



HAL
open science

Modeling and analysis of organizational network analysis graphs based on employee data

Abdel-Rahmen Korichi, Hamamache Kheddouci, Taha Tehseen

► To cite this version:

Abdel-Rahmen Korichi, Hamamache Kheddouci, Taha Tehseen. Modeling and analysis of organizational network analysis graphs based on employee data. INTERNATIONAL CONFERENCE ON OPTIMIZATION AND LEARNING (OLA2023), May 2023, Malaga, Spain. 10.1007/978-3-031-34020-8_27. hal-04131575

HAL Id: hal-04131575

<https://hal.science/hal-04131575v1>

Submitted on 16 Jun 2023

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.

Modeling and analysis of organizational network analysis graphs based on employee data

Abdel-Rahmen Korichi^{1,2}, Hamamache Kheddouci¹, and Taha Tehseen²

¹ Université de Lyon - LIRIS UMR 520, Bâtiment Nautibus 43, bd du 11 novembre 1918, 69622 Villeurbanne - France abdelrahmen.korichi@gmail.com

<https://www.univ-lyon1.fr/>

² Panalyt Pte. Ltd. - Singapore

<https://www.panalyt.com>

Abstract. Business leaders have access to a new set of tools to monitor and improve employees' productivity and wellness as well as to retain talents: People or Human Resource Analytics. However, while People Analytics helps understanding employees at the individual level, it doesn't reflect the social interactions happening in the workplace. For that, you would rather look at Organizational Network Analysis to examine how people communicate across an organization and give a more accurate and realistic understanding of employees' interactions. In this paper, we propose a framework for organizations to supplement their People Analytics with Organizational Network Analysis. We address the different challenges associated to that and provide different case-studies based on real clients' data to demonstrate how Organizational Network Analysis can be used to solve practical business applications.

Keywords: Relational Analytics · Organizational Network Analytics · People Analytics · 4th Industrial Revolution.

1 Introduction

The Human Resource (HR) industry is experiencing a revolution led by the profusion of employee data and new possibilities to take data-driven decisions and discover new insights about how organizations function [1]. Business leaders are looking for ways to back up their hypotheses, understand better their employees, manage them better, and improve internal processes. They want to have real-time access to their company's turnover - where it happens and why? They want to know their gender pay gap, see how diverse their organization is, retain their talent, and more. This is called People Analytics (PA) or Human Resource Analytics: the collection and application of talent data to improve critical talent and business outcomes [2]. As a matter of fact, more than 70% of companies now say that they consider PA to be a high priority [3]. PA focuses mostly on employee attribute data, of which there are two kinds:

- Static attributes (or traits): facts about individuals that don't change, such as ethnicity, gender, hire date, etc.
- Dynamic attributes (or state): facts about individuals that do change, such as age, company tenure, compensation, etc.

PA is great to check that numbers are in order, benchmark data across different teams, and set Key Performance Indicators (KPIs) for managers. A limitation of PA, as Paul Leonardi and Noshir Contractor argue [4], is that it limits itself to individuals attributes and neglect 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. Organizational Network Analytics (ONA), also called Relational Analytics, 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.

In this paper, we present a framework for not only exploring organizations' key information based on employees' static attributes as we see in standard HR analytics systems, but we will also present how to factor in communication metadata and unlock more insights. At Panalyt [5], we created a Software as a Service (SaaS) tool where we plug in our clients' data and allow them to perform PA as well as ONA using our suite of pre-built dashboards 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 PA and ONA can help to deduce transformative insights by performing queries on a partial subgraph. Finally, we will share advanced real case-studies from our client base.

2 Challenges

2.1 Ethical concerns

The accessibility of employee communication data raises important ethical concerns. Beyond legal considerations, employers should be mindful of the ethical standards that they adhere to while utilizing this information. One study estimated that 81% of people analytics projects are jeopardized by ethics and privacy concerns [7]. At Panalyt, we purposefully refuse to ingest and analyse the content of the communication between employees, due to the sensitivity of the information. According to a recent study, transparency was identified as being one of the most critical considerations for PA projects [8]. The use of aggregated, non-identifiable data is recommended where possible, to demonstrate to employees that the purpose behind PA projects is to capture larger organizational trends.

