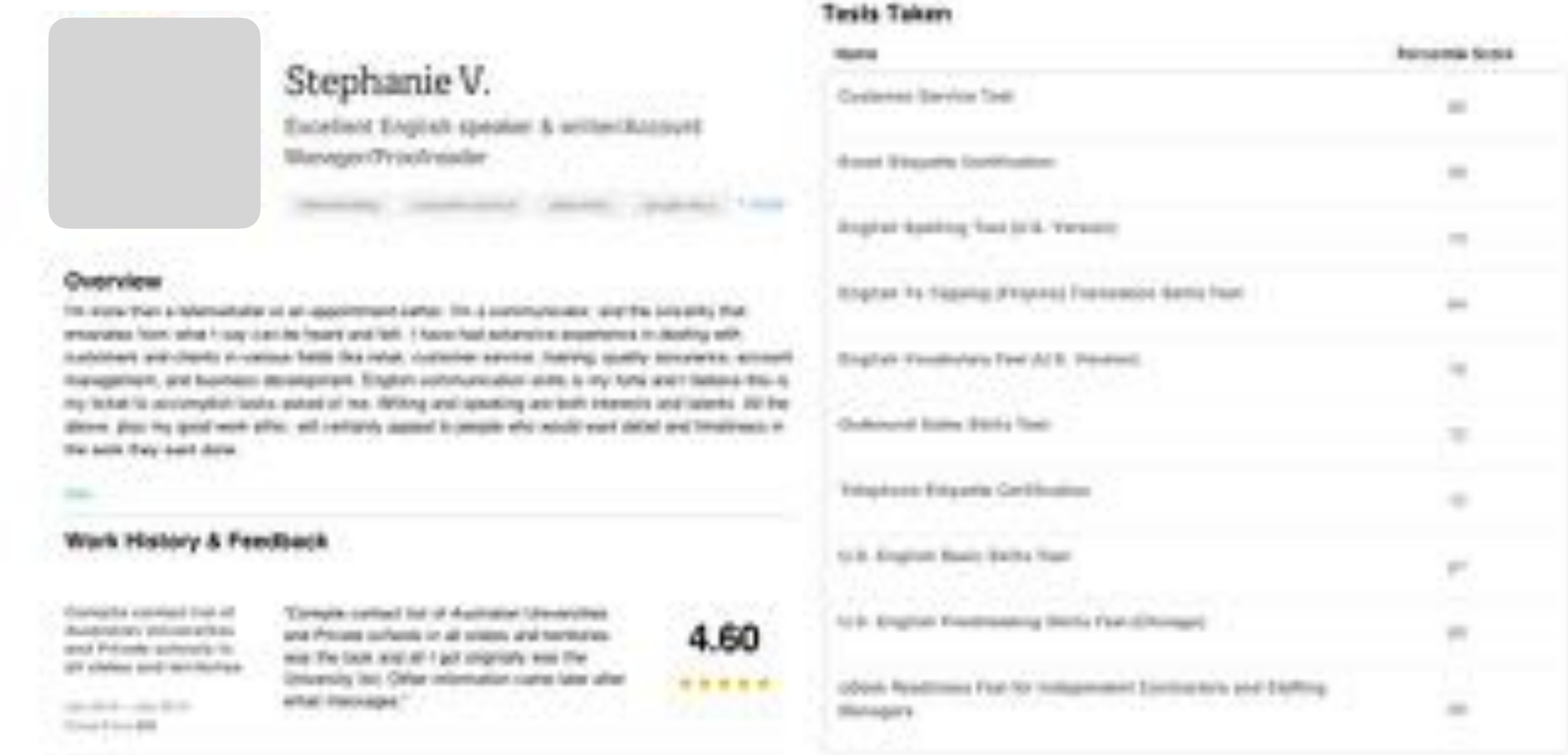
- Substitute "low bandwidth" information for "high bandwidth" information
- Employers choose to use low bandwidth search methods, even when they produce worse matches!
- Improvements to algorithms would produce better labor market outcomes

upwork

Worker profiles:

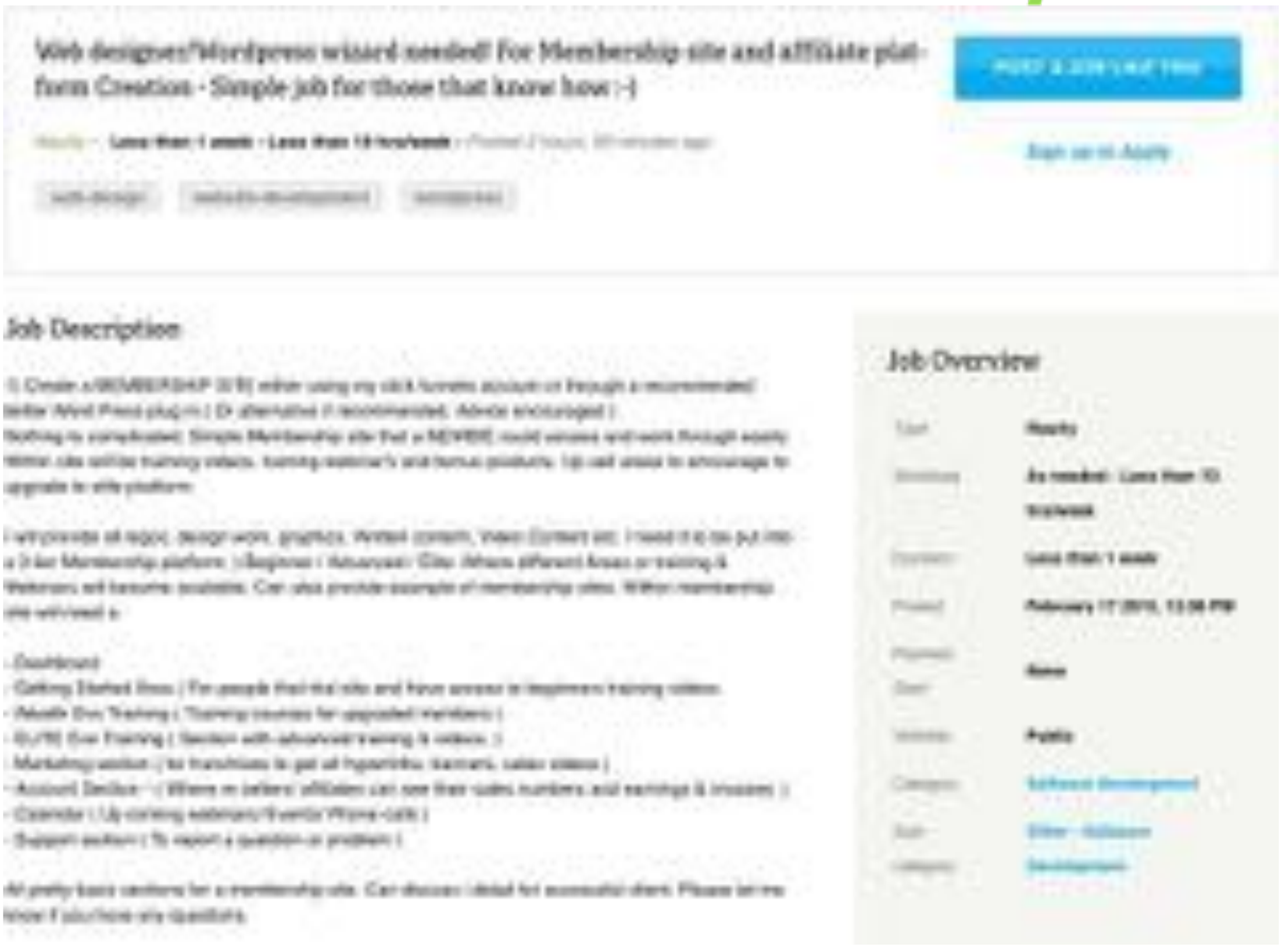


There are a wide range of worker types…



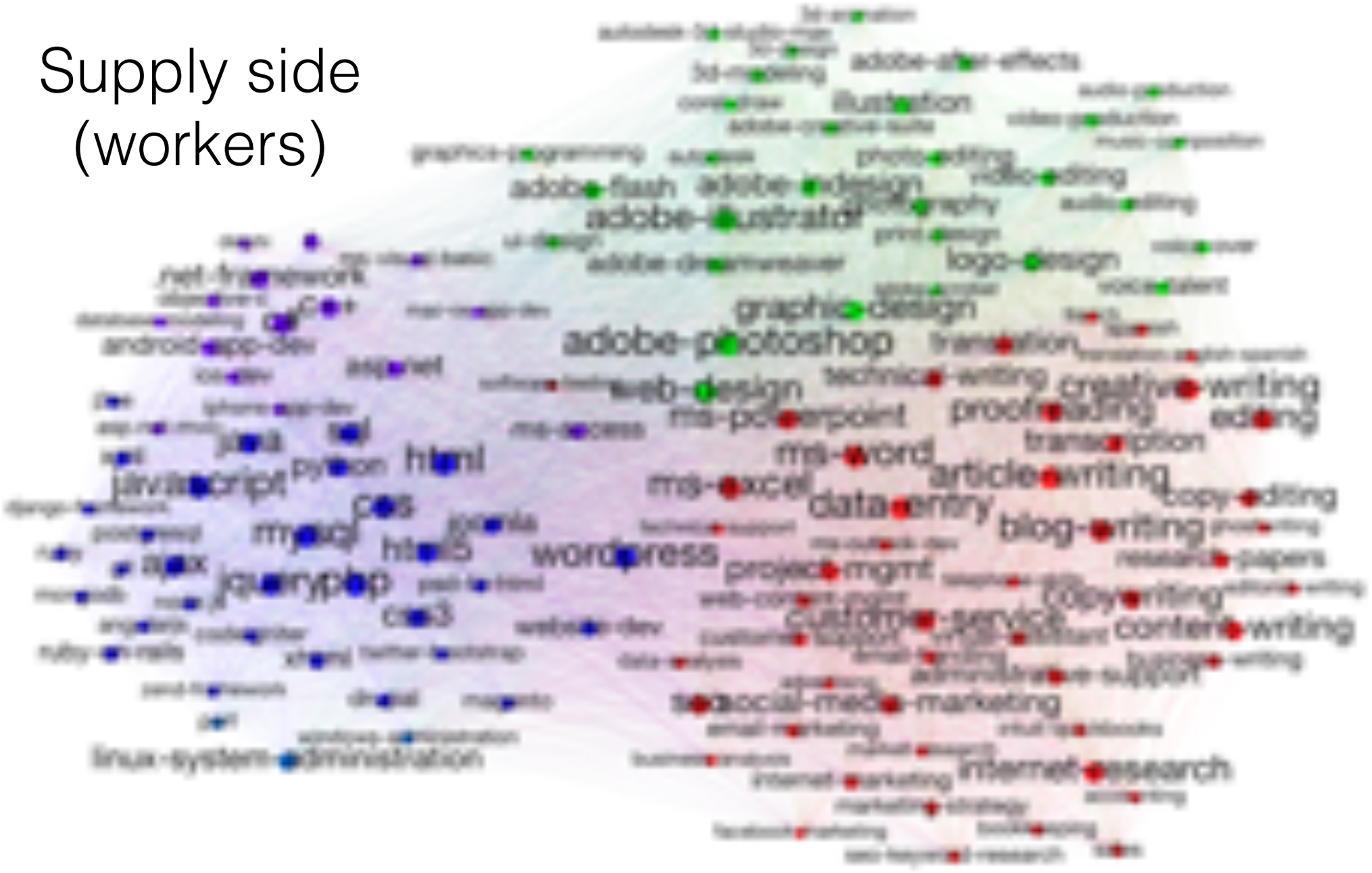
Job Postings:

Upwork



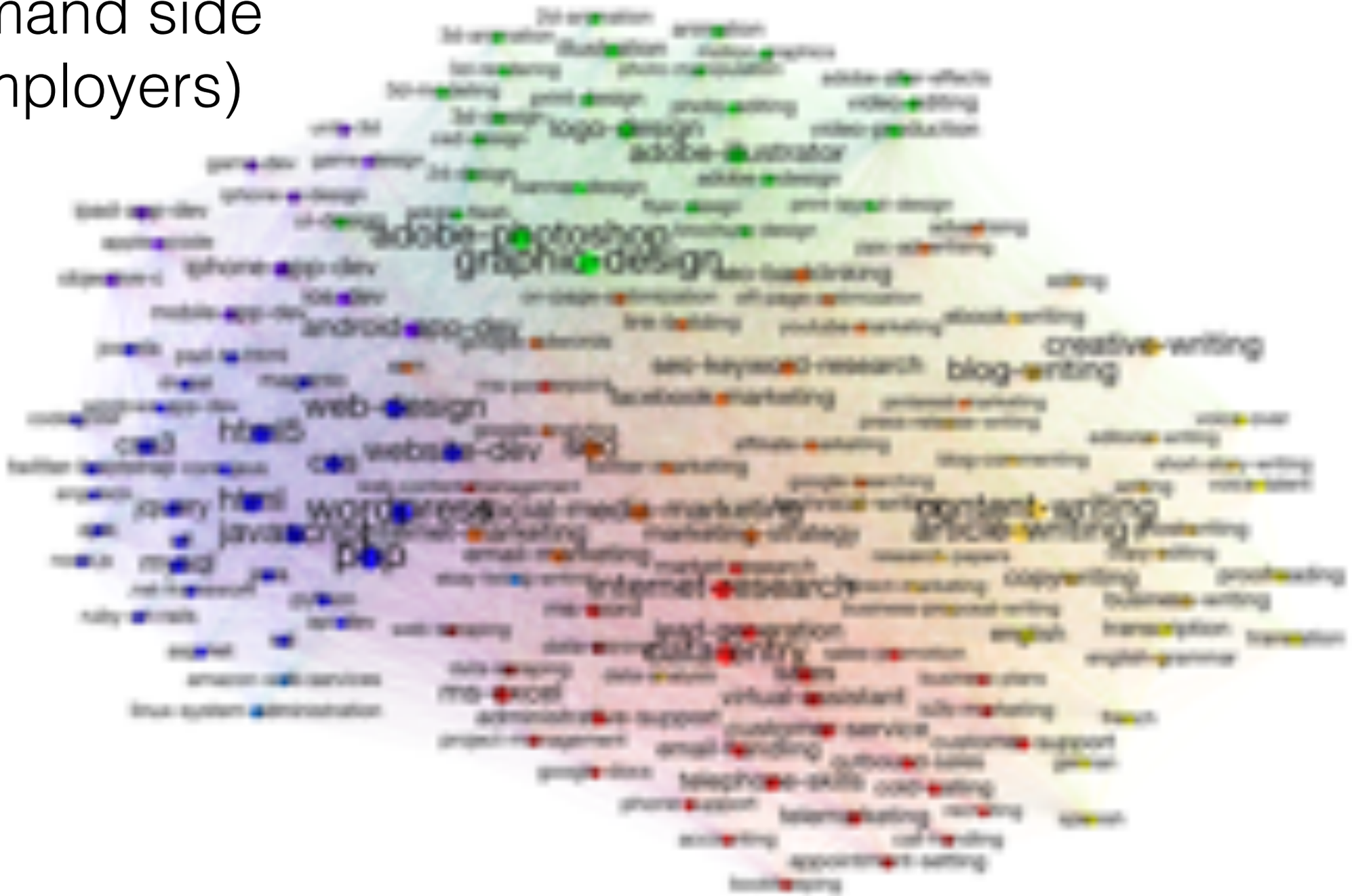
Human Capital Network

Supply side
(workers)



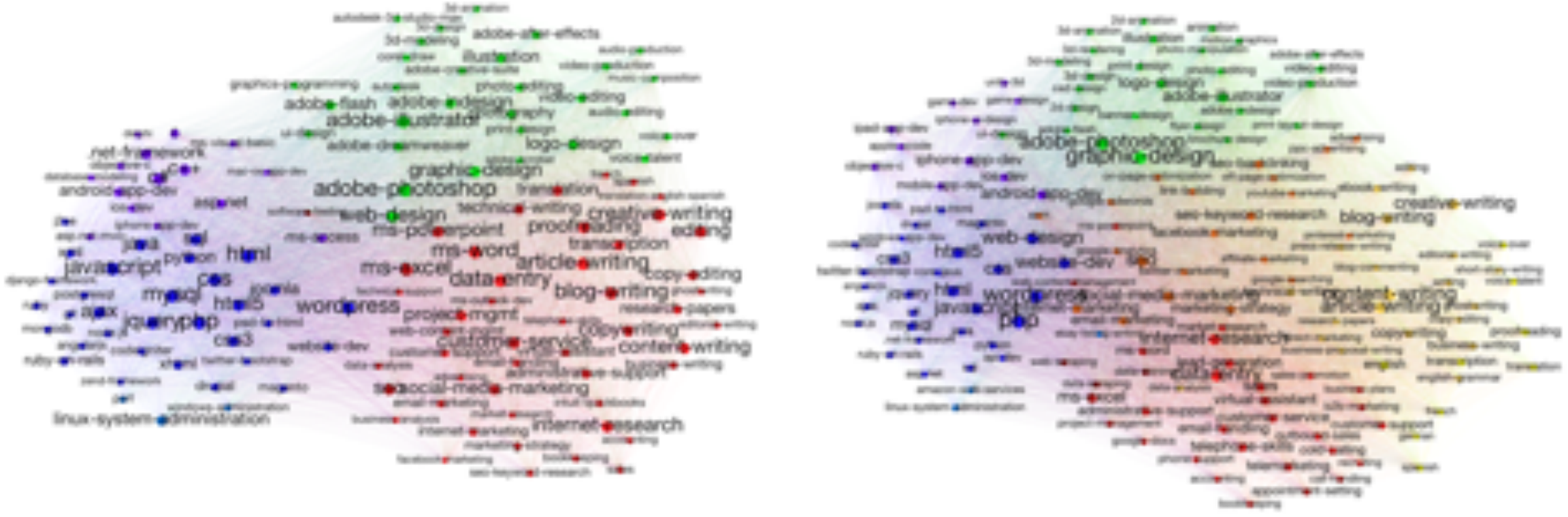
Human Capital Network

Demand side
(employers)



Communities in the Network

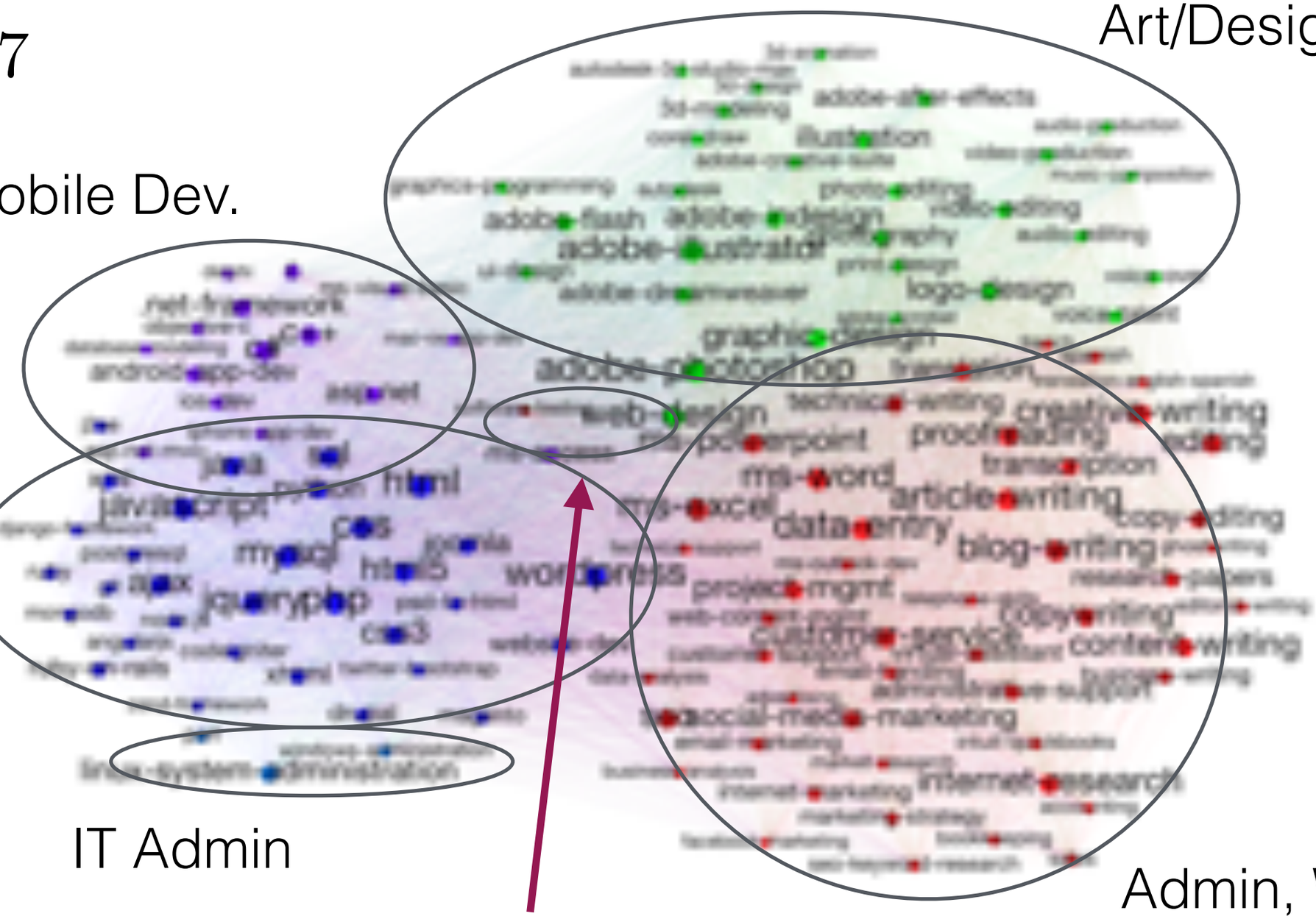
These two networks summarize a huge amount of information about the labor market



