



**HAL**  
open science

# People Analytics and Organizational Network Analysis for Better Decision Making and Human Resource Management

Abdel-Rahmen Korichi

► **To cite this version:**

Abdel-Rahmen Korichi. People Analytics and Organizational Network Analysis for Better Decision Making and Human Resource Management. Computer Science [cs]. Université Claude Bernard - Lyon I, 2023. English. NNT : 2023LYO10153 . tel-04561724

**HAL Id: tel-04561724**

**<https://theses.hal.science/tel-04561724>**

Submitted on 27 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE de DOCTORAT DE**  
**L'UNIVERSITÉ CLAUDE BERNARD LYON 1**

**Ecole Doctorale N° 512**  
**Mathématiques et Informatique (InfoMaths)**

**Spécialité/Discipline de doctorat :**  
Informatique

Soutenue publiquement le 18/09/2023, par :  
**Abdel-Rahmen KORICHI**

---

**People Analytics and Organizational  
Network Analysis for Better Decision-  
Making and Human Resource  
Management**

---

Devant le jury composé de :

Mme Zahia GUESSOUM, Maître de conférence - HDR, Université de Reims Champagne-Ardenne, Rapporteur  
M. Allel HADJALI, Professeur, Université de Poitiers, Examineur  
Mme. Angela BONIFATI, Professeure, Université Lyon 1, Examinatrice  
M. Hocine CHERIFI, Professeur, Université de Bourgogne Franche-Comté, Président  
M. Hamamache KHEDDOUCI, Professeur, Université Lyon 1, Directeur de thèse

# Résumé

Dans le paysage commercial en constante évolution et au rythme effréné d'aujourd'hui, les organisations doivent s'adapter et innover en continu pour maintenir leur avantage concurrentiel. L'analytique des ressources humaines et l'analytique des réseaux organisationnels sont devenues des outils essentiels pour exploiter les données afin de soutenir les processus de prise de décision et améliorer l'efficacité organisationnelle. L'analytique des ressources humaines se concentre sur l'analyse des données de la main-d'œuvre pour obtenir une compréhension plus profonde, gérer et optimiser le capital humain. L'analytique des réseaux organisationnels, en revanche, étudie les modèles de communication, les interactions et les relations entre les employés pour favoriser une organisation hautement collaborative et efficace.

De nombreux avantages découlent de l'utilisation de l'analytique des ressources humaines et l'analytique des réseaux organisationnels, tels que l'amélioration de l'engagement des employés, l'augmentation de la productivité et une meilleure prise de décision, entre autres.

L'attrition est un problème particulièrement coûteux et perturbateur pour de nombreuses organisations, et l'utilisation d'analyses basées sur les données pour prévoir et atténuer ce phénomène peut offrir des avantages significatifs.

L'objectif principal de cette thèse est d'apporter une contribution importante au domaine en développant un cadre pour créer des outils analytiques et des visualisations axés sur l'analytique des ressources humaines et l'analytique des réseaux organisationnels. Nous compléterons cela en présentant des utilisations pratiques où nous réaliserons des analyses avancées sur des études réelles avec les données clients de Panalyt, afin de démontrer les bénéfices qui en découlent.

## CHAPTER 0. RÉSUMÉ

---

De plus, cette thèse abordera les défis inhérents et les préoccupations éthiques associés à l'analyse des informations sensibles des employés. Il est crucial de traiter ces problèmes pour maintenir la confiance et assurer une utilisation responsable des données.

Un autre aspect critique de cette recherche implique d'examiner et de développer des modèles de prédiction d'attrition qui utilisent les informations historiques du parcours des employés au sein d'une organisation. Une étude de cas montrera également que les mesures déduites des métadonnées de communication sont des prédicteurs efficaces de l'attrition.

La méthodologie de recherche englobera l'analyse des données clients réelles de la base de clients de Panalyt, auprès de laquelle j'ai réalisé ma thèse, et l'expérimentation de ces outils d'analyse et modèles sera dans le contexte d'organisations clientes de taille moyenne à grande. Cette approche aidera à mieux comprendre leur applicabilité et leur efficacité dans des situations réelles.

# Abstract

In today's fast-paced and ever-changing business landscape, organizations must continuously adapt and innovate to maintain their competitive edge. People Analytics and Organizational Network Analytics have emerged as vital tools for harnessing data to support decision-making processes and enhance organizational effectiveness. People Analytics focuses on analyzing workforce data to gain a deeper understanding, manage, and optimize human capital. In contrast, Organizational Network Analytics explores communication patterns, interactions, and relationships among employees to foster a highly collaborative and efficient organization.

Numerous benefits arise from employing People Analytics and Organizational Network Analytics, such as improved employee engagement, increased productivity, and better decision-making, among others.

Attrition is a particularly costly and disruptive issue for many organizations, and using data-driven analysis to predict and mitigate it can provide significant advantages.

The primary objective of this thesis is to make a meaningful contribution to the field by developing a framework for creating analytic tools and visualizations focused on People Analytics and Organizational Network Analytics. This will be complemented by practical use cases in which we perform advanced analysis on client case studies using real data from Panalyt to demonstrate their benefits.

Additionally, this thesis will address the inherent challenges and ethical concerns associated with analyzing sensitive employee information. It is crucial to address these issues to maintain trust and ensure responsible data usage.

Another critical aspect of this research involves examining and developing attrition

## CHAPTER 0. ABSTRACT

---

prediction models that utilize the historical information of employees' journey within an organization. One case study will also showcase that metrics derived from communication metadata are effective predictors of attrition.

The research methodology will involve analyzing real client data from Panalyt's client base, the company where I conducted my thesis, and the experimentation with these analytic tools and models will be within the context of medium to large-sized client organizations. This approach will help to better understand their applicability and effectiveness in real-world settings.

# Acknowledgements

First and foremost, I would like to express my deepest gratitude to God, who has granted me the strength, perseverance, and wisdom to face the different challenges encountered throughout the course of my PhD and life journey over the last 3 years.

I would like to extend my sincere appreciation to my supervisor, Prof. Hamamache Kheddouci, for his invaluable guidance, support, and encouragement. His expertise and insights have significantly contributed to the development of my research skills and the quality of my thesis. I am truly grateful for the opportunity to work under his mentorship.

To my loving family, I owe an immense debt of gratitude. To my wife, whose unyielding support and understanding have been present at every stage of this journey, I offer my most heartfelt thanks. To my parents, who have instilled in me a love for learning and the importance of hard work, your unwavering belief in my abilities has been a constant source of motivation.

I would also like to thank my colleagues at Panalyt, who have made my time there both intellectually stimulating and enjoyable. A special mention goes to our CEO, Daniel West, for his trust and confidence in me to carry out this challenging research. I am very grateful for the opportunities afforded to me as a result of his vision and leadership. Additionally, I would like to acknowledge my colleague Taha Tehseen, whose support and camaraderie during this process have been truly precious.

Finally, I would like to express my gratitude to all members of the People Analytics and Organizational Network Analytics community who have inspired, challenged, and supported my research endeavors. It is my hope that the findings presented in this

## CHAPTER 0. ACKNOWLEDGEMENTS

---

thesis will contribute to the ongoing advancement of our collective understanding and serve as a valuable resource for future explorations in this field.

# Contents

<b>Résumé</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Context and motivations . . . . .	3
1.2 Contribution . . . . .	4
1.3 Structure of the thesis . . . . .	5
<b>2 People Analytics, ONA and Attrition Prediction: State of the Arts</b>	<b>7</b>
2.1 People Analytics & ONA . . . . .	8
2.1.1 People Analytics: Literature Review . . . . .	8
2.1.2 Organizational Network Analytics: Literature Review . . . . .	11
2.1.3 Components of People Analytics . . . . .	15
2.1.4 Challenges . . . . .	16
2.1.5 Practical Applications . . . . .	20

## CONTENTS

---

2.2	Attrition Prediction . . . . .	24
2.2.1	Models . . . . .	24
2.2.2	Features . . . . .	26
2.2.3	Sampling . . . . .	28
2.3	People Analytics & ONA Analytic Tools . . . . .	29
2.4	Conclusion . . . . .	33
<b>3</b>	<b>Designing of a Framework for People Analytics and ONA Tools and Dashboards</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	Identified challenges and Panalyt's Solution . . . . .	36
3.2.1	3.1.1 Challenges in implementing Panalyt's People Analytics solutions . . . . .	36
3.3	Ethical concerns . . . . .	43
3.3.1	Privacy and Consent . . . . .	43
3.3.2	Bias and Discrimination . . . . .	43
3.3.3	Autonomy and Surveillance . . . . .	43
3.3.4	Data Ownership and Control . . . . .	44
3.3.5	Transparency and Accountability . . . . .	44
3.3.6	Conclusion . . . . .	44
3.4	Data Ingestion and Storage . . . . .	45
3.5	Panalyt Analytic Tools . . . . .	47
3.5.1	Recruitment Analytic Tool . . . . .	47
3.5.2	Attrition Analytic Tool . . . . .	49

CONTENTS

---

3.5.3	Overtime Analytic Tool . . . . .	50
3.5.4	Evaluation Analytic Tool . . . . .	52
3.5.5	Median Base Analytic Tool . . . . .	54
3.5.6	PBS Analytic Tool . . . . .	56
3.5.7	ONA Analytic Tool . . . . .	57
3.6	When ONA Meets People Data . . . . .	59
3.6.1	ONA Metrics . . . . .	59
3.6.2	Use Cases . . . . .	61
3.6.3	Disengagement . . . . .	66
3.6.4	Effective Onboarding . . . . .	68
3.6.5	Sales Effectiveness . . . . .	69
3.6.6	Conclusion . . . . .	70
<b>4</b>	<b>Case Studies of Panalyt Clients</b>	<b>71</b>
4.1	introduction . . . . .	71
4.2	Understanding Employee Attrition Using Communication Behavior Analysis . . . . .	71
4.2.1	Introduction . . . . .	72
4.2.2	Data Collection and Processing . . . . .	72
4.2.3	Description of the Communication Data . . . . .	73
4.2.4	Descriptive Analysis . . . . .	75
4.2.5	Channels Activity . . . . .	79
4.2.6	Number of Collaborations . . . . .	79
4.2.7	Conclusion of the Analysis . . . . .	82

## CONTENTS

---

4.3	Improving Manager Effectiveness and Innovation . . . . .	83
4.4	Identifying the Root Cause of Employee Turnover and Improving Post-Merger Integration . . . . .	85
4.4.1	Introduction . . . . .	85
4.4.2	Problem 1: A Lack of Visibility into HR Data . . . . .	86
4.4.3	Problem 2: HR Data Quality and Cleanliness . . . . .	86
4.4.4	Solution . . . . .	87
4.4.5	Insights Uncovered with Panalyt . . . . .	87
4.5	Conclusion . . . . .	88
<b>5</b>	<b>Attrition Prediction</b>	<b>89</b>
5.1	Introduction . . . . .	89
5.2	Method . . . . .	91
5.3	Experiments . . . . .	92
5.3.1	Dataset . . . . .	92
5.3.2	Results . . . . .	94
5.4	Explainability . . . . .	97
5.5	Attrition Analytic Tool . . . . .	97
5.6	Conclusion . . . . .	99
<b>6</b>	<b>Conclusions</b>	<b>101</b>
<b>7</b>	<b>Furthermore</b>	<b>103</b>

# List of Figures

2.1	People Analytics Tech Market Solution Matrix   RedThread Research, 2022. . . . .	31
3.1	Some of the Panalyt Integrations from Different Data Sources . . . . .	46
3.2	Recruitment Analytic Tool . . . . .	48
3.3	Attrition Analytic Tool . . . . .	50
3.4	Overtime Analytic Tool . . . . .	51
3.5	Evaluation Analytic Tool . . . . .	53
3.6	Median Base Analytic Tool . . . . .	55
3.7	People Balance Sheet Analytic Tool . . . . .	57
3.8	ONA Analytic Tool . . . . .	58
3.9	Collaboration between Two Teams . . . . .	62
3.10	Inter-Team Collaboration with Manager Layer . . . . .	63
3.11	New Hires Communications . . . . .	65
3.12	Gender Diversity . . . . .	66
3.13	Network Graph Layered by Gender . . . . .	67
3.14	Example of an Employee Disengaging . . . . .	67

## LIST OF FIGURES

---

3.15 Organizational Graph Used to Highlight New Joiners and their Connections . . . . .	68
3.16 Client Responsiveness . . . . .	69
4.1 Analysis of Volume of Messages Sent . . . . .	76
4.2 Analysis of Communication Times . . . . .	78
4.3 Analysis of Active Channels . . . . .	80
4.4 Analysis of Collaborations . . . . .	81
4.5 Surveys vs ONA . . . . .	84
5.1 Method Using One Observation per Employee . . . . .	90
5.2 Method Using Multiple Observations per Employee . . . . .	92
5.3 Distribution of the Predictions by Class . . . . .	95
5.4 Performance of the Model for Different Probability Buckets . . . . .	96
5.5 Performance of the Model at T-6, at T-4, and T-2 . . . . .	96
5.6 Top 3 Drivers for a Population of Employees at Risk of Leaving . . . . .	97
5.7 Attrition Risk Analytic Tool . . . . .	98

# Chapter 1

## Introduction

### 1.1 Context and motivations

Organizations face a variety of challenges related to their workforce, including employee turnover, low engagement, skills gaps, and productivity issues. In order to address these challenges, organizations need to have a deep understanding of their employees, their skills, their interests, their work patterns, and their networks. This is where People Analytics and Organizational Network Analytics come into play.

People Analytics is the use of data and analytics to inform and improve people-related decisions in organizations. It involves collecting and analyzing data from various Human Resources (HR) systems in order to gain insights into employee behavior, performance, and engagement. With People Analytics, organizations can make more informed decisions about hiring, performance management, learning and development, and employee retention.

Organizational Network Analytics is a subset of People Analytics that focuses on the analysis of social networks within organizations. It involves mapping and analyzing the relationships and communication patterns between employees in order to understand how information flows, how decisions are made, and how teams collaborate. By leveraging insights from Organizational Network Analytics, organizations can improve collaboration, innovation, and decision-making, identify and address potential

communication gaps or bottlenecks.

According to recent studies, the use of People Analytics can have a significant impact on various aspects of HR management, such as recruitment, employee engagement, and retention [IH16]. Over 70% of companies use People Analytics to improve their performance, according to a 2019 Deloitte report [Del19]. A 2018 study found that 65% of those who work in an organisation with a strong People Analytics culture said that their business performance was stronger than competitors [CIP18].

Moreover, with the growing availability of big data and advances in machine learning techniques, the potential for People Analytics and Organizational Network Analytics to drive strategic decision-making in HR is increasing. For example, predictive models based on employee data have been developed to identify employees at risk of leaving the organization, allowing for targeted retention efforts. Additionally, the use of communication metadata from email and chat systems has shown promise in predicting employee engagement and job satisfaction.

Together, People Analytics and Organizational Network Analytics can provide powerful insights into the workings of an organization and help address a wide range of challenges related to the workforce. However, despite their potential benefits, many organizations struggle to implement these tools effectively. They may lack the necessary data infrastructure, analytical skills, or understanding of how to turn data into actionable insights.

Given these potential benefits, there is a growing interest in developing tools and dashboards for People Analytics and Organizational Network Analytics, as well as exploring new approaches for predicting attrition and improving employee retention.

## 1.2 Contribution

As mentioned before, many organizations struggle to implement People Analytics and Organizational Network Analytics effectively.

This thesis aims to address this gap by developing tools and dashboards that are practical and user-friendly, and by providing real-world case studies that demonstrate the benefits of these approaches.

---

### 1.3. STRUCTURE OF THE THESIS

We will explore how to create People Analytics and Organizational Network Analytics tools and dashboards, display different use cases regarding their benefits, and propose a model to predict and understand attrition based on employee data.

Our methodology will involve studying real client data and experimenting with the effectiveness of our tools and models with medium to large size clients.

By using real client data and experimenting with different models and techniques, we aim to give practical insights that can be applied by medium to large size organizations in a variety of industries.

Overall, this thesis aims to provide practical guidance for organizations looking to leverage People Analytics and Organizational Network Analytics to improve their workforce management and drive better business outcomes.

## 1.3 Structure of the thesis

The remainder of this thesis is organized as follows:

The literature review in chapter 2 will provide a comprehensive overview of the prevailing research in the domains of People Analytics and Organizational Network Analytics. The intention is to lay the foundation for the subsequent chapters and establish a thorough understanding of the current state of knowledge.

In chapter 3, we will outline the conceptual framework to be employed for the construction of People Analytics and Organizational Network Analytics tools and dashboards. The chapter will also explore ethical considerations pertaining to data collection and analysis in order to address privacy and security concerns that may arise. This section will detail the process of data collection, handling, and related queries. Additionally, several use cases will be given to demonstrate the functionalities of these tools. Finally, we will present some of the main dashboards that we were able to develop over the last 3 years.

Chapter 4 will present various case studies to demonstrate the advantages that People Analytics and ONA tools and dashboards bring to the decision-making process within organizations. By discussing practical applications and techniques, this chapter aims

## CHAPTER 1. INTRODUCTION

---

to highlight their relevance and effectiveness in solving real-world management problems.

Chapter 5 will present our attrition prediction model, applied to real client data. We will also display Panalyt's attrition dashboard which was developed for this model.

The concluding sections will summarize the significant findings of this thesis, highlighting the advancements and contributions made in the realm of People Analytics and Organizational Network Analytics. Additionally, it will emphasize the areas that warrant further investigation and propose possibilities for future research endeavors to continue refining and enhancing these tools for more effective organizational decision-making.

## Chapter 2

# People Analytics, ONA and Attrition Prediction: State of the Arts

In this chapter, our goal is to present a comprehensive review of the current knowledge and recent developments in the fields of People Analytics and Organizational Network Analytics, as well as the aspect of attrition prediction models. We will begin our journey by examining the growth and rising importance of People Analytics and ONA in modern businesses, exploring how it has become an essential tool for understanding, managing, and optimizing human capital.

We will also explore the role of attrition prediction models in this context, which can help organizations better anticipate and address employee turnover, thereby minimizing disruptions and costs associated with it.

Next, we will provide a detailed analysis of various notable analytic tools and platforms that enable the successful implementation of People Analytics. We will also pay attention to Panalyt's distinct solutions and distinguishing factors in comparison to its competitors.

## 2.1 People Analytics & ONA

### 2.1.1 People Analytics: Literature Review

People Analytics is the practice of collecting and transforming human resources (HR) data and organizational data into actionable insights that improve the way you do business [Visnd]. Combining elements of behavioral science, business management, and data analysis, this approach helps optimize HR decisions, employee engagement, and overall organizational performance.

Over the past decade, People Analytics has become an essential aspect of modern HR management, thanks to the increasing availability of data and the rise of powerful analytical tools [VB17].

Tracing the origins of People Analytics back to the development of HR processes that aimed at measuring employee performance and identifying areas for improvement, modern technology has played a significant role in this field. Organizations began to collect more data on their employees, enabling data-driven approaches to HR management. This was facilitated by the continuous advances in technology, which allowed HR professionals to accumulate and process vast amounts of data, thereby adopting more sophisticated analytical methods in their decision-making processes.

Figure 2.1 gathers some of the most up to date literature reviews in the field of People Analytics.

Amongst the top most cited articles in the HR/People Analytics field, all with more than 100 citations, we find:

Angrave et al. (2016) [Ang+16], who's study highlights the need for HR professionals to enhance their analytical thinking in order to improve the development of HR analytics. They suggest that HR professionals should focus on developing their skills and knowledge to become "champions" of this approach, which will ultimately enable HR to have a strategic influence at the managerial level, benefiting both the organization and its employees.

Ulrich & Dulebohn (2015) [UD15] suggest that HR should adopt an external-internal approach, in which HR responds to the organization's challenges and participates

Table 2.1: Summary of the studies

Authors	Publication	Context
An Evidence-Based Review of HR Analytics [MB17]	2017	The study presents an evidence-based review using an integrative synthesis of published peer-reviewed literature on Human Resource Analytics (HR Analytics). The authors aim to provide a comprehensive understanding of the field and its developments.
Human Resources Analytics: A Systematic Review from a Sustainable Management Approach [Aiv+22]	2022	In this work, the authors analyze the current state of Human Resource Analytics (HRA) and propose a framework for its development as a sustainable practice. This study contributes to the understanding of HRA's potential and offers guidance for its practical implementation.
Human Resources Analytics: A Systematization of Research Topics and Directions for Future Research [Mar22]	2022	The study focuses on a systematization of research topics and directions for future research in Human Resources analytics. It uses a systematic literature review process to deconstruct the concept of HRA as presented in the literature. The authors identify 106 key research topics associated with three major areas: enablers of HR analytics (technical and organizational), applications (descriptive and diagnostic/prescriptive), and value (employee value and organizational value).
People Analytics Enabling HR Strategic Partnership: A Review [SL22]	2022	In this paper, a literature review is conducted to identify and synthesize existing literature on People Analytics and its conceptualized efficacy. The objective of this study is to explore and understand the efficacy of People Analytics, enabling the HR function to become a strategic partner within organizations.
Promise versus Reality: A Systematic Review of the Ongoing Debates in People Analytics [MF22a]	2022	The study conducts a systematic literature review of peer-reviewed articles focused on People Analytics in the Association of Business School (ABS) Index. The aim is to investigate the current reality of People Analytics and uncover the debates and challenges that are emerging as a result of its adoption. This research provides insights into the ongoing discussions and potential future directions of People Analytics.
Systematic Bibliometric Review of Artificial Intelligence Technology in Organizational Management, Development, Change and Culture [Bil+22]	2022	This study presents a bibliometric review of People Analytics using VOSviewer, a tool that constructs and visualizes bibliometric networks. The study includes various publication types, such as articles, conference papers, book chapters, editorials, notes, and reviews. This research offers an overview of the existing literature in the field and helps to identify key authors, institutions, and research trends.
Exploring the Evolution of Human Resource Analytics: A Bibliometric Study [Bon23]	2023	This paper is a bibliometric analysis of the scientific literature on HR analytics with the goal to understand the evolution of the conceptual, intellectual, and social structure of this topic.

## CHAPTER 2. PEOPLE ANALYTICS, ONA AND ATTRITION PREDICTION: STATE OF THE ARTS

---

more fully in strategy and value-adding development. They argue that HR analytics should focus on addressing the right problems.

Sivathanu & Pillai (2018) [SP18] present a case study using the Smart HR 4.0 analysis methodology to identify employees at risk of attrition and emphasize the importance of linking the concept of Smart HR 4.0 to the digital transformation of HR functions based on a "science of people."

Davenport et al. (2010) [DHS10] report that leading companies such as Google, Best Buy, P&G, and Sysco utilize sophisticated data collection and analysis technology to maximize the value of their talent. The paper provides six key ways to track, analyze, and use employee data.

Marler & Boudreau (2017) [MB17] and Aral et al. (2012) [ABW12] develop a model to examine HR analytics practices alongside an incentive system that yields higher productivity when implemented collectively rather than separately. They also include detailed data on HR software adoption.

Elia Rigamonti (2021) [Rig21] conducted a scoping review using Natural Language Processing and Topic Modelling techniques to explore the state of the art and future directions of People Analytics/HRA . The study analyzed 1,057 documents related to HRA and identified six topics based on the top terms extracted through topic modeling.

Topic 1: Big data and analytics in employee management and skill evaluation This topic covers the interaction between big data and people management, discussing the effects of digital technologies on HR needs and how data and analytical techniques can be used to assess employee KSAs.

Topic 2: HRA strategic development, implementation, and impact This topic is the "heart" of HRA literature, discussing analytics from a strategic and managerial viewpoint.

Topic 3: Advanced statistical and analytics techniques to solve employee-related problems This topic focuses on advanced statistical and analytical techniques in HRA solutions applied to employee turnover, retention, and other HR processes.

Topic 4: AI and analytics to support employee recruitment This topic is about optimizing personnel recruitment through analytics and AI.

Topic 5: Miscellaneous HRA-related topics This cluster includes a variety of HRA-related topics, such as ethical implications of HRA, key ingredients for HRA teams, and HR opportunities related to IoT (Internet of Things) and wearable devices.

Topic 6: HRA technological infrastructure This topic focuses on HR Information Systems (HRIS) and digital technologies as precursors to HRA, investigating HRIS adoption and outcomes.

The study also identified challenges and opportunities in the field of HRA, such as the need for greater collaboration between HR and data science teams, increased use of machine learning techniques, and the need to address ethical and privacy concerns.