Organizations should communicate their reasons for pursuing PA projects and the kind of benefits employees should expect from them. We also recommend establishing clear governance around data collection, access and storage, consent, and anonymity.

2.2 Data and storage

To perform our analysis, we need 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. In many cases, organizations use Software as a service (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 they can be retrieved, and used to perform analysis. At Panalyt, after being collected, the data is stored, transformed and queried using SQL, 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. If a record is updated for an employee during a month, we forward propagate the change to the future versionized rows, as we can see in Table 1. We then make query on a partial subgraph to visualize a defined population of the network, where nodes are employees (or their external contact) and where edges represent an information about their interactions. We chose to aggregate the data by month as we realize this granularity is large enough deducing trends, and small enough to notice significant changes quickly in the scale that companies operate in.

Table 1: Employee table

<i>empID</i>	<i>versionId</i>	<i>companyemail</i>	<i>department</i>	<i>salary</i>	<i>JobTitle</i>	...
0001	30/09/2022	emp1@email.com	Technology	\$6,000	Data Engineer	
0001	31/10/2022	emp1@email.com	Technology	\$6,000	Data Engineer	
0001	30/11/2022	emp1@email.com	Technology	\$6,000	Data Engineer	
0001	31/12/2022	emp1@email.com	Technology	\$7,000	Sr Data Engineer	
0002	31/08/2022	emp2@email.com	Marketing	\$4000	UI Designer	
0002	30/09/2022	emp2@email.com	Marketing	\$4,5000	UI Designer	
...						

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
- *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.
- 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 to 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 send 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 [9], and others suggests to use the geometric mean of sent-received counts, [10], 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 use simple metrics such as those described above as well as more complex ones that we can't disclose as they are confidential. We have also created a score 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. 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.

For the purpose of this study, we use a metric called *collaboration score* which is defined as follows: For each version (month), we count the number of groups/channels where both employees are active. We define being active by sending at least 5 message in the same group/channel and the same month. For instance, if Mizuki and Hong have both sent more than 5 messages in the channels "Client A" and "Engineers", they would have a collaboration score equal to 2.

2.3 Dashboards

In order to efficiently use network data to understand their teams, managers need access to interactive and easy to use dashboards. Most companies rely on data analysts to get insights related to talent and performance management using BI tools such as Tableau or Power BI. That often creates a bottleneck for multiple reasons:

1. It requires a dedicated data team and decision makers to help them put the data into context, which is very costly.
2. There are not enough data analysts to address all management queries in a timely manner.

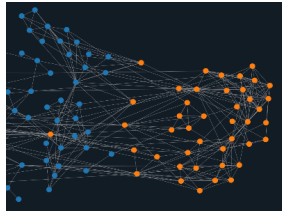
- Employee data is sensitive, and business leaders often hesitate to share this information with regular employees. The BI tools also just do not have a permission system granular enough to give appropriate permissions to users.

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 4 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 dashboards where data from different sources can interact with each other.

We also have a permissions system that allows an admin to give access a user to specific data, aggregated or not, based on the line hierarchy of the user or on defined attributes (e.g. a manager can be limited to see aggregated or individual’s values only for the team he manages, or for a specific department/location/etc.)

In our application we use d3.js [6] 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.



(a) Nodes from two different teams coloured based on their teams.



(b) Nodes from the same team connected by an attribute



(c) Nodes layered with different properties of the data.

