
Challenges in the Organizational Network Analysis: A Data Collection Process Perspective

Submitted 02/07/21, 1st revision 22/07/21, 2nd revision 14/08/21, accepted 10/09/21

Piotr Andrzejak¹, Łukasz Wawrzynek²

Abstract:

Purpose: Organizational network analysis is usually a long and complex process, consisting of many steps and generally using many analytical tools.

Design/Methodology/Approach: Numerous papers have been written about the analysis itself, especially about network measures, concepts, and visualizations used in research. This is unfortunately not the case for the data collection process.

Findings: The practice of conducting research shows that effective data collection is often a barrier to network analysis in organizations. The current situation related to COVID-19 has become an additional difficulty in assessing the occurrence of relationships between employees based on observations.

Practical Implications: An analysis of the literature and especially years of conducted studies in different organizations show that researchers still lack deeper insights into that kind of issue, such as identifying the main challenges and potential ways to improve the data collection process. This is a research gap we strive to address within this paper.

Originality/Value: In the addition to the effort made towards structuring the entire process of data collection in the organizational network analysis research contribution of this paper include our proposals of potential improvements that could be made to the data collection process in analyzing organizational networks.

Keywords: Social networks analysis, data collection.

JEL Classification: L10, L20, L22, M10.

Research Type: Research article.

Acknowledgment: The project is financed by the Ministry of Science and Higher Education in Poland under the program "Regional Initiative of Excellence" 2019 - 2022 project number 015/RID/2018/19 total funding amount 10 721 040,00 PLN.

¹Doctoral candidate, Wrocław University of Economics and Business, Faculty of Management, Department of Management Systems Design, Poland. Corresponding author, piotr.andrzejak@ue.wroc.pl

²Associate Prof., Wrocław University of Economics and Business, Faculty of Management, Department of Management Systems Design, Poland. Corresponding author (2nd) lukasz.wawrzynek@ue.wroc.pl

1. Introduction

Organizational Network Analysis (ONA), or managerial Social Network Analysis (SNA), has been a proven tool for people analytics for nearly 30 years, yet its adoption rate is still relatively tiny (Kutik, 2019). Nevertheless, due to the new ways of working, changing business requirements, and continuous developments in technology, the interest in SNA is proliferating. People analytics leaders indicated this analytics technique as one they most wanted to learn more about (Green, 2018). This supports recent research results that show that the adoption of diverse data sources and various people analytics solutions (such as SNA) allows organizations to conduct more sophisticated yet more efficient analyses (Visier, 2018). Nevertheless, successful organizational network analysis in each case requires gathering as complete and high-quality data as possible. This became an exceptionally demanding process in the age of COVID-19, where managers and researchers lost their natural way of concluding by conducting observations. They have the limited possibility of motivating employees to participate in the study by different channels than virtual ones.

There are at least three ways of collecting data in the described process. Despite the growing importance of the digital footprint in collecting data, most of this type of research is still based on questionnaires with questions about relationships (Kawa, 2013), while investigations using behavioral observations are endangered due to the outbreak coronavirus pandemic. This study combines literature review and case studies performed in several organizations from different industries. Results include our proposals of potential improvements that could be made to the data collection process in analyzing organizational networks.

2. Data Collection Process

Network data is more complex and requires more time to collect than other social science data (Borgatti, Brass, and Halgin, 2014). Quantitative data needed to build a network can be ordered using a questionnaire, behavioral observations, or what is becoming more and more popular, using the so-called digital footprint, i.e., a term referring to any information about a person left in digital form (Internet, intranet, database, communicator, etc.) (Borgatti *et al.*, 2009).

Based on the declarative definition of relationships with other people, the first of the mentioned possibilities is the traditional and most frequently used collection of data on social and organizational networks (Monge and Contractor, 2003). The benefits of using it certainly include the accuracy of the obtained mapping of relations significant from the point of view of the conducted research and, consequently, the interpretation of the received data with relatively high precision. In addition, the possibility of adjusting the questionnaires depending on the context of the conducted research is convenient for the researcher. The disadvantages of this approach include a large amount of work involved in the preparation and process of data collection, inadequate

response of the questionnaires, subjectivism, different interpretations of the same question by different people, or possible concealment or other manipulation of data respondents.

The advantage of the second possibility of collecting data, the one based on the observations of the actors' behavior, is the potential ability to automate the process as well as the greater objectivity of the researcher in comparison to the declarative answers made by the employees. On the other hand, incomplete or lacking knowledge of the researcher about the nature of interactions between individuals and limited possibilities of interpreting the resulting data are among the disadvantages of this approach (Wasserman and Faust, 1994).

The last of the methods mentioned above of collecting data, leading to a network image of the social structure, assumes concentration on data without engaging respondents (Kawa, 2013). For this purpose, information on online interactions (Zbieg *et al.*, 2012), data on the exchange of e-mail messages (Bird *et al.*, 2006), records of telephone calls (Eagle, Macy, and Claxton, 2010), or data on meetings entered in for a corporate calendar (Haythornthwaite, Wellman, and Mantei, 1995). The advantage of the approach is undoubtedly the excellent potential for automating data collection and processing. At the same time, the necessity to analyze the semantic of the correspondence remains a big challenge. The mere information about the frequency of exchanging e-mails does not answer the question about its nature.