### 2.1.2 Organizational Network Analytics: Literature Review

Organizational Network Analysis (ONA), also called Relational Analytics (RA), can be assigned a specific subset of People Analytics that focuses on relationships, communication patterns, and collaboration among individuals within an organization. ONA tools visualize and analyze these connections, giving HR professionals insights into the social network structure of their organization. ONA is also a subset of methods developed for Social Network Analysis (SNA), and most of the research-based paper will refer to SNA instead of ONA or RA, probably as the term is better referenced and established. In practice, all of our clients at Panalyt and other vendors refer to ONA and/or RA [23a; 23b; 23c; 23d; 23e; 23f].

ONA refers to the systematic study of communication and collaboration patterns within an organization to aid decision-making processes and optimize overall performance [CP04]. Drawing from network science, social network theory, and organizational behavior, ONA identifies informal networks, collaborative relationships, and information flows within an organization. In recent years, ONA has garnered attention as a valuable tool for organizations to gain insights into the dynamics of their workforce and enhance productivity.

As Paul Leonardi and Noshir Contractor argue [LC], People Analytics without ONA limits itself to individuals attributes and neglects the interdependence between employees. Indeed, what drives an organization or a team performance is not only the individual attributes, but also the interplay among people and how it evolves over

time. ONA addresses that and captures communication between people to give insights about the nature and the quality of their relationships. It makes it possible to evaluate a collaboration in the context of the surrounding network, which is a critical criterion for success in any collaborative project.

The application of ONA offers a shift from traditional hierarchy-based organizational decision-making to a more dynamic, network-driven approach. Organizations typically establish supervisory hierarchies to streamline management, decision-making, and collaboration processes [Rob04; Gal08]. However, routine interpersonal communication within workplaces tends to emerge more organically due to supervisory relationships and organizational needs. As Krackhardt and Hanson describe it [KH93], the formal organization is like the skeleton of a company, while the informal organization is the central nervous system that drives its collective thoughts, actions, and relationships. Understanding the interplay between these two networks is vital for effective organizational management.

Organizational behavior and management scholars have proposed theories linking social network features to business performance metrics [TTF79; KS88] and developed strategies to measure and enhance communication networks within organizations [NO16]. Some studies explicitly suggest how communication network features relate to organizational effectiveness. For instance, Krackhardt and Stern [KS88] argue that effective organizations have a higher rate of between-team communication than within-team communication. In contrast, Hatch and Schultz [HS97] argue that decentralized organizations with dispersed authority have better communication. Several authors also claim that top-performing employees have more communication network connections [CP02; NO16].

To assess these theories, researchers from various fields have proposed frameworks and methodologies for examining latent hierarchical structures within interaction or communication networks [PBK11; FP16]. One objective is to use communication network features to classify individuals by rank or job title, thereby identifying network correlates of institutional positions. Common approaches include computing network centrality measures for employees and ranking them by importance [SA04; NGD06; Row+07; Don+15; NM20]. Additionally, some methods in the literature attempt to infer tree-like structures (ideally, the entire organizational hierarchy) using inter-employee communication data [Gup+11; CMN08].

Nakonečný (2020) emphasized the importance of the recipient's understanding of the message and the need for clear communication objectives, as well as the selection of appropriate channels for information transmission and feedback [Nak20]. Ove (2019) explored the use of social networks in internal company communication and the potential for increased efficiency in corporate environment management [Leo19]. Ragozini and Vitale (2020) provided an overview of recent Social Network Analysis (SNA) research, highlighting advanced methods, theoretical development, and applications in various fields [RV20].

Mengbin (2019) investigated advanced mathematical SNA methods to study how individuals develop social strength during network interactions [Men19]. Dickinson, Magnani, and Rossi (2016) discussed modeling, data recovery, and dynamics of social networks, aiming to unify existing practical and theoretical knowledge about multilayer networks [DMR16]. Freeman (2004) presented a historical overview of SNA development, while Arnaboldi et al. (2015) analyzed the importance of major online social networks for corporations [Fre04; Arn+15].

Kaya and Alhajj (2019) examined the role of psychological factors in social media, focusing on the analysis of influence and behavior in the context of social networking and social media applications [KA19]. Brennecke (2019) conducted research on employees seeking help in solving problems with coworkers using "dissonant links" and found that seeking coworkers' problem-solving assistance can bring performance benefits [Bre19]. Lastly, Maiorescu-Murphy (2020) discussed the growing importance of social media in large American corporations, emphasizing the potential of social tools to impact the wider economy as significantly as other business information technologies [Mai20].

Much of the methodology used to study the relationship between organizational hierarchy and communication patterns originates from analyses of the Enron email corpus [KY04]. However, relying solely on this dataset has its limitations, as the Enron email corpus may not represent more successful organizations. To overcome these limitations, researchers have conducted empirical analyses of other organizations, but each study has its own scope limitations [Gui+06].

### A word about active vs passive ONA

ONA can be classified into two primary categories: active and passive ONA [Leo15]. Understanding the differences between these two approaches is crucial for determining their suitability and effectiveness in addressing organizational issues. This section will discuss the differences between active and passive ONA and argue in favor of passive ONA, with support from relevant references.

Active ONA involves the direct participation of organizational members in the data collection process. It typically relies on surveys or questionnaires to gather information on the relationships, communication patterns, and collaboration within the organization. Active ONA allows participants to self-report their interactions, providing insights into their perceptions of the organizational network.

However, active ONA is subject to several limitations. First, it relies on the accuracy and honesty of participants' self-reporting, which can be biased or incomplete. Second, active ONA can be time-consuming and resource-intensive, requiring organizations to devote significant resources to data collection and processing.

Passive ONA, on the other hand, employs unobtrusive methods to collect data on the organization's network. It leverages existing digital traces of communication and collaboration, such as email exchanges, calendar invitations, or project management tools, to map-out the network. Passive ONA can provide a more comprehensive and objective view of the organizational network, as it does not rely on self-reported data.

There are several reasons why passive ONA is considered superior to active ONA:

1. **Objectivity:** Passive ONA reduces the risk of biased or incomplete data by relying on digital traces rather than self-reported information [Leo15]. This allows for a more accurate representation of the organizational network.
2. **Efficiency:** Passive ONA is less resource-intensive and time-consuming than active ONA, as it capitalizes on existing data sources [BH11]. This allows organizations to obtain valuable insights with minimal disruption to their operations.
3. **Longitudinal analysis:** Passive ONA enables the analysis of data over time, providing a dynamic view of the organization's network evolution [CCE16]. This can help organizations identify trends and emerging issues more effectively.

4. **Privacy concerns:** While passive ONA raises potential privacy concerns due to the use of digital traces, these can be mitigated through proper anonymization and aggregation of data, thus ensuring compliance with privacy regulations [Pen10].

In summary, while both active and passive ONA provide valuable insights into organizational networks, passive ONA offers several advantages over its active counterpart. By leveraging existing digital traces, passive ONA offers a more objective, efficient, and dynamic view of the organization's network, making it a superior choice for organizations seeking to understand and improve their communication and collaboration patterns. This is why at Panalyt [23a], we mainly focus on passive ONA.

### 2.1.3 Components of People Analytics

Over time, People Analytics has experienced significant evolution due to its growing strategic importance for organizations and the rise of digital technologies. In the broader context of business analytics, three primary components or stages can be identified [Mar22], each with distinct levels of difficulty, value, and intelligence. These include:

- (1) Descriptive analytics: which addresses questions about what happened, why it happened, and current events.
- (2) Predictive analytics: which seeks to answer questions about future events and their causes.
- (3) Prescriptive analytics: which aims to determine optimal actions and their rationale. [KWL16; Lep+20; Siv+17].

The "descriptive" stage of People Analytics involves the utilization of internal and external organizational data, as well as HR administrative and workplace information, to produce ratios, metrics, analytic tools, and reports on human resources, primarily focused on the past. It focuses on aggregating historical employee data to understand the current state of the workforce, and it provides insights into why specific workforce patterns exist, such as high attrition rates or low employee engagement.

Predictive analytics, on the other hand, emphasizes data-driven insights and decision-making, employing statistical techniques, data mining, and advanced algorithms to analyze process and workflow data, as well as generate predictions and scenarios. The focus is shifted to the future, aiming to forecast the likelihood of specific workforce events, such as employee turnover or skill gap.

This has given rise to prescriptive analytics, which leverages extensive and diverse HR data to provide HR decision-makers with options for optimizing performance and transforming the entire HRM decision-making process [FJ14; SDY16]. It seeks to guide decision-makers in taking appropriate actions to achieve desired workforce outcomes.

Thus, the development of People Analytics should not be seen as a simple chronological evolution. Instead, it should be understood as a maturity progression. Different forms or "types" of analytics (i.e., descriptive, predictive, prescriptive) can be adopted and are indeed being utilized by organizations within more complex and integrated People Analytics initiatives and processes.

#### 2.1.4 Challenges

The field of People Analytics has seen considerable progress in recent years, as demonstrated by the expanding body of research. This increase in literature, along with the implementation and adoption of People Analytics by organizations for strategic workforce decisions, has given rise to various debates and challenges currently faced by HR and business professionals.

Five major issues have emerged in these debates [MF22b]: (1) lack of consistency in the concept and definition of People Analytics, (2) insufficient empirical evidence supporting the performance impact of People Analytics, (3) organizations' unpreparedness to perform People Analytics, (4) the question of People Analytics ownership, and (5) ethical and privacy concerns associated with using People Analytics.

##### **Lack of consistency in the concept and definition of People Analytics**

Various terms have been used to refer to People Analytics, including of course People Analytics [Fin17; LE21], Human Resources Analytics [LML03; MB17; VB17], Tal-

ent Analytics [DHS10], Workforce Analytics [HBB18; MLL18], and Human Capital Analytics [LF17; Min18].

Initially, Human Capital and Workforce Analytics were the most commonly used terms, while in the last decade, People Analytics and HRA have become more popular [FG20].

The term "People Analytics" was popularized by Google, as it represents their data-driven approach to Human Resource Management (HRM) [MB17]. This concept has gained traction as a best practice in HRM, as it is employed by leading companies to enhance their competitive advantage [ZI20]. Google's Project "Oxygen", initiated in 2010, has been a successful example of incorporating data analytics into daily decision-making processes, which has facilitated the acquisition of valuable insights on people operations [Ang+16; MS19].

Despite the various labels used, they all refer to the same talent management approach based on data collection and analysis [Mar22]. However, the use of different labels over time has made it challenging for researchers to identify relevant contributions.

### **Insufficient empirical evidence supporting the performance impact of People Analytics**

With the growing adoption of People Analytics, consulting firms such as Deloitte have promoted its benefits through publications [MB17]. Numerous case studies demonstrate how HR departments can leverage People Analytics to address particular challenges [HCL11b; BT18; MLL18]. However, scholars remain skeptical due to the lack of empirical evidence supporting its impact on strategic decision-making [LF17; MB17; LE21]. Few organizations can determine the business value that People Analytics provides [Ang+16; And17; MB17; HBB18; MLL18].

For example, Hussain et al. [HL18] states that *"Despite the recent popularity of workforce analytics, there is much that we do not yet know about the processes through which analytics affects the strategy execution process in organizations and, ultimately, firm success"*.

This debate reflects the tension between practice and academia, with practitioners focusing on case studies' evidence and academics seeking empirical evidence and gen-

eralizable findings. Until future research empirically links People Analytics to organizational outcomes, its impact remains uncertain.

### **Organizations are unprepared**

To effectively enable organizations to make informed, data-driven decisions through people analytics, HR professionals must possess a diverse skillset [HCL11b; RU15; Ang+16; MB17; Kry+18; MF22b]. [MB17] argue that HR employees need the right knowledge and skills to collect and analyze data, and communicate results meaningfully. Harris et al. [HCL11b] contend that human capital analytics success depends on HR's ability to find and nurture analytical talent.

Academic literature has proposed and debated various skills and competencies required for HR professionals to effectively conduct people analytics [And17; Kry+18; MLL18; Min18]. HR professionals must have the ability to analyze, identify appropriate data, and translate results into actionable insights for managers. McCartney et al. [MLL18] emphasize the need to ask the right questions, determine the right metrics, and provide evidence for strategic decision-makers. [MF22b] have developed a competency model for HR analysts, outlining six distinct competencies required in the field: consulting, technical knowledge, data fluency and data analysis, HR and business acumen, research and discovery and lastly storytelling and communication.

In addition to that, high-quality data is essential for successful people analytics, but many organizations struggle with data quality and confidence [And17; MLL18; Min18]. A study conducted by Deloitte showed that only 8% of HR leaders were confident using their data [Del17; MLL18]. Workforce data is often vast, messy, and stored in various sources, which can impede analytics and increase their cost.

Finally, although HR technology has advanced over the past 15 years, it often fails to provide solutions for advanced and predictive analytics [Ang+16; And17; FG20]. The lack of analytical capabilities, high-quality data, and adequate technology represent real challenges for people analytics. Efforts have been made to address the skills gap, improve data quality, and enhance HR technology.

### People Analytics' ownership

The question of whether people analytics should remain within the HR domain or be moved to a centralized analytics team has been explored by various scholars [RU15; And17; MB17; Min18; FG20]. Van den Heuvel and Bondarouk [VB17] found that 50% of respondents believe people analytics should be part of a centralized team, while 35% think it should stay in HR. Advocates for removing people analytics from HR argue that keeping it within HR limits collaboration opportunities, access to rich data sources, and strategic potential [RU15; And17; FG20].

However, those who support keeping people analytics in HR argue that other departments will not be as committed to acting on workforce data insights and may build models without HR interests in mind [And17; MB17].

A balance between both sides could involve keeping people analytics within HR while addressing the main limitations, such as adopting a highly collaborative approach and building boundary-spanning relationships with other departments such as marketing, finance or IT.

### Ethical and privacy concerns

Employee monitoring and data collection has extended beyond the work environment into personal and private lives, creating ethical and privacy concerns [KT17; JC20]. Despite the benefits of using employee data, such as better recruitment and selection decisions, efficient identification of training opportunities, and reduced workplace stress [HCL11b; RU15; Tur+21], collecting such data still involves significant risks for HR professionals. In fact, employees may feel threatened if they are unsure about what type of data is gathered, which collection methods are applied, and how their data are being used [Cha+21].

Countries have implemented regulations like GDPR in Europe to mitigate the misuse of employee personal information. Organizations have also started creating new positions and bodies such as Chief Data Officers and data governance committees to ensure ethical standards concerning organizational-wide data [She+21]. Being mindful of employees' privacy concerns, acting transparently, and maintaining open communication between employees and management can offer more pros than cons, but HR

professionals need to be included in these committees to review the boundaries around data collection and continuously question which data is too invasive to collect.

### 2.1.5 Practical Applications

People Analytics leverages data-driven insights to enhance human resource strategies and improve overall business performance. By harnessing the power of data analytics and statistical techniques, HR practitioners can now tailor people management practices to specific organizational contexts, thereby tackling unique challenges faced by their company.

**Workforce Planning:** People Analytics has proven instrumental in several use cases, such as identifying workforce trends and predicting staffing requirements. By using advanced analytics techniques, organizations can assess and forecast their workforce requirements to ensure that they have the right mix of people and skills to meet their business objectives. This enables companies to make data-informed decisions about hiring, talent development, and workforce allocation, thereby optimizing their human capital investments.

**Talent Acquisition:** People Analytics has supported HR teams in refining their recruitment processes by enabling the use of data-driven insights to evaluate candidate fit and potential for future success [Riv17]. Employers can now identify, attract, and retain top talents by leveraging analytical tools that inform their talent sourcing, assessment, and selection methodologies. People Analytics also facilitates the identification of potential biases in the recruitment process, promoting a more inclusive and diverse workforce.

**Employee Engagement:** Employee engagement is another domain in which People Analytics use cases have flourished. Organizations have started to analyze employee feedback, sentiment data, and workplace trends to inform employee-centric initiatives and foster an organizational culture that supports work satisfaction and productivity. Consequently, by leveraging these powerful data-driven insights, HR professionals and leaders can improve overall employee engagement, reduce turnover rates, and enhance organizational performance.

**Performance Management:** In the sphere of performance management, People Analytics plays a pivotal role in the design and implementation of fair, unbiased, and

---

## 2.1. PEOPLE ANALYTICS & ONA

employee-centric performance evaluation systems [Dav17]. By employing data analytics techniques, organizations can enhance their performance management systems to identify individual strengths and areas for improvement, as well as provide personalized development plans for their employees. This data-driven approach enables companies to foster a continuous learning environment that supports employee growth and development.

**Employee retention:** When it comes to employee retention, People Analytics has proven useful in identifying key factors contributing to staff attrition, allowing employers to develop targeted interventions that address these drivers. Consequently, organizations can now design initiatives that focus on employee career development, upskilling, and reskilling, ensuring their workforce remains competitive and satisfied. By proactively addressing the root causes of employee turnover, companies can enhance their ability to retain top talent and maintain a high-performance workforce.

**Diversity & Inclusion:** People Analytics helps organizations identify and address potential biases in their workforce management practices. By analyzing demographic data, HR professionals can ensure that diversity and inclusion initiatives are effectively implemented and monitored, leading to a more inclusive work environment. Moreover, People Analytics enables organizations to track the impact of diversity and inclusion efforts on various organizational outcomes, such as employee engagement, innovation, and financial performance. By leveraging these insights, HR professionals can make data-driven decisions to refine and enhance their diversity and inclusion strategies, promoting a more equitable and inclusive workplace culture.

**Learning and Development:** People Analytics also enables organizations to tailor learning and development programs to the specific needs of their employees. By analyzing data on skill gaps, employee performance, and career progression, HR teams can create customized training programs that target the most relevant areas for improvement and growth. This data-driven approach to learning and development enables organizations to optimize their investments in employee training and development, ensuring that their workforce remains competitive in the rapidly evolving business landscape. Furthermore, People Analytics can help organizations identify the most effective learning modalities and delivery methods, enhancing the overall effectiveness and efficiency of their learning and development initiatives.

**Succession Planning:** Through the use of People Analytics, organizations can iden-

## CHAPTER 2. PEOPLE ANALYTICS, ONA AND ATTRITION PREDICTION: STATE OF THE ARTS

---

tify high-potential employees and create targeted development plans to prepare them for future leadership roles. By analyzing factors such as performance, engagement, and leadership potential, HR professionals can make data-driven decisions about succession planning and leadership development [SD09]. This proactive approach to succession planning helps organizations build a robust leadership pipeline, ensuring that they have a steady supply of capable and qualified leaders to fill critical positions. Moreover, by leveraging People Analytics, organizations can monitor the progress of their succession planning initiatives, making timely adjustments to address potential challenges and ensure the successful transition of leadership.

**Employee Wellness:** By analyzing data on employee health, well-being, and engagement, People Analytics can help organizations design and implement targeted wellness initiatives that improve employee satisfaction and productivity. This can lead to better work-life balance, reduced stress levels, and increased overall well-being for employees. Additionally, People Analytics can be used to identify specific employee populations that may be at higher risk for health issues or burnout, enabling organizations to provide targeted support and resources to promote employee well-being. By prioritizing employee wellness, organizations can also reduce absenteeism, lower healthcare costs, and enhance overall workforce performance.

If we focus specifically on ONA applications, use cases are numerous:

**Identifying Influencers and Emerging Leaders:** ONA can be utilized to understand employees' collaboration patterns, interpersonal relationships, and subsequently, to identify key players and emerging leaders within a company. These individuals often serve as influencers or bridges that connect different groups, demonstrating remarkable communication and decision-making skills. Organizations can benefit from leveraging their abilities and networks to disseminate information, drive change, or promote best practices.

**Enhancing Collaboration and Cross-functional Teams:** Organizational Network Analysis can be employed to enhance collaboration and facilitate the creation of cross-functional teams. This is achieved through analyzing the network of relationships among employees from different departments and functional areas, and identifying overlaps, gaps, or opportunities for collaboration. By using ONA data, organizations can promote cross-functional engagement by forming strong connections and identifying barriers to collaboration. In this way, ONA acts as a catalyst for improving

## 2.1. PEOPLE ANALYTICS & ONA

---

collaboration and fostering a more cohesive organization.

**Facilitating Talent Management:** By leveraging ONA tools, organizations can develop more effective talent management strategies based on the insights obtained from network analysis. These strategies can involve identifying high-potential employees, addressing skill gaps, refining performance evaluations, and executing successful succession planning.

**Innovation and Knowledge Flow:** ONA can also reveal insights on the flow of knowledge and innovation across an organization. By analyzing how employees interact and share information, the organization can better understand the channels that support innovation and knowledge transfer. Additionally, this analysis can pinpoint employees who act as gatekeepers or key sources of knowledge in the organization. By using ONA to optimize and strengthen these networks, the organization can elevate its ability to innovate and maintain a competitive edge.

**Employee Engagement and Retention:** ONA serves as a tool for predicting employee engagement and retention by assessing individuals' ties within organizational networks, their level of job satisfaction, and workloads. By identifying employees who feel isolated or possess weak links within the network, targeted interventions can be employed to improve work relationships, develop trust among team members, and ensure a sense of belonging within the organization.

**Mimicking Successful Behaviors:** ONA can be employed to study top-performing employees and their connectivity patterns within the organization. By identifying the behaviors and network position of these individuals, companies can foster similar attributes in others and offer training programs to enhance performance. Moreover, coaching employees to develop their networks can also aid in spreading innovative and successful practices across the organization.

These use cases illustrate ONA's potential in guiding organizational leaders and human resources professionals to make informed decisions around workforce optimization.

## 2.2 Attrition Prediction

We mentioned earlier that three primary stages are identified in People Analytics: Descriptive, predictive and prescriptive analytics. In terms of predictive and prescriptive analytics, one topic stands out in comparison to others.

Employee attrition, the voluntary or involuntary departure of employees from an organization, is a significant concern for businesses worldwide. High attrition rates can lead to increased recruitment and training costs, reduced productivity, and a loss of valuable knowledge and skills within an organization. Data from Work Institute's 2020 Retention Report estimates that it (conservatively) costs as much as 30% of a worker's annual salary to replace him and that 78% of the reasons employees quit could be prevented by the employer [Ins20].

Therefore, understanding the factors that contribute to employee attrition and developing effective strategies to predict and mitigate it is crucial for maintaining a stable and high-performing workforce.

One of the most effective ways to tackle this issue is to build a predictive model - based on the wide variety of human resources (HR) data available [HCL11a] - to identify employees who are showing an intention to leave voluntarily. These models can vary in how they are built and operated but essentially follow the same basic principle - take a sample of employees that have already quit the company and use their data to build a model to highlight with an adequate advance who is likely to leave the company in the near future. By identifying the employees at risk and by understanding why employees choose to end their relationship with the company, the latter can make improvements to reduce existing attrition problems. In recent years, numerous machine learning algorithms and techniques have been applied to this problem.

### 2.2.1 Models

Predicting off-boarding and factors affecting employee attrition has been a significant trend in Human Resources Management literature. Various studies have employed machine learning models to predict employee attrition. For instance, Jain et al. (2020) argue that the RF model is the most suitable for predicting employee attrition [JJP20]. Khera and Divya (2018) used the SVM model to predict employee attrition, testing

the model with archived data from three Indian IT companies[KD18]. Fallucchi et al. (2020) employed the Gaussian Naïve Bayes (GNB) classifier model to analyze how objective factors affect employee attrition and tested the predictive accuracy of their model using IBM analytics data[Fal+20].

Yadav et al. (2018) compared the accuracy of various models, including Random Forest (RF), AdaBoost, Decision Trees (DT), Logistic Regression (LR), and SVM, to develop accurate and reliable models that optimize hiring and retaining quality staff costs [YJS18]. They found that DT outperforms other models in their dataset for predicting attrition. This study’s findings revealed that salaries, financial aspects, and promotions are insufficient incentives for retaining employees. Considering the high employee turnover costs in organizations, Zhao et al. (2018) developed a model for predicting employee turnover using the Extreme Gradient Boost (XGBoost) model[Zha+18]. They analyzed the model using data from an IBM database and a bank database.

Yahia et al. (2021) [YHC21] provide a comprehensive benchmark of different classification techniques, emphasizing the importance of communication patterns in predicting attrition.

Ziwei Jin et al. (2020) [Jin+20] design a hybrid model based on survival analysis and machine learning, and propose a turnover prediction algorithm named RFRSF, which combines survival analysis for censored data processing and ensemble learning for turnover behavior prediction.

Liu et al. (2018) compared the accuracy of LR, Support Vector Classifier (SVC), RF, and Adaboost in predicting employee turnover[Liu+18]. Additionally, Anh et al. (2020) combined SVM, LR, and RF models to develop a model for predicting future employee move [Anh+20]. They tested their model’s accuracy on the data of 1470 employees in an organization. These findings indicate that AI models can effectively predict employee attrition and turnover. HR managers can use these models to identify the factors causing employee attrition and avoid the high costs associated with human capital outflow.

