The curse of self-presentation: Looking for career patterns in online CVs

Johanna M. Werz
IMA, RWTH Aachen University
Aachen, Germany
johanna.werz@ima-ifu.rwth-aachen.de

Valerie Varney
IMA, RWTH Aachen University
Aachen, Germany
valerie.varney@ima-ifu.rwth-aachen.de

Ingrid Isenhardt
IMA, RWTH Aachen University
Aachen, Germany
ingrid.isenhardt@ima-ifu.rwth-aachen.de

*Abstract***— Climbing the career ladder to a senior executive position is a long and complex process that, nevertheless, many people are trying to master. Over the last decades, the number of people providing their CVs on professional online social networks, such as LinkedIn is growing. New methods of pattern detection raise the question of whether online CVs provide insights into career patterns and paths. The respective hypothesis is that online CVs map people's careers and therefore build the ideal data set to detect career patterns. To test this hypothesis, 100.006 online CVs were downloaded and preprocessed. This paper presents initial results of one educational and one internship variable. Whereas a higher degree positively predicts career level, having made an internship negatively relates to career level. These results reveal that rather than objectively mirroring people's career trajectories, online career platforms provide selective information. The information of online CVs and the respective career level is intermingled, i.e. people with a high career level present different parts of their careers than people on lower levels. Furthermore, self-presentational effects might have an impact. The effect on similar research and possible implications are discussed.**

Keywords— ***Career research, professional online social network, data mining, human generated data***

I. INTRODUCTION

Since the beginning of the 20th century and the time of individualization, we no longer have to take up our parents' profession[1]. Instead, people can choose, plan and follow their own careers and career mobility has been growing [2]. Nevertheless, as much as the predestination of vocational choice has diminished, individual decisions have become more important [3]. To be successful, people feel they have to take the right steps at the right time [1], [4]. Besides the feeling of freedom, insecurity and the pressure on individuals to come to the right decisions increase [5], [6]. Nowadays, coaching and consulting services have become increasingly popular. To optimize their careers, to find success and eventually happiness, people long for career tricks or disclosed career patterns.

At the same time and due to the evolvement of data science methods, pattern detection is on the rise[7], [8]. The growing number of available data sets, especially in human produced data, leads researchers to employ text mining as a method to extract knowledge from diverse, semi-structured text data [7], [9]. Companies have discovered the new possibilities of data science and machine learning, too. In the recruiting process, algorithms filter appropriate candidates and machines support human resource departments in conducting job interviews [10]. Online CVs as published on LinkedIn or similar services are nothing but semi-structured data. They can be aggregated, clustered and applied to algorithmic models. Nowadays, users of LinkenIn or the like produce these data in great number. New data science methods allow to look for patterns in online CVs using text mining applied to a data set of online CVs. In this vein, the aim of our research (funded by the German Federal Ministry of Education and Research) was to detect factors that promote or hinder people's career development, eventually to unveil steps leading to a CEO position. This intention bases on the assumption that online CVs provide the opportunity to detect insights into career paths. We want to answer the following question: Do online CVs provide appropriate information to detect career patterns?

II. STATE OF THE ART

Using career platforms is considered crucial for both employers and applicants. Practitioners and researchers state that active participation on LinkedIn is a key determinant of success [11]. Nevertheless, many students and young professionals do not use LinkedIn to its full potential yet [12] although studies show that when exposed, they see the benefits of those platforms, especially compared to rather informal social media pages like Facebook [13]. This is why [14] claims, that if higher education institutions want to prepare their students for a successful career they should teach them how to use LinkedIn and career platforms properly.

