A Network Data Science Approach to People Analytics

Nan Wang, Deepmacro LLC, USA
Evangelos Katsamakas, Gabelli School of Business, Fordham University, USA

ABSTRACT

The best companies compete with people analytics. They maximize the business value of their people to gain competitive advantage. This article proposes a network data science approach to people analytics. Using data from a software development organization, the article models developer contributions to project repositories as a bipartite weighted graph. This graph is projected into a weighted one-mode developer network to model collaboration. Techniques applied include centrality metrics, power-law estimation, community detection, and complex network dynamics. Among other results, the authors validate the existence of power-law relationships on project sizes (number of developers). As a methodological contribution, the article demonstrates how network data science can be used to derive a broad spectrum of insights about employee effort and collaboration in organizations. The authors discuss implications for managers and future research directions.

KEYWORDS

Collaboration, Data Science, Employee Effort, Employee Engagement, Knowledge Work, Network Analysis, Network Science, People Analytics, Power Laws, Productivity, Talent Analytics

1. INTRODUCTION

The best companies compete on people, or talent analytics (Davenport et al., 2010). They maximize performance, engagement and retention of top talent, and they get the most value from their talent, their most important asset for gaining competitive advantage. Professional sports teams have been leading users of analytics, but the talent analytics trend spans across all industries, with examples such as Google, Best Buy, Sysco, Starbucks and many others. Companies seek to establish a data-driven employee management, which includes employee data management, transparent HR processes, HR data segmentation, and linking employee performance to organizational objectives.

This paper proposes a network data science approach to people analytics. Network science has grown into one of the most important interdisciplinary fields of research (Newman, 2010), which has led to novel insights across many problem domains. We will use the term network data science to emphasize that our approach is a data science approach, and maybe differentiate clearly from pure network theory. We define network data science as the use of data science algorithms, techniques and tools in the modeling and analysis of network data. Our main focus is the estimation, analysis and
visualization of employee performance and collaboration patterns. Estimating the performance of information (knowledge) workers is an important consideration in all organizations (Aral et al., 2012). The goal is to understand what the actual patterns of knowledge work in organizations are and what can we learn to improve organizing and management for superior talent engagement and performance.

While the proposed approach could be used in any organization in which work is organized based on projects, the data we use to illustrate an application of the methodology come from a software development organization. In this context, evaluation of developers’ achievement and contribution has been proved to be critical for the long-term development of companies as it is directly related to employee morale, overall productivity and creativity. Poor capability to discover and monitor coding performance prevents technical companies from transforming information on network activity and infrastructural capabilities into strategic knowledge (McGarry et al., 2001). The questions we attempt to answer include identification of developers’ contribution and role, project membership, corporate cohesion, collaboration patterns and software development productivity. We demonstrate how a network data science approach can help companies gain rich insights about these issues.

We model and analyze developer effort and contribution through a weighted bipartite graph. Many real-life networks are widely recognized to be intrinsically weighted. Thus, characterizing features and structures in weighted networks, has important practical significance. We show how graph properties can be used to gain insight about productivity, and roles. We model collaboration through a developer-developer graph. We characterize graph centrality properties, and we validate a power-law distribution in the graph. Then we use and evaluate the performance of a number of algorithms for community detection. Lastly, we model network dynamics via a statistical approach and provide insights about the network over time.

The rest of the paper is structured as follows. Section 2 provides a background on people analytics, and describes the dataset and its preparation. Section 3 presents analysis of the developer-repository graph, and section 4 focuses on the developer-developer network. Section 5 analyzes complex network dynamics. Section 6 discusses the results and contribution of the paper and concludes with suggestions for future research.

2. BACKGROUND: PEOPLE ANALYTICS AND DATASET

This research proposes a network data science approach to people analytics. In this section, we present a background on people analytics, and briefly describe the research dataset.

2.1. People Analytics

Organizations compete on people analytics (Davenport et al., 2010). They define human-capital investment metrics, mine data and examine “what-if” scenarios to anticipate workforce needs, and maximize talent retention. They make more real-time talent related decisions. As an example, Google created a people analytics function, and through its Project Oxygen determined the attributes of successful managers. They also found that Googlers stay at Google for three reasons: mission, quality of people, and the chance to build their skill set, especially around leadership and entrepreneurship.

Much of the academic literature uses the term HR analytics, instead of people, or talent analytics. Aral et al. (2012) find empirical support for three-way complementarities between information technology, performance pay and HR analytics practices. They develop an economic model explaining how these three practices work together, and analyze data on human capital management software adoption over 11 years and across 189 firms.

Angrave et al. (2016) caution that HR analytics will not deliver the promised transformational change, unless current pitfalls are recognized and better methods and approaches are developed. Structured data from HRIS (Human Resources Information Systems) must be combined with external unstructured data in order to maximize business value from HR data. A key issue is the use of analytics to create, capture, leverage and protect value from HR data. Another important issue is measuring
Education in Conflict Resolution Using ICT: A Case Study in Colombia
Ana Dolores Vargas Sánchez and Luis Eduardo Veloza Chamucero (2017). Teaching Cases Collection (pp. 29-43).
www.igi-global.com/article/education-in-conflict-resolution-using-ict/181072?camid=4v1a

Utilization of Information Resources for Business Success: The Knowledge Sharing Model
www.igi-global.com/article/utilization-information-resources-business-success/1306?camid=4v1a