Here is a brief description of some of the most widely-used algorithms:

- **Decision Trees:** A simple yet powerful classification algorithm that recursively

splits the data based on feature values to form a tree-like structure. Decision trees are highly interpretable but prone to overfitting if not pruned properly.

- **Random Forests:** An ensemble method that constructs multiple decision trees and combines their predictions to improve generalization and reduce overfitting. Random forests introduce randomness in both feature selection and data sampling, resulting in diverse and robust models.
- **Gradient Boosting Trees (GBT):** Another ensemble technique that iteratively builds decision trees while minimizing a loss function, thereby correcting errors made by previous trees. GBT often provides better performance than random forests but requires careful tuning of hyperparameters.
- **Logistic Regression:** A linear model that estimates the probability of an event (such as attrition) by fitting a logistic function to the input features. Logistic regression is simple, interpretable, and can be easily extended to include regularization for better generalization.
- **Support Vector Machines (SVM):** A powerful classification algorithm that constructs hyperplanes in a high-dimensional space to separate classes with the maximum margin. SVMs can handle non-linear classification using kernel functions but may require substantial computational resources for large datasets.
- **Artificial Neural Networks (ANN):** Flexible and powerful models that consist of multiple layers of interconnected nodes or neurons that can learn complex patterns and representations in the data. ANNs can be customized with various architectures and activation functions but may require extensive computational resources and tuning.

These algorithms differ in terms of complexity, interpretability, and performance, making it crucial to select the appropriate algorithm based on the specific requirements and constraints of the problem.

### 2.2.2 Features

The accuracy and effectiveness of attrition prediction models largely depend on the quality and relevance of the input features. Researchers have identified various fac-

tors that influence employee attrition, which can be broadly categorized as follows [ACG20]:

- Demographic features: age, gender, ethnicity and education level.
- Job-related features: job role, salary, promotion history, department and years of experience.
- Organizational features: company size, industry, and organizational culture.
- Psychological features: job satisfaction, work-life balance, engagement and commitment.
- Social features: interpersonal relationships and communication patterns.

Conventional methods in human resource management for addressing employee turnover and retention have mainly concentrated on individual-level factors. These factors often include job stress, perceived unfair treatment, insufficient compensation, and the availability of alternative job opportunities [Hom+17].

Regarding the social features, Younis et al. conducted a systematic review to develop an employee retention model using SNA constructs for Human Resource Management and Knowledge Management practices [YAC23]. The study aimed to identify knowledge gaps by analyzing recent contributions on SNA's application to employee turnover. The review found that research in this area is still underexplored, with multiple factors affecting social network parameters and employee turnover.

Four trends emerged: network position, network centrality, types of networks and the snowball effect of turnover (i.e. impact of an influential employee leaving an organization). The research showed that centrality measures in SNA can help identify key individuals in a network. However, the correlation between network position and turnover decision was found to be contradictory. Contrary to the traditional belief that central employees have lower turnover propensity, many studies countered this notion, arguing that unfavorable conditions could push a well-connected central employee to leave.

The most prevalent network measures in the literature were degree, closeness, betweenness, k-shell scores, and eigenvector centrality, with friendship and advice networks being the most studied. Turnover was found not to be random but occurring

in groups, especially when influential employees resign. Some studies hypothesized that turnover intention could shape employees' future friendship and advice networks [Glo+17; Trö+19].

### 2.2.3 Sampling

Typically, classification algorithms used to predict churn or attrition assume that the dataset used to create the classifier is balanced. However, many datasets are considerably imbalanced. An imbalanced dataset is one where the target class variable has one class significantly outnumbering the others. For instance, a dataset where 1000 active employees, and only 30 have left, is an example of a highly imbalanced dataset. Similar examples can be found in areas such as customer churn prediction [ZBB17], predicting freshmen student attrition [Tha+14] and insurance fraud detection [HA15]. Classifiers built using highly imbalanced datasets tend to be biased towards the majority class cases, meaning they perform better at predicting majority class cases than minority class cases. However, in real-world applications, a classifier should be unbiased and equally adept at predicting both minority and majority cases. Balancing imbalanced datasets is crucial for reducing bias in model predictions and improving classifier performance.

To tackle the issue of imbalanced datasets in classification tasks, several solutions have been proposed [Gal+12; Hai+17; KKP06], broadly categorized as data-preprocessing-level solutions, cost-sensitive learning methods, algorithm-level solutions, and ensemble methods. Data-preprocessing-level solutions involve resampling the original data before constructing the classifier, making them easy to implement and classifier-independent. Cost-sensitive learning methods consider the importance of misclassifying majority and minority class instances. Algorithm-level solutions propose new algorithms or modify existing ones, requiring a thorough understanding of the algorithm for implementation and making them less popular than resampling techniques. Ensemble solutions combine ensemble (bagging and boosting) models with resampling techniques or a cost-sensitive approach.

However, traditional methods for attrition prediction typically involve the use of a single observation per employee during a specific period [Lee+18; Fry+18], often spanning a few months, does not account for variations in employees' features over time and is limited to a particular time period. This limitation raises several issues:

exclusion of many employees from training datasets, insufficient information captured by a single observation, inability to predict the timeframe within which employees are at risk of leaving and lack of consideration for seasonality and its impact on attrition risk.

The approach to predicting employee attrition is very similar to that of customer churn [DK17] and studying customer churn can help deduce other methods. Ali et al. [GA14] showed that using multiple training observations per customer from different time periods (Multiple Period Training Data) increases churn model predictive accuracy compared to traditional approaches that only use the most recent observation.

However, their framework has certain limitations, such as the potential lack of independence introduced by multiple observations with minimal change, leading to duplicate data and imbalanced training datasets. Another limitation is the restriction of multiple observations to a specific period which excludes many customers.

In this thesis, we will focus particularly in proposing a comprehensive framework to develop a dynamic model that adapts to the changing behavior of clients and the evolving nature of the market. This dynamic approach allows the model to remain relevant and accurate over time, unlike static models that may become outdated.

We will leverage multiple observations per employee, capitalizing on the longitudinal nature of the data without the drawbacks of the aforementioned methods, ultimately allowing organizations to better predict and address the issue of attrition.

## 2.3 People Analytics & ONA Analytic Tools

Various People Analytics and ONA tools have emerged in the market to cater to the increasing demand for data-driven decision-making in HR. These tools facilitate effective workforce management, employee engagement, and productivity improvement.

Some popular People Analytics and ONA tools in the market include:

- **Visier:** a People Analytics and workforce planning tool that offers comprehensive insights into key HR metrics like workforce demographics, attrition, and diversity. It features built-in predictive analytics models for turnover predictions helping organizations plan for and reduce employee churn [23f].

- **TrustSphere:** an ONA tool that captures communication metadata and measures social graphs to identify key influencers, collaboration patterns and connectivity within an organization. TrustSphere’s LinksWithin technology maps the flow of organizational communication to provide insights into collaboration, knowledge sharing and engagement [23e].
- **Workday:** Workday offers a suite of enterprise applications, including financial management, human capital management, and planning software. Workday’s People Analytics tool provides insights to improve decision-making in HR, such as talent and performance management, diversity and inclusion and employee engagement [23h].
- **SAP SuccessFactors:** a cloud-based human resource management (HRM) system that encompasses a comprehensive range of human capital management applications. Its People Analytics tool enables organizations to acquire insights on issues related to HR, such as workforce cost analysis, employee turnover rate, and workforce planning [23c].
- **Swoop Analytics:** An ONA tool that focuses on improving collaboration and engagement in organizations by providing insights into employees’ digital interactions. From email exchanges to internal communication platforms like Microsoft Teams, Swoop Analytics offers easy-to-understand dashboards and real-time analytics on individuals’ network behavior and performance [23d].
- **Polinode:** An ONA tool that offers network analysis and visualization capabilities, enabling organizations to analyze and uncover insights within their workforce. Through employee survey data and integration with communication platforms, Polinode can identify key influencers, measure engagement, and provide data-driven insights into organizational culture and effectiveness [23b].
- **Panalyt:** it differentiates itself by combining People Analytics and ONA into one comprehensive platform, integrating data from various HR systems and communication platforms, like emails and intranets. This holistic approach allows organizations to better understand their workforce, evaluate collaboration efficiency and understand their attrition. It also provides ready-to-use and customizable dashboards, making analytical insights accessible to HR professionals without advanced analytical skills [23a].

### 2.3. PEOPLE ANALYTICS & ONA ANALYTIC TOOLS

RedThread Research in its *People Analytics Technology 2022: report* [GM22] has also developed a People Analytics Technology Quadrant (Figure 2.1) to help organizations understand the positioning of various PA technology vendors based on their capabilities, breadth of offering, and market presence. The quadrant consists of four categories:

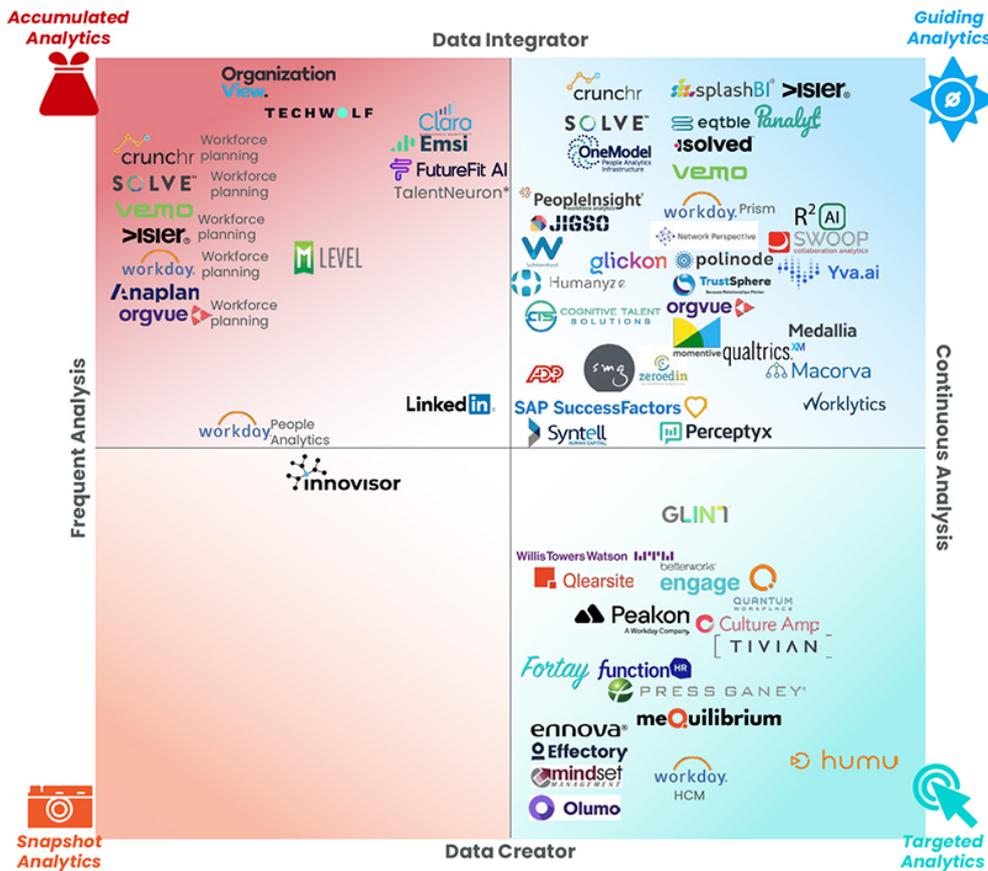


Figure 2.1: People Analytics Tech Market Solution Matrix | RedThread Research, 2022.

1. **Pioneers:** These vendors are characterized by a strong focus on innovation and possess advanced PA capabilities that push the boundaries of what is possible in the PA/ONA field. They may cater to niche markets or have specialized solutions that address specific organizational challenges. Examples of such vendors include Visier, Syndio, and TrustSphere.

## CHAPTER 2. PEOPLE ANALYTICS, ONA AND ATTRITION PREDICTION: STATE OF THE ARTS

---

2. **All-Stars:** These vendors have established themselves as market leaders with a comprehensive set of tools and solutions for PA and ONA. They have a strong market presence and are often considered go-to options by organizations seeking a robust and reliable PA platform. Notable examples in this category include SAP SuccessFactors, Workday, and Microsoft Viva Insights.
3. **Challengers:** Vendors in this category have demonstrated impressive growth and are working on refining their PA capabilities. They may not have as strong a market presence as All-Stars, but they show promise in delivering innovative PA solutions that can potentially disrupt the market. Examples of Challengers include Lattice, ChartHop, and Qlearsite.
4. **Specialists:** These vendors focus on a particular aspect of PA or ONA and deliver best-in-class solutions for that specific area. Specialists may provide complementary tools to organizations that have an existing PA infrastructure and are looking to address a particular challenge or tap into a specific insight. Examples of Specialists include Humanyze, Polinode, and Innovisor.

The report by RedThread Research serves as a valuable resource for organizations looking to adopt, refine or expand their PA and ONA capabilities. It provides a detailed overview of the current state of PA technology, highlights potential use cases for PA and discusses key trends shaping the market.

In the RedThread Research's People Analytics Tech 2020 Report, it is mentioned that *"Panalyt is the only multi-source analytics platform in our study that also offers ONA to analyze digital collaboration and communication based on passive data. The company further integrates sales and business data, allowing line managers to identify and focus on areas critical to them."*

In conclusion, People Analytics and ONA tools equip organizations with actionable insights to optimize workforce performance and improve decision-making. Panalyt differentiates itself from other players by offering a comprehensive solution that integrates both People Analytics and ONA, delivering a holistic view of an organization's workforce and communication patterns.

## 2.4 Conclusion

The literature review offered a comprehensive understanding of People Analytics and ONA, emphasizing their potential to optimize HR decisions and predict employee attrition. Furthermore, we highlighted popular analytic tools available, emphasizing the differentiating factors of Panalyt.

In the following chapter, we will have a closer look at the framework used at Panalyt to not only explore organizations' key information based on employees' attributes as we see in standard HR analytics systems, but also understand how to factor in communication metadata to unlock more insights.

We will also explore how to use analytic tools and visualization techniques to simplify complex data and create actionable insights for decision-makers.



## Chapter 3

# Designing of a Framework for People Analytics and ONA Tools and Dashboards

### 3.1 Introduction

In this chapter, we embark on a comprehensive exploration of the design framework for People Analytics and ONA used in Panalyt. We begin by delving into the challenges encountered during the implementation of Panalyt's cutting-edge People Analytics solutions.

Next, we address ethical considerations, including privacy, consent, bias, autonomy, data ownership, transparency, and accountability.

Following that, we delve into the critical components of data ingestion and storage, laying the foundation for understanding Panalyt's analytic tools, which are specifically designed to tackle various HR challenges, ranging from recruitment and attrition to overtime and evaluation.

Finally, we explore the synergy between ONA and people data, showcasing ONA met-

rics and real-world use cases that highlight its impact on issues such as disengagement, effective onboarding, and sales effectiveness.

## 3.2 Identified challenges and Panalyt’s Solution

At Panalyt [23a], we created a software as a service (SaaS) analytic platform where we plug in our clients’ data and allow them to perform People Analytics as well as ONA using our suite of pre-built analytic interfaces which cover the end-to-end employee lifecycle.

Based on our experience building this product and interacting with dozens of clients, we have identified multiple challenges. We will present the main ones in the following sections, and how we overcame them. Later, we will see how the association of People Analytics and ONA can help to deduce transformative insights by performing queries on a partial subgraph.

### 3.2.1 3.1.1 Challenges in implementing Panalyt’s People Analytics solutions

Despite the numerous benefits associated with People Analytics, companies still face considerable challenges when it comes to implementation. This section will explore the most common obstacles we faced with Panalyt and present our solutions to help organizations overcome these barriers.

#### Limited Resources

Organizations often struggle with limited resources, which can significantly hinder the successful implementation of People Analytics initiatives. These limitations may include:

1. Financial constraints: the adoption of People Analytics often requires investments in technology, tools and training. Organizations with tight budgets may find it difficult to allocate the necessary funding for these elements.

### 3.2. IDENTIFIED CHALLENGES AND PANALYT'S SOLUTION

---

2. Lack of appropriate technology: the successful implementation of People Analytics relies on the availability of advanced data analytics tools, software and technology infrastructure. Companies that lack these resources may struggle to analyze and interpret the data effectively.
3. Insufficient staffing: implementing People Analytics initiatives requires dedicated personnel to manage data collection, analysis and interpretation. Organizations with limited staffing resources may face difficulties in successfully deploying and maintaining People Analytics programs.

Panalyt's platform provides organizations with access to pre-built platforms that cover the end-to-end employee lifecycle. This enables companies to leverage People Analytics without the need for extensive investments in technology, tools and training. The SaaS model also allows for scalability, accommodating organizations of various sizes and budgets.

#### **Lack of Expertise**

Many HR professionals lack the analytical skills necessary to interpret complex data sets and derive actionable insights. This skills gap can prevent the effective implementation of People Analytics as organizations may struggle to translate raw data into meaningful information. This challenge can manifest itself in several ways:

1. Inadequate data literacy: HR professionals may lack sufficient data literacy to understand and interpret the results of People Analytics, hindering their ability to make informed decisions based on the insights generated.
2. Limited knowledge of analytical techniques: the implementation of People Analytics requires a strong foundation in statistical techniques and methodologies. HR professionals without this knowledge may struggle to effectively analyze and interpret the data.
3. Insufficient understanding of domain-specific issues: People Analytics often involves addressing domain-specific issues related to human resources, such as employee engagement, retention and performance. HR professionals may lack the necessary understanding of these issues to effectively leverage People Analytics in addressing them.

## CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS

---

4. Difficulty in integrating with existing systems: many organizations have legacy HR systems that may not be compatible with modern People Analytics tools. Integrating these systems can be time-consuming, costly, and technically challenging, limiting the ability of organizations to effectively implement People Analytics.

Panalyt simplifies the process of analyzing complex data sets by providing user-friendly analytic tools and visualizations. This empowers HR professionals with limited analytical skills to derive actionable insights from their data, closing the skills gap and facilitating more data-driven decision-making.

### **Permission Systems**

Permission systems are essential in People Analytics tools for several reasons, notably because they play a critical role in managing access to sensitive information, ensuring data privacy and security, and maintaining compliance with regulations. Implementing permission systems in People Analytics tools can be challenging due to the following factors:

1. Complexity of data and user roles: organizations often have complex structures with multiple levels, departments and roles, each requiring different levels of access to data and tools. Designing a permission system that caters to this complexity can be difficult and time-consuming.
2. Balancing accessibility and security: striking the right balance between providing access to data for informed decision-making and maintaining data privacy and security is a challenge. Overly restrictive permission systems may hinder collaboration and efficiency, while overly permissive ones can compromise security and compliance.
3. Scalability: as organizations grow, the number of users and roles may increase, requiring the permission system to be scalable and adaptive. Implementing a permission system that can easily accommodate growth and change is challenging.

### 3.2. IDENTIFIED CHALLENGES AND PANALYT'S SOLUTION

---

4. Compliance with regulations: organizations must adhere to various data protection and privacy regulations, which may have different requirements depending on the jurisdiction. Implementing a permission system that meets all these requirements can be a complex task.
5. Maintaining and updating permissions: over time, user roles and access requirements may change, requiring updates to the permission system. Keeping the system up-to-date and managing these changes can be resource-intensive.

Panalyt provides customizable permission settings that allow organizations to define different levels of data access based on roles, departments or other relevant factors. This flexibility helps create a permission system tailored to the organization's specific needs.

Moreover, Panalyt is designed to scale with the organization's growth, making it easier to manage permissions as the company evolves. It simplifies the process of adding new roles, updating permissions or modifying access levels as needed.

The platform is designed with data privacy in mind and helps organizations comply with relevant regulations such as GDPR and CCPA. The platform includes features that support secure data storage and handling, ensuring data privacy while maintaining accessibility for authorized users.

Finally, Panalyt offers an intuitive user platform that balances data security with user experience. The platform makes it easy for users to access the information they need while still maintaining appropriate data access controls.

#### **Data Collection and Management**

One of the most significant challenges in implementing People Analytics is collecting and managing the vast amounts of data required. Organizations need to gather data from various sources, such as employee records, surveys, performance evaluations, among others. Ensuring data accuracy and completeness can be arduous. Additionally, organizations must establish processes for updating and maintaining data ensuring compliance with privacy regulations, and safeguarding sensitive information.

Panalyt's platform assists organizations in managing their data more effectively by providing a central repository for collecting, storing, and analyzing data from various

## CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS

---

sources. This streamlines the data management process and ensures data accuracy and completeness. Panalyt complies with international data protection regulations, including the General Data Protection Regulation (GDPR), to ensure the privacy of its clients' data.

### **Data Quality Issues**

Accurate and meaningful insights depend on high-quality data. However, many organizations face difficulties in collecting, storing and managing data effectively. These challenges can include:

1. **Incomplete data:** organizations may lack comprehensive data on their employees, leading to incomplete or biased insights.
2. **Inaccurate data:** data inaccuracies can arise from various sources such as data entry errors or outdated information. These inaccuracies can negatively impact the insights derived from People Analytics.
3. **Inconsistent data:** inconsistencies in data collection and storage practices can make it difficult to compare and analyze data across different departments or time periods.

Panalyt addresses data quality issues by offering tools to cleanse, validate and standardize data. By identifying and rectifying incomplete, inaccurate or inconsistent data, Panalyt helps organizations improve the quality of their People Analytics insights.

### **Resistance to Change**

Implementing People Analytics may meet resistance from both managers and employees who are skeptical of data-driven decision-making or fear that it may lead to invasive monitoring. Implementing People Analytics often requires a significant shift in organizational culture and mindset. This resistance can create roadblocks in the adoption of People Analytics and limit its potential benefits. Resistance to change can manifest in several ways:

### 3.2. IDENTIFIED CHALLENGES AND PANALYT'S SOLUTION

1. Skepticism of data-driven decision-making: some employees may be skeptical of the value of data-driven decision-making, preferring to rely on their intuition or experience.
2. Fear of job security: the implementation of People Analytics may lead to concerns among employees about potential job loss or the devaluation of their skills and expertise.
3. Inertia and comfort with existing practices: some organizations may be reluctant to change their existing HR practices and decision-making processes, even in the face of evidence supporting the benefits of People Analytics.

Panalyt helps organizations overcome resistance to change by demonstrating the benefits of data-driven decision-making through case studies, success stories, and best practices. This fosters a more data-driven culture and encourages adoption of People Analytics.

#### **Organizational Silos**

People Analytics initiatives often require cross-functional collaboration and data sharing among various departments within an organization. However, organizational silos can hinder this collaboration, leading to difficulties in implementing People Analytics effectively:

1. Limited data sharing: departments may be reluctant to share data due to concerns about privacy or competitiveness, resulting in an incomplete view of the organization and limiting the effectiveness of People Analytics insights.
2. Lack of cross-functional collaboration: organizational silos can impede communication and collaboration between departments, making it difficult to align People Analytics initiatives with broader organizational goals and strategies.
3. Conflicting priorities and goals: different departments may have competing priorities and goals, which can create challenges in implementing People Analytics initiatives that require cross-functional support and cooperation.

## CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS

---

Panalyt's platform encourages cross-functional collaboration and data sharing by providing a centralized platform accessible to various departments within an organization. This breaks down organizational silos, fostering communication and collaboration, and enabling People Analytics initiatives to align with broader organizational goals and strategies.

### **Demonstrating ROI**

People Analytics initiatives often require significant investment in technology, training and personnel. As a result, stakeholders may demand concrete evidence of return on investment (ROI) before fully supporting these initiatives. Demonstrating ROI can be difficult, particularly when the benefits of People Analytics may not be immediately apparent or quantifiable.

Panalyt's pre-built platform enable organizations to quickly and effectively analyze their People Analytics data, revealing insights that can lead to cost savings, improved efficiency and enhanced employee engagement. By showcasing the tangible benefits of People Analytics, Panalyt helps organizations demonstrate the ROI of their initiatives and secure stakeholder support.

### **Conclusion**

Because of the complexity behind obtaining the data, as well as the technical resources and cost needed to process large amounts of sensitive data, it is difficult to build such a solution in-house for most companies. We had cases where clients told us that they liked our product but wanted to build an in-house solution. Many came back more than a year later with no progress. The resources and amount of work required to build such a tool is often underestimated.

We have spent more than 5 years building Panalyt's product at the time of writing this paper. We have a product that efficiently extracts, processes and stores people data and communication data every day and displays interactive visualizations in pre-built platform where data from different sources can interact with each other.

### 3.3 Ethical concerns

The increased prevalence of employee communication data in PA and ONA has given rise to significant ethical concerns that organizations must address. Beyond legal considerations, employers should be mindful of the ethical standards that they adhere to while using this information.