At the same time, companies are "missing out on a huge opportunity to gain access to top talent" if they do not use LinkedIn [15]. In what McKinsey has called a "War for

Federal Ministry of Education and Research, Germany

ASONAM '19, August 27–30, 2019, Vancouver, BC, Canada

ACM ISBN 978-1-4503-6868-1/19/08...$15.00
https://doi.org/10.1145/3341161.3343681

Talent" already twenty years ago [16], companies struggle with talent acquisition and expand their activities beyond traditional recruiting practice. That is, many recruiters already use LinkedIn to look for talents. In addition, they have developed new methods using big data and data intelligence for recruiting purposes [17]. Such increasingly smart systems span from umbrella platforms that allow recruiters to look for candidates across all big career networks [18], recruiting robots that support at both screening and interviewing candidates [10] or algorithms that parse CVs to look for required skills [19]. However, up to our knowledge these tools do not use any deep learing methods but simply match the skills the employers require with the skills an applicant offers. What is more, they very often still need human control. Amazon experienced the danger of automation in 2014, when their recruiting algorithm structurally discriminated against women [20]. When this became public, John Jersin, Senior Director of LinkedIn Talent Solutions, was very clear about the issue of using AI in recruiting: "I certainly would not trust any AI system today to make a hiring decision on its own. The technology is just not ready yet" [20].

In 2012 LinkedIn filed a patent for a tool they developed that matched users' experience with different jobs, providing information about possible jobs and nicely presenting them in a "career path" [21]. Nevertheless, this tool is also not in use anymore – the reasons remain unclear.

For the last decade, career researchers have also been using the new emerging means of mining techniques and the increasing number of digital data for new perspectives on their topic. Heinze used online job advertisements to investigate on employer's requirements [22], [23] predicted scientists' movement between institutions based on their scientific profile, and Chopra et al. looked for gender differences in science and engineering by analyzing 60,000 application documents [24]. As online CVs are available in large number, text mining analysis of CVs has become a popular source of research. Nevertheless, the results are still far from generalizable: They concentrate on specific domains (e.g., the academic sector [25]) or remain limited in their focus (e.g., women in Saudi Arabia; [26]; the effect of international mobility on career success [27]. As far as our literature review has shown us, until now, nobody has yet analyzed large data sets of online CVs with deep learning methods or the like.

This is why, in the current research we use a large data set of online CVs to detect career patterns. In a very elementary approach, we focus on factors that promote or interfere with high hierarchy. In line with the aforementioned research, the plan was to use online CVs to shed light on career patterns and, eventually, ways to a senior executive position. The hypothesis was the following:

Hypothesis: Online CVs illustrate people's careers and therefore, they build an ideal data set for detecting career patterns.

III. DATA

100,006 online CVs in semi structured JSON format served as the basis of the analysis. The data stem from a large German CV platform and were downloaded via a respective API. The only condition for selecting a CV was that it contained words such as "manager", "leader", or "leadership", as the analysis required advanced CVs. We did not set any other preconditions for the data collection to gather a broad range of online CVs.

The data was structured into arrays. On the first level, these are "education" and "job", with subarrays such as "company" or "education station", that contain objects like "size of company", "industry", "title of job" or "school name", "degree", "subject", respectively. Users can extend their online CV by specifying as many education and job stations etc. as they wish. About half of the data fields consist of free text (e.g. "company name" or "job description"), the other half are categories (e.g. "career level" or "industry") or strings (beginning and end date of the respective stations).

IV. METHOD

The data set of CVs was loaded into MongoDB and analyzed with Python. During preprocessing, existing text and categorical information was transformed into either Boolean values or integers, resulting in more than 200 variables (such as "highest degree", "duration of education" or, as an example of a detailed variable, "job in same company as internship"). This process of preprocessing is actually indefinite, as there are always possible variables that could be specified in more detail than previous ones.

Within all the data fields, the variable indicating hierarchy is a categorical one. When filling out their CVs, users were able to indicate a career level for each of their career stations. They could choose one of five levels: "student/intern", "entry level", "professional experienced", "executive", and "senior executive". The career level of the current position served as the variable "current level" whereas the highest position ever reached was labeled as "highest level". The two variables showed similar effects as for many people the current position is also their highest yet. For the analysis of career success, those two variables served as the dependent variables.

To test the hypothesis of whether online CVs illustrate peoples careers and can therefore be used to detect career patterns, this paper will concentrate on two aspects of the CVs: educational information and internships. Most career trajectories contain information about those variables and other studies have already investigated on the impact of them on career success. Therefore, we will use these two aspects to demonstrate what kind of information one can retrieve from online CVs. To start with, we ran simple linear regression analyses on the dependent variables current level and highest level with educational and internship information as predictors.

