A network analysis of personnel exchange and companies' relevant sector: the LinkedIn case study.

Tommaso Cavalieri[1], Andrea Fedele[1], Federica Guiducci[1], Valentina Olivotto[1], and Giulio Rossetti[2]

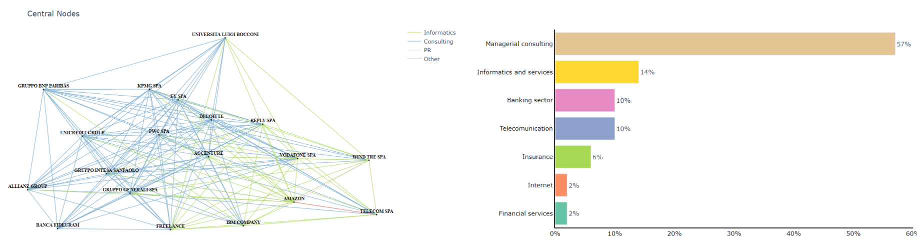
[1] University of Pisa, Italy,
cavalieri.tommaso@gmail.com, andrea-fedele@hotmail.it,
guifede3@gmail.com, valentina.olivotto19@gmail.com,
[2] KDD Lab, ISTI-CNR, Italy,
giulio.rossetti@isti.cnr.it

1 Introduction

In nowadays fast world, people tend to change workplace very often, carrying along with them the old company's know-how and, possibly, some strategic information. Analysing how companies exchange personnel is interesting in both social and economic terms and, moreover, it could be a useful tool for every company that does not want to lose its best employees. In fact, over the last few years, one of the major problems managers had to deal with was employee retention [3, 7]. To gain insight into this problem we built a professional network using data retrieved from the **LinkedIn** social network and we analysed how employees from different sectors moved between companies. The final goal of the study was to understand whether diverse sectors presented some peculiar characteristics in terms of turnover, with a follow-up focus on those companies exchanging a significant number of employees. A further objective was to investigate whether it was possible to infer companies' prevalent sector from employees' exchange information. A sample of 221,782 users was gathered filtering by location (Italy), industrial sector and the number of employees. Such sample was later exploited to obtain a network connecting various companies, which were represented as nodes in the graph. An edge between two companies was built when at least two persons have been employed in both of them at some point in time. Each of these links was weighted according to the total number of people that had worked in both firms and characterized by an attribute representing the industrial filter. The building of such edges did not consider the chronological order of the employee's exchanges, resulting in an undirected graph of 14,875 nodes and 43,932 weighted edges. During the scraping phase we selected 23 of the industrial sectors available on LinkedIn, that were later aggregated into 4 macro-sectors: **Consulting**, **Informatics (IT)**, **Public Relations (PR)** and **Others**.

2 Experimental Results

To investigate the correlations between companies through the exchanged personnel we analysed how sectors, communities and central nodes influenced the structure of



(a) Central nodes by macro-sector (b) *Accenture*'s sectors percentage distribution

Fig. 1. Central nodes subgraph and *Accenture* sectors' distribution

the graph. The communities analysed were discovered via the Louvain algorithm [1] to maximize the modularity measure [5]. Our analysis underlines that each of the identified community is mainly composed by companies belonging to the same sector. Moreover, we observed the sectors composing the IT macro-sector to be uniformely distributed - being spread all over the communities - while the sectors composing the Consulting and the PR macro-sector were confined in fewer communities. However, these last two macro-sectors appeared to behave differently: in fact PR sectors covered most of the communities in which they were presented - revealing a weak connection between them - while the Consulting ones followed a better balance distribution.

We then carried out a sector distribution analysis on those nodes that were found to be central in the network (by means of classic centrality indexes). The first thing to notice was that these nodes represent well-known companies in their industry as shown in figure 1(a). A common pattern was indeed found in these central nodes' sector distribution: each of them was mainly composed by one sector with a value between 60% and 80% and several others with a value smaller than 10% separately. Figure 1(b) highlights this phenomenon for the *Accenture* node. Depending on the company, the major sector of its edges changes as expected: while in nodes representing well-known banks (eg. Gruppo BNP Paribas) it is the banking industrial sector, in IT companies (eg. Amazon) it is the IT sector and in Consulting firms (eg. Accenture) it is the consulting one. We observed interesting characteristics about the *Freelancer* node, which presented the most sparse sector's distribution. This might be due to the nature of the node that does not represent a specific company, but a temporary status (if not a life choice).

To examine the correlation between sectors we created a co-presence matrix that stores in each cell (i,j) the number of nodes that had both a link of sector i and a link of sector j attached to them. We observed, as expected, high co-presence values between pairs of sectors belonging to the same macro-sector. However, pairs composed of different macro-sectors were also observed: the highest value was reached by the couple *(Managerial consulting - Informatics and services)*, followed by *(Managerial consulting - Human resources)* and *(Financial services - Informatics and services)* reflecting what the authors concluded in [4]. Moreover, it emerged that *Managerial consulting* was the most present sector in such pairs and this result could be associated with the multidisciplinarity of the Consulting macro-sector [6], capturing people with various backgrounds and moving from different sectors.

Finally, we constructed a derivative dataset using both semantic and topological information characterising the network nodes and we used it in a Machine Learning classification problem with the aim of predicting the industrial macro-sector. The selected classifier to address such task was the Random Forest (RF). We achieved satisfactory results reaching an overall accuracy of 76% and a F1-score value of the Consulting, IT, PR and Others macro-sectors labels respectively of [0.83, 0.75, 0.70, 0.50]. It is important to notice that the lowest performance was obtained for those labels with lower support due to the unbalance in the dataset. Finally, to identify, explain and interpret some classification rules, we applied a local rule based explainator, namely LORE [2]. While eigenvector and pagerank centralities appear to be the features of uttermost importance for the RF classifier (considering their prediction weights), the local classification rule explainator highlighted something different. In fact, by analysing several random records with LORE, it emerges that another feature has the most discriminative power: community belonging. The classifier decides the classification label according to the community to which the nodes connected by the link belong. Moreover, the final prediction is set to the class of the most common industrial sector in the node community.

These kind of LinkedIn networks' analysis could be exploited to study the knowledge exchange development between companies and discover what are the social phenomena underneath these changes. As a future work, we plan to collect a larger amount of data and store more information in the network such as users' education, location and job's role. A different network could also be explored taking into account the chronological order of employees' movements to better investigate people turnover, migration and knowledge exchange between companies.

References

1. Blondel, Vincent D., et al. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment 2008.10 (2008): P10008.
2. Guidotti, R., Monreale, A., Ruggieri, S., Pedreschi, D., Turini, F. and Giannotti, F. 2018. Local rule-based explanations of black box decision systems. arXiv preprint arXiv:1805.10820
3. James, L and Mathew, L. (2012). Employee Retention Strategies: IT Industry. SCMS Journal of Indian Management, July - September, 2012.
4. Miele, M.G. and Scognamiglio, A. (december 2019). *Indagine FinTech nel sistema finanziario italiano*. Divisione Editoria e stampa della Banca d'Italia.
5. Newman, M.E.J. and Girvan, M. Finding and evaluating community structure in networks. arXiv:cond-mat/0308217
6. Pinza, A. and Zanchi, R. (April 2020). *La domanda di servizi di consulenza in Italia I risultati dell'indagine*. Confindustria, ASSOCONSULT
7. Singh, D. A Literature Review on Employee Retention with Focus on Recent Trends. International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 6 Issue 1, pp. 425-431, January-February 2019. Available at doi : https://doi.org/10.32628/IJSRST195463Journal URL : http://ijsrst.com/IJSRST195463