For one, by using a community-find algorithm, we can learn something about how the labor market is arranged

Supply side (workers)

$Q = 0.47$

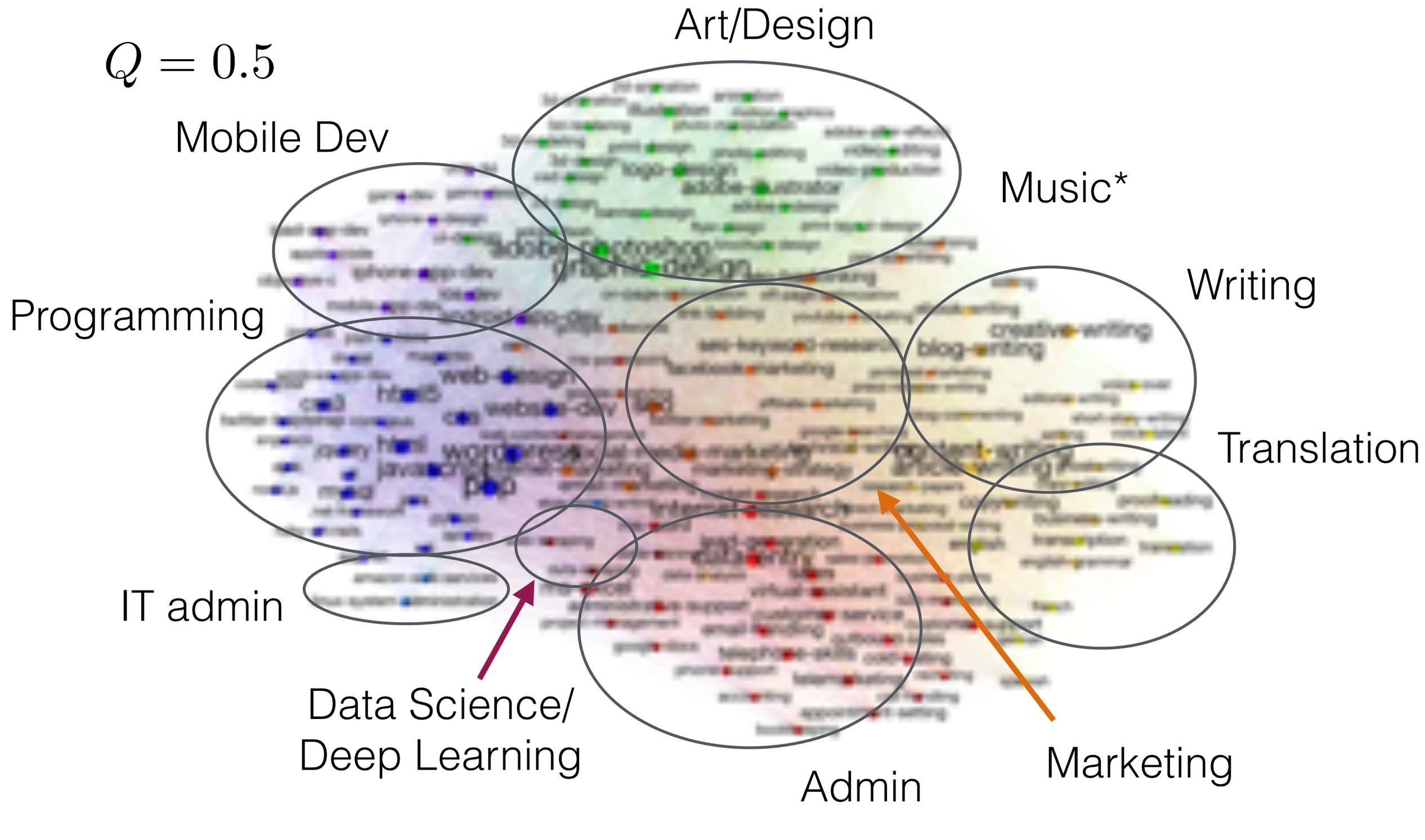


A Taxonomy of Skills

Admin/Writer/Sales	Artist/Designer	Programmer/Technical			Testing
		General	Mobile/Stats	IT/Network Admin	
microsoft excel	adobe photoshop	php	c#	linux systems admin	software testing
data entry	adobe illustrator	javascript	c++	perl	software qa testing
article writing	web design	html	.net framework	windows admin	manual testing
microsoft word	graphic design	mysql	asp.net	apache admin	functional testing
blog writing	adobe indesign	css	android app dev	network admin	usability testing
creative writing	logo design	wordpress	c	lamp admin	regression testing
internet research	adobe flash	jquery	microsoft access	amazon web services	atlassian jira
customer service	illustration	java	ios dev	unix shell	black box testing
proofreading	adobe dreamweaver	ajax	objective c	computer networking	automated testing
editing	photography	html5	ms visual basic	unix system admin	web testing

Demand side (employers)