V. Results

Education, in this data set, was described by multiple variables such as "has bachelor", "studied STEM subject" or "number of bachelors". In the following, one very general, basic variable "maximum degree" is presented. The variable represents the highest one of the following degrees someone has, from low to highest: "bachelor", "diploma", "master" or "PhD". It has to be noted that the analysis only considered the CVs when the degree could be identified as one of the above mentioned. Running a linear regression, the variable "maximum degree" showed a small but significant effect on the current career level, $b = 0.35$, $t(11136) = 13.22$, $p < .001$, $R^2 = .015$, and a similar effect on highest career level, $b = 0.40$, $t(14387) = 16.10$, $p < .001$, $R^2 = .018$.

Regarding internships the variables span from where ("internship in STEM"), to how long or with which effect ("same company after internship") someone made an internship. Here, again, one very general and straightforward variable is presented: whether someone made an internship or not. One linear regression was run on each of the two career level variables with "has internship" coded as 0 and 1 (no and yes, respectively) as predictor variable. The internship variable shows a significant and comparatively strong relation to the current career level, $b = -0.11$, $t(4181) = -8.10$, $p < .001$, $R^2 = .015$, and similarly to the highest career level someone ever had, $b = -0.09$, $t(5523) = 6.66$, $p < 0.001$, $R^2 = 0.008$. As having an internship is coded as 1, both relations were negative.

VI. Discussion

Looking at career research, both someone's academic degree and whether he or she has made an internship increase career prospects and employability [28], [29]. Regarding the first aspect, the association of degree and career level supports this assumption in this study. Concerning internships, the picture is different. Nowadays employers urge universities to put a focus on practical experience. Since Bologna, many curricula require an internship in the industry [30]. That is, the effects presented here should not lead us to the conclusion that internships negatively influence careers. One other explanation might be that twenty or thirty years ago, internships were not as relevant as they are today. That could have led to the fact, that people that hold senior executive positions today indeed did not do any internships. Another explanation would be a confounding of information and career level. More precisely, the higher someone climbed the career ladder, the more probably his or her information about internships is missing. Putting it the other way, only people that are comparatively low in career level insert information about internships. Degrees, in comparison, represent one of the most important information of one's career, which is why most CVs contain this information.

To investigate on the effect of career level and CVs of different detail, we created a variable called "cv quality". This variable spans from 0 to 1. As the scores of the sections job and education were multiplied with each other, a lack of information in either section led to a score of 0 whereas two stations in each section already led to a score of 1. The decision to use this very low threshold bases on the fact that most CVs were lacking information somewhere. Running a regression on current career level with CV quality as determinant, showed a negative effect, $b = -0.3$, $t(63511) = -33.00$, $p < .001$, $R^2 = .017$. That is, people that are currently in a high position include less information in their CVs than people that are in a lower position. One reason might be that the proposed assumption of what online CVs show is wrong. It might be true that someone who is just starting the career today fills out the online form and keeps adding information. However, at the time current senior executives filled out their online CV they were already high up and did not include information about their young years. Their online CVs did not develop along with their careers, as it might be the case with online CVs of younger people nowadays.

These results make us reason that our hypothesis has to be declined: Online CVs do not objectively illustrate people's careers and therefore do not serve as an ideal data set to detect career patterns. Rather career level and information provided in online CVs are confounded and systematically depend on each other.

When analyzing online data, we "need to be careful perceiving communication over social networks and digital footprints as 'authentic'. People's posts, tweets, uploaded photographs, comments, and other types of online participation are not transparent windows into their selves; instead, they are usually carefully curated and systematically managed" [9, p. 465]. People know they have to present themselves favorably in their online CVs [31]. Furthermore, the kind of information considered favorable differs depending on the career level. Having little experience, any internship is better than none. A senior manager, in contrast, should present recent successes than first steps.

VII. Limitations and other approaches

In comparison to the total number of LinkedIn users, which is currently 630 million (LinkedIn 2019), a sample size of 100.000 CVs is relatively small. Having the opportunity to use meta data such as CV quality for the analysis could diminish the problem of structural differences between career levels. Unfortunately, we did not have access to a larger set of data. Nevertheless, self-presentation persists to be a problem as long as the data relies on public self-disclosure. To handle this, neutral sources of career trajectories are necessary. E.g., German DAX-listed companies have to publish the CVs of the members of their boards of managements [32]. Nevertheless, these are only few data sets (n = 170) and thus might be too few for means of pattern detection.

