What Do Engineers Do All Day? Using LinkedIn Profiles to Study Engineering Work

Christopher Gewirtz  
Dpt. of Engineering Education  
Virginia Tech  
Blacksburg, VA, USA  
chrage@vt.edu

Marie Paretti  
Dpt. of Engineering Education  
Virginia Tech  
Blacksburg, VA, USA  
mparetti@vt.edu

Jennifer Case  
Dpt. of Engineering Education  
Virginia Tech  
Blacksburg, VA, USA  
jencase@vt.edu

Lee Vinsel  
Dpt. of Science and Technology in Society  
Virginia Tech  
Blacksburg, VA, USA  
leev@vt.edu

Jessica Deters  
Dpt. of Engineering Education  
Virginia Tech  
Blacksburg, VA, USA  
jdeters@vt.edu

Omer Zulfiqar  
Dpt. of Computer Science  
Virginia Tech  
Fairfax, VA, USA  
omer95@vt.edu

Abstract—This full paper examines the potential for using LinkedIn profile analysis to understand engineers’ everyday work activity. Calls for change in engineering education are often framed around perceptions of the changing nature of engineering work in present times. Existing research on engineering work, which draws on interviews, ethnography, case studies, or surveys, consistently highlights the social and contextual dimensions of practice, but also suggests that engineering work may have more to do with maintaining existing systems than with creative development of new systems. To add to this growing body of research, in this exploratory study we turn to a novel data source: publicly available descriptions of engineering work available on social networks and job posting sites. Such data sources, if found to be useful, could significantly expand the reach of research on engineering work and provide access to much larger data sets than those obtained through current – predominantly qualitative – methods. This study is a preliminary look at using LinkedIn for these purposes. Specifically, we ask the question, “How does data about engineering work from LinkedIn compare to findings from in-depth interviews?” To answer this question, we use interviews with 15 new engineers at six and 12 months of work, and compare them to participants’ LinkedIn profiles. The interviews, collected as part of a separate study, focused on newcomers’ workplace responsibilities and challenges. Analysis focused on inductive coding to categorize engineers’ work experiences and environments. The findings were then compared to participants’ LinkedIn profiles to identify gaps and overlaps.

Keywords—Public Datasets, Engineering Work, Natural Language Processing

I. INTRODUCTION

In 2006, when the Journal of Engineering Education published its special report, “The Research Agenda for the New Discipline of Engineering Education” [1], the first of its five proposed research areas was “Engineering Epistemologies.” The journal identified the need for research “describing and defining the nature of engineering work as a professional enterprise and articulating the roles of engineers in that work” (p. 259). Such work, it argued, is central to our ability to develop educational experiences that effectively prepare students for practice. In 2014, Trevelyan [2], as part of the edited volume Engineering in a Global Context called for a theory of engineering practice, built from empirical research, that could provide intellectual leadership for researchers and practitioners alike to inform not only how we effectively educate future engineers, but how current engineers operate effectively within their organizational context. In 2018, the National Academy of Engineering (NAE), in their report Understanding the Educational and Career Pathways of Engineers [3], found that “data gaps hinder understanding of the engineering educational and career paths” (p. 8). Focusing on large scale “administrative” data sets (i.e. those from educational, government, and other organizations), they called on “researchers and policymakers … to identify and build on administrative data resources to establish a better empirical foundation for research on the educational and career paths of engineers using a wide variety of definitions of what it means to be an engineer” (p. 9).

In other words, for much of the 21st century, voices from across the engineering education research community have been calling for increased attention to engineering work. And while research on practicing engineers and engineering work has grown over that time, it still lags behind research on student experiences in terms of both the number of publications and the breadth of the data sources.

One untapped resource in this arena, we suggest, is the data available through professional networking and job
posting sites that are amenable to newer data scraping and analysis techniques such as natural language processing. Such data could, in particular, address the NAE’s call for better use of large-scale data sets and inform the kind of theory that Trevelyan and others have called for. But the value of such data remains unknown: what, that is, can such sites tell us about engineers, the work they do, and the skills they need?

To that end, we present an exploratory study of LinkedIn data that compares participants’ representations of themselves on this site to descriptions of their workplace practices and experiences elicited in interviews.

II. BACKGROUND LITERATURE

While research on engineering work and the experiences of practicing engineers is more limited than research on student experiences, it has been a persistent thread both within and beyond engineering education in recent decades, as captured in the review chapter on “Professional Engineering Work” by Stevens et al. in the 2014 Cambridge Handbook of Engineering Education Research [4].