$Q = 0.5$



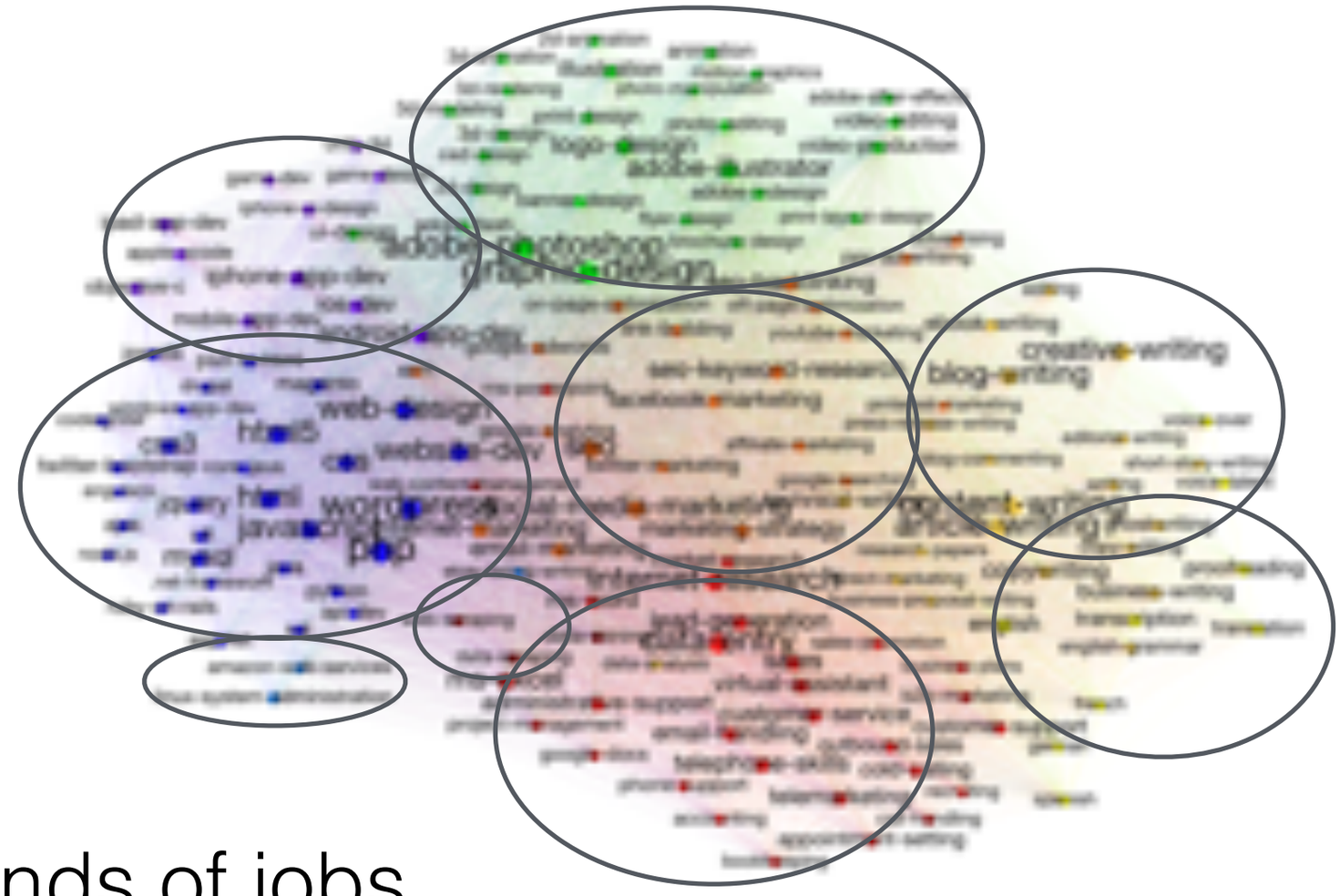
A Taxonomy of Skills

Administrative	Writing	Translation	Art/Design	Music/Audio	Marketing
data entry	content writing	english	graphic design	audio editing	seo
internet research	article writing	transcription	adobe photoshop	audio production	social media market
microsoft excel	blog writing	proofreading	adobe illustrator	audio mixing	internet marketing
lead generation	creative writing	translation	logo design	audio post-product	seo keyword research
sales	copywriting	voice over	illustration	music composition	email marketing
telephone skills	technical writing	voice talent	video editing	audio mastering	marketing strategy
virtual assistant	ebook writing	french	UI design	audio engineering	seo backlinking
customer service	business writing	spanish	video production	music producer	facebook marketing
email handling	ghostwriting	german	print design	music arrangement	twitter marketing
telemarketing	english grammar	english to french	animation	sound editing	link-building

Programming/Technical					Testing
General	Mobile Dev	IT/Network Admin	Engineering	Data/Statistics	
php	android app dev	linux system admin	c	data mining	software testing
wordpress	iphone app dev	amazon web services	electrical engineering	data scraping	QA testing
javascript	ios dev	ebay listing writing	electronics	web scraping	web testing
web design	mobile app dev	ebay marketing	matlab	web content mgmt	manual testing
html5	ipad app dev	network admin	pcb design	scripting	functional testing
html	objective c	cpanel	arduino	web crawler	automated testing
css	game dev	email deliverability	microcontroller prog	website wireframing	usability testing
website dev	apple xcode	amazon ec2	circuit design	salesforce apex	database testing
css3	game design	network security	embedded systems	salesforce.com	localization
mysql	iphone UI design	voip software	electrical drawing	salesforce app dev	selenium

A Taxonomy of Skills

These categories are endogenous: they reflect the way that skills are used in this particular labor market

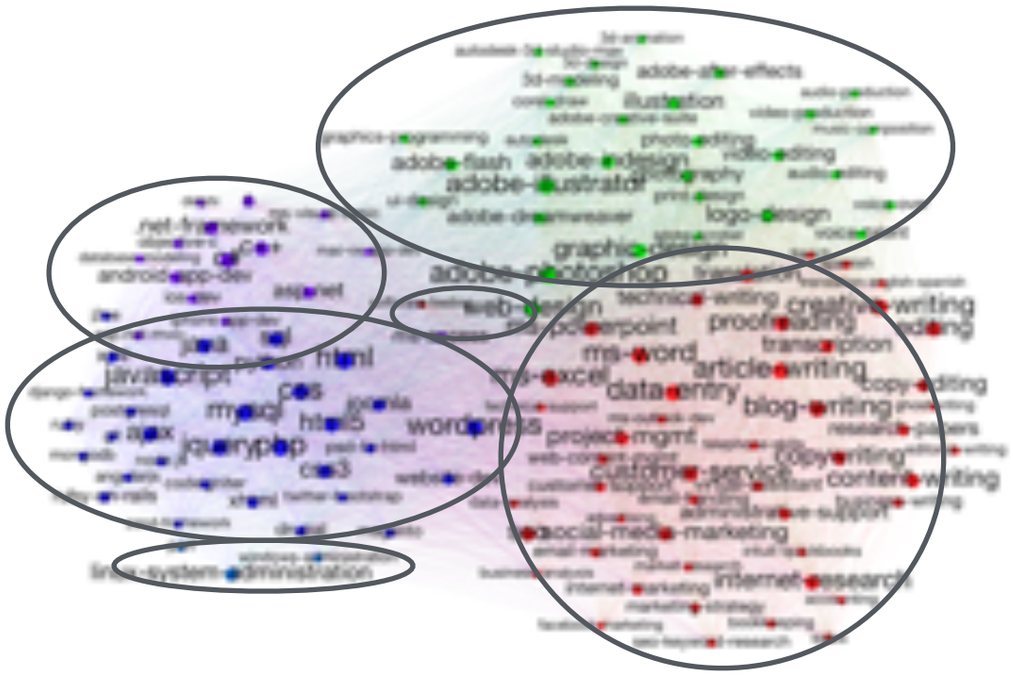


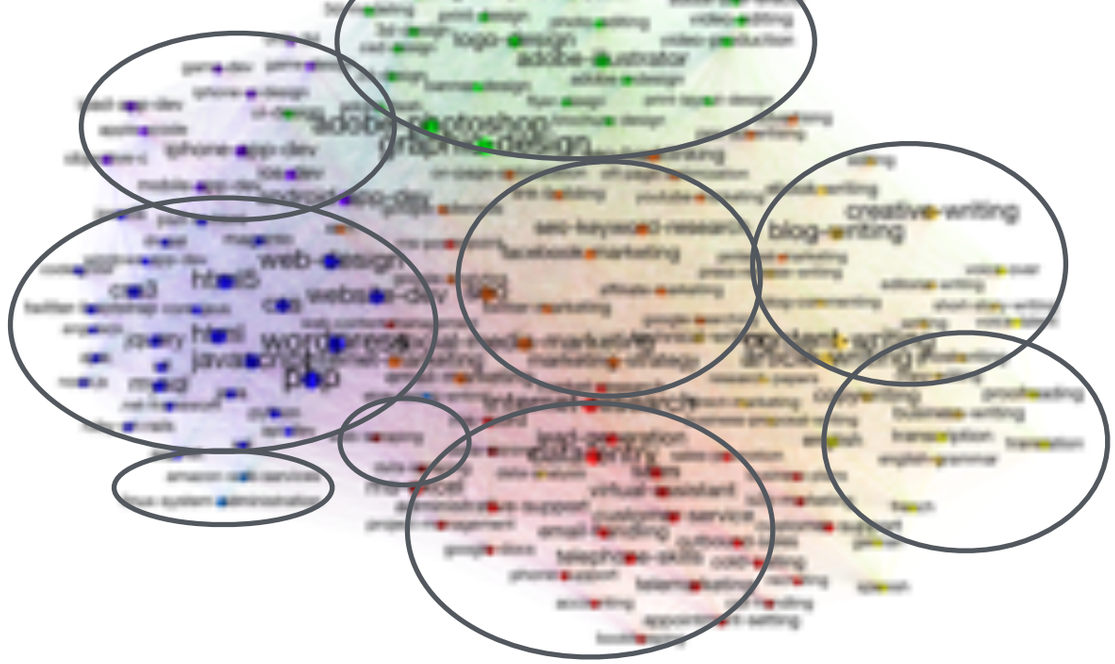
Demand side: kinds of jobs
Supply side: sub-pools of labor

A Taxonomy of Skills

Some interesting notes:

- Jobs are more narrowly-defined than workers
- There are categories of job for which there is no dedicated labor pool





Complex Measures of Human Capital

The position of a workers skills on the two networks characterizes their skill set

Specialist

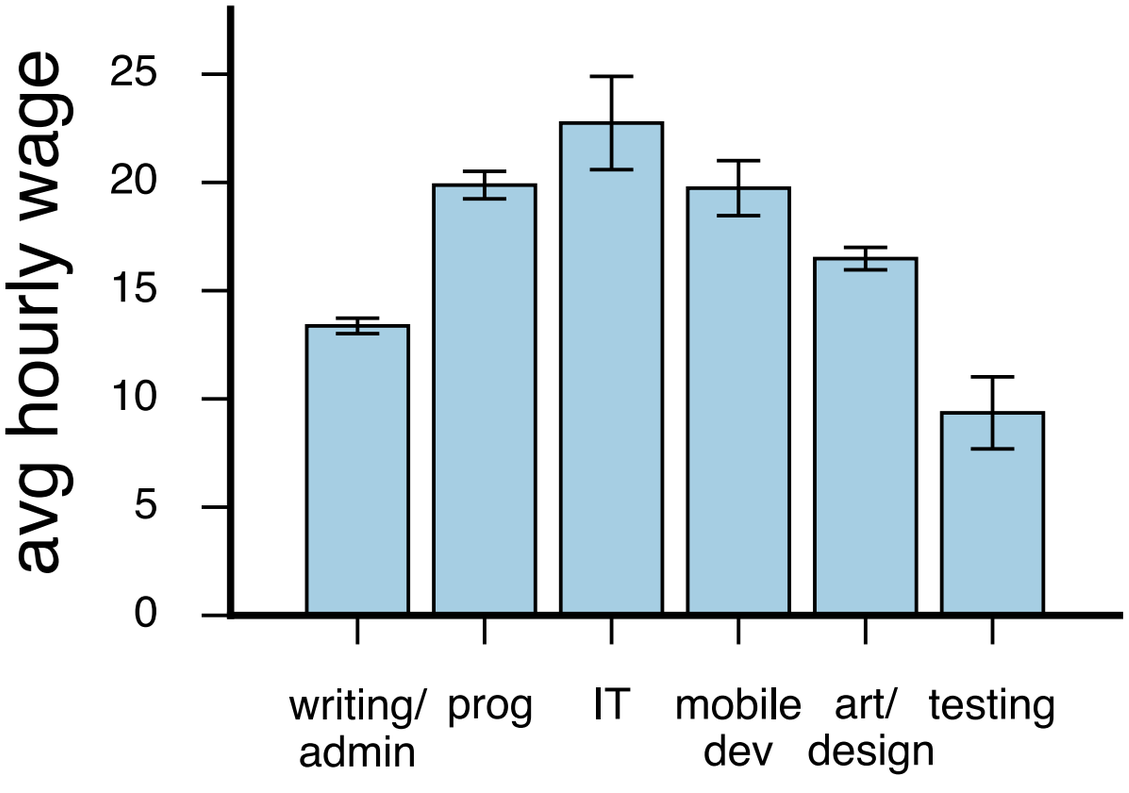
Diverse Skills

Measuring these differences opens up new questions

Complex Measures of Human Capital

As an illustration…consider workers whose skills fall into a single category

Wages vary dramatically by area of specialization



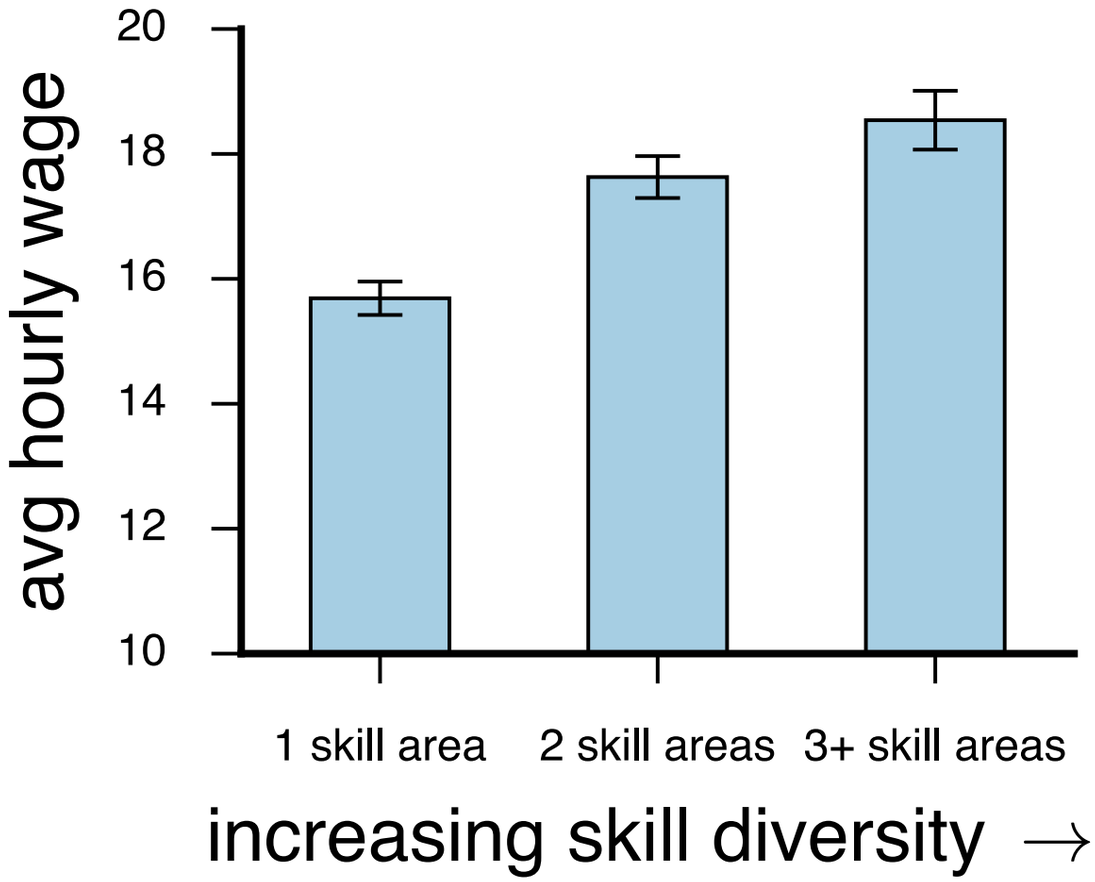
NB: for now, I control for literally nothing

Human Capital and Wages

Now compare specialists to those who span multiple categories:

Complex Measures of Human Capital

Workers with more diverse skills tend to earn more



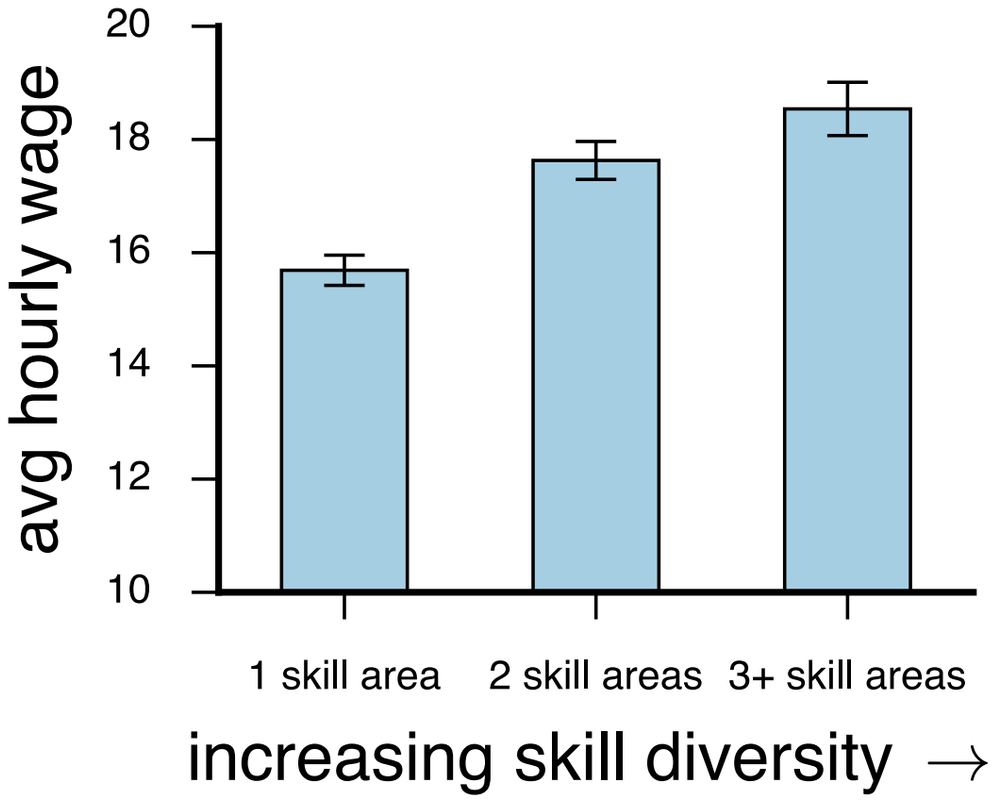
Complex Measures of Human Capital

But then the question is, *why* do workers with diverse skills earn more?

Two possible advantages:

1. diverse skills =
   larger pool of jobs

2. diverse skills =
   unique gap-filler



Complex Measures of Human Capital

Difference between the two can be seen in the job network

Skills tightly clustered on the job network are often required in combination

Complex Measures of Human Capital

Worker Network

Job Network

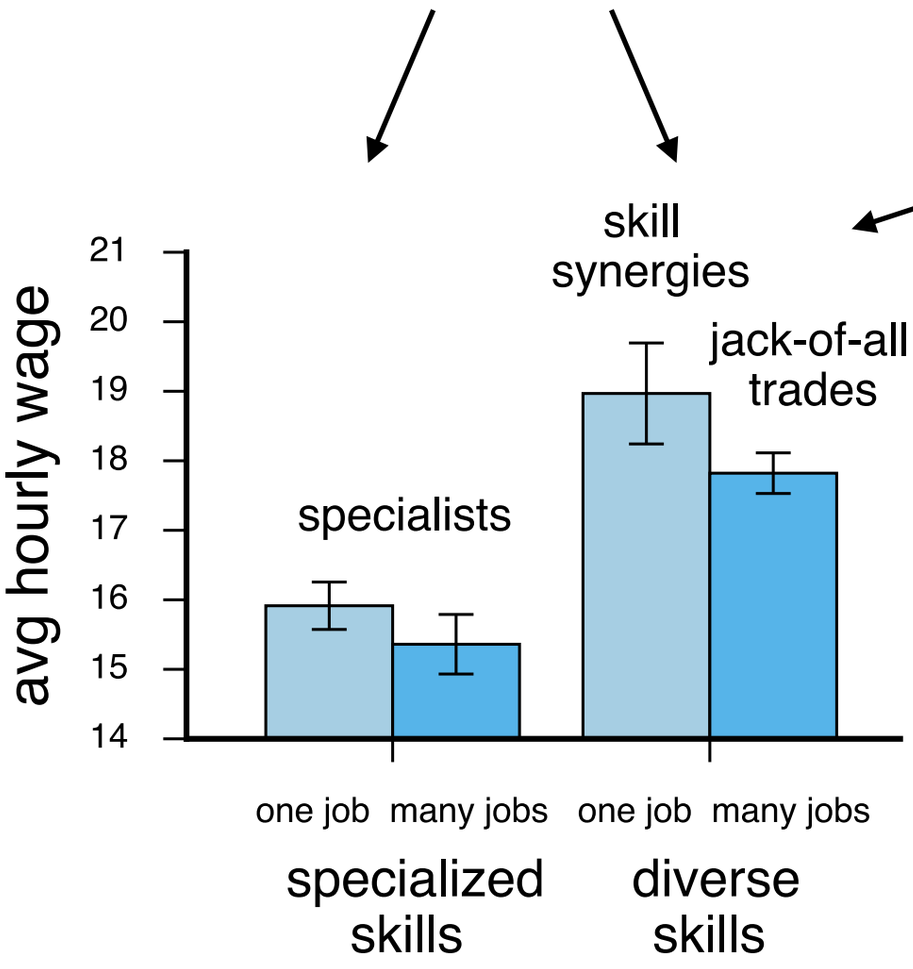
Fit Multiple Job Areas

Fit One Job Area

Complex Measures of Human Capital

Skill diversity = higher wages

Diverse skills are better when used together



People who use their diverse skills synergistically tend to earn higher wages

The Value of Skill Interactions

A totally reasonable question:

Do the network-based measures of skill interactions actually do any better than just looking at a workers skills independently?

The Value of Skill Interactions

Take an incredibly flexible model specification:

$$ln(w_{it}) = \beta_0 + \beta_1 S_{it} + \epsilon_{it}$$

dummies for the 60 most common skills in the labor pool

Compare that to a model including the independent skills *and* network-based measures of human capital

$$ln(w_{it}) = \beta_0 + \beta_1 S_{it} + \beta_2 X_{it} + \epsilon_{it}$$

The Value of Skill Interactions

	Model 1	Model 2	Model 3	Model 4
N	18165	8859	18165	18165
Adj R^2	0.178	0.190	0.180	0.183
constant	2.6***	2.36***	2.54***	2.51***
program.		0.51***		
IT admin.		0.59***		
mobile dev.		0.47***		
art & design		0.26***		
testing		-0.3***		
number of skill areas			0.05***	
number of job categories				0.05***
html	-0.08***	-0.08**	-0.09***	-0.09***
illustration	0.12***	0.15***	0.12***	0.13***
social media marketing	0.04*	0.12***	0.04	-0.0
mysql	0.08***	0.01	0.07***	0.07***
project management	0.18***	0.24***	0.16***	0.14***
adobe photoshop	-0.13***	-0.17***	-0.14***	-0.15***
microsoft word	-0.15***	-0.11***	-0.15***	-0.16***

The network-based measures are significant even when you control for the skills individually

The Value of Skill Interactions

Effect Sizes:

- Programmers earn 51% (~$5.40) more than Administrative Workers
- QA testers earn 30% (~$4.00) less
- Workers with an skill in an additional category earn $0.65 more

Human Capital and Wages: Skill "Coherence"

So some skill diversity seems to improve outcomes. But is there a limit to that?

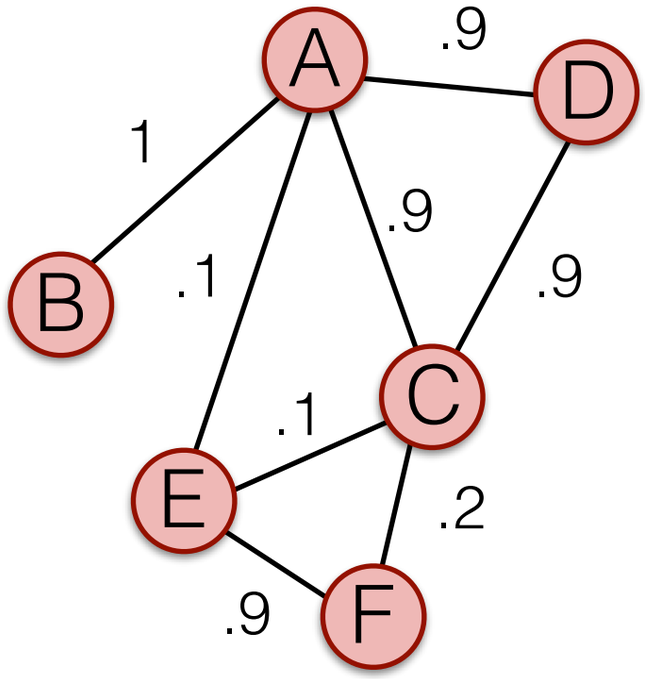
- Sometimes breadth of experience is seen as a positive sign (Renaissance Person)
- Sometimes it is seen as a lack of focus (Dilettante)

So what distinguishes a Renaissance Person from a Dilettante?

> Leung 2014: it is not just the *breadth* of experience, but the *coherence* of that experience

Skill “Coherence"

Measure of skill set coherence: average distance between a workers skills in a particular labor market



A B =1

E F C =.4

weights = Jaccard Similarity Index

$$w_{AB} = \frac{A \cap B}{A \cup B} = \frac{n_{AB}}{n_A + n_B - n_A B}$$

Call this *JI similarity*: how well your skills fit in with those of your peers

Worker Skill Similarity

High JI Specialization

High JI specialization:
Diverse, but coherent

Low JI specialization

Skill “Coherence"

Note: There is a difference between “diversity” and “coherence”

Set Designer + Screen Play Writer + Director = Coherent

Graphic Design + UIX + Mobile Development = Coherent

Underwater Salvage + Creative Writing = Incoherent

Worker Skill Similarity

$$ln(w_{it}) = \beta_0 + \beta_1 S_{it} + \beta_2 X_{it} + \epsilon_{it}$$

Workers benefit from adhering to "type"

Workers with many common skills earn lower wages

Dummies for skills in different categories and pairwise interactions →

Dependent variable = log wages				
	(1)	(2)	(3)	(4)
	Round 1 data		Round 2 data	
JI similarity	0.699***	0.466***	0.479***	0.276***
	(0.0763)	(0.0796)	(0.0617)	(0.0633)
Number of skills	0.0306***	0.0359***	0.0375***	0.0372***
	(0.00302)	(0.00378)	(0.00262)	(0.00307)
Number of rare skills	0.0819**	0.0153	0.0284	0.00859
	(0.0379)	(0.0365)	(0.0234)	(0.0221)
Number of common skills	-0.144***	-0.125***	-0.0901***	-0.0763***
	(0.00874)	(0.00918)	(0.00688)	(0.00704)
Number of jobs worked	0.00387***	0.00419***	0.00113***	0.00138***
	(0.000287)	(0.000276)	(0.000141)	(0.000134)
Average rating	0.0679***	0.0719***	0.136***	0.146***
	(0.0122)	(0.0117)	(0.0167)	(0.0158)
Average test score	0.00119***	0.00135***	-0.00693	0.0101**
	(0.000257)	(0.000247)	(0.00518)	(0.00494)
Lives in USA	0.0670***	0.0713***	0.0895***	0.0831***
	(0.0184)	(0.0182)	(0.0171)	(0.0168)
GDP per capita (logs)	0.0641***	0.0916***	0.0812***	0.110***
	(0.00792)	(0.00771)	(0.00654)	(0.00627)
Skill fixed effects	No	Yes	No	Yes
Centroid fixed effects	Yes	Yes	Yes	Yes
Observations	8507	8507	7971	7971
Adjusted R^2	0.149	0.222	0.156	0.255

Summing up

Skill combinations matter!

- Employers care about the whole package that an employee brings to the table
- The difference between workers may come down to synergies between skills
- The benefit of acquiring a skill depends on the skills the employee already has

Online job market markets rely on algorithms to match employees with jobs.

- Making those matches requires knowledge of synergies