In a next step, machine learning methods could be used to further investigate on career pattern. Nevertheless, even the most advanced methods will not lead to better results if the underlying data is biased and confounded.

VIII. Conclusion

Analyzing CVs to make predictions about different aspects of careers, such as the path to a management position, seems reasonable. At the same time, our brief results show that when analyzing human produced online data, one has to be very careful about what this data really contains. Although a higher degree positively predicted career level, having an internship showed a negative relation with career level. Rather than assuming that people should not make internships, it seems plausible to ascribe this negative association to other factors. Although the presented results need further exploration, the results show that online CVs do not depict career trajectories in an objective way. What is more, the kind of information someone presents and his or her career level interrelate. Other researchers have pointed towards the fact that on the internet, users deliberately choose the information they publish. Further research should back data science approaches with alternative sources such as meta data or qualitative studies. The presented results shows that when using human generated data, it is important to control for and think of biases and constraints they might contain– such as self-presentational intentions.

References

[1] U. Beck and E. Beck-Gernsheim, ‘Institutionalized individualism and its social and political consequences’, *Ann Arbor Univ. Mich. Press*, 2002.

[2] S. T. Lyons, L. Schweitzer, and E. S. W. Ng, ‘How have careers changed? An investigation of changing career patterns across four generations’, *J. Manag. Psychol.*, vol. 30, no. 1, pp. 8–21, Jan. 2015.

[3] A. Hirschi, ‘The Fourth Industrial Revolution: Issues and Implications for Career Research and Practice’, *Career Dev. Q.*, vol. 66, no. 3, pp. 192–204, Sep. 2018.

[4] E. S. W. Ng, L. Schweitzer, and S. T. Lyons, ‘New Generation, Great Expectations: A Field Study of the Millennial Generation’, *J. Bus. Psychol.*, vol. 25, no. 2, pp. 281–292, Jun. 2010.

[5] S. R. Madara, P. Maheshwari, and C. P. Selvan, ‘Future of millennial generations: A review’, in *2018 Advances in Science and Engineering Technology International Conferences (ASET)*, 2018, pp. 1–4.

[6] M. L. Savickas, ‘Life Design: A Paradigm for Career Intervention in the 21st Century’, *J. Couns. Dev.*, vol. 90, no. 1, pp. 13–19, 2012.

[7] X. Wu, X. Zhu, G. Wu, and W. Ding, ‘Data mining with big data’, *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97–107, Jan. 2014.

[8] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, ‘Deep learning applications and challenges in big data analytics’, *J. Big Data*, vol. 2, no. 1, pp. 1–21, Feb. 2015.

[9] L. Manovich, ‘Trending: The Promises and the Challenges of Big Social Data’, in *Debates in the Digital Humanities*, M. K. Gold, Ed. University of Minnesota Press, 2012, pp. 460–475.

[10] ‘Robot Vera’. [Online]. Available: https://ai.robotvera.com/static/newrobot_en/index.html. [Accessed: 12-Mar-2019].

[11] M. B. McCabe, ‘Social media marketing strategies for career advancement: An analysis of LinkenIn’, *J. Bus. Behav. Sci.*, vol. 29, no. 1, pp. 85–99, 2017.

[12] A. R. Slone and A. L. H. Gaffney, ‘Assessing students’ use of LinkedIn in a business and professional communication course’, *Commun. Teach.*, vol. 30, no. 4, pp. 206–214, Oct. 2016.

[13] W. Kim and K. Malek, ‘Social networking sites versus professional networking sites: Perceptions of hospitality students’, *J. Hum. Resour. Hosp. Tour.*, vol. 17, no. 2, pp. 200–221, Apr. 2018.

[14] R. S. Bridgstock, ‘Educational practices for employability and career development learning through social media: Exploring the potential of LinkedIn’, in *Practice Futures For the Common Good*, J. Higgs, D. Horsfall, S. Cork, and A. Jones, Eds. Rotterdam, The Netherlands: Sense-Brill Publishers, 2018.