Here are some of the salient features of the ONA dashboard:

- Colouration: By default, nodes are coloured based on the department of the employee they belong to. In Figure 1a, the nodes belong to different departments - Sales and Marketing. The nodes in Figure 1b belong to Engineering.
- Layering: Each node can also be recoloured based on user-selected properties. For instance, in Figure 1c, the nodes are not coloured based on departments - yellow is applied to all employees that are working part-time and red to all employees with less than one year of tenure. There is one employee that is both part time and has less than one year of tenure. On the other hand, there are three employees that have neither of those properties (in blue).
- A-B Filtering: The graph can itself be filtered based on employee data attributes. For instance, in Figure 1a we have filtered the data such that both the senders (A) and receivers (B) are from two departments only.

2.4 Datasets

For the upcoming analyses, we will be using 3 datasets from 3 different clients:

- Client 1: the Slack communication data of 486 internal users of a Japanese company from January 2019 to March 2022. This data includes their public and private channels, but does not include any direct messages. We do not include direct messages as their company’s Slack subscription does not allow access to direct messages for extraction due to privacy reasons.
- Client 2: the Slack communication data of 845 internal users of an Australian company from August 2022 to December 2022. Again, the data does not include any direct messages for the same reasons.
- Client 3: we will share a case study with Panasonic Operational Excellence Co., Ltd. that partnered with Panalyt in March 2022 and applied ONA to identify organizational issues. The size of the dataset is confidential.

3 Basic Analysis

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.

3.1 Inter-Team collaboration

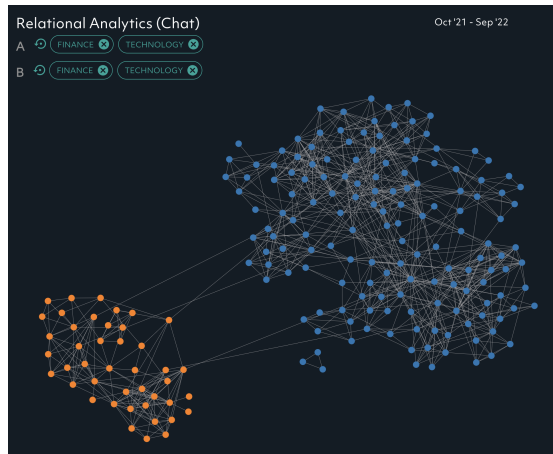


Fig. 2: Collaboration between two teams

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.

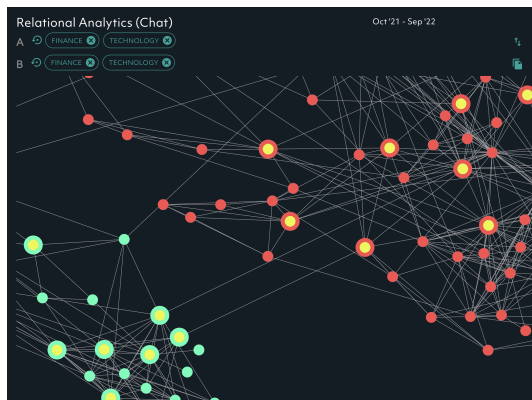


Fig. 3: Inter-Team Collaboration with Manager layer

In Figure 2 (client 2), 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 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.

3.2 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 Client 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 4. 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.

3.3 Gender diversity in teams

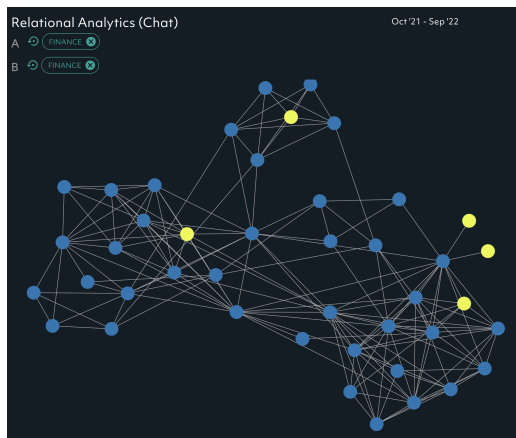


Fig. 4: New Hires Communications

communication gaps between the different groups. We see that companies already spend time and effort on streamlining their hiring 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.