One study estimated that 81% of People Analytics projects are jeopardized by ethics and privacy concerns [Vis]. This section will elaborate on these concerns and propose solutions to minimize potential risks.

#### 3.3.1 Privacy and Consent

Respecting employee privacy is crucial when handling communication data. Organizations should provide clear guidelines on data collection and usage, and obtain informed consent from employees [ONe+20]. In addition, incorporating privacy-enhancing technologies, such as differential privacy, can help protect individual-level data while maintaining analytical value [DR14].

#### 3.3.2 Bias and Discrimination

People Analytics projects can unintentionally perpetuate and exacerbate existing biases in the workplace. To mitigate this risk, organizations must employ rigorous data auditing processes to detect and correct biased algorithms, as well as collaborate with diverse stakeholders to ensure the fairness of their analytics models [Fri+19].

#### 3.3.3 Autonomy and Surveillance

Employee autonomy can be undermined when communication data is used for excessive monitoring or surveillance. Companies should establish clear policies on data usage and avoid overreach that may lead to a "Big Brother" culture, which can negatively impact employee morale and trust [Zub19].

Panalyt purposefully avoids analyzing sensitive communication data, ensuring privacy and protecting employees from intrusive surveillance.

The use of aggregated, non-identifiable data is recommended where possible to demonstrate to employees that the sole purpose behind People Analytics projects is to capture larger organizational trends.

### 3.3.4 Data Ownership and Control

Employees should have the right to access, correct or delete their personal data, as well as exercise control over how their data is used [Com18]. Implementing data management practices that empower employees will foster trust and promote ethical People Analytics practices.

Panalyt recommends setting up well-defined policies and processes for data collection, access, storage, consent, and anonymity. These guidelines help maintain ethical standards and ensure responsible data handling.

### 3.3.5 Transparency and Accountability

According to a recent study, transparency was identified as being one of the most critical considerations for People Analytics projects [Tur+21]. Organizations should strive for transparency in their People Analytics projects, informing employees of the project's goals, methodologies, and potential implications. This includes sharing the outcomes and any decisions made based on the analyses.

Panalyt advocates for organizations to clearly communicate the reasons for pursuing People Analytics projects and the benefits employees can expect from these initiatives. Such transparency helps build trust and improve employee buy-in for People Analytics initiatives.

### 3.3.6 Conclusion

Addressing the ethical concerns of People Analytics is crucial for ensuring responsible use of sensitive employee data by organizations.

By implementing the abovementioned measures, Panalyt addresses the ethical concerns related to privacy, autonomy, transparency and data governance, ensuring responsible use of People Analytics and ONA in organizations.

## 3.4 Data Ingestion and Storage

Before building our platform, we needed data of 2 kinds: people data and communication metadata.

People data is collected from HR systems, such as human resources information systems (HRIS), payroll tools, absence management tools, performance management tools, and recruitment tools (Figure 3.1).

In many cases, organizations use SaaS tools to store their data, and most of them would have API connections. In other cases, the data is stored in Microsoft Excel or CSV documents.

On the other hand, communication metadata can be retrieved from chat/email tools. Over the last few years, online communication and management tools such as Microsoft Teams/Outlook, Slack, Gmail and others have started providing API access for their customers, making it easier for organizations to collect communication metadata.

Having the experience of working with many different clients, we have a good overview of how people data and communication data is stored, and how it can be retrieved, and used to perform analysis. At Panalyt, we first collect the data using APIs or CSV/Excel files. The data is then transformed and stored in the database and queried from the analytic tool.

We use SQL to query our data, a structured query language that allows fast retrieval and transformation of data.

We recommend to versionize the data, at least monthly, before storing it in one or multiple tables, to make it easier to query snapshots of the data for any particular month and perform join operations across tables (Table 3.1).

Versioning HR data helps to analyze trends over time. We can compare data from

CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS

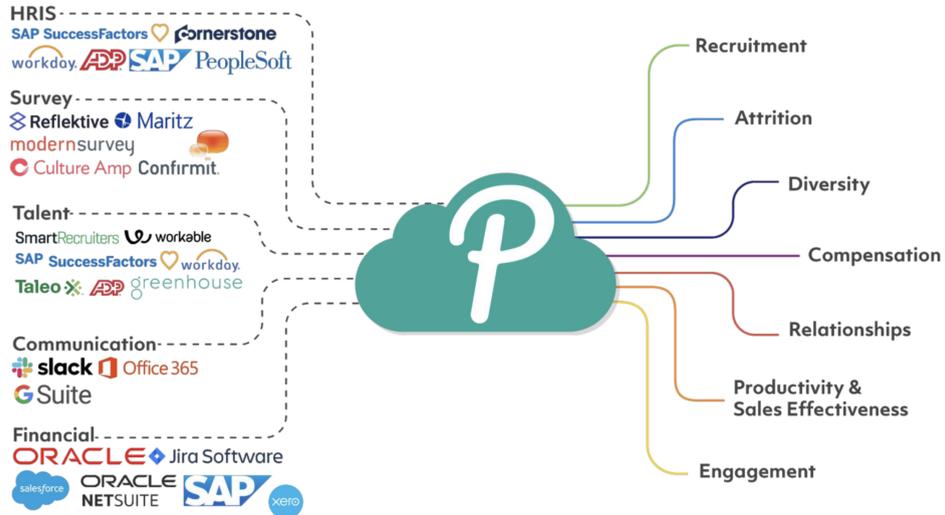


Figure 3.1: Some of the Panalyt Integrations from Different Data Sources

Table 3.1: Employee table

<i>empID</i>	<i>versionId</i>	<i>department</i>	<i>salary</i>	<i>jobTitle</i>	...
0001	30/09/2022	Technology	\$6,000	Data Engineer	
0001	31/10/2022	Technology	\$6,000	Data Engineer	
0001	30/11/2022	Technology	\$6,000	Data Engineer	
0001	31/12/2022	Technology	\$7,000	Sr Data Engineer	
0001	31/01/2023	Technology	\$7,000	Sr Data Engineer	
0002	31/08/2022	Marketing	\$4000	UI Designer	
0002	30/09/2022	Marketing	\$4,5000	UI Designer	
0002	31/10/2023	Marketing	\$4,5000	UI Designer	
...					

different months to identify patterns, changes and trends in your HR data. For example, we may notice that attrition rates are increasing over time, or that certain departments have higher attrition rates than others.

We then make a query on a partial subgraph to visualize a defined category of the population. We generally prefer to aggregate the data by month as we realized that this granularity is large enough to deduce trends, and small enough to notice significant changes quickly in the scale that companies operate in.

### 3.5 Panalyt Analytic Tools

In this section, we will be presenting some of the main analytic tools Panalyt has to offer and that we have built over the past few years.

#### 3.5.1 Recruitment Analytic Tool

Recruiting the right talent is crucial for the success of organizations. However, many organizations struggle to track their recruitment progress and identify areas for improvement. This is where the recruitment Analytic Tool of Panalyt can help. It provides a comprehensive view of an organization's recruitment funnel, allowing to track key metrics and identifying areas of improvement.

##### **Tracking Recruitment Metrics**

The recruitment Analytic Tool of Panalyt (Figure 3.2) allows to track a variety of recruitment metrics, such as the number of candidates at each stage of the recruitment funnel, the time to fill positions and the cost per hire. By tracking these metrics, organizations can identify bottlenecks in their recruitment process and take steps to improve recruitment outcomes.

For instance, if many candidates drop out of the recruitment process at the screening stage, the organization may need to revisit its screening criteria to ensure that it is attracting the right candidates. Alternatively, if the time to fill positions is longer

## CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS

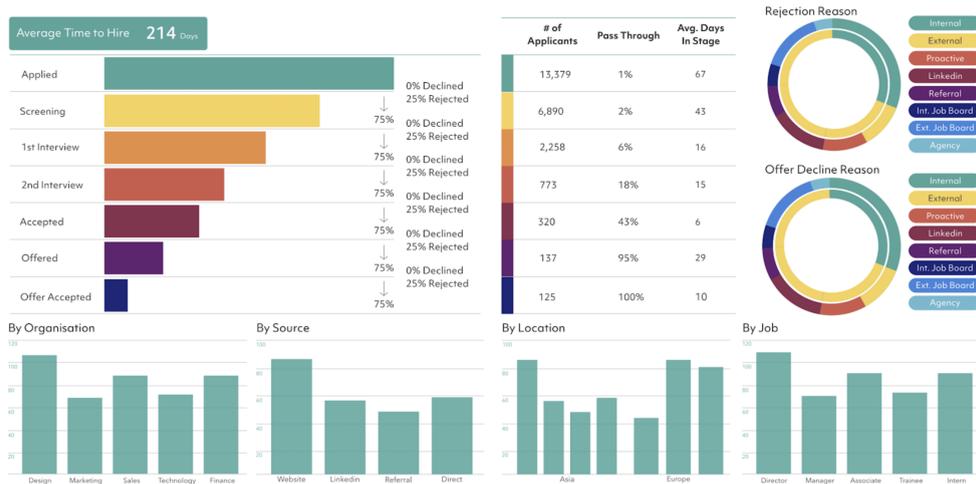


Figure 3.2: Recruitment Analytic Tool

than industry benchmarks, the organization may need to streamline its recruitment process or invest in additional recruiting resources.

### Analyzing Diversity in Recruitment

In addition to tracking recruitment metrics, the recruitment platform of Panalyt allows for analyzing diversity in recruitment. This is important because a diverse workforce has been shown to improve organizational performance and innovation.

Panalyt's recruitment platform allows to track diversity metrics, including the percentage of female candidates at each stage of the recruitment funnel and the percentage of candidates from different ethnic and racial backgrounds. By tracking these metrics, organizations can identify areas where their recruitment process may be unintentionally biased and take steps to improve diversity in their organization.

For example, if the percentage of female candidates drops off significantly after the initial application stage, the organization may need to revisit its job postings to ensure that they are inclusive and appealing to a diverse candidate pool. Alternatively, if the percentage of candidates from certain ethnic or racial backgrounds is low, the organization may need to broaden its recruiting sources to reach a more diverse candidate pool.

### 3.5.2 Attrition Analytic Tool

Employee turnover, or attrition, is a major concern for organizations. High levels of attrition can be costly and disruptive, leading to decreased productivity, increased recruiting costs and reduced employee morale. The attrition platform of Panalyt provides a comprehensive view of an organization's attrition rates and allows to track key metrics to identify areas of improvement.

#### Tracking Attrition Metrics

The attrition platform of Panalyt (Figure 3.3) allows to track a variety of attrition metrics, such as the number of employees leaving the organization, the reasons for their departure and the cost of attrition. By tracking these metrics, organizations can identify patterns in attrition rates and take steps to address underlying issues.

For instance, if a particular department or team has a consistently high attrition rate, the organization may need to investigate further to understand the underlying reasons. It may be that employees in that department or team are dissatisfied with their work or lack opportunities for career growth. Alternatively, it may be that the organization is not effectively communicating its values and goals to employees, leading to a lack of engagement.

#### Analyzing Reasons for Attrition

In addition to tracking attrition metrics, the attrition platform of Panalyt allows for analyzing attrition costs. This is important because attrition can be costly for organizations, both in terms of direct costs such as recruiting and training new employees, and indirect costs such as decreased productivity and reduced morale.

The attrition platform of Panalyt allows to track attrition costs by department, team and individual employee. By analyzing these costs, organizations can identify areas where attrition is driving up costs and take steps to address underlying issues.

For example, if a particular department is experiencing high attrition costs, the organization may need to investigate further to understand the reasons for attrition

## CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS



Figure 3.3: Attrition Analytic Tool

and take steps to improve retention. This may involve improving management practices, offering additional training and development opportunities, or providing greater opportunities for employee engagement.

### 3.5.3 Overtime Analytic Tool

Tracking employee overtime can be a challenge for organizations, as it can be difficult to identify employees who are consistently working longer hours and the reasons behind the overtime. This is where the overtime platform of Panalyt can help. It provides a comprehensive view of an organization's employees' overtime hours, allowing to track key metrics and identifying areas of improvement.

#### Tracking Overtime Metrics

Panalyt's overtime platform of Panalyt (Figure 3.4) allows to track a variety of overtime metrics, such as the number of employees working overtime, the total number of overtime hours worked and the reasons of such overtime. By tracking these metrics, patterns in overtime hours can be identified and steps can be taken to address

### 3.5. PANALYT ANALYTIC TOOLS



Figure 3.4: Overtime Analytic Tool

underlying issues.

For example, if certain teams or departments are consistently working longer hours, further investigation can be done to understand the underlying reasons. It may be that employees in those teams or departments are overloaded with work or lack the necessary resources to complete their tasks within regular working hours. Alternatively, it may be that those teams or departments are experiencing higher levels of absenteeism or turnover, which is causing other employees to work longer hours.

#### Analyzing Overtime Costs

In addition to tracking overtime metrics, the overtime platform of Panalyt allows for analyzing overtime costs. This is important because excessive overtime can be costly for organizations, both in terms of direct overtime pay and indirect costs such as decreased productivity and increased burnout.

The overtime platform of Panalyt allows to track overtime costs by employee, team and department. By analyzing these costs, areas where overtime is driving up costs can be identified and steps can be taken to address underlying issues.

For example, if a particular team is consistently incurring high overtime costs, further investigation can be done to understand the reasons for the overtime and take steps to reduce it. This may involve hiring additional staff, redistributing workloads, or improving processes to increase efficiency.

### 3.5.4 Evaluation Analytic Tool

Employee performance is essential for driving an organization's success. Monitoring and analyzing employee performance enables organizations to identify areas for improvement, allocate resources effectively, and ensure employees are engaged and productive. Panalyt's performance analytic tool offers a comprehensive view of an organization's performance metrics, helping to identify trends and drive better decision-making.

#### Monitoring Performance Metrics

Panalyt's performance platform enables organizations to monitor a wide range of performance metrics, such as individual and team performance, performance by department, goal completion rates and employee engagement. By tracking these metrics, organizations can identify trends and patterns in performance and take targeted steps to improve employee performance.

For example, if a specific department or team is consistently underperforming, the organization can investigate the underlying reasons. It may be due to inadequate resources, lack of training, or poor management practices. Identifying the root cause allows the organization to implement targeted interventions, such as providing additional training, reallocating resources or adjusting management strategies.

While most companies don't segment performance data by a demographic basis due to the complexity of the analysis, looking at the distribution of evaluation grades across groups, for example men vs women, might help uncover that a certain population has a performance distribution curve disadvantageous to the average baseline or more privileged demographic groups (Figure 3.5).

In addition to monitoring performance metrics, the performance platform of Panalyt enables organizations to analyze employee engagement. High levels of employee

### 3.5. PANALYT ANALYTIC TOOLS



Figure 3.5: Evaluation Analytic Tool

engagement are linked to increased productivity, better job satisfaction and lower attrition rates. By tracking engagement levels, organizations can identify areas where employee engagement is lacking and implement strategies to improve it.

#### Analyzing Performance Costs

Understanding the costs associated with employee performance is crucial for organizations to make informed decisions about resource allocation and improvement initiatives. The performance platform of Panalyt allows organizations to track performance-related costs, such as the cost of training, employee development programs and performance incentives.

By analyzing performance costs at the department, team, and individual employee levels, organizations can identify areas where resources may be allocated inefficiently or where additional investment is required to drive better performance. This information allows organizations to make data-driven decisions about resource allocation, ensuring they are investing in the most effective strategies for improving employee performance and engagement.

For instance, if a particular department has high performance costs but is not delivering the expected results, the organization can investigate the reasons for the discrepancy and take targeted actions to improve performance. This may involve adjusting training programs, reallocating resources, or modifying performance incentives to better align with organizational goals.

### 3.5.5 Median Base Analytic Tool

Understanding the median base salary across the organization is essential for evaluating and maintaining competitive compensation practices. The median base salary provides a useful benchmark for comparing employee compensation within the organization and against industry standards. Panalyt's median base analytic tool offers a comprehensive view of the organization's median base salary data, enabling effective monitoring and informed decision-making.

#### Monitoring Median Base Metrics

The median base platform of Panalyt enables organizations to track median base salary metrics across various dimensions, such as departments, teams, job levels, and employee demographics. By monitoring these metrics, organizations can identify trends, ensure equitable compensation practices, and make informed decisions about their compensation strategy.

For example, if a specific department or job level consistently has a lower median base salary compared to the rest of the organization or industry benchmarks (Figure 3.6), the organization can investigate the underlying reasons. It may be that the compensation strategy is not aligned with market trends or that the organization's pay structure is not effectively rewarding employees in those roles. Identifying these issues allows the organization to adjust its compensation strategy to remain competitive and ensure employee satisfaction.

In addition to monitoring median base salary metrics, the median base platform of Panalyt allows organizations to track the impact of their compensation strategy on key performance indicators, such as employee retention, performance and engagement. By correlating median base salary data with these indicators, organizations can gain

### 3.5. PANALYT ANALYTIC TOOLS

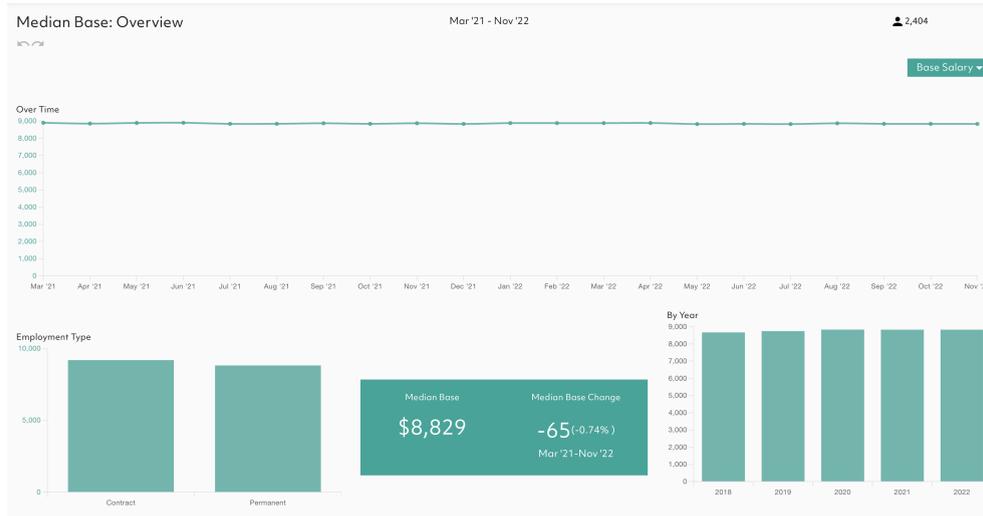


Figure 3.6: Median Base Analytic Tool

valuable insights into the effectiveness of their compensation practices in motivating and retaining employees.

#### Analyzing Median Base Trends

Understanding trends in median base salary data can help organizations make informed decisions about their compensation strategy and ensure that it remains competitive and equitable. The median base platform of Panalyt enables organizations to analyze trends in median base salary metrics over time and across different dimensions, such as departments, teams, job levels, and employee demographics.

By analyzing these trends, organizations can identify areas where median base salaries may be lagging behind market rates or where pay disparities may exist within the organization. This information allows organizations to make data-driven decisions about their compensation strategy, ensuring that they are investing in the most effective practices for attracting and retaining top talents and promoting pay equity.

For instance, if an organization identifies a trend of decreasing median base salaries in a particular department or job level, it can investigate the reasons for this change and take targeted actions to address the issue. This may involve adjusting salary ranges,

modifying compensation policies or implementing pay equity initiatives to ensure that all employees are compensated fairly and competitively.

### **3.5.6 PBS Analytic Tool**

Managing a company's human capital is a crucial aspect of ensuring long-term success. However, tracking and managing an organization's workforce can be a daunting task. This is where the People Balance Sheet (PBS) platform of Panalyt can help. It provides a comprehensive view of an organization's human capital, allowing to track key metrics and identifying areas of improvement.

#### **Tracking Human Capital Metrics**

Panalyt's PBS platform allows to track a variety of human capital metrics, such as headcount, employee turnover, employee demographics and employee costs (Figure 3.7). By tracking these metrics, organizations can gain insights into their workforce and identify areas of improvement.

For instance, if a particular department has high employee turnover, the organization may need to investigate further to understand the underlying reasons. It may be that employees in that department lack the necessary resources or support to perform their jobs effectively. Alternatively, it may be that the department is not effectively communicating its values and goals to employees, leading to a lack of engagement.

#### **Analyzing Workforce Demographics**

In addition to tracking human capital metrics, the PBS platform of Panalyt allows to analyze workforce demographics. This is important because it provides a comprehensive view of an organization's workforce diversity and allows to identify areas for improvement.

The PBS platform of Panalyt allows to analyze workforce demographics by gender, age, ethnicity and other factors. By analyzing these demographics, organizations can identify areas where their workforce may be lacking in diversity and take steps to improve diversity and inclusion.

### 3.5. PANALYT ANALYTIC TOOLS

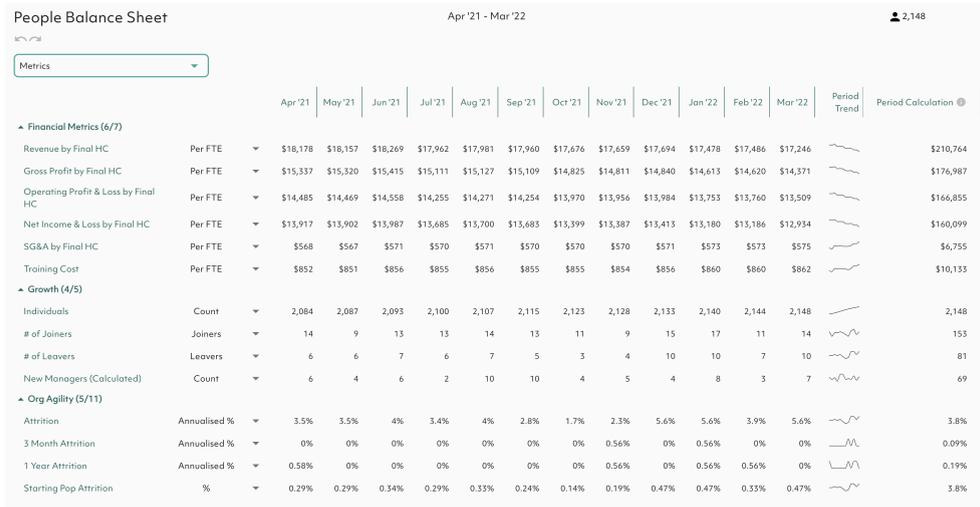


Figure 3.7: People Balance Sheet Analytic Tool

For example, if the organization’s workforce is predominantly male, the organization may need to revisit its recruitment and retention strategies to ensure that they are attracting and retaining female talents. Alternatively, if the organization’s workforce lacks diversity in terms of ethnicity, the organization may need to broaden its recruiting sources to reach a more diverse candidate pool.

#### 3.5.7 ONA Analytic Tool

Understanding the collaborations and connections between employees is a crucial aspect of managing an organization. Organizational Network Analysis (ONA) provides insights about how employees interact and collaborate, allowing organizations to identify areas of improvement and optimize their workforce. The ONA platform of Panalyt provides a comprehensive view of an organization’s network and allows to track key metrics to identify areas of improvement.

In our application, we use `d3.js` for our network data visualizations. The library takes in source-target pairs to create nodes and the edges between them. In addition to that, we can also add other attributes like colour to each node based on features like tenure, organization, function, etc.

## CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS

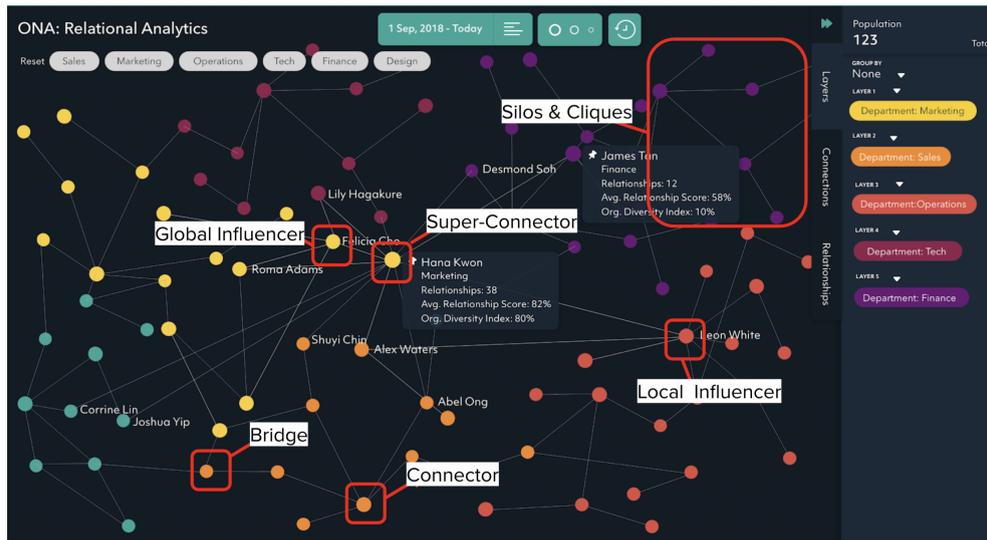


Figure 3.8: ONA Analytic Tool

### Tracking ONA Metrics

The ONA platform of Panalyt allows to track a variety of ONA metrics, such as the strength of connections between employees, the size of an employee’s network, and the centrality of an employee within a network. By tracking these metrics, organizations can gain insights into how their employees collaborate and identify areas of improvement (Figure 3.8).

For instance, if a particular department or team has a low network density, the organization may need to investigate further to understand the underlying reasons. It may be that employees in that department or team lack opportunities to collaborate effectively, leading to siloed information and decreased productivity. Alternatively, it may be that the organization needs to implement more cross-functional projects to encourage collaboration across different departments.

### Analyzing Network Structure

In addition to tracking ONA metrics, the ONA platform of Panalyt allows to analyze network structure. By analyzing network structure, organizations can identify areas

where employees may benefit from increased connections and take steps to improve collaboration.

For example, if a particular employee has a low degree centrality score, the organization may need to encourage that employee to collaborate more with his/her colleagues or provide such employee with additional training on collaboration and communication skills. Alternatively, if a particular department or team has low network density, the organization may need to implement team-building exercises or cross-functional projects to encourage collaboration.

## 3.6 When ONA Meets People Data

### 3.6.1 ONA Metrics

Regarding communication data, the main fields that we retrieve are:

- *messageId*: unique identifier of the message;
- *sendingFrom*: sender's email;
- *sendingTo*: receiver's email;
- *sentDatetime*: time when the message was sent;
- *threadId*: thread/conversation the message belongs to; and
- *channelId*: channel where the message has been sent - only for chat tools.

We process the data to get one line per sender and receiver pair for each month. We then compute various aggregated metrics, from person 1 to person 2 and person 2 to person 1, including, but not limited to:

- count of messages sent: the volume of messages sent by either or both correspondents.
- average response time: the time taken for a correspondent to reply to the other correspondent.
- response rate: the percentage of messages received by either correspondent that they replied to.
- the reciprocity: the ratio of the volume of communication between a correspondent and another.

### CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS

---

- average time first/last message sent: the average time when a correspondent sends his first/last message to another person.
- average time first/last message received: the average time when a correspondent receives his first/last message from another person - emails only.
- channels in common (chat tools only): the number of channels or groups where both correspondents are active (i.e. have both sent more than X number of messages during that month).

We then create a graph based on the analysis. This could be a direct graph, where arrows point from the sender to the recipient, or an undirected graph. Although some authors propose to use the number of e-mails sent by a node  $i$  to a node  $j$  divided by the total number of e-mails sent by member  $i$  [MPK16], and others suggest to use the geometric mean of sent-received counts, [Cho+15], there is no unique way to generate the attributes to connect people in an ONA graph. It really depends on the use case.

At Panalyt, we have created a score - which we can't disclose - that includes the information of some of the metrics defined above, especially the count of messages sent, the average response time, the response rate, and the reciprocity.

$$R(t_1, t_2)((\alpha_1, M_1), (\alpha_2, M_2), \dots, (\alpha_n, M_n)) \quad (3.6.1)$$

Where  $M_1, M_2, \dots, M_n$  are measures of the communication network (deduced from employees' logs),  $\alpha_1, \alpha_2, \dots, \alpha_n$  are weights of importance of these measures, and  $t_1, t_2$  are two timestamps where  $t_1 < t_2$  that define the time range of the communications.

Ideally,  $M_1, M_2, \dots, M_n$  should come from a discussion between managers and data experts.

To take a concrete example, let's consider communications in an organisation. A good collaboration could be defined as having:

- a high reactivity ( $M_1$ ).
- a high volume of exchange ( $M_2$ ).
- a balanced communication ( $M_3$ ).
- no communication outside working hours ( $M_4$ ).

---

## 3.6. WHEN ONA MEETS PEOPLE DATA

Additionally, the organisation could decide that it considers that  $\alpha_1$  and  $\alpha_3$  are twice as important as  $\alpha_2$  and  $\alpha_4$ .

In general, our recommendation is to define a metric based on the use case. For example, if we want to highlight employees and managers who have an asymmetric collaboration, we should query the pairs of employees where the number of direct messages sent one way is at least  $X$  times greater than the other way.

We can deduce important information about an organization using its communication data. However, on its own the network data can be imposing and unwieldy. It is when we add attributes about the employees (e.g. their department, age, grade, etc.) and make queries on a partial subgraph that we can get a deeper and more precise understanding of the organization.

### 3.6.2 Use Cases

In this section, we will look at a few use cases on how ONA data can be used with people data to deduce meaningful insights for an organization.

We will be using an anonymized Slack communication dataset from one of our client, an Australian company with 845 internal users, between August 2022 to December 2022.

#### **Inter-Team Collaboration**

Many companies are divided into multiple teams that often work together. Communication within the teams as well as across teams is very important for good and timely results. It is difficult to ascertain the quality of communication between two teams except through the use of surveys or interviews with managers. Not only can they be difficult to arrange, it is not feasible to have every member of the team take a survey and companies often have to defer to managers for their perspective on the communication level between teams. Even if managers are unbiased in their responses, surveys can be prone to human errors of judgement as we will see later in a case-study. A faster, more accurate and reproducible way to judge communication health between two teams is through the use of their ONA data.

CHAPTER 3. DESIGNING OF A FRAMEWORK FOR PEOPLE ANALYTICS AND ONA TOOLS AND DASHBOARDS

---

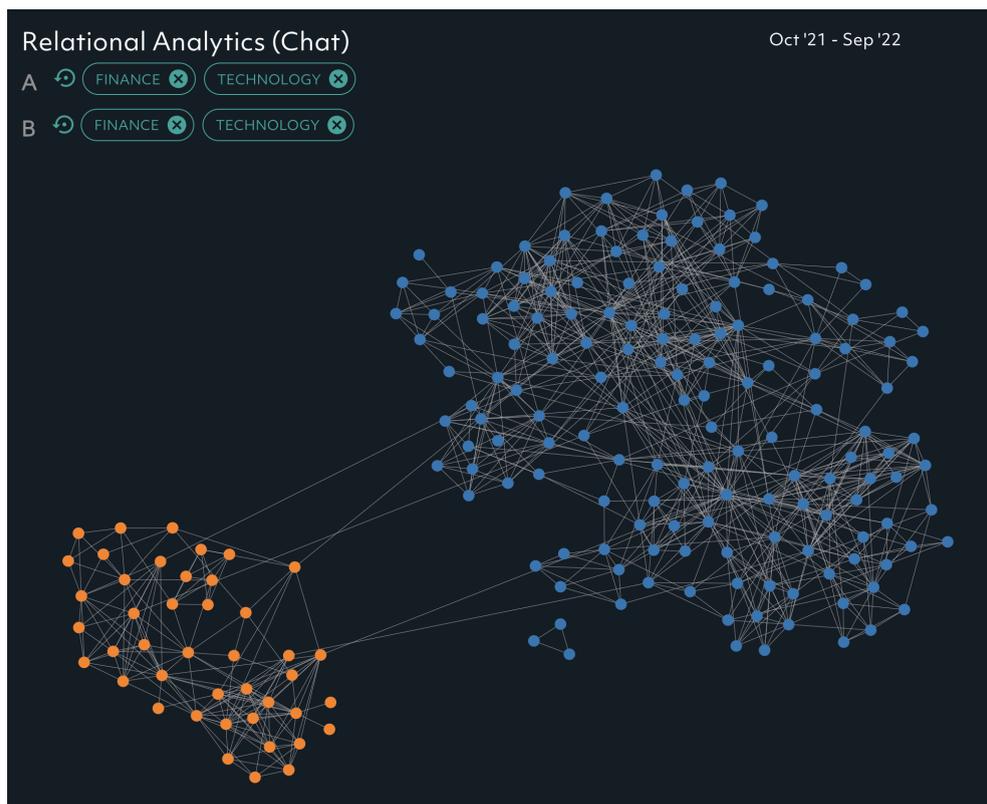


Figure 3.9: Collaboration between Two Teams

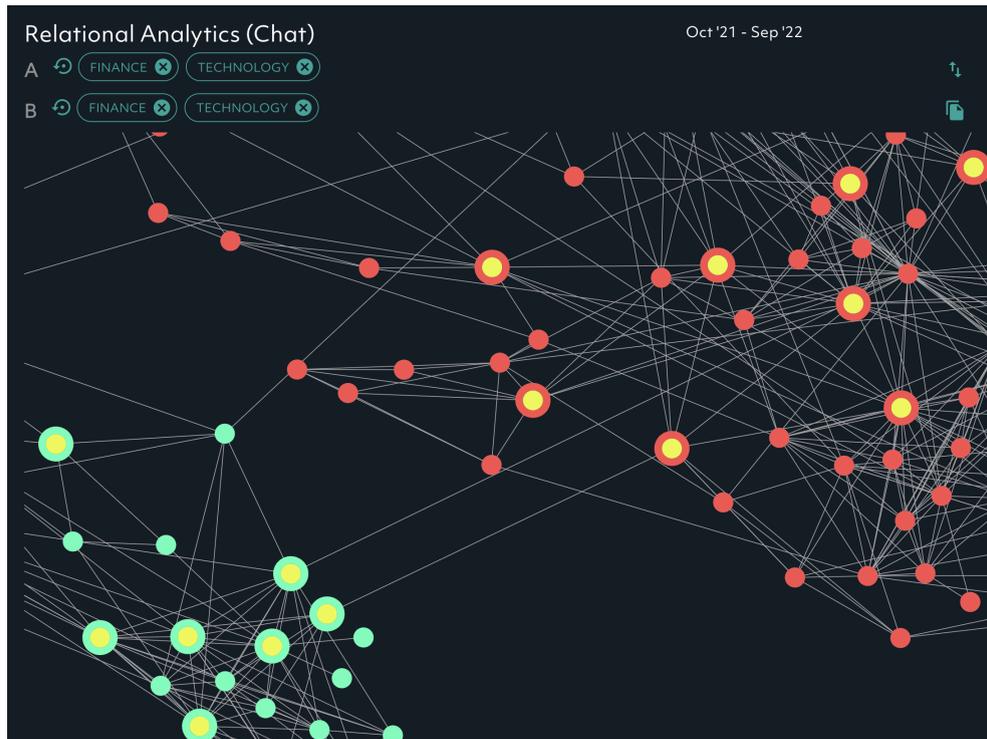


Figure 3.10: Inter-Team Collaboration with Manager Layer

In Figure 3.9, we look at ONA data for the Finance and Technology teams (by applying the A-B filters to those teams only) and we see that while there is an abundance of connections within each team, the flow of communication between the two teams is limited to only 5 channels, i.e, there are 5 pairs of employees that are responsible for the communication between the two teams. This can be crucial information for the organization.

If the two teams are supposed to be mostly independent of each other, then this communication model will work fine. However, in a technology company, where the product is the technology itself, the communication between the two teams is rather slim. It would make sense to open up more channels of communication to prevent a backlog of information. Moreover, if we use the layering function to add a layer for managers(yellow), we uncover more detail on the health of communication between the two teams.

We see in Figure 3.10 that even among the five channels, only one of them is between managers. This means that there are only two decision makers between the two teams that are connected. This can be an additional concern.

### **Better Onboarding of New Employees**

Many companies spend a lot of resources on getting the best talents. The cost of hiring does not end with a signed contract, rather it continues for the next few months with companies spending money and months on onboarding the new hires. At this stage, it is important to make sure that the new hires are sufficiently connected to their team members to have a smooth onboarding process.

They need people to teach them the team's business practices and the tools they use. In addition, they also need to be able to get along with their teammates and get to know their working styles.

Managers can make use of ONA data to pinpoint new hires that might not be getting an adequate onboarding experience or who seem to be left out. They can do that by looking at their communication within the team and see if they are connected to their manager and team members.

As an example, we can use A-B filters to restrict clients 2's data to the Finance team and use layering to pinpoint the employees with less than six months of tenure. The resultant graph is displayed in Figure 3.11. We can see that among the five new hires for the Finance team, three seem to be adequately integrated. They have multiple connections/collaborations with team members.

However, we also see two new hires that are only in communication with one other team mate. This can be worth looking into as these two hires might not be getting adequate resources to transition into their roles and this can prolong the time required to get them fully on board.

In cases where new employees are replacing existing team members, HR managers can utilize Relational Analytics to strategically build the new joiners' networks. By connecting them with the same individuals their predecessors were connected to, the transition process can be streamlined, ensuring a smooth handover and integration into the existing team dynamic. This approach not only fosters a sense of belonging for

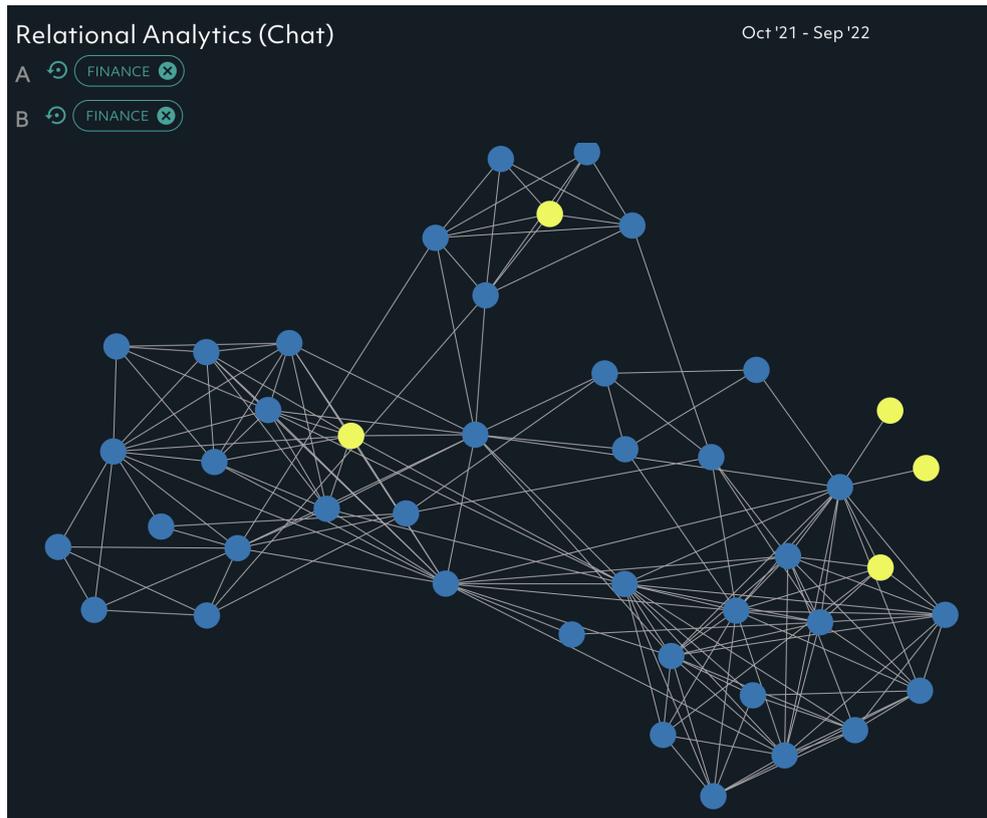


Figure 3.11: New Hires Communications

new employees but also helps maintain continuity and preserve valuable connections within the organization.

### Gender Diversity in Teams

Companies strive for an inclusive work culture where employees from various genders and demographics are not only represented but also included readily in conversations. There benefits in having diversity in workplace such as plurality of ideas owing to diverse experiences and the likelier retention of employees. But they also note that there can be significant communication gaps between the different groups.

We see that companies already spend time and effort on streamlining their hiring

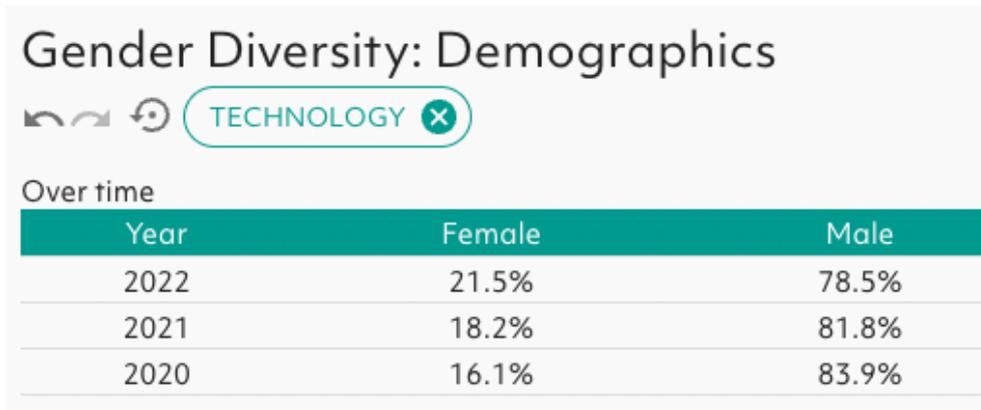


Figure 3.12: Gender Diversity

practices to make them more inclusive. For instance, there has been a sustained effort, spanning many decades, to include more women in the workplace, especially in the higher positions and male-dominated sectors like technology and finance.

In this use case, we focused on looking at the gender diversity in the Technology department. From our Panalyt dashboard, we can see that for a particular department the female headcount has been steadily rising by more than 13% per year to reach 21.5% of women from January 2020 to November 2022 (Figure 3.12). However, when we look at the communication graph (Figure 3.13) for the same department, we found that many females (in red) cluster together, and there are entire subsections of the graph where there are no collaborations with females - like in the bottom right, with just male employees communicating with each other.

Gaps like this in communication between demographics can impede the company from fully reaping the benefits of diversity. Our clients also benefit from similar use cases by studying how populations with different attributes (ages, tenure, etc.) work together.

### 3.6.3 Disengagement

People disengaging from their organization often exhibit a pattern of isolation. Relational Analytics can quickly reveal if an employee is losing connections over time and with whom (3.14). This information is valuable in determining whether an em-

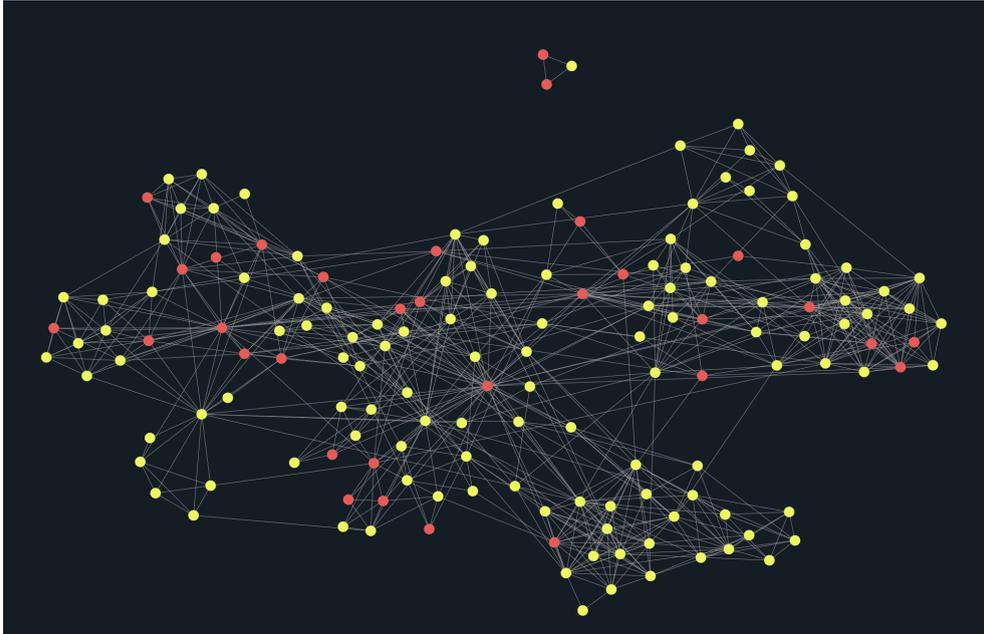


Figure 3.13: Network Graph Layered by Gender

ployee is becoming disengaged and disconnected from his/her team, department, or the organization as a whole.

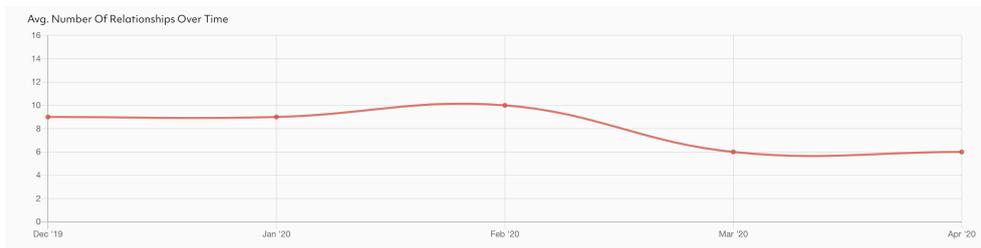


Figure 3.14: Example of an Employee Disengaging

Furthermore, Relational Analytics can provide insights into an employee's overall engagement levels by comparing the number of his/her connections with his/her peers within the same department, location, or other groupings. This comparison allows organizations to assess whether an employee has a generally low engagement level or if he/she is experiencing a decline in interactions, potentially signaling disengagement.

By identifying employees with declining connections or consistently low engagement levels, organizations can take proactive steps to address the issue, such as providing additional support, offering opportunities for professional development, or opening lines of communication to address concerns. In turn, these actions can help improve employee engagement, strengthen workplace collaborations, and contribute to a positive work environment that fosters productivity and growth.

### 3.6.4 Effective Onboarding

Relational Analytics can also be used to improve inclusion for new joiners. In Figure 3.15, we highlighted in orange people who joined the company for less than 6 months to find isolated new joiners. The company can then develop strategies to intervene and connect people by arranging 1 on 1s or group projects for example, and have a better onboarding.



Figure 3.15: Organizational Graph Used to Highlight New Joiners and their Connections

HR managers can help new joiners replacing an existing employee by building their networks strategically and by connecting them to the same people the employee they are replacing was connected to.

### 3.6.5 Sales Effectiveness

Relational Analytics can be utilized for both internal analysis within a company and external analysis involving collaborations between sales or customer success teams and their clients. This multifaceted approach allows organizations to optimize not only employee engagement but also client interactions, which can impact overall business success.

One key aspect of external Relational Analytics is analyzing responsiveness over time, such as the speed at which clients and sales executives reply to each other. Responsiveness is often correlated with the likelihood of closing a deal, and long periods without communication can signal potential issues.

In Figure 3.16, the chart displays the responsiveness and percentage of non-response over time for the sales team, as broken down by clients. The table on the right presents the same information, focusing on the latest month and further segmented by individual sales team members.



Figure 3.16: Client Responsiveness

Relational Analytics can also support the onboarding of new sales personnel. As they take over collaborations from previous team members, it is possible to examine the history of the predecessor’s responsiveness, providing insights into typical client expectations. This information enables organizations to assess how well the new salesperson is managing both inherited and newly established collaborations.

By leveraging Relational Analytics for both internal and external collaboration management, organizations can make data-driven decisions that foster employee engagement, optimize client interactions, and ultimately drive business success.

### 3.6.6 Conclusion

As collaboration platforms such as Slack, Teams and others become prominent and are increasingly used in virtual teams, organizations can now better understand what drives group or organizational performance.

This is confirmed by Keith G. Provan when he says: *"only by examining the whole network can we understand such issues as how networks evolve, how they are governed, and, ultimately, how collective outcomes might be generated."* [PFS07]

## Chapter 4

# Case Studies of Panalyt Clients

### 4.1 introduction

Chapter 4 of this thesis delves into a comprehensive examination of three distinct case studies involving Panalyt clients. Each case study provides a unique lens through which we analyze and address pertinent issues in the realm of human resources and organizational effectiveness. These case studies demonstrate practical applications of Panalyt's innovative solutions in real-world scenarios. In this chapter, we will explore how Panalyt's data-driven approach has been employed to understand employee attrition patterns, enhance managerial effectiveness, and tackle complex HR challenges, such as post-merger integration and HR data quality. Through these case studies, we aim to showcase the transformative power of data analytics and technology in addressing critical HR and organizational challenges.

### 4.2 Understanding Employee Attrition Using Communication Behavior Analysis

The following case study allowed us to win the third place of the 2023 Wharton People Analytics White Paper Competition [23g].

The Wharton People Analytics Conference is one of the most well-known and respected conferences in the field of People Analytics. It is organized by the Wharton School of the University of Pennsylvania and brings together leading experts, researchers and practitioners in the field.

### 4.2.1 Introduction

One of the most common ways for organizations to measure their employees' engagement is through surveys. However, the rise of digitalization and remote working, accelerated by the COVID-19 crisis, has created an environment where millions of users from thousands of organizations now use tools like Slack or Microsoft Teams for their daily communication. This offers new ways to measure employee engagement and preempt employee attrition.

In this case study, we aimed to leverage all this untapped information and get meaningful insights about employee engagement. We explored Slack channel data from two of Panalyt's clients and tried to deduce conclusions about employee engagement, especially for those who left the organization, in the last 15 months before they quit.

### 4.2.2 Data Collection and Processing

The population under study is from two Japanese tech companies. Dataset A is composed of the Slack communication data of 479 employees, from January 1st 2019 to March 31st 2022, including 50 employees who quit during that period. Dataset B is composed of the Slack communication data of 587 employees from January 1st 2020 to October 31st 2022, including 94 employees who quit during that period.

The datasets include public and private channels, but do not include any direct messages. We do not include direct messages as the companies' respective Slack subscription does not allow access to direct messages for extraction due to privacy reasons. For the purpose of this study, we look at three populations:

1. *Leavers*:

These are employees who have a valid termination date in the time frame we are studying. In addition, we only choose leavers that have a minimum tenure

## 4.2. UNDERSTANDING EMPLOYEE ATTRITION USING COMMUNICATION BEHAVIOR ANALYSIS

---

of 12 months i.e., they stayed employed for at least 12 months before their termination. This is to make sure that the employees have had enough time to get used to the company's business and practices and that we have enough data points to see their behavior change over time.

2. *The entire population:*

The total population, including the leavers we have defined above. This population helps to identify each company's wide communication trends over time.

3. *The active population:*

The total population, excluding the leavers we have defined above. This population serves as a benchmark to compare leavers versus active employees.

### 4.2.3 Description of the Communication Data

The data is obtained directly from Slack in the form of a series of files for each channel. Each file has the channel name as its filename and contains a series of text files for every day of communication in the channel. Each of these text files contains a list of messages sent where each message has the following fields:

- *type*: to indicate that the record is a message.
- *user*: an alphanumeric ID of the employee that sent the message.
- *text*: the content of the message.
- *ts*: UNIX timestamp for when the message was sent.
- *channelID*: the ID of the channel where the message has been posted.

As explained earlier, we do not include any text analysis on the message content. Thus, all content is redacted from the data before using it. For this purpose, we created an anonymizer application that reads in each file and iteratively removes any content from it. This includes the body of the message, as well as any links, URLs or descriptions included within. The application then exports the redacted files which we concatenate together and transform into a single database table using a Python script.

With the data imported, we begin our analysis on the outcome. We define the following groups of metrics:

1. Volume of messages sent:

The metric we measure here is the total number of unique messages sent by an employee in a month. This is the most basic way of measuring communication volume with more connected employees tending to have more messages sent.

2. Working hours:

This group of metrics measures the timing of communication. In particular we capture 3 metrics:

- (a) When employees are starting their communication. This is the time they send their first message of the day.
- (b) When employees are ending their communication. This is the time they send their last message of the day.
- (c) The range of communication. This is the total span of their communication and is the difference in hours and minutes of the previous two metrics.

This set of metrics is another angle to look at engagement levels among employees. We wanted to observe if communication spans changed as employees neared termination with our intuition suggesting that they would shrink for employees on the way out.

3. Channels Activity:

The metric we measure here is how many channels is the employee actively a part of. An employee being active in a channel means that they sent at least 5 messages in it in the same month. This threshold is set to avoid situations where an employee sends a one-off message in a channel without the intention of actively contributing to it. This metric is a good measure for diversity of communication and how involved the employee is in separate communication groups.

4. Number of collaborations:

The metric we measure here is how many frequent collaborators does the employee have. This is essentially a measure of the communication sphere of an employee. We define two employees being active collaborators if they are both

## 4.2. UNDERSTANDING EMPLOYEE ATTRITION USING COMMUNICATION BEHAVIOR ANALYSIS

---

active in at least one channel, i.e. if they both have sent at least 5 messages in the same channel.

Each of the above-mentioned metric is calculated in two different ways. Firstly, we calculate the metric for the leavers for each month in the past 15 months leading to their termination. Secondly, we calculate a benchmark as follows: we take the calculated metric for each terminated employee and subtract the same metric calculated for the active population in that particular month. This helps us observe each metric for leavers against the active population at that point in time.

### 4.2.4 Descriptive Analysis

#### Volume of Messages Sent

We start by studying the volume of messages sent, which we calculate by looking at the average count of messages sent by month for each employee.

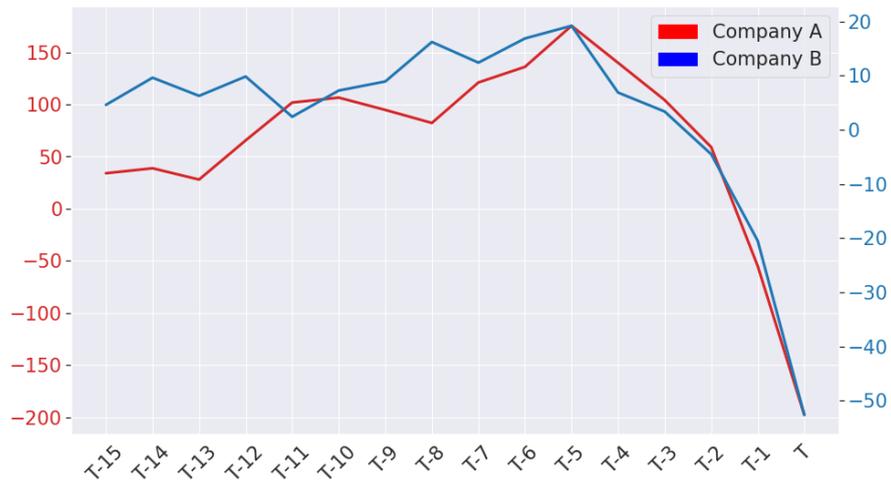
In Figure 4.1a, we focus only on the terminated employee and especially the last 15 months before they quit. We see that the average messages sent increases slightly for both companies. For Company A, it increases from 550 messages sent 15 months before their eventual termination to 625 messages sent 5 months before their eventual termination. Similarly for Company B, it increases from 85 messages sent 15 months before termination to 95 messages sent 6 months before termination. At this point, the average number of messages sent constantly decreases at a sharp rate as they get closer to their termination date. This indicates a gradual disengagement from 5 to 6 months before employees quit.

When benchmarked against the active population (Figure 4.1b), we see that the leavers sent a higher number of messages from 15 to 5 months before their eventual termination compared to their counterparts. However, similar to the previous graph, the average number of messages benchmarked against the active population starts rising at 8 months before termination, reaching a peak at 5 months before termination.

At this point the leavers have sent an average of around 160 messages more than



(a) Average count of messages sent by leavers for each month before their eventual termination date.



(b) Average count of messages sent by leavers for each month before their eventual termination date benchmarked against average messages sent by the active population for that month.

Figure 4.1: Analysis of Volume of Messages Sent

## 4.2. UNDERSTANDING EMPLOYEE ATTRITION USING COMMUNICATION BEHAVIOR ANALYSIS

---

the company average in Company A and 20 more in Company B. After this point we see a sharp decreasing trend, similar to the previous figure, and the average for leavers dips below the average for the active population eventually decreasing to 200 lesser messages for Company A and 50 lesser messages for Company B at the time of termination.

What we understand from this graph is that, on average, leavers tend to be more active than the active population and the gap widens until it reaches its peak 5 months before their termination. There is a similar inflection point in Figure 4.1a and Figure 4.1b that could indicate a pattern of overworking individuals.

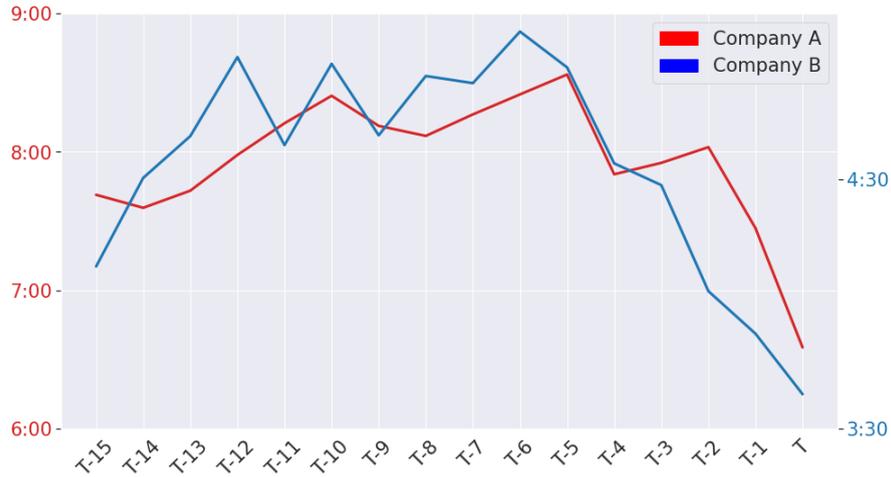
### **Working Hours**

In the followings graphs, we study the range of communication times for messages sent in the active population and for messages sent by leavers.

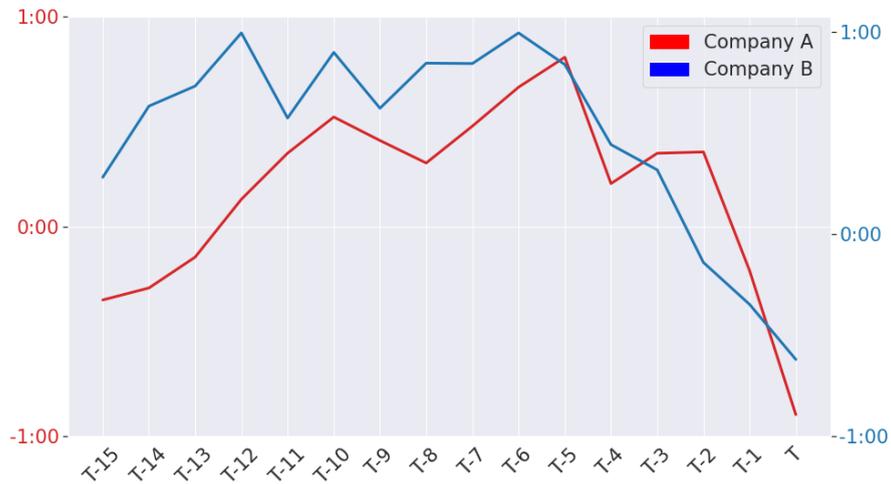
The average total communication (Figure 4.2a) for leavers starts increases steadily from 15 months before termination to around 5 months before the termination date. The average total communication then shows a decreasing trend for the last 5 months, having a small spike around 2-3 months before termination. The average total communication falls to the lowest value towards the end of their employment. This is 6 hours 30 minutes hours for Company A and 3 hours 45 minutes for Company B.

We can observe that on average, before employees quit, they start communicating more until they reach a point around 5 months before termination where the trend starts to reverse.

In Figure 4.2b, we observe that leavers are communicating lesser overall 15 months before their termination. Both these averages increase compared to the benchmark, reaching highs in the window of 5-7 months before termination. The highest average positive gap in total communication between leavers and the active population is 5 months before termination and is about 40 minutes. The highest positive difference in last message sent is 7 months before termination with leavers on average sending their last message 15 minutes later than the active population average. Both these averages then decrease to fall below the active population average. Leavers were on average communicating an hour lesser than the active population average.



(a) Average total hours of communication for each terminated employee for each month leading up to their eventual termination date.



(b) Average total hours of communication for terminated employee benchmarked against the averages for the active population for that month for each month leading up to their eventual termination date.

Figure 4.2: Analysis of Communication Times

## 4.2. UNDERSTANDING EMPLOYEE ATTRITION USING COMMUNICATION BEHAVIOR ANALYSIS

---

We find again a pattern of employees overworking compared to the average employee, especially 5 to 10 months before their termination.

### 4.2.5 Channels Activity

Here we study the number of channels in which employees are active.

The average count of active channels per leaver (Figure 4.3a) starts from around 10 for Company A and 9.5 for Company B, increasing to till 7 months before termination where it stabilizes at around 10 for Company A and 5 for Company B. It then starts a downward trend through the rest of the time period reaching the lowest count at the time of termination. The average count of channels increases to under 7 for Company A and under 3 for Company B.

The average count of channels actively used by leavers benchmarked against the active population average (Figure 4.3b) starts at 0.5-1 more channels 15 months before termination and starts widening slightly until 5 months before termination. At this point the gap begins shrinking rapidly and eventually we see leavers being involved in lesser channels. At the time of termination, the leavers are involved in 1.5 channels lesser than the active population.

Employees being active in more channels indicates that not only are they communicating more but there are also likely to be involved in more projects for the organization. This suggests again that employees who quit tend to overwork, or are over tasked compared to the rest of the company, and this is apparent a few months before they leave.

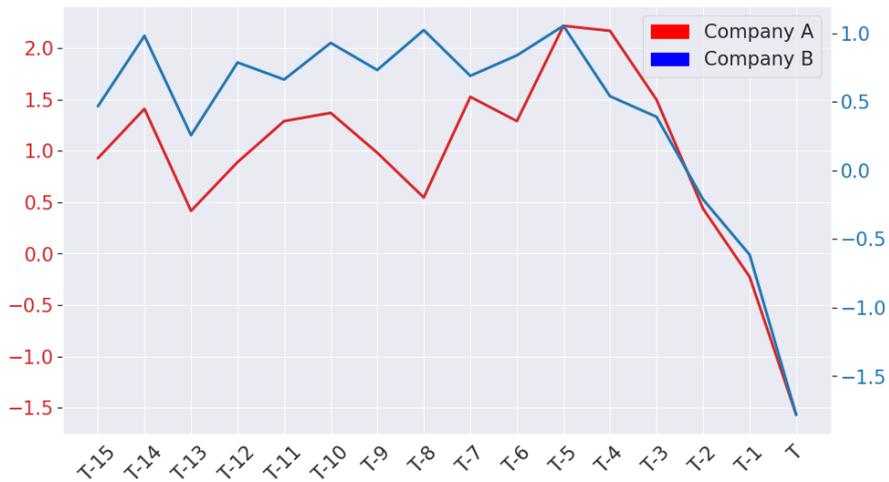
### 4.2.6 Number of Collaborations

In this last section, we analyze the average number of collaborations for the employees in the organization.

The average count of close collaborators for each leaver (Figure 4.4a) shows an increasing trend until 5 months before termination. From there the average goes on a rapidly decreasing trend reaching the lowers count at the time of termination.



(a) Average count of active channels used by leavers for each month leading up to their eventual termination date.



(b) Average count of active channels used by leavers benchmarked against the average for the active population for that month for each month leading up to their eventual termination date.

Figure 4.3: Analysis of Active Channels

## 4.2. UNDERSTANDING EMPLOYEE ATTRITION USING COMMUNICATION BEHAVIOR ANALYSIS

---



(a) Average count of active collaborators for each terminated employee for each month leading up to their eventual termination date.



(b) Average count of active collaborators for each terminated employee benchmarked against the average for the active population for that month for each month leading up to their eventual termination date.

Figure 4.4: Analysis of Collaborations

For leavers, the average count of collaborators benchmarked against the active population starts with a positive difference of 5 for Company A and 3 for Company B 15 months before termination (Figure 4.4b). It increases to 10 for Company A and 4 for Company B at 5 months before termination. For the remaining months the number of collaborators decreases rapidly and each terminated employee is involved in active conversation with 20 lesser collaborators for Company A and 6 lesser collaborators for Company B. We see again an inflection point at 4 months before the termination date.

#### 4.2.7 Conclusion of the Analysis

The analysis of the four metrics - volume of messages sent, working hours, channels activity and collaboration - all show that there is a clear trend between 5 to 6 months before the termination and the termination date.

The average trend is monotonic in most cases (volume of messages sent, channels activity and collaboration). In the last 5 to 6 months, on average, employees send less messages (monotonic downward trend), see their communication range reducing, are active in less channels (monotonic downward trend), and communicate with less people (monotonic downward trend). What we notice is that people disengaging from their organization tend to have a pattern of isolation in their last months before they leave.

A second observation is that we can see an inflection point for all metrics. The inflection point is 5 months before termination for volume of messages sent, working hours and collaboration, and 7 months before termination for channels activity. This suggests that those employees tend to disengage after a high peak of activity, and this might be part of the reason why they want to leave.

Finally, when compared to the active population, we see that employees about to quit tend to be on average more active and engaged in more channels. This could indicate that those employees are hard workers and/or are involved in more projects for the company. This insight along with the behavior mentioned in the earlier paragraphs might be a pattern of gradual overwork for employees about to quit, followed by a strong reduction of engagement in the company.

### 4.3. IMPROVING MANAGER EFFECTIVENESS AND INNOVATION

To conclude, this study is helpful to deduce very insightful conclusions with a limited amount of metrics.

Especially as a business leader, these insights can help prevent or reduce attrition by making sure employees are not being overworked either with too many collaborators and projects or that they are not working too many extra hours.

## **4.3 Improving Manager Effectiveness and Innovation**

Our second case study has been done with Panasonic Operational Excellence Co., Ltd., that partnered with Panalyt in March 2022 and applied ONA to identify organizational issues that could not be discovered through conventional methods such as surveys.

The detailed case-study is accessible online [[Pan](#)]. Following a merger and a group restructuring, the client's goal was to further strengthen collaboration between different departments that used to historically work independently.

Initially, the team had introduced a pulse survey to understand and improve the collaboration between employees with a relatively short tenure and their managers, as well as their motivation.

They quickly found that pulse surveys alone were not effective in identifying from where the communication issues were coming from, presumably because the survey answers are subjective and that bias could not be completely eliminated.

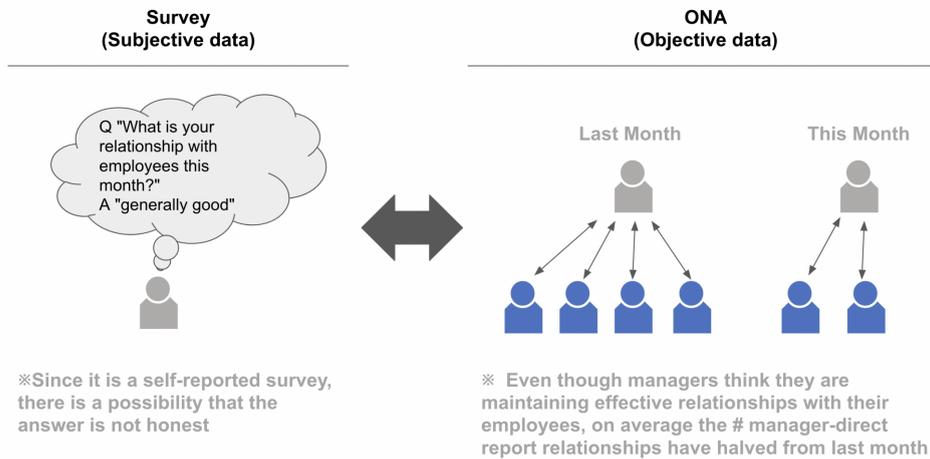


Figure 4.5: Surveys vs ONA

Some of their initial assumptions were that *"an employee with low motivation in a pulse survey may not be able to interact with other employees and may be isolated."* and that *"employees who interact with many other employees have high motivation."*

After analyzing their ONA data with Panalyt, they observed that some employees were *"isolated but maintain high motivation"*, or had *"low motivation despite being in the center of the organization and interacting with many other employees including those from other departments"*. The initial assumptions were therefore proven to be false.

Another assumption was that the higher the collaboration between an employee and his manager, the more likely the employee would meet the expectations relating to his/her role and responsibilities.

However, the client realized that managers were only able to have a grasp on their relationships with their direct reports; but they were unaware of the work their team members were doing in collaboration with members from other departments - making it impossible to notice signs of collaboration overload that can lead to burnout unless the employees speak out themselves (Figure 4.5).

#### 4.4. IDENTIFYING THE ROOT CAUSE OF EMPLOYEE TURNOVER AND IMPROVING POST-MERGER INTEGRATION

---

This was especially true at the time of the imposed working from home period during the COVID-19 pandemic.

## 4.4 Identifying the Root Cause of Employee Turnover and Improving Post-Merger Integration

### 4.4.1 Introduction

CAPLAN Corporation [CAPLAN] started as a recruitment agency, an affiliated company of ITOCHU Corporation in the year 1982, and has been expanding its services in human resource-related businesses since. The company is now part of the Pasona Group, providing education and training, as well as HR tech services.

CAPLAN faced a major hurdle in identifying the reasons for employee turnover - HR had no visibility into people data collected across the employee lifecycle, as a result of which significant people decisions were being made solely based on intuition.

Within a month of implementation, CAPLAN Corporation was able to identify that a major cause of turnover was the lack of employee transfers/cross-pollination of talent between merged/acquired companies, as a result of which the diversity of employee experience and skills were significantly affected, along with silos forming across the merged entity.

In this case study, Hirama Yoshikazu-sama, General Manager of the Talent Management Division shares how CAPLAN leverages Panalyt's solutions to embed data-driven HR decision making at the heart of its talent processes.

He says that leveraging Panalyt's solutions to analyze the company's HR data has allowed it to discover the real talent issues affecting the organization, and has allowed the employees to validate hypotheses made on intuition and suggest data-driven approaches to improve talent management.

#### 4.4.2 Problem 1: A Lack of Visibility into HR Data

CAPLAN is a company that has experienced mergers, acquisitions, and business transfers in the past. CAPLAN believes that its strength lies in the diversity of its employees' ways of thinking, the culture they have experienced, and the differences in their careers and the skills associated with them. However, as CAPLAN operated as a group of merged entities, it was of utmost importance to get a holistic view into the employee journey across the group.

CAPLAN faced a pressing issue of a lack of visibility into HR-related problems, as a result of which significant people decisions were being made solely on intuition without being backed by data.

Companies that have been using talent management tools such as SAP's Success Factors for a long time are beginning to seek People Analytics as to the next level.

Some of CAPLAN's customers are also advanced in their efforts to embrace People Analytics not just as a means to identify individual problems but identify and address wider organizational issues. As a result, CAPLAN was on the lookout for a People Analytics tool to give a bird's eye view into the organization's talent across the employee journey from hire to fire.

CAPLAN tried HR core systems and BI tools to find the real right issues to focus on at a macro level in the organization. This was because all the preexisting tools were focused on the individual talent with a bug's eye view and were not designed to provide a bird's eye view into the organization's employee management system.

They had a strong feeling that it would be impossible to determine the exact issues faced by CAPLAN without taking a bird's eye view of the wider organizational HR management system rather than looking at each employee individually

#### 4.4.3 Problem 2: HR Data Quality and Cleanliness

As a company that supports the implementation of HR tech services, CAPLAN is more aware of the difficulties of implementing and utilizing HR data than anyone else.

#### 4.4. IDENTIFYING THE ROOT CAUSE OF EMPLOYEE TURNOVER AND IMPROVING POST-MERGER INTEGRATION

---

Even as a mid-sized organization, CAPLAN had to deal with the classic issue of HR data not being neat and tidy. When CAPLAN's HR team first saw 20,000 CSV data files not in order, their initial reaction was that it felt like it would take a year just to manually clean up the HR data.

Hirama Yoshikazu-sama mentioned that the first step that companies often take when trying to implement and utilize HR data was to hire a data scientist, have the HR department learn statistics, or hire and train a specialist and they take a detour. Also, in most cases, HR data is not in order, and even if they can hire and train data experts, the cost and effort of data cleansing is enormous.

##### 4.4.4 Solution

Panalyt integrates and analyzes people data scattered across disparate HR, Productivity, Finance and Digital Communication systems and files to provide actionable and holistic people insights to the C-suite, HR/HRBPs and Line Managers through its suite of pre-built analytic tools covering commonly requested insights across the end-to-end employee lifecycle.

The deciding factors for CAPLAN to choose Panalyt as their trusted People Analytics partner were the accuracy and speed of the data cleansing process, comprehensive Japanese compatibility despite being a foreign product, along with the ability to start uncovering insights within a week of starting implementation.

As a result, people data and insights are delivered to users without any omissions, and Panalyt's overwhelmingly easy-to-use interface allows people managers to validate their own hypotheses about issues affecting the wider organization.

##### 4.4.5 Insights Uncovered with Panalyt

Hirama Yoshikazu-sama's feedback is that Panalyt has helped them clarify and align on the wider organizational issues that need to be addressed for improving CAPLAN's people strategy and processes. By taking into account the analysis of people (the entire organization) instead of the analysis of talent (individual), they were able to gain a multi-dimensional understanding of their talent issues and allow their HR team and people managers at CAPLAN to provide suggestions and solutions at an early stage.

At the beginning of the implementation, only static data such as employee attributes were connected, but the trend of attributes of employees who were leaving the company discovered was very different from CAPLAN's former hypotheses.

CAPLAN further discovered that there were almost no personnel transfers between the merged companies, as a result of which the diversity of employee experience and skills was not being fully exploited.

The next step is to implement measures to improve the issues that have been clarified. In the future, CAPLAN would like to further evolve the PDCA cycle (Planning, Doing, Checking, and Acting) for organizational improvement from issue discovery to solution implementation by adding insights from more dynamic sources of people data, such as analyzing communications and collaboration patterns using Panalyt's Relational Analytics (Organizational Network Analysis) capabilities.

## 4.5 Conclusion

In summary, this chapter has provided an in-depth exploration of three case studies featuring Panalyt clients, demonstrating the practical application of data-driven HR solutions.

We started by examining how communication behavior analysis sheds light on employee attrition, showcasing Panalyt's ability to uncover hidden insights within organizational communication data. We then witnessed the tangible improvements in managerial effectiveness and innovation driven by Panalyt's interventions. Finally, we addressed the intricate challenges of post-merger integration and HR data quality, illustrating Panalyt's capacity to bring clarity and transformative change to complex scenarios.

These case studies underscore the vital role technology and data analytics play in addressing HR and organizational challenges, emphasizing Panalyt's success in driving positive change and enhanced organizational performance. In an era where data-driven decision-making is paramount, these use-cases highlight the potential for innovation and improvement through advanced HR analytics solutions.

## Chapter 5

# Attrition Prediction

### 5.1 Introduction

In this chapter, we will present the attrition model we built at Panalyt.

Employee attrition is a reality that catches the attention of every business, no matter what it does. One of the most effective ways to tackle this issue is to build a predictive model - based on the wide variety of HR data available - to identify employees who are showing an intention to leave voluntarily.

These models can vary in how they are built and operate but essentially follow the same principle - take a sample of employees that have already quit the company and use their data to build a model to highlight with an adequate advance who is likely to leave the company in the near future.

By identifying the employees at risk and by understanding why employees choose to end their employment with the company, the latter can make improvements to reduce existing attrition problems.

To date, attrition prediction research based on HR data typically focuses on constructing classification models that are trained using only one observation per employee over a limited period of time (Fig. 5.1), and do not consider its variation over time, discarding employees who have left earlier [Aji16].

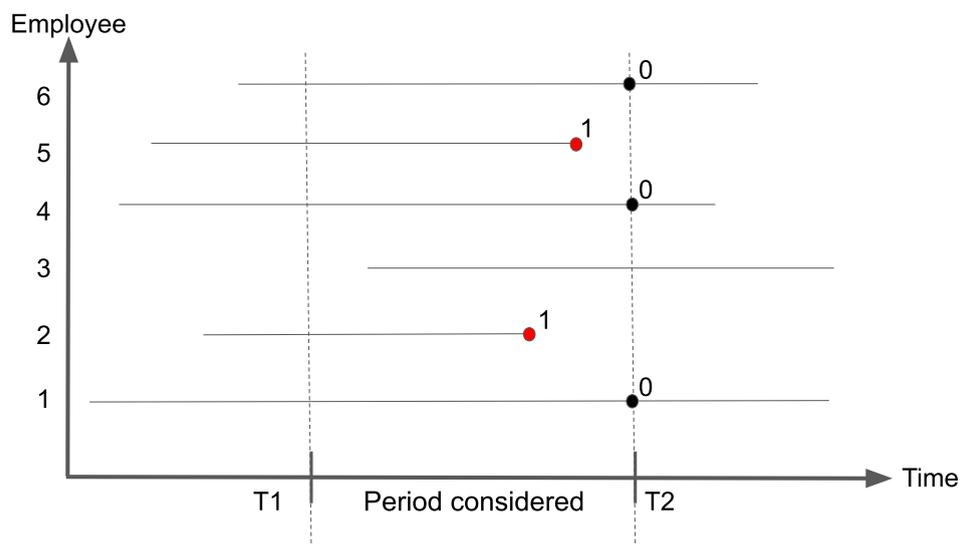


Figure 5.1: Method Using One Observation per Employee

Historical records that change over time like the employee salary or the grade are transformed into non-time-series ones by applying statistical operations like the sum, the average or looking at the latest values. Such approaches result in information loss and therefore less effective predictions.

In this chapter, we introduce a dynamic method for generating training data from employee records and we use it to predict who is at risk of leaving the company within the next 6 months. By categorizing the employees in different ‘buckets of risk’ from *low risk* to *high risk*, our goal is to identify employees with high attrition risk to prevent them from leaving.

A few studies have considered different approaches for attrition prediction. However, to the best of our knowledge, none of them applies our framework to solve the employee attrition prediction problem.

The obtained model for the prediction of employee’s attrition is tested on an anonymized dataset on a real company provided by Panalyt Pte. Ltd. [23a], which includes HR information from 1650 terminated employees.

## 5.2 Method

We define *Employees at risk* as a discrete variable as follows. Records where the value *number of months before the termination date* is smaller than 6 are labelled as 1 (at risk), and records where the value *number of months before the termination date* is greater than 6 are labelled as 0 (not at risk).

$$\text{Employees at risk} = \begin{cases} 1 & \text{if } \# \text{ months before the termination date} \leq 6 \\ 0 & \text{if } \# \text{ months before the termination date} > 6 \end{cases}$$

One of the issues with the above labelling is that employees with a long tenure in the company can have many times more records compared to others.

Another issue is that it usually takes at least a few months to see a noticeable change in the employee records, which makes the records very redundant and the dataset imbalanced.

Thus, we use a sampling method to tackle those issues and we will compare the performance of the model using our sampling method and using only one observation per employee.

For the records labelled as 0 (not at risk) and for each employee, we keep up to 4 equidistant records between the earliest record and T-6, with a minimum distance of 3 months between the records. For the records labelled as 1 (at risk), we keep up to 3 records : 2, 4, and 6 months before the termination date ( $T - 2$ ,  $T - 4$  and  $T - 6$ ).

We need to keep different records for the model to be able to capture any change at different stages of the employees' employment in the company.

Figure 5.2 helps us to visualize how the sampling works. The black dots represent the observations labelled as at risk, and the white dots represent the observation when the employees are not at risk.

Regarding the model, Ajit et al. [Aji16] shows that to predict attrition, the XGBoost classifier performs better compared to Logistic Regression, Random Forest, KNN and others.

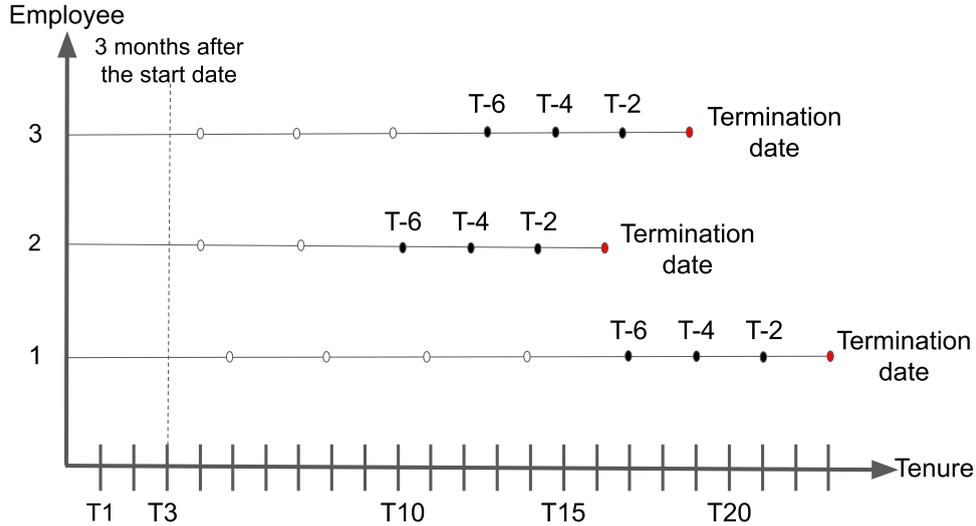


Figure 5.2: Method Using Multiple Observations per Employee

XGBoost is capable of handling the noise and the null values of the dataset as compared to the other classifiers, overcoming important challenges [Kak+20]. In light of these reasons, we trained our model using XGboost.

## 5.3 Experiments

### 5.3.1 Dataset

For our experiment, we use an anonymized dataset provided by Panalyt[23a]. The dataset contains monthly records of 1,650 permanent employees, from their start date to their termination date. As the probation period is too unpredictable (many people leave during their onboarding because there is simply no good fit), we decide to exclude the first 3 records of each employee.

For the first dataset (named D1), we implement the method described by Ali et al. [GA14], which entails utilizing multiple training observations per customer with a consistent time interval of 1 month (Multiple Period Training Data). The resulting dataset contains data from 1650 employees that have already left the company,

with a total of 56,464 monthly records from 4 months after their start date to their termination date.

The reduced dataset (named D2) contains 8755 records, from which 4948 are labelled 'at risk' and 3807 'not at risk'. 3338 out of 8755 records contains missing values, which is very common in HR dataset, and XGBoost is particularly convenient as it can handle missing values.

For the purpose of our study, the dataset is reduced to 15 attributes:

- *employee id*: unique identifier of the employee
- *effective date*: date of the record
- *start date*: date on which the employee started
- *termination date*: date on which the employee left
- *tenure*: time since the employee joined the organization
- *month of starting date*: month when the employee joined
- *month of the observation*: month of the record
- *salary*: current salary of the employee
- *time since last salary*: number of months since no salary increase
- *grade*: current grade of the employee
- *time since last grade*: number of months since no grade increase
- *median salary*: median salary of the employee's team
- *gender*: gender of the employee
- *age*: age of the employee
- *months before the termination date*: number of months between the effective date and the termination date

The features *number of months since no salary increase* and *number of months since no grade increase* are created using the historical records.

To train the model, we exclude the variables employee id, effective date, start date, termination date and months before the termination date. The training dataset contains variables that we separate in 3 types:

- Traits, or variables that never change: *gender, month of starting date*
- States, or variables that can change overtime: *age, tenure, salary, etc.*
- Lags: *time since last salary, time since last grade*

States and lags are particularly useful to organizations as it can tell them what to do to keep an employee at risk of leaving. For instance, if a manager knows that an employee is at risk of leaving mainly because he/she hasn't been promoted for a long time, the manager can take action.

### 5.3.2 Results

We apply a 10-fold cross-validation to provide a robust estimate of the performance of our model (Table 5.1). To avoid any data leakage, we make sure that different records from the same employee are not used in the training datasets and the testing datasets. Apart from the accuracy, all the other metrics are higher for the reduced dataset (D2). With a recall of 0.2, the model trained on D1 fails to identify the employees at risk compared to the model trained on D2.

Table 5.1: Model Results

Dataset	Accuracy (std)	Precision (std)	Recall (std)	AUC (std)
D1	0.83 (0.00)	0.64 (0.03)	0.20 (0.03)	0.59 (0.1)
D2	0.72 (0.01)	0.74 (0.01)	0.79 (0.02)	0.71 (0.1)

If we pay attention to the model trained on D2 and the distribution of the probabilities (Fig. 5.3), we see that the model's predictions seem very accurate when the probabilities are smaller than 0.20 and greater than 0.80.

When we take a closer look at the proportion of correct predictions for different probabilities (Fig. 5.4), we see that the higher the probability, the more confident the

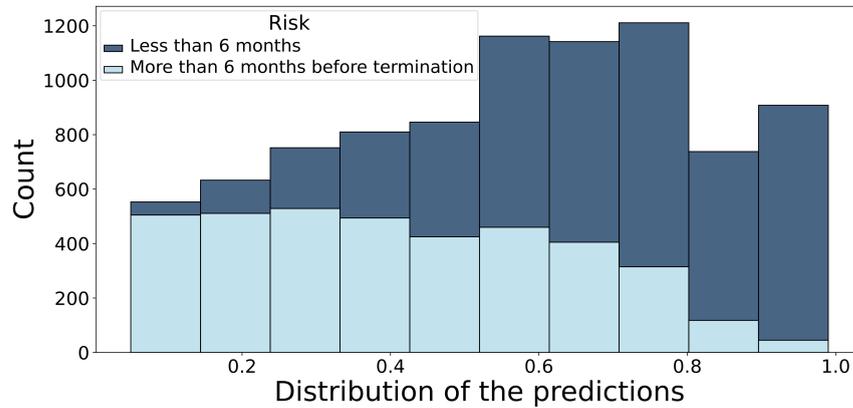


Figure 5.3: Distribution of the Predictions by Class

model is that an employee is at risk of attrition - it reaches 96.2% of accuracy when the probabilities are greater than 0.9. On the other hand, the lower the probability, the more confident the model is that an employee is not at risk.

The model is less accurate when the probabilities are between 0.4 and 0.6, which makes sense.

We propose to have different ‘buckets of risk’ depending on the probabilities: very low Risk ( $[0,0.2)$ ), low risk ( $[0.2,0.4)$ ), intermediate risk ( $[0.4,0.6)$ ), high risk ( $[0.6,0.8)$ ) and very high risk ( $[0.8,1]$ ). Using this classification can help organizations to highlight where to focus their attention.

Another question that we asked ourselves is how does the model perform for the records at T-2, T-4 and T-6? This time, we trained the model on 80% of the dataset and we kept track of which record were at T-6, at T-4, and at T-2 to see the distribution of the probabilities returned for these records (Fig. 5.5).

With an average score of 0.64 at T-6, 0.68 at T-4 and 0.72 at T-2, we notice that as expected, the closest we are to the termination date, the more confident the model is that the employee is at risk of leaving.

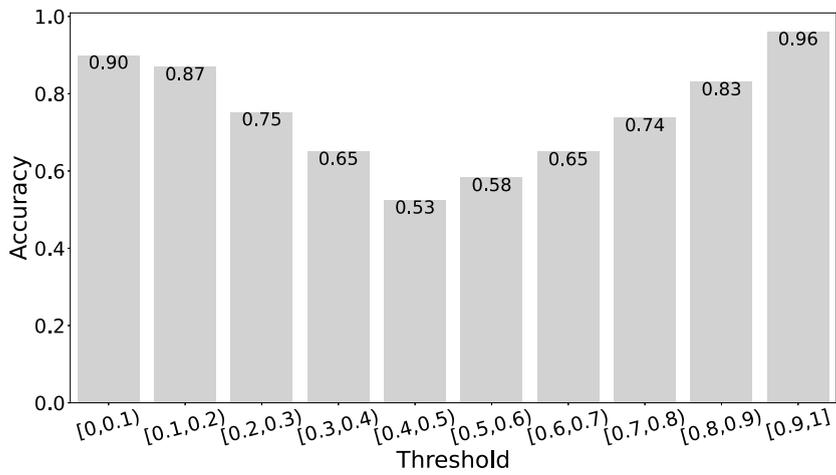


Figure 5.4: Performance of the Model for Different Probability Buckets

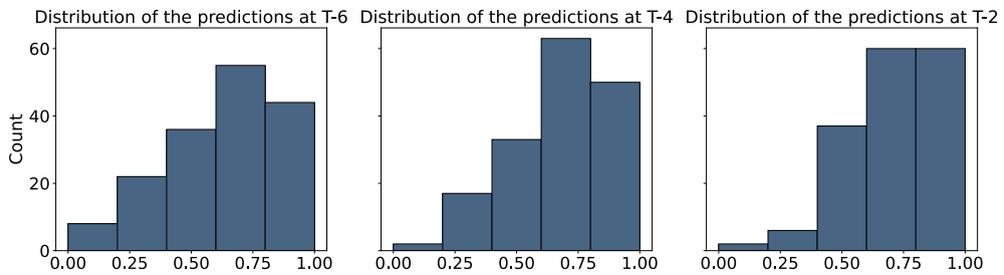


Figure 5.5: Performance of the Model at T-6, at T-4, and T-2

## 5.4 Explainability

Following our prior results, the main request that we had from clients was to present the drivers of attrition for a population, which belongs to the realm of explainability.

To solve that problem we used a well-known library called SHAP (SHapley Additive exPlanations) [LL21]. SHAP is a popular and powerful open-source framework in Python for interpreting and explaining the output of machine learning models. It is based on the concept of Shapley values, which is a game theory technique used to fairly distribute the contributions of different features to a prediction.

Using this library, we are able to identify the top driving factors of any population at risk of leaving the company (Figure 5.6).



Figure 5.6: Top 3 Drivers for a Population of Employees at Risk of Leaving

## 5.5 Attrition Analytic Tool

Following this study, we built an analytic tool at Panalyt to help business leaders identify the populations at risk of leaving.

We measure attrition risk by analyzing various demographic-based data points - more than those used in the prior study - such as age, gender, tenure, who's the manager, how many managers the employee had in the past year, the manager's attrition rate, etc., in conjunction with Organizational Network Analysis on how connected employees are with the organization, their manager and their teams, allowing us to predict how likely someone is to leave the organization over the next six months and why. This model is fine-tuned for each client's historical employee data.

This allows us to predict the attrition risk, the likelihood of an employee leaving in the next 6 months, and then visualize this information, both as a risk % and headcount

## CHAPTER 5. ATTRITION PREDICTION



Figure 5.7: Attrition Risk Analytic Tool

at risk, across multiple dimensions (Figure 5.7)) such as by organization, department, tenure groups, etc.

Attrition risk, as we define it at Panalyt, can also be seen as a proxy for engagement risk/ loss of engagement over the next six months.

A deeper analysis of an organization’s attrition risk data allows it to see employee attrition risk profile by team.

Once the organization has identified the level of risk, it can then drill down to what are the key drivers of the risk and identify leading indicators of the risk of people leaving. In most organizations we work with, there is usually a dominant driver related to specific organizational issues around diversity, compensation/lack of change in compensation, particular managers, etc.

HR teams can then interpret these drivers in the context of their organization to improve their understanding of possible risk factors for each employee/group of employees and then come up with a practical approach to mitigating risk based on their knowledge of what’s possible to implement. For example, coming up with initiatives to give top performers more facetime with senior leadership or designing an innovation program that lets people take time to do their own projects.

It is important to not confuse drivers with recommendations. For example, if the % of women in a team is a major risk driver, it may or may not be possible to move employees around, change the team structure, or hire externally to improve the % of women or team depending on your organizational context. Similarly, if compensation is a major risk driver, you cannot realistically do much if someone is at the top of their compensation and the compensation hasn't changed in the past 12 months.

The ability to predict the attrition risk for each employee, along with a playbook for using this data in an ethical manner, can help empower HR departments and managers to change from a reactive mindset to a proactive approach for preventing key talent from leaving the company.

## 5.6 Conclusion

In this section, we proposed an effective approach that takes advantage of the longitudinal nature of the data to solve the employee attrition prediction problem. Our sampling method reduced the dataset and we demonstrated that the model performs significantly better after sampling.

We also showed that on average, the risk score gets higher and higher during the last 6 months before an employee leaves.

We then used the probabilities returned by the model, and we proposed a bucketization method to highlight different risk profiles for different employees.

We also explained our solution to provide the main drivers of population at risk of quitting using the SHAP library.

Finally, we showcased Panalyt's attrition analytic tool that was built based on this research.

Further research would allow us to explore other methods of sampling, looking at the impact of training the data with more or less records.

Another research could be to study more deeply the probabilities returned by the model and if, for each employee, a trend could be identified.



## Chapter 6

# Conclusions

This thesis has provided a comprehensive analysis of People Analytics and Organizational Network Analytics, their benefits, and practical applications in medium to large-sized organizations. Our research has demonstrated the potential of these tools in driving improved employee engagement, increased productivity, and better decision-making processes.

By developing a practical platform and user-friendly analytic tools, we aim to make these concepts more accessible to organizations and contribute to their successful implementation in the workplace.

We had a particular focus on attrition, and studied how communication metadata were good predictors for attrition. Later we developed an attrition prediction models, with a particular sampling method taking into account the historical information about employees.

Furthermore, we have addressed the challenges and ethical concerns associated with analyzing sensitive employee data. Our research highlights the importance of responsible data usage, ensuring that employees' privacy and security are protected at all times. As the field of People Analytics and Organizational Network Analytics continues to evolve, it is critical to maintain a focus on ethical considerations to foster trust and credibility among stakeholders.

## CHAPTER 6. CONCLUSIONS

---

Moreover, the practical client case studies based on real data provided further validation for the proposed methods, emphasizing their real-world applicability and impact.

Finally, this thesis has contributed to the field by introducing a dynamic method for generating training data from employee records and developing attrition prediction models. This innovation will enable organizations to proactively address potential employee turnover and mitigate its negative consequences.

## Chapter 7

# Furthermore

The field of People Analytics and Organizational Network Analytics continues to evolve rapidly, with new advances and opportunities emerging at a staggering pace. This research opens the door for additional exploration and development in the field of People Analytics and Organizational Network Analytics.

Future studies may delve deeper into the integration of artificial intelligence and machine learning techniques to improve predictive models and provide more tailored recommendations for organizations. This could lead to more accurate and efficient decision-making processes and further enhance the value of these tools in organizational settings.

Additionally, researchers could investigate the applicability of these analytics tools in various industries and organizational structures, uncovering new insights and best practices for diverse contexts.

The potential for cross-disciplinary research also exists, as the principles of People Analytics and Organizational Network Analytics could be applied to other domains such as education, healthcare, and public administration. This would broaden the understanding of human interactions and collaborative networks in different settings, ultimately contributing to more effective, data-driven decision-making across a wide range of fields.

By continuing to expand upon the findings of this thesis, the future of People Analytics

## CHAPTER 7. FURTHERMORE

---

and Organizational Network Analytics holds immense promise for unlocking the true potential of human capital and fostering a culture of continuous improvement and innovation.

# Bibliography

- [TTF79] Noel M. Tichy, Michael L. Tushman, and Charles Fombrun. “Network analysis in organizational settings”. In: *Human Relations* 32.11 (1979), pp. 923–965.
- [KS88] David Krackhardt and Robert N. Stern. “Informal networks and organizational crises: An experimental simulation”. In: *Social Psychology Quarterly* (1988), pp. 123–140.
- [KH93] David Krackhardt and Jeffrey R. Hanson. “Informal networks: The company behind the chart”. In: *Harvard Business Review* 71.4 (1993), pp. 104–111.
- [HS97] Mary Jo Hatch and Majken Schultz. “Relations between organizational culture, identity and image”. In: *European Journal of Marketing* 31.5/6 (1997), pp. 356–365.
- [CP02] Rob Cross and Laurence Prusak. “The people who make organizations go - or stop”. In: *Harvard Business Review* 80.6 (2002), pp. 104–112.
- [LML03] Edward E. Lawler, Susan A. Mohrman, and Gerald E. Ledford. “Strategies for high-performance organizations: The CEO report”. In: *Jossey-Bass* (2003).
- [CP04] Rob Cross and Andrew Parker. “The hidden power of social networks: Understanding how work really gets done in organizations”. In: *Harvard Business School Press* (2004).
- [Fre04] L. C. Freeman. *The development of social network analysis: a study in the sociology of science*. Vancouver, B. C.: Empirical Press, 2004.

## BIBLIOGRAPHY

---

- [KY04] Bryan Klimt and Yiming Yang. “The Enron corpus: A new dataset for email classification research”. In: *European Conference on Machine Learning* (2004), pp. 217–226.
- [Rob04] Stephen P. Robbins. “Organizational Behavior”. In: (2004).
- [SA04] Jitesh Shetty and Jafar Adibi. “The Enron email dataset database schema and brief statistical report”. In: *European Conference on Machine Learning*. Springer, 2004, pp. 217–226.
- [Gui+06] Roger Guimerà et al. “Team assembly mechanisms determine collaboration network structure and team performance”. In: *Science* 308.5722 (2006), pp. 697–702.
- [KKP06] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas. “Handling imbalanced datasets: A review”. In: *GESTS Int. Trans. Comput. Sci. Eng.* 30 (2006), pp. 1–12.
- [NGD06] Galileo Namata, Lise Getoor, and Christopher Diehl. “Inferring formal titles in organizational email archives”. In: *Proc. of the ICML Workshop on Statistical Network Analysis*. 2006.
- [PFS07] Keith G. Provan, Andrew Fish, and Jörg Sydow. “Interorganizational networks at the network level: A review of the empirical literature on whole networks”. In: *Journal of Management* 33.3 (2007), pp. 479–516.
- [Row+07] Robert Rowe et al. “Automated social hierarchy detection through email network analysis”. In: *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*. 2007, pp. 109–117.
- [CMN08] Aaron Clauset, Cristopher Moore, and M. E. J. Newman. “Hierarchical structure and the prediction of missing links in networks”. In: *Nature* 453.7191 (2008), pp. 98–101.
- [Gal08] Jay R. Galbraith. *Designing organizations*. John Wiley & Sons, 2008.
- [SD09] Rob Silzer and Ben E Dowell. *Strategy-driven talent management: A leadership imperative*. Jossey-Bass, 2009.
- [DHS10] T. H. Davenport, J. G. Harris, and J. Shapiro. “Competing on talent analytics”. In: *Harvard Business Review* 88.10 (2010), pp. 52–58.
- [Pen10] Alex Pentland. “The new science of building great teams”. In: *Harvard Business Review* 88.4 (2010), pp. 60–69.