Much of this work comes from smaller-scale qualitative studies using ethnographic approaches. For example, Anderson et al. [5] conducted observations and interviews with engineers at six companies in the Midwest, highlighting the highly social nature of engineering work and the simultaneous tendency of participants to discount much of this social activity as “not engineering.” Similarly, Vinson and colleagues recently conducted ethnographic studies of early engineers at six different companies, examining the types of problems new engineers encounter in their transition [6]. The latter study, as with earlier work by other scholars, highlighted the importance of interpersonal relationships and networking in engineering work – a practice often noted as disconnected from the ways engineers learn at school. More directly, in studies comparing design experiences at school and at work, Lauff et al. draw on observation and interview data to highlight the critical impact of context, including spatial and temporal practices as well as workplace cultures, that play key roles in shaping design and pose challenges for educators seeking to prepare students for moving from school to work [7].

Several larger-scale studies have also emerged. Most notably, the Engineering Pathways Study (EPS) [8] and the Professional Engineering Pathways Study (PEPS) [9] (both following from the Academic Pathways Study (APS) [10]), which included surveys and interviews conducted longitudinally to track participants’ career trajectories. Drawing on data from the APS and EPS, for example, Brunhaiver examined interviews from 57 working engineers at four different organizations, along with longitudinal interviews with 9 participants from 3 institutions [11] to explore the gaps between school and work, surfacing a full range of professional (non-technical) skills that participants found central to their success. More recently, the Capstone to Work (C2W) project followed approximately a hundred graduates from four institutions through their first year of work using a combination of weekly surveys and periodic interviews [12]. The findings identified challenges in self-directed learning and interpersonal communication as most common for new engineers in this transition.

In addition, Trevelyan has amassed a data set of some 300 interviews with practicing engineers across 3 continents. That research has led to rich descriptions of engineering practice that move well beyond the social/technical binary to consider the ways in which engineers manage themselves, their teams, and their work in complex process of technical coordination [13] and the myriad of ways beyond innovation that they bring value to the organizations that employ them [14].

Beyond such research, national census and survey data provide some indicators of the demographics of engineers in today’s workforce in the U.S. [15], including numbers of graduates, median salaries, industry sectors, demographics, and related data. But as the NAE report points out, such data is necessarily limited [3], and can be hampered by response rates and general survey fatigue.

Such work has provided useful insights into the things engineers do each day and the paths they follow through their careers. Still, given that the U.S. alone graduated more than 130,000 engineers in 2018 [16], and these engineers move into an incredibly diverse array of companies and positions, the existing research represents only a small fraction of the field and leaves many questions about the work of engineers – and the skills needed for that work – unanswered.

To that end, and particularly in response to the NAE’s calls for large-scale data sets, our team has been exploring the potential of public data to contribute to our understanding of engineering work. While such datasets hold significant potential for studying engineering work at a very large scale, many questions remain about exactly what such public representations tell us and how they compare to the kind of rich data obtained through interviews and ethnographies. As part of that work, this study compares data from one such data set, LinkedIn profiles, to information about engineering work provided through longitudinal in-depth interviews.

III. DATA COLLECTION AND ANALYSIS

A. Data Collection

For this study, participants were recruited from a pool of individuals who had participated in a prior engineering education study, the Capstone to Work (C2W) project [12]. Participants were graduates from three engineering institutions in the United States. Of the 15 participants who consented to the use of their data for this study, 8 were women, 7 were men; 3 worked at small companies and startups, 2 at medium sized companies, and 9 at multinational companies (one company size was unreported); 13 were White, 2 were Asian. From these participants, two forms of data were used to approach the research questions: interviews and LinkedIn profiles.

Dataset 1: Interview Data

As already mentioned, interviews were collected as part of the C2W study concerning engineering students’ transitions into work [12]. Each participant had previously completed
interviews at both six months and 12 months of work. Among other questions, the interviews asked participants about their challenges and accomplishments at work, and the learning they needed to navigate their workplace tasks. These interviews were recorded in person or over the phone, transcribed, and then cleaned of identifying information.

Dataset 2: LinkedIn Profiles

For the second source of data, participants were contacted to ask for their consent to pair their previously collected interviews with their publicly available LinkedIn profile. Once participants gave their consent, their LinkedIn profiles were downloaded, turned into a text file, cleaned of identifying information, and then assigned an identifier to pair the participant’s profile with their previously completed interviews. At the time of collection of profiles, participants had graduated 2.5 years prior, and had been at work for roughly 2 years.