It should be noted that attempts to integrate the methods mentioned above constitute an important area of network research (Quintane and Kleinbaum, 2011; Johnson, Kovács, and Vicsek, 2012). Regardless of the data collection method used in the study, the entire process ends once networks such as cooperation, communication, knowledge, or information flow are constructed (Wawrzynek, 2019).

3. Challenges Identification

Existing tools: The most available and popular network analysis tools include: Ucinet system with built-in NetDraw data visualization tool, The Pajek or NodeXL system in the PRO edition, allowing a slightly more comprehensive set of indicators from the analyzed networks, as well as Gephi. All those systems were thoroughly verified in the process of conducting network analysis. None of the mentioned systems has a tool for information collection that would support collecting information used to identify relationships and reflect networks in such a way that no further transformations leading to logical networks creation are necessary. Appropriate, clearly defined procedure of data collection tools selections are highly demanded. Alternatively, a system or platform that integrates all phases of ONA with the explicit inclusion of the data collection process would be of much value. It could contribute to the removal of some barriers that block the broader adoption of ONA in organizations.

Organization context: Conducting research based on network analysis requires preparation and adaptation to the analyzed organization. Once management has identified the issue, case boundaries, time, and spatial scope should be established. In the proposed ONA approach, the boundaries of the network are defined as part of the endogenous approach, for which the credibility of the network boundaries will depend on the substantive knowledge of the research problem. Defining network boundaries belongs to critical challenges in social network analysis (Heath, Fuller, and Johnston, 2009).

When using network generators to obtain base information about relationships, which is one of the methods of obtaining data for the network, the attributes of people related to the defined issue are determined based on substantive knowledge. The research is directed to people with specific attributes values. The use of the network generator based on the attributes indicated by the owner of the analyzed issue defines the original network boundary, showing the spatial scope and administrative areas analyzed in each case. The details refer both to the analyzed region of the organization, to individual processes, and to the entire organization, which allows building a network of different dimensions for separate functions. The identification of primary respondents does not, however, exhaust the problem of defining case boundaries. The use of a network generator, as opposed to indicating relations based on a selection from a list, means that the edges of the analyzed case may be changed due to indications of the original respondents.

Using the network generator, which works on the principle of free recall, the respondent can indicate a relationship with a person who does not belong to the original group of respondents, and even more - may not belong to the analyzed area or the analyzed organization at all. The person conducting the research, together with the person responsible for defining the issue, decides whether to include people or not indicated by the original respondents. Suppose the decision to add new respondents from the people shown because of using a network generator is favorable. In that case, the survey is also extended to the indicated people, thus changing the original boundaries but shaping the final borders of the case.

Data correctness and completeness: The issue of defining respondents is related to the network boundaries determination. As such, it results from the problem identification in the organization for ONA and the research preparation phase as well. What is essential, informal research, it often happens that lists of people existing in the organization and submitted for analysis, determined following the scope of the boundaries of the study, are incomplete and out of date, which in turn hits the issue of ensuring the correctness and completeness of the attribution of attributes (Huisman, 2009). Lack of proper attribute assignment may cause the displacement of actors between networks and, in exceptional circumstances, completely disturb the image of the network structure.

Different, low level, the technical issue is related to the recognition process of people selected by the respondents. There are two ways of handling answering questions about relations:

- a. Predefined list of employees to be chosen from,
- b. Free text typing.

In most cases, the first possibility is preferred. Nevertheless, sometimes researchers or managers are not sure what the boundaries are for the given analysis. Alternatively, they would like to determine them. In such scenarios, respondents cannot be constrained to the predefined list of potential answers. That answering questions about relations could lead to the possible mistakes or ambiguities of the names provided. Therefore, an additional mechanism to match answers with the employees within the organization might be needed.

Data confidentiality: Due to the specific nature of network analysis, the actors in the network must be both unequivocally identified and confidential. The respondent, when asked about the existence of any relationship type, must indicate a specific person. In most cases, people are identified by their first and last names, and sometimes - for unambiguous indications - additionally by their position, function, etc. Due to the large number of people remaining in relationships, the probability of repeating the primary data of specific people increases. Another way to identify data is to use a unique code associated with a person who is generally known within the organization. This ensures uniqueness but does not ensure anonymity.

However, there are two different issues here. One is related to the anonymity of indications of a specific person, and the other is related to the processing of personal data in general. If data confidentiality can be ensured in the information gathering system by giving respondents certain symbols, then protecting the processing of personal data should be considered separately. Nonetheless, the data collection process in the organizational network analysis has become a critical and challenging issue for European researchers after the General Data Protection Regulation (GDPR), which came into force on the 25th of May 2018 as a piece of European legislation regulation (Kotsios *et al.*, 2019).

Identified challenges: The data collection procedure is designed to ensure the triangulation of sources and respondents. Analysis of the case studies performed in several differently profiled organizations indicates that challenges faced by the researchers in collecting data relating to the respondents focus on the following issues:

- determine the respondents (according to the network boundary framework),
- ensuring the correct and complete assignment of respondent attributes,
- providing an appropriate response rate level,
- ensuring the completeness of the scope of collected information to build network models,

- guaranteeing the confidentiality of data.

Improvements proposal: To ensure the consistency of the collected data, to flexibly adjust the measures used in the SNA, to properly select network visualization algorithms to the problems constantly occurring in the organizations, and finally, to prepare managerial interventions aiming at improving widely understood company's efficiency Netwisor system was developed based on the cooperation of a research team associated with the Department of Management Systems Design at the University of Economics in Wrocław. This system is used for the entire ONA research. It allows building graphical representations of the network about research questions and issues submitted for the analysis of specific cases in a more readable way, also for people without experience in this type of research. Integrally, the Netwisor_WebSurvey system (see Figure 1) related to Netwisor was the basis for collecting data on employee relationships.

Figure 1. Netwisor_WebSurvey tool with question used to form cooperation network. Authors' own work.

A.1. My everyday work and its effects depend to the greatest extent on my cooperation with:

Importance:	First name:	Last name:	Remarks:
1 - high ▼	Adam	Nowak	test remark
▼			
▼			

[+ More rows](#)

[Next question](#)

Source: Own creation.

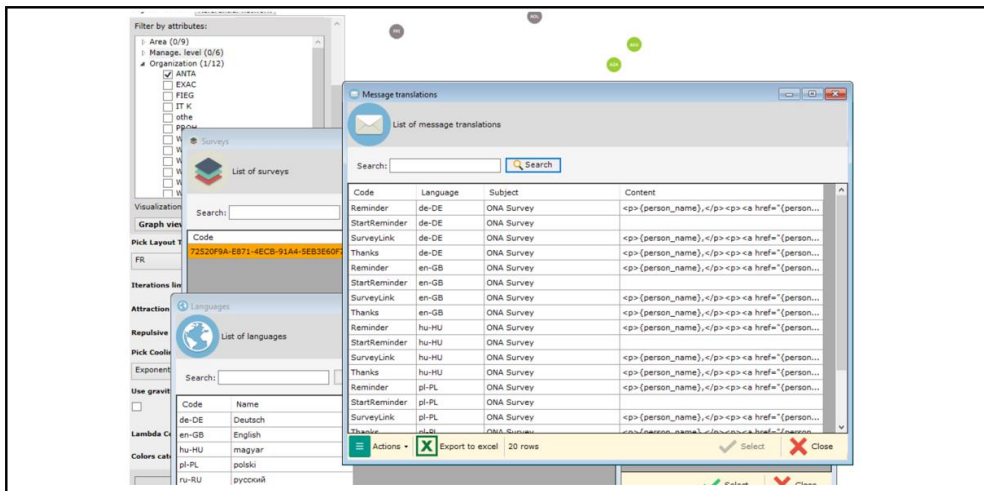
The Network system is proprietary, original software created by identified difficulties in collecting data in the ONA approach and problems in the quick interpretation of challenges and management problems based on understandable and unambiguous to interpret network visualizations. Further part of this section presents an improvements proposal for the identified challenges, with some already implemented mechanisms in the Network that eventually lead to the more straightforward and better-conducted data collection process for ONA.

Data collecting support mechanisms: The essential tool for collecting data, which is the basis for building logical networks, is a questionnaire examining relationships between the respondent (ego) and other participants in the organization or Network (alter). The survey questionnaire is directed to the predefined respondents (for example, via the Netwisor_WebSurvey system) based on the set of 15 essential questions (e.g., questions visible in Figure 1). On its basis, the respondents indicate people (their surnames and forenames or their unambiguous equivalents functioning in the organization in the form of a symbol resulting from the organization's findings

and remaining in everyday use) whom they have specific relationships. The questions were constructed based on literature analysis, especially about the research presented by Cross and Parker and the Cross, Ehrlich, Dawson, and Helferich team, which in turn are based on previous research by people such as Borgatti, Prusak, and Parker, and because of a large-stage process verification (Cross and Parker, 2004; Cross *et al.*, 2008).

To ensure an appropriate response rate level, it is necessary to conduct activities playfully, use the statistics of the current survey completion, return to specific, selected groups of respondents, including their immediate superior in the communication, indicate comparative values for other groups, ask for reasons and clarify doubts (Ferrin, Dirks, and Shah, 2006). To address those issues, Netwisor_WebSurvey has several automatic (active) or manual (passive) mechanisms for analyzing and sending messages to specific groups of respondents at different stages of the data collection process. Examples of tools are presented in Figure 2.

Figure 2. Netwisor_WebSurvey examples of data collection mechanisms. Authors' own work.



Source: Own creation.

Respondents and attribute assignment: As a possible remedy for this kