Are occupations “bundles of skills”?

Identifying latent skill profiles in the labour market using topic modeling

Marie Labussière
Thijs Bol
University of Amsterdam

Are standard occupational classifications a good proxy for the skill content of jobs?

Sociologists and economists often use **standard occupational classifications** (e.g., ISCO) to make inferences about the task or skill content of jobs, assuming that occupations are made up of **distinct and well-defined “bundles of skills”**. Yet, this assumption requires empirical examination:

- ➟ the content of jobs is found to vary considerably *within* occupations
- ➟ workers transfer a substantial part of their skills portfolios when they change occupations, indicating overlaps *between* occupations

From skills to skill *profiles*

Previous work tends to focus on narrow sets of skills and analyze them in isolation from one another. We argue that the relevance and meaning of a particular skill depend on the other skills required for the job.

- ✓ Need to look at **skill profiles**, i.e., the combination of different (types of) skills that workers need to master to perform their job effectively
- ✓ Need for a **bottom-up approach**, with a flexible methodology that allows skills to be part of different skill sets simultaneously

Data

Dataset of 6.9 million job ads collected by Burning Glass Technologies in the **UK** in **2019**

↪ Sample of 600 000 job ads
Stratified by occupation (2-digit SOC codes) to under-sample professional occupations and favour elementary occupations

Identifying the skill profiles of occupations

1. Topic modelling using the skill requirements of job ads

Using the **Biterm topic model**, we identify **19 latent topics of skills**: sets of skills that tend to be mentioned together in job ads.
Each **topic** is characterised by a probability vector over the whole corpus of skills contained in the job ads.

Skill set (or topic)	*Skills, in decreasing order of probability*
Digital Marketing	*social media; marketing; digital marketing; creativity; marketing management; Google Analytics; market strategy; content management; copy writing; editing...*
...	...

Each **job ad** is defined by a probability vector over the 19 topics, representing its **skill profile**:

Social Media Account Executive
Required skills: Time Management, Content Management, Creative Writing, Social Media, Creativity, Business-to-Business, Social Media Platforms.

⇒

Skill set	*probability*
Digital Marketing	*0.75*
Graphic Design	*0.08*
Effective Collaboration	*0.07*
Other skill sets	*< 0.05*

2. Calculating the distance between occupations

For each 4-digit SOC occupation, we define an empirical distribution over the 19-dimension probability vectors of the job ads categorised in that occupation. We use the **Maximum Mean Discrepancy (MMD)** distance to compare the empirical distributions of occupations: how much do they overlap or differ according to the skill profiles of their associated job ads?

Mapping occupations according to their skills profile

Based on the MMD distance matrix, we construct a network where each node represents a 4-digit SOC occupation. Two nodes are connected when the pairwise MMD distance is small, i.e., that their skill distributions are similar (threshold: MMD distance < 0.1).

A top-down classification of 4-digit occupations

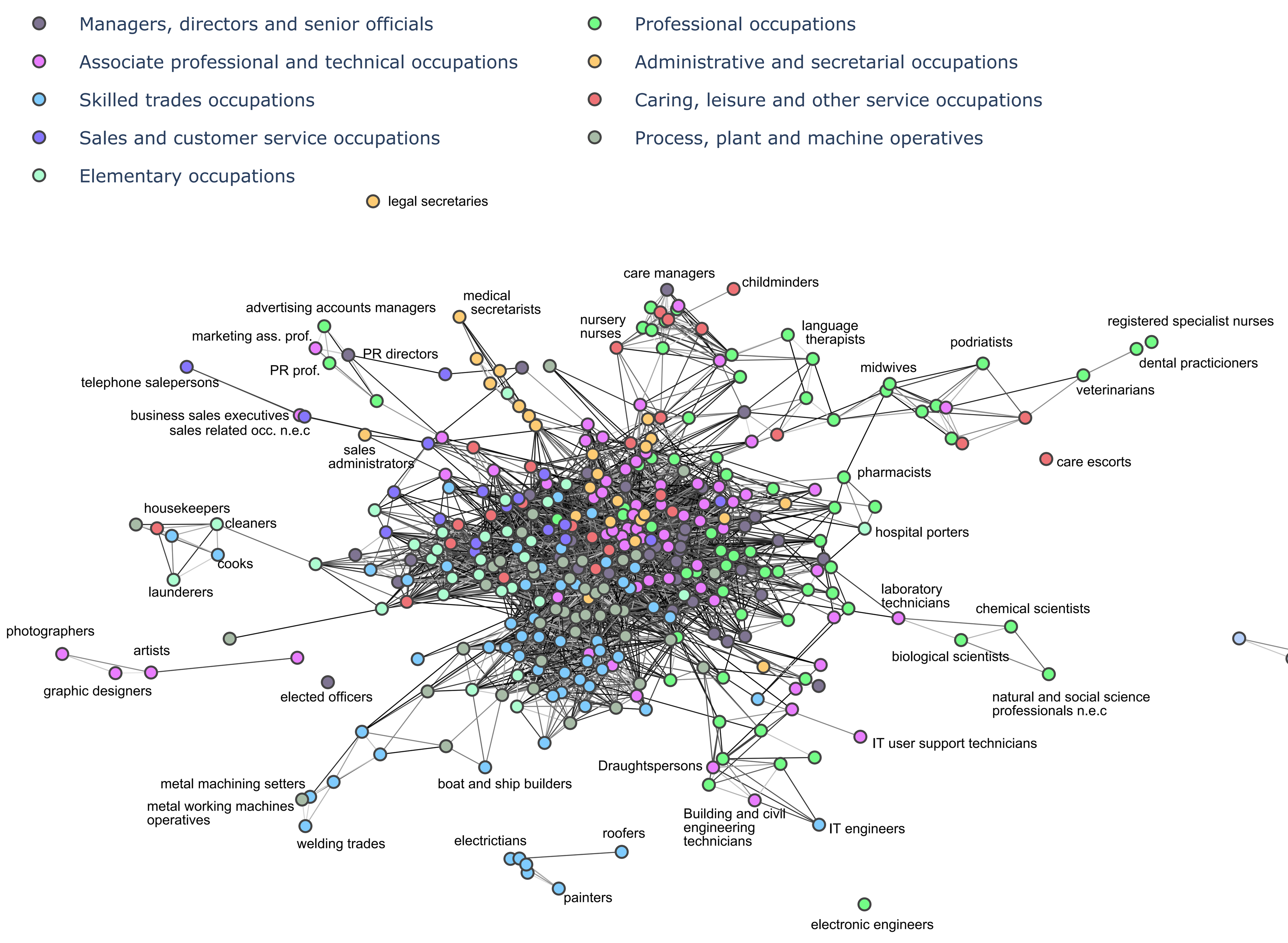

Figure 1: Graph of 4-digit SOC occupations, clustered in SOC major groups (1-digit)

A bottom-up classification of 4-digit occupations

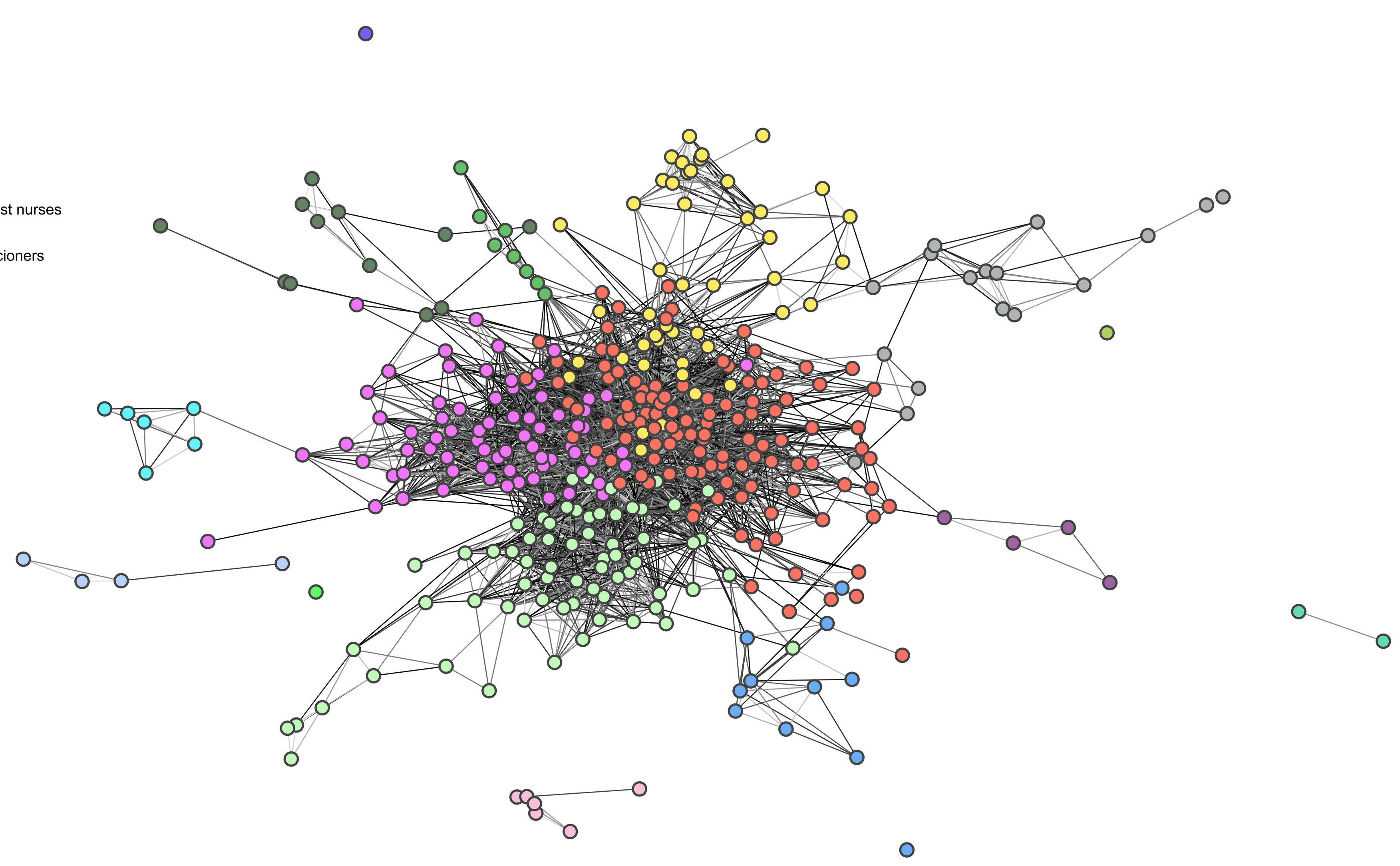

Figure 2: Graph of 4-digit SOC occupations, clustered in data-driven communities (Louvain algorithm)

Conclusions and Future Work

Data-driven typology based on skill profiles does not coincide with standard occupational categories
↪ Occupations are not a good proxy for the skill content of jobs
Many occupations have similar skill profiles – or at least, they are not distinctive!
⇒ **Are there more opportunities for mobility than we usually think?**

Questions, suggestions?

m.labussiere@uva.nl   LabussiereMarie

career