[15] F. Holland, ‘Recruitment of talent: Social media – friend or foe?’, *HR*, 24-Oct-2012. [Online]. Available: https://www.hrmagazine.co.uk/article-details/recruitment-of-talent-social-media-friend-or-foe. [Accessed: 25-Feb-2019].

[16] E. G. Chambers, M. Foulon, H. Handfield-Jones, S. M. Hankin, and E. G. Michaels, ‘The war for talent’, *McKinsey Q.*, no. 3, pp. 44–57, 1998.

[17] T. Chamorro-Premuzic, D. Winsborough, R. A. Sherman, and R. Hogan, ‘New Talent Signals: Shiny New Objects or a Brave New World?’, *Ind. Organ. Psychol.*, vol. 9, no. 3, pp. 621–640, Sep. 2016.

[18] ‘Home - Talentwunder’. .

[19] A. Cernian, D. Carstoiu, and B. Martin, ‘Semi-automatic Tool for Parsing CVs and Identifying Candidates’ Abilities and Competencies’, in *International Conference on Education, Management and Applied Social Science (EMASS 2016)*, Beijing, China, 2016, vol. 0.

[20] J. Dastin, ‘Amazon scraps secret AI recruiting tool that showed bias against women’, *Reuters*, San Francisco, 10-Oct-2018.

[21] R. Jurney *et al.*, ‘Methods and systems for exploring career options’, US20120226623A1, 06-Sep-2012.

[22] U. Heinze, *Web content mining for analyzing job requirements in online job advertisements*. Aachen: Apprimus, 2015.

[23] C. James, L. Pappalardo, A. Sirbu, and F. Simini, ‘Prediction of next career moves from scientific profiles’, Feb. 2018.

[24] S. Chopra, M. Mirsafian, A. Khan, and L. Golab, ‘Gender differences in science and engineering: A data mining approach’, in *Workshop Proceedings of the EDBT/ICDT 2019 Joint Conference*, London, 2019.

[25] A. Geuna *et al.*, ‘SiSOB data extraction and codification: A tool to analyze scientific careers’, *Res. Policy*, vol. 44, no. 9, pp. 1645–1658, Nov. 2015.

[26] W. Bawazeer and H. M. Gunter, ‘Using the curriculum vitae in leadership research’, *Manag. Educ.*, vol. 30, no. 2, pp. 74–78, Apr. 2016.

[27] N. Schiffeler, T. Sommer, V. Stehling, A. Richert, and S. Jeschke, ‘Internationalisation in career development – A data science analysis of manager CVs’, in *Proceedings of the ICERI 2017 10thInternational Conference of Education, Research and Innovation*, Seville (Spain), 2017, pp. 5482–5490.

[28] K. U. Mayer, W. Müller, and R. Pollak, ‘Germany: Institutional change and inequalities of access in higher education’, in *Stratification in Higher Education: A Comparative Study*, Y. Shavit, Ed. Stanford: Stanford University Press, 2007, pp. 240–263.

[29] E. L. Shoenfelt, N. J. Stone, and J. L. Kottke, ‘Internships: An Established Mechanism for Increasing Employability’, *Ind. Organ. Psychol.*, vol. 6, no. 1, pp. 24–27, Mar. 2013.

[30] M. Winter, ‘Praxis der Studierenden und Praxisbezug im Studium. Ausgewählte Befunde der Hochschulforschung zum “neuen” und “alten” Studieren’, in *Nach Bologna: Praktika im Studium - Pflicht oder Kür? Empirische Analysen und Empfehlungen für die Hochschulpraxis*, W. Schubarth, K. Speck, and S. Andreas (Hrsg.), Eds. Potsdam: Universitätsverlag Potsdam, 2011, pp. 7–44.

[31] J. van Dijck, ‘“You have one identity”: performing the self on Facebook and LinkedIn’, *Media Cult. Soc.*, vol. 35, no. 2, pp. 199–215, Mar. 2013.

[32] T. Neuscheler, ‘Manager-Lebensläufe: Jeder fünfte Dax-Chef begann seine Karriere als Unternehmensberater’, *Frankfurter Allgemeine Zeitung*, 07-Jun-2018.