B. Comparative Analysis

For the exploratory analysis presented in this paper, we chose to compare the two datasets by determining the following, case by case:

A. what aspects of work discussed in the interviews were represented in the LinkedIn profile,
B. what aspects of the interviews were missing from the LinkedIn profile, and
C. what aspects were present in the LinkedIn profile but missing in the interviews.

To facilitate this comparison, the interview data was condensed using a coding scheme that included both deductive and inductive elements. Starting with a broad deductive approach sensitized the researchers to the activities that comprised participants’ work, and inductive sub-codes allowed for the particularities of engineering work to be compared between profile and interview [17]. For a given participant, either the 6 month or 12 month interview was coded, depending on whether participants mentioned their work activities having changed in their 12 month interview.

The analysis presented in this paper involved individually comparing participants’ LinkedIn profiles to their coded interviews. Profiles were coded by the first author. During this process, the author also simultaneously wrote memos to speculate on theoretical connections and patterns between profiles [18]. Once all individual interview/profile papers were coded, second cycle coding was done to determine patterns across profiles.

IV. Results

This section presents the results of the analysis organized according to the three aspects of this comparative analysis outlined above. Looking across profiles, we summarize our findings in each category with themes. Throughout this section, the direct quotes from participants’ LinkedIn profiles are limited or not present because they were found to threaten participants’ anonymity.

A. Represented in both Interview and LinkedIn Profile

In this section we consider information that was generally found in both the interview data and the LinkedIn profiles.

A key part of all LinkedIn profiles was a fairly standard form of representations of participants’ titles and qualifications, and this was information matched the information participants provided in the interview data. Participant 4A’s profile is used as an example below:

[Company]
1 year 5 months
Sales Engineer
April 2019 - Present

Beyond this introductory information, we noted the text that followed for many profiles tended to be structured around keywords that elaborated on key aspects of the participant’s role. These were elements that were also noted in the interview data, for example:

...lean principles...optimization...optimize productivity
and maximize... [Participant 1A, Profile]

...Customer Service... strong consulting professional ... [Participant 6A, Profile]

In the case of participant 6A, their profile mentioned “Microsoft Word” and “Excel” as software they were skilled with, but these skills were not a relevant aspect of the work they discussed in their interview, nor did those skills set them apart from other participants. However, their mentions of “Customer Service” and “consulting” did correspond to work they discussed during the interview. These keywords were unique to participant 6A’s profile among the 15 profiles examined and reflected the ways in which their work was unique among participants as evident in interviews. Thus, these keywords provide a way to categorize participants as engaged in a certain type of work.

Overall, of the 15 profiles examined, 4 were judged to fully capture the descriptions of work that participants had provided in their interviews, including both technical and professional dimensions (e.g. Documentation and “customer service”). These four profiles, usually including a paragraph starting with something like “My responsibilities include....” gave a good indication of the breadth of participants’ work and the specific ways that they interacted with the technologies and people that surrounded them. They also described their work in enough detail as to differentiate their work from engineering work “in general.” Perhaps most importantly for this study, the descriptions provided on their profiles closely aligned with the ways in which they described their responsibilities in the interviews.
B. Present in Interview but missing in LinkedIn Profile

The types of information missing from LinkedIn Profiles varied to some extent based on the type of profile. While all profiles had information about participants’ titles and qualifications, for some participants, this was practically all the information yielded by the profile, as in the case below for participant 11A:

[Participant 11A], EIT LEED GA
Mechanical Engineer at [Company]

In these cases, the overall profile was referred to as a “stub”; the profile technically existed but provided minimal information. Of the 15 profiles examined, 5 had almost no information beyond participant’s title, degree and current place of work – occasionally these profiles also listed skills or software the participant was proficient in. Among participants who did have “stub” profiles, there seemed to be an indication in their interviews that they had plans to stay at the company for a long time, in some instances the rest of their career, as Participant 11A explains:

I’m planning on staying and [Company] for a long time or at least the next three to five years so I’m trying to figure out how I can become the best employee possibly here, and then how I can change [Company] in the future and how I think would be fit. [11A, Interview]

Such long-term plans may lessen participants’ incentives for updating their profiles, and could mean that LinkedIn data is not useful for capturing engineers who hope to remain at a single company throughout their career.

Aside from the 5 “stub” profiles and the 4 full description profiles, the other 6 profiles had descriptions of work that fell somewhere in the middle. In 4 of these profiles, participants did not describe their current position beyond title, though they did have a summary from which information could be drawn about their current work. One profile (12A) had a one sentence description of their work. One profile (8A) had a detailed, paragraph-length description of their work but did not mention significant work activities concerning budgeting and learning how systems work.