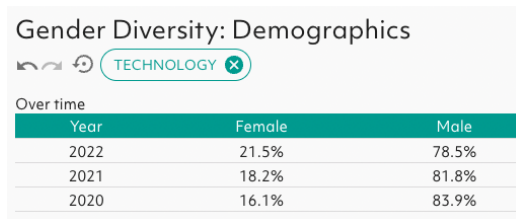


Fig. 5: Gender Diversity chart

more than 13% per year to reach 21.5% of women from January 2020 to November 2022 (Figure 5). However, when we look at the communication graph (Figure 6) 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

Companies strive for an inclusive work culture where employees from various genders and demographics are not only represented but also included readily in conversations. Foma [11] details the benefits of workplace diversity such as plurality of ideas owing to diverse experiences and the likelier retention of employees. But they also note that there can be significant

Looking at client 2, 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

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.

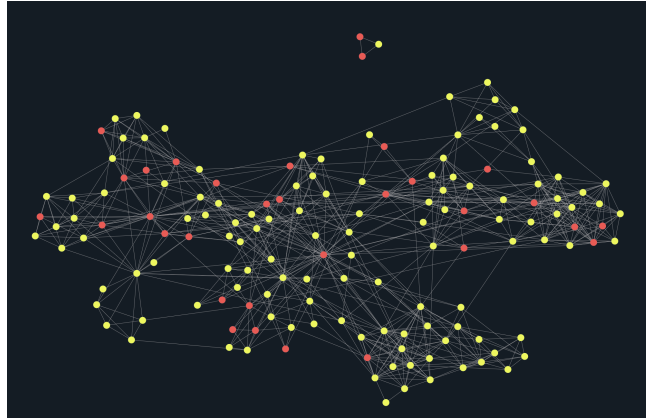


Fig. 6: Network Graph layered by gender

4 Advanced case studies

4.1 Attrition vs amplitude of communication

For client 1, to understand how to help them promote a healthier company culture and retain talents, we studied their historical communication data to understand the impact of the amplitude of communication in attrition. We started by looking at the count of messages sent by active employees by month (Figure 7a). We then averaged the results for all active employees by month. The average time of the first message sent was roughly between 10am and 11am, while the average time of the last message sent was between 6pm and 7pm. After that, we focused our attention on the employees who quit (37 terminated employees with at least 12 months tenure) to see if we could find any pattern in their amplitude of work in the last 15 months before they left. We uncovered from Figure 7b and Figure 7c that on average before employees quit, they start communicating earlier, and send their last message later, until they reach a point around 5 months before termination where the trend starts to reverse. In Figure 7d, we looked at the total amplitude of communication (average time of last message sent - average time of first message sent). As expected, we saw a similar pattern: the amplitude of communication keeps increasing until it reaches a pic and start decreasing again. Finally, in Figure 7e, we wanted to make sure that this trend was observable in comparison to their peers. We benchmarked the average first/last time of messages sent by terminated employees, as well as the amplitude, against the averages for the active population (380+ active employees) for that month for each month leading up to their eventual termination date.

We found again a pattern of employees that start communicating earlier and finish later compared to the average employee, especially between 5 to 7 months

before their termination where it peaks to 40 minutes more on average compared to the active employees.



(a) Average time of the day for the first and last message sent by each employee across the entire population for each month.

(b) Average time of the day for the first message sent by each terminated employee for each month leading up to their eventual termination date.

(c) Average time of the day for the last message sent by each terminated employee for each month leading up to their eventual termination date.

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

(e) Average time of the day for first and last message as well as 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.

Fig. 7: Analysis of communication times

Our conclusions, based on those findings, was that the amplitude of communication is positively correlated to the overwork of employees and factors greatly in the attrition. Interestingly, employees gradually disengage from 5 to 6 months before they quit. This study is a good example that shows that it is possible to uncover insightful information based on aggregated, non-identifiable data.

Companies can then use it to create new policies and limit overwork by discouraging communication after a certain hour when possible, and to identify groups, departments or talents where employees tend to overwork and disincentivize it.

4.2 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 [12]. Following a merger and a group restructuring, their 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.

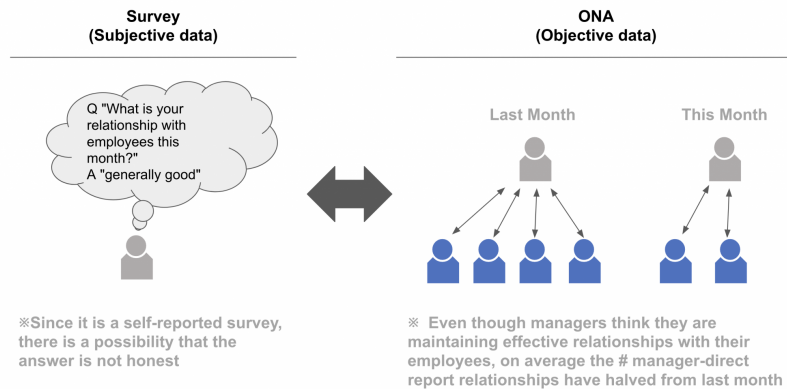


Fig. 8: Surveys vs ONA

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 expectations of their roles and responsibilities. However, they 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 8). This was especially true at the time of the imposed working from home period during the COVID-19 pandemic.

5 Discussions and conclusion

People analytics has introduced new ways to make evidence-based decisions to improve organizations' processes. But most companies have been focused on studying the attributes of their employees, forgetting the reality of interdependence. 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." [13]

In future work, we would like to develop ways to take communication metrics from different sources and formulate a combined score that explains the overall communication behaviour across multiple communication platforms.

References

1. DiClaudio, Michael. People analytics and the rise of HR: how data, analytics and emerging technology can transform human resources (HR) into a profit center. Strategic HR Review (2019).
2. HR Analytics, [https://www.gartner.com/en/human-resources/glossary/hr-analytics#:~:text=HR%20analytics%20\(also%20known%20as,and%20promote%20positive%20employee%20experience](https://www.gartner.com/en/human-resources/glossary/hr-analytics#:~:text=HR%20analytics%20(also%20known%20as,and%20promote%20positive%20employee%20experience). Last accessed 3 Dec 2022
3. People analytics: Recalculating the route, <https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2017/people-analytics-in-hr.html>. Last accessed 3 Dec 2022
4. Better People Analytics, <https://hbr.org/2018/11/better-people-analytics>. Last accessed 3 Dec 2022
5. Panalyt, <https://panalyt.com/>. Last accessed 4 Dec 2022
6. D3.js, <https://d3-graph-gallery.com/network.html>. Last accessed 4 Dec 2022
7. Data ethics: 6 steps for ethically sound people analytics, Visier, <https://www.visier.com/clarity/six-steps-ethically-sound-people-analytics>.
8. Tursunbayeva, A., Pagliari, C., Di Lauro, S. and Antonelli, G., 2021. The ethics of people analytics: risks, opportunities and recommendations. Personnel Review.
9. R. Michalski, S. Palus, P. Kazienko, Matching Organizational Structure and Social Network Extracted from Email Communication
10. Munmun De Choudhury, Winter A. Mason, Jake M. Hofman, Duncan J. Watts, Inferring Relevant Social Networks from Interpersonal Communication
11. Foma, Elizabeth. "Impact of workplace diversity." Review of Integrative Business and Economics Research 3.1 (2014): 382.
12. How Panasonic is Improving Manager Effectiveness and Innovation with Organizational Network Analysis <https://panalyt.com/case/panasonic-ona/#panalyt>.
13. Provan, K.G., Fish, A. and Sydow, J., 2007. Interorganizational networks at the network level: A review of the empirical literature on whole networks. Journal of management, 33(3), pp.479-516.