- [BH11] Stephen P Borgatti and Daniel S Halgin. “On network theory”. In: *Organization science* 22.5 (2011), pp. 1168–1181.
- [Gup+11] Manish Gupte et al. “Finding hierarchy in directed online social networks”. In: *Proceedings of the 20th international conference on World wide web* (2011), pp. 557–566.
- [HCL11a] J. Harris, E. Craig, and D. Light. “Talent and analytics: new approaches, higher ROI”. In: *Journal of Business Strategy* 32 (2011), pp. 4–13.
- [HCL11b] J.G. Harris, E. Craig, and D.A. Light. “Talent and analytics: new approaches, higher ROI”. In: *Journal of Business Strategy* 32.6 (2011), pp. 4–13. DOI: [10.1108/02756661111180087](https://doi.org/10.1108/02756661111180087).
- [PBK11] Sebastian Palus, Piotr Brodka, and Przemyslaw Kazienko. “Evaluation of organization structure based on email interactions”. In: *International Journal of Knowledge Society Research (IJKSR)* 2.1 (2011), pp. 1–13.
- [ABW12] S. Aral, E. Brynjolfsson, and L. Wu. “Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology”. In: *Manage. Sci.* 58 (2012), pp. 913–931. DOI: [10.1287/mnsc.1110.1460](https://doi.org/10.1287/mnsc.1110.1460).
- [Gal+12] M. Galar et al. “A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches”. In: *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 42 (2012), pp. 463–484.
- [DR14] Cynthia Dwork and Aaron Roth. “The Algorithmic Foundations of Differential Privacy”. In: *Foundations and Trends® in Theoretical Computer Science* 9.3–4 (2014), pp. 211–407.
- [FJ14] Jac Fitz-Enz and II John Mattox. *Predictive analytics for human resources*. John Wiley & Sons, 2014.
- [GA14] Özden Gür Ali and Umut Aritürk. “Dynamic churn prediction framework with more effective use of rare event data: The case of private banking”. In: *Expert Systems with Applications* 41 (2014), pp. 7889–7903.
- [Tha+14] D. Thammasiri et al. “A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition”. In: *Expert Syst. Appl.* 41 (2014), pp. 321–330.
- [Arn+15] V. Arnaboldi et al. *Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs*. Amsterdam: Elsevier, 2015.

## BIBLIOGRAPHY

---

- [Cho+15] Munmun De Choudhury et al. “Inferring Relevant Social Networks from Interpersonal Communication”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 22.2 (2015), pp. 1–27.
- [Don+15] Yuxiao Dong et al. “Inferring user demographics and social strategies in mobile social networks”. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2015), pp. 15–24.
- [HA15] A.K.I. Hassan and A. Abraham. “Modeling insurance fraud detection using imbalanced data classification”. In: *Proceedings of the 7th World Congress on Nature and Biologically Inspired Computing (NaBIC2015)*. Pietermaritzburg, South Africa. 2015, pp. 117–127.
- [Leo15] Paul M Leonardi. “Ambient awareness and knowledge acquisition: Using social media to learn “who knows what” and “who knows whom””. In: *MIS Quarterly* 39.4 (2015), pp. 747–762.
- [RU15] T. Rasmussen and D. Ulrich. “Learning from practice: how HR analytics avoids being a management fad”. In: *Organizational Dynamics* 44.3 (2015), pp. 236–242. DOI: [10.1016/j.orgdyn.2015.05.008](https://doi.org/10.1016/j.orgdyn.2015.05.008).
- [UD15] D. Ulrich and J. Dulebohn. “Are we there yet? What’s next for HR?”. In: *Hum. Resour. Manag. Rev.* 25 (2015), pp. 188–204. DOI: [10.1016/j.hrmr.2015.01.004](https://doi.org/10.1016/j.hrmr.2015.01.004).
- [Aji16] Pankaj Ajit. “Prediction of employee turnover in organizations using machine learning algorithms”. In: *algorithms* 4.5 (2016), p. C5.
- [Ang+16] D. Angrave et al. “HR and Analytics: Why HR Is Set to Fail the Big Data Challenge”. In: *Hum. Resour. Manag. J.* 26 (2016), pp. 1–11. DOI: [10.1111/1748-8583.12090](https://doi.org/10.1111/1748-8583.12090).
- [CCE16] Rob Cross, Jonathon N Cummings, and Amy C Edmondson. “Tie and network correlates of individual performance in knowledge-intensive work”. In: *Academy of Management Journal* 59.5 (2016), pp. 1705–1727.
- [DMR16] M. Dickinson, M. Magnani, and L. Rossi. *Multilayer social networks*. New York: Cambridge University Press, 2016.
- [FP16] Michael Fire and Rami Puzis. “Organization mining using online social networks”. In: *Networks and Spatial Economics* 16.1 (2016), pp. 55–81.

---

## BIBLIOGRAPHY

- [IH16] Jean Paul Isson and Jesse S Harriott. *People analytics in the era of big data: Changing the way you attract, acquire, develop, and retain talent*. John Wiley & Sons, 2016.
- [KWL16] Julian Krumeich, Dirk Werth, and Peter Loos. “Prescriptive control of business processes: new potentials through predictive analytics of big data in the process manufacturing industry”. In: *Business & Information Systems Engineering* 58 (2016), pp. 261–280.
- [MPK16] R. Michalski, S. Palus, and P. Kazienko. “Matching Organizational Structure and Social Network Extracted from Email Communication”. In: *International Journal of Knowledge Society Research* 7.2 (2016), pp. 13–28.
- [NO16] Aaron K. Nielsen and P. Orton. *Leading with Strategic Thinking: Four Ways Effective Leaders Gain Insight, Drive Change, and Get Results*. John Wiley & Sons, 2016.
- [SDY16] Mishra Sujeet, Lama Dev, and Pal Yogesh. “Human resources predictive analytics (HRPA) For HR management in organizations”. In: *Int J Sci Technol Res* 5.5 (2016), pp. 2277–8616.
- [And17] Morten Kamp Andersen. “Human capital analytics: the winding road”. In: *Journal of Organizational Effectiveness: People and Performance* 4.2 (2017), pp. 133–136. DOI: [10.1108/JOEPP-03-2017-0024](https://doi.org/10.1108/JOEPP-03-2017-0024).
- [Dav17] Thomas H. Davenport. *People analytics reveals three things HR may be getting wrong*. <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/people-analytics-reveals-three-things-hr-may-be-getting-wrong>. 2017.
- [Del17] Deloitte. *Rewriting the Rules for the Digital Age: 2017 Deloitte Global Human Capital Trends*. New York, NY: Deloitte, 2017.
- [DK17] Sepideh Dolatabadi and Farshid Keynia. “Designing of customer and employee churn prediction model based on data mining method and neural predictor”. In: *Proceedings of the xth International Conference*. IEEE, 2017, pp. 74–77.
- [Fin17] David R. Fineman. “People analytics: Recalculating the route”. In: *Deloitte Review* (21 2017), pp. 102–117.

## BIBLIOGRAPHY

---

- [Glo+17] Peter A. Gloor et al. “Forecasting managerial turnover through e-mail based social network analysis”. In: *Computers in Human Behavior* 71 (2017), pp. 343–352. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2017.02.017>. URL: <https://www.sciencedirect.com/science/article/pii/S0747563217300900>.
- [Hai+17] G. Haixiang et al. “Learning from class-imbalanced data: Review of methods and applications”. In: *Expert Syst. Appl.* 73 (2017), pp. 220–239.
- [Hom+17] PW Hom et al. “One hundred years of employee turnover theory and research”. In: *J Appl Psychol* (2017). DOI: [10.1037/apl0000103](https://doi.org/10.1037/apl0000103).
- [KT17] S. Khan and J. Tang. “The paradox of human resource analytics: being mindful of employees”. In: *Journal of General Management* 42.2 (2017), pp. 57–66.
- [LF17] Alec Levenson and Alexis Fink. “Human capital analytics: too much data and analysis, not enough models and business insights”. In: *Journal of Organizational Effectiveness: People and Performance* 4.2 (2017), pp. 145–156.
- [MB17] J.H. Marler and J.W. Boudreau. “An Evidence-Based Review of HR Analytics”. In: *Int. J. Hum. Resour. Manag.* 28 (2017), pp. 3–26. DOI: [10.1080/09585192.2016.1244699](https://doi.org/10.1080/09585192.2016.1244699).
- [Riv17] Lauren A Rivera. “When two bodies are (not) a problem: Gender and relationship status discrimination in academic hiring”. In: *American Sociological Review* 82.6 (2017), pp. 1111–1138.
- [Siv+17] Uthayasankar Sivarajah et al. “Critical analysis of Big Data challenges and analytical methods”. In: *Journal of business research* 70 (2017), pp. 263–286.
- [VB17] Sjoerd Van den Heuvel and Tatiana Bondarouk. “The rise (and fall?) of HR analytics: A study into the future application, value, structure, and system support”. English. In: *Journal of organizational effectiveness* 4.2 (2017), pp. 127–148. ISSN: 2051-6614. DOI: [10.1108/JOEPP-03-2017-0022](https://doi.org/10.1108/JOEPP-03-2017-0022).
- [ZBB17] B. Zhu, B. Baesens, and S.K.L.M. Broucke. “An empirical comparison of techniques for the class imbalance problem in churn prediction”. In: *Inf. Sci.* 408 (2017), pp. 84–99.

---

## BIBLIOGRAPHY

- [BT18] H. Buttner and W. Tullar. “A representative organizational diversity metric: a dashboard measure for executive action”. In: *Equality, Diversity and Inclusion: An International Journal* 37.3 (2018), pp. 219–232. DOI: [10.1108/GM-12-2013-0140](https://doi.org/10.1108/GM-12-2013-0140).
- [CIP18] CIPD. *People data: driving performance*. [Accessed: April 22, 2023]. 2018. URL: <https://www.cipd.co.uk/knowledge/strategy/analytics/people-data-driving-performance#gref>.
- [Com18] European Commission. “General Data Protection Regulation (GDPR)”. In: (2018). URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
- [Fry+18] Alex Frye et al. “Employee Attrition: What Makes an Employee Quit?” In: *SMU Data Science Review* 1.1 (2018), p. 9.
- [HBB18] M. A. Huselid, B. Becker, and R. Beatty. *The workforce scorecard: Managing human capital to execute strategy*. Harvard Business Press, 2018.
- [HL18] Mazhar Hussain and Yu Lei. “Big data analytics in human resource management: a review”. In: *Journal of Organizational Computing and Electronic Commerce* 28.1 (2018), pp. 62–83.
- [KD18] S. N. Khera and Divya. “Predictive modelling of employee turnover in Indian IT industry using machine learning techniques”. In: *Vision* 23 (1 2018), pp. 12–21. DOI: [10.1177/0972262918821221](https://doi.org/10.1177/0972262918821221).
- [Kry+18] D. Kryscynski et al. “Analytical abilities and the performance of HR professionals”. In: *Human Resource Management* 57.3 (2018), pp. 715–738. DOI: [10.1002/hrm.21854](https://doi.org/10.1002/hrm.21854).
- [Lee+18] Thomas W Lee et al. “Managing employee retention and turnover with 21st century ideas”. In: *Organizational dynamics* 47.2 (2018), pp. 88–98.
- [Liu+18] J. Liu et al. “Analyzing employee turnover based on job skills”. In: *Proceedings of the International Conference on Data Processing and Applications*. 2018. DOI: [10.1145/3224207.3224209](https://doi.org/10.1145/3224207.3224209).
- [MLL18] D. McIver, M.L. Lengnick-Hall, and C.A. Lengnick-Hall. “A strategic approach to workforce analytics: integrating science and agility”. In: *Business Horizons* 61.3 (2018), pp. 397–407. DOI: [10.1016/j.bushor.2018.01.005](https://doi.org/10.1016/j.bushor.2018.01.005).

## BIBLIOGRAPHY

---

- [Min18] Dana B Minbaeva. “Building credible human capital analytics for organizational competitive advantage”. In: *Human Resource Management* 57.3 (2018), pp. 701–713.
- [SP18] B. Sivathanu and R. Pillai. “Smart HR 4.0—Hindustry 4.0 is disrupting HR”. In: *Hum. Resour. Manag. Int. Dig.* 26 (2018), pp. 1–4. DOI: [10.1108/HRMID-04-2018-0059](https://doi.org/10.1108/HRMID-04-2018-0059).
- [YJS18] S. Yadav, A. Jain, and D. Singh. “Early prediction of employee attrition using data mining techniques”. In: *2018 IEEE 8th International Advance Computing Conference (IACC)*. 2018. DOI: [10.1109/IADCC.2018.8692137](https://doi.org/10.1109/IADCC.2018.8692137).
- [Zha+18] Y. Zhao et al. “Employee turnover prediction with machine learning: A reliable approach”. In: *Proceedings of SAI Intelligent Systems Conference*. 2018. DOI: [10.1007/978-3-030-01057-7\\_56](https://doi.org/10.1007/978-3-030-01057-7_56).
- [Bre19] J. Brennecke. “Dissonant Ties in Intraorganizational Networks: Why Individuals Seek Problem-Solving Assistance from Difficult Colleagues”. In: *Academy of Management Journal* 63.3 (2019), pp. 743–778.
- [Del19] Deloitte. *People Analytics Solutions Market Primer*. [Accessed: April 22, 2023]. 2019. URL: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/human-capital/us-human-capital-bersin-lt-people-analytics-solutions-market-primer.pdf>.
- [Fri+19] Sorelle A. Friedler et al. “A comparative study of fairness-enhancing interventions in machine learning”. In: *Annual Review of Computer Science* 2 (2019).
- [KA19] M. Kaya and R. Alhajj. *Influence and Behavior Analysis in Social Networks and Social Media*. Cham: Springer, 2019.
- [Leo19] Helmut Leopold. “Social media and corporate innovation management: Eight rules to form an innovative organisation”. In: *Elektrotechnik Informationstechnik* 136.3 (2019), pp. 241–253. DOI: [10.1007/s00502-019-0729-5](https://doi.org/10.1007/s00502-019-0729-5).
- [MS19] D.Ø. Madsen and K. Slaten. “An Examination of the Current Status and Popularity of HR Analytics”. In: *Int. J. Strateg. Manag. Econ.* 19 (2019), pp. 17–38.

---

## BIBLIOGRAPHY

- [Men19] Y. Mengbin. *Opinion Dynamics and the Evolution of Social Power in Social Networks*. Cham: Springer, 2019.
- [Trö+19] Christian Tröster et al. “The coevolution of social networks and thoughts of quitting”. In: *Academy of Management Journal* 62.1 (2019), pp. 22–43.
- [Zub19] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.
- [Anh+20] N. T. Anh et al. “Integrating employee value model with churn prediction”. In: *International Journal of Sensors Wireless Communications and Control* 10 (4 2020), pp. 484–493. DOI: [10.2174/2210327910666200213123728](https://doi.org/10.2174/2210327910666200213123728).
- [ACG20] Olabode Adekunle Ayodele, Alice Chang-Richards, and Vicente González. “Factors affecting workforce turnover in the construction sector: A systematic review”. In: *Journal of construction engineering and management* 146.2 (2020), p. 03119010.
- [Fal+20] F. Fallucchi et al. “Predicting employee attrition using machine learning techniques”. In: *Computers* 9 (4 2020), p. 86. DOI: [10.3390/computers9040086](https://doi.org/10.3390/computers9040086).
- [FG20] V. Fernandez and E. Gallardo-Gallardo. “Tackling the HR digitalization challenge: key factors and barriers to HR analytics adoption”. In: *Competitiveness Review: An International Business Journal* (2020). DOI: [10.1108/CR-12-2019-0163](https://doi.org/10.1108/CR-12-2019-0163).
- [Ins20] Work Institute. *2020 Retention Report*. <http://info.workinstitute.com/en/retention-report-2020>. Accessed: April 17, 2023. 2020.
- [JJP20] P. K. Jain, M. Jain, and R. Pamula. “Explaining and predicting employees’ attrition: a machine learning approach”. In: *SN Applied Sciences* 2 (4 2020), pp. 1–11. DOI: [10.1007/s42452-020-2519-4](https://doi.org/10.1007/s42452-020-2519-4).
- [JC20] D. Jeske and T. Calvard. “Big data: lessons for employers and employees”. In: *Employee Relations* 42.1 (2020), pp. 248–261. DOI: [10.1108/ER-06-2018-0159](https://doi.org/10.1108/ER-06-2018-0159).
- [Jin+20] Ziwei Jin et al. “RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis”. In: *Web Information Systems Engineering – WISE 2020*. Ed. by Zhisheng Huang et al. Cham: Springer International Publishing, 2020, pp. 503–515.
- [Kak+20] Shital Kakad et al. “Employee Attrition Prediction System”. In: *Int. J. Innov. Sci., Eng. Technol* 7.9 (2020), p. 7.

## BIBLIOGRAPHY

---

- [Lep+20] Katerina Lepenioti et al. “Prescriptive analytics: Literature review and research challenges”. In: *International Journal of Information Management* 50 (2020), pp. 57–70.
- [Mai20] R. D. Maioreescu-Murphy. *Corporate Diversity Communication Strategy: An Insight into American MNCs’ Online Communities and Social Media Engagement*. Cham: Palgrave Macmillan, 2020.
- [Nak20] M. Nakonečný. *Online communication within a company: Case study of small company*. Prague: Stanislav Juhňák-Triton, 2020.
- [NM20] Marcin Nurek and Radoslaw Michalski. “Combining machine learning and social network analysis to reveal the organizational structures”. In: *Applied Sciences* 10.5 (2020), p. 1699.
- [ONe+20] Cathy O’Neil et al. “The Ethical Algorithm: The Science of Socially Aware Algorithm Design”. In: (2020).
- [RV20] G. Ragozini and M. P. Vitale. *Challenges in Social Network Research: Methods and Applications*. Cham: Springer, 2020.
- [ZI20] S. Zeidan and N. Itani. “HR Analytics and Organizational Effectiveness”. In: *Int. J. Emerg. Technol.* 11 (2020), pp. 683–688.
- [Cha+21] S. Chatterjee et al. “Examining the dark side of human resource analytics: an empirical investigation using the privacy calculus approach”. In: *International Journal of Manpower* (2021). DOI: [10.1108/IJM-02-2021-0087](https://doi.org/10.1108/IJM-02-2021-0087).
- [LE21] A.S. Larsson and M.R. Edwards. “Insider econometrics meets people analytics and strategic human resource management”. In: *International Journal of Human Resource Management* (2021), pp. 1–47. DOI: [10.1080/09585192.2020.1847166](https://doi.org/10.1080/09585192.2020.1847166).
- [LL21] Scott Lundberg and Su-In Lee. *SHAP: SHapley Additive exPlanations*. <https://shap.readthedocs.io/en/latest/index.html>. Accessed: April 19, 2023. 2021.
- [Rig21] Elia Rigamonti. “HR Analytics State of the Art and Future Directions: A Scoping Review Based on Natural Language Processing Technique”. In: *Sustainability* 13.9 (2021), p. 5255. DOI: <https://doi.org/10.5093/jwop2022a18>.

- [She+21] S.V. Shet et al. “Examining the determinants of successful adoption of data analytics in human resource management – a framework for implications”. In: *Journal of Business Research* 131 (2021), pp. 311–326. DOI: [10.1016/j.jbusres.2021.03.054](https://doi.org/10.1016/j.jbusres.2021.03.054).
- [Tur+21] A. Tursunbayeva et al. “The ethics of people analytics: risks, opportunities and recommendations”. In: *Personnel Review* (2021). DOI: [10.1108/PR-12-2019-0680](https://doi.org/10.1108/PR-12-2019-0680).
- [YHC21] Nesrine Ben Yahia, Jihen Hlel, and Ricardo Colomo-Palacios. “From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction”. In: *IEEE Access* 9 (2021), pp. 60447–60458. DOI: [10.1109/ACCESS.2021.3074559](https://doi.org/10.1109/ACCESS.2021.3074559).
- [Álv+22] F.J. Álvarez-Gutiérrez et al. “Human Resources Analytics: A Systematic Review from a Sustainable Management Approach”. In: *Rev. Psicol. Trab. Las Organ.* 38 (2022), pp. 129–147.
- [Bil+22] S. Bilan et al. “Systematic Bibliometric Review of Artificial Intelligence Technology in Organizational Management, Development, Change and Culture”. In: *Bus. Theory Pract.* 23 (2022), pp. 1–13.
- [GM22] Stacia Garra and Priyanka Mehrotra. “People Analytics Technology 2022: Executive Summary”. In: (2022). URL: <https://redthreadresearch.com/pat-2022/>.
- [Mar22] A. Margherita. “Human Resources Analytics: A Systematization of Research Topics and Directions for Future Research”. In: *Hum. Resour. Manag. Rev.* 32 (2022), p. 100795.
- [MF22a] S. McCartney and N. Fu. “Promise versus Reality: A Systematic Review of the Ongoing Debates in People Analytics”. In: *J. Organ. Eff. People Perform.* 9 (2022), pp. 281–311.
- [MF22b] Steven McCartney and Na Fu. “Promise versus reality: a systematic review of the ongoing debates in people analytics”. In: *Journal of Organizational Effectiveness: People and Performance* 9.2 (2022), pp. 281–311.
- [SL22] N. Suri and P. Lakhapal. “People Analytics Enabling HR Strategic Partnership: A Review”. In: *South Asian J. Hum. Resour. Manag.* 1 (2022), p. 35.

## BIBLIOGRAPHY

---

- [Bon23] Eithel F. Bonilla-Chaves. “Exploring the Evolution of Human Resource Analytics: A Bibliometric Study”. In: (2023).
- [23a] *Panalyt - People Analytics and Workforce Planning*. <https://www.panalyt.com/>. 2023.
- [23b] *Polinode - Network Analysis Made Easy*. <https://www.polinode.com/>. 2023.
- [23c] *SAPSuccessFactors - Human Experience Management (HXM) Suite*. <https://www.sap.com/products/human-experience-management.html>. 2023.
- [23d] *Swoop Analytics - Social Network Analysis and Collaboration Analytics*. <https://www.swoopanalytics.com/>. 2023.
- [23e] *TrustSphere - People Analytics and Relationship Analytics*. <https://trustsphere.com/>. 2023.
- [23f] *Visier People - People Analytics Software*. <https://www.visier.com/people/>. 2023.
- [23g] “Wharton People Analytics Conference 2023”. In: Wharton School of the University of Pennsylvania. Philadelphia, PA, 2023. URL: <https://wpa.wharton.upenn.edu/2023-conference-competitions>.
- [23h] *Workday - Human Capital Management and Financial Management Software*. <https://www.workday.com/en-us/homepage.html>. 2023.
- [YAC23] Sundus Younis, Ali Ahsan, and Fiona M Chatteur. “An employee retention model using organizational network analysis for voluntary turnover”. In: *Social Network Analysis and Mining* 13.1 (2023), p. 28.
- [] *D3.js*. <https://d3-graph-gallery.com/network.html>.
- [LC] Paul Leonardi and Noshir Contractor. *Better People Analytics*. <https://hbr.org/2018/11/better-people-analytics>. Last accessed 12 June 2023.
- [Pan] Panalyt. *Panasonic’s ONA Journey*. <https://panalyt.com/case/panasonic-ona/>. Accessed: April 17, 2023.
- [Vis] Visier. *Data ethics: 6 steps for ethically sound people analytics*. <https://www.visier.com/clarity/six-steps-ethically-sound-people-analytics>.

## BIBLIOGRAPHY

---

- [Visnd] Visier. *What is People Analytics and How Do I Get Started?* n.d. URL: <https://www.visier.com/blog/what-is-people-analytics-and-how-do-i-get-started/>.