In general, these profiles without full descriptions underrepresented the social aspects and professional skills required for participants’ work, even when these aspects were emphasized as essential during the interview. This was the case for participant 2A, for example, who made no mention of his drive to be a reliable teammate in his LinkedIn profile, but spent most of the interview discussing that as an approach to make up for what he felt was his lack of technical expertise.

…I think it's nice for them to have someone else to help get the work done. But sometimes you kinda have to ask for it, say "Hey, I don't have a lot on my plate. I can afford to take another project or two. Why don't you let me help you with this or that?"[…] I don't have skills that a lot of people in the office [have]. So I'm trying to work around ... That's why I try to be reliable. I try to do all the other things right because I don't have a lot of experience yet... [2A, Interview]

These profiles with partial representations often (but not always) contained keywords to indicate general technical activities, such as “structural analysis,” but had no indications of activities or practices aligned with collaboration, management, interpersonal skills, or related professional practices.

C. Present in LinkedIn Profile but Missing in Interview

One aspect that seemed present in multiple LinkedIn profiles but not the corresponding interviews was the use of highly specific technical language, such as:

…RFIs and RFPs…
…Security and user profile management…
…using Ansys…

In contrast to the keywords discussed in a previous section, this language was not used in the interviews and does not necessarily indicate the work activities participants discussed. The specific technical language does indicate the field and technologies that participants were working with, but does not indicate how the participant interacted with that technology or practice, how they fit into their community of practice, nor what activities they engaged in at work.

D. Summary

The table below summarizes the coding of the 15 profiles used for this study. The table also provides examples of both the keywords (which might be used to distinguish a participant’s work) and specific technical language (which indicates field of work, but only rarely work activity).

Table 1: Summary of Coded LinkedIn Profiles

<table>
<thead>
<tr>
<th>ID</th>
<th>Profile Type</th>
<th>Keyword</th>
<th>Specific Technical Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>Full Description</td>
<td>Optimization</td>
<td>Forklift workload</td>
</tr>
<tr>
<td>2A</td>
<td>Summary Only</td>
<td>N/A</td>
<td>Signal Processing</td>
</tr>
<tr>
<td>3A</td>
<td>Stub</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>4A</td>
<td>Full Description</td>
<td>Sales</td>
<td>RFI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interactions</td>
<td></td>
</tr>
<tr>
<td>5A</td>
<td>Summary Only</td>
<td>Cost-Benefit</td>
<td>Business</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Analyses</td>
<td>Intelligence</td>
</tr>
<tr>
<td>6A</td>
<td>Summary Only</td>
<td>Customer</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Service</td>
<td></td>
</tr>
<tr>
<td>7A</td>
<td>Stub</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
The table below summarizes the features of each of the three types of profiles present in this dataset, compared to the interviews that each participant completed. Categorizing each of the profiles included in the study, 5 were stubs, 6 were general summaries of participants’ past work or partial descriptions, and 4 were full descriptions of participants' current position.

Table 2: Features of each type of profile

<table>
<thead>
<tr>
<th>Job Title &amp; qualifications</th>
<th>Technical Activities</th>
<th>Social/Professional Activities</th>
<th>Specialized technical language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stub</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary or partial description</td>
<td>x x x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Full description</td>
<td>x x x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Interviews</td>
<td>x x x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


All forms of data included mention of participants’ current titles and their qualifications.

For profiles which included a discussion of work, either in a summary or a full description, there were generally keywords that corresponded to the technical work activities discussed in interviews (with the exception of profiles for participants 2A and 13A, whose summaries did not include any work activities).

Generally, summary profiles overemphasized technical aspects of work in comparison to social aspects when compared to interviews (although 6A did mention that he was skilled in customer service among other technical skills). Full description profiles all had sentences and keywords that corresponded with the social activities of work which were discussed in interviews.

Finally, summary profiles and full description profiles both included specific technical language which was connected with the technology participants worked with, or the field they worked in (with the exception of profiles for participants 6A and 13A, whose summaries did not include specific technical language). These instances of specific technical language were generally missing in participants’ interviews concerning their work activity.

V. DISCUSSION

This exploratory study sought to assess the potential value of public data by examining LinkedIn profiles to answer research questions about engineers, the work they do, and the skills they need. To that end, we compared participants’ representations of themselves on their LinkedIn profiles to descriptions of their workplace practices and experiences elicited in interviews.

The results show strong correspondence between these two datasets on the very basic matters of job titles and the keyword descriptors for skills. Moreover, profiles that include more than this basic information typically also captured the technical dimensions of engineering work. However, consistently with findings of Trevelyan [13] and others [e.g., 5], most LinkedIn profiles excluded what we term the “social” or “professional” aspects of engineering work. That is, the more limited the information on the profile, the more likely it was to eliminate the social dimensions that participants in their C2W project interviews. At the same time, the LinkedIn profiles were more likely to use highly specialized technical language and industry-specific jargon. Overall, these LinkedIn profiles are in line with the normative “ideology” of engineering, as participants tended to separate and downplay (or completely omit) the social aspects of their work [4, 19]. However, there are those few full description profiles which better represent the scope of the sociotechnical work that participants discussed in their interviews.

This brings attention to the issue that, in order to use public datasets to understand engineering work, it is important to understand the wide variation in types of information participants present for public consumption. Profiles in this study ranged from minimal to extended, and the scope of the profile corresponded to the correlations between the profile and the interview data. Notably, the scope of the profile may be heavily influenced by the author’s purpose. Previous work investigating the purposes of LinkedIn profiles found that these profiles are (unsurprisingly) used for purposes such as self-promotion and networking [20, 21]. Such uses would also explain the presence of “stub” profiles among participants who hoped to remain with their current company for an extended period. Participants who did not intend on re-entering the job market would have less incentive to update.
their LinkedIn profile. At the same time, when the profiles moved beyond the stub to present summaries of work, these representations seemed to align with stereotypical representations of engineering workplaces, emphasizing individualized work and specific technical expertise. This is the case even when interviews with these participants placed greater importance on other traits (such as being a reliable team member), or other activities (writing design reports, following strict protocols). One potential explanation for this emphasis is that profiles represent the parts of their work that participants most want to continue doing, the practices they believe are most marketable, or are most valuable. Again, the less descriptive profiles seem to comply with the norm of social-technical dualism in engineering, which frames technical work as separate from and prized above social work [4, 19].

Thus, in interpreting the results from an analysis of LinkedIn profiles, or any publicly available representation of engineering work, the findings need to be considered in light of the potential scope and purposes of the representation itself. In that sense, LinkedIn profiles, especially briefer ones, may only confirm normative conceptions of engineering work.

At the same time, though the context of LinkedIn suggests that the profiles were intended for purposes other than accurate portrayals of engineering work activities, they still demonstrate potential as a new method for understanding engineering work. The small sample examined in this analysis did include multiple profiles that were more expansive and that captured a broader range of skills and activities. Examining a larger data set could yield automated analytic approaches that would first categorize profiles (e.g. as stub, summary, or full) prior to using methods such as Natural Language Processing (NLP) to analyze work representations. Given the presence of representative keywords, scaling up the analysis to many more profiles with computational methods such as NLP may be able to reveal meaningful details about engineering work. Such details include the variation in activities and categories of work that engineers engage in, the percentage of engineering profiles that fit those categories, and the educational degrees that lead to each category of work. Even the analysis of variation in engineers’ titles and qualifications may prove valuable in understanding how education prepares engineering students for positions in and outside of engineering.

Ultimately, though the profiles may not have the primary purpose of being an accurate portrayal of engineering work, they can still provide information about engineering work, especially information regarding how work may vary among participants.

VI. CONCLUSION

In conclusion then, this exploratory study delivers a fairly cautious positive assessment on the value of publicly available data such as that on LinkedIn, for informing research into the nature of engineering work, thus responding to calls from the NAE [3] and others. Scaling up the study of online resumes or other publicly available data can fill in the gaps of current methodological approaches to studying engineers at work. The approach avoids the sample size issues that limit the generalizability of smaller scale ethnographic studies. Because the data is publicly available and has no recruitment or solicitation process, the approach also avoids some limitations of large-scale surveys. Additionally, the approach is more exploratory and allows for findings to emerge inductively from the data in ways that are more similar to ethnography or interview-based studies.

With more research, the analysis of public datasets may serve to complement existing approaches which use surveys and interviews, forming a “middle ground” which preserves the exploratory, inductive nature of interviews, while also being grounded in large datasets similar to those used in surveys.

A key consideration, which is, of course, also pertinent to the surveys and interviews that have been the mainstay of such research to date, is that an individual’s intent strongly influences how they represent their work. Further research will seek to elucidate more clearly informed methods for analyzing and working with such data, including exploring the potential for using Natural Language Processing to automate some of these analyses, and other potential publicly available datasets.

ACKNOWLEDGMENT

This material is based upon data collection from the Capstone to Work project supported by NSF under Grant Number 1607811. Further research was supported by Virginia Tech’s ISCE Scholars 2019-2020 program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, nor Virginia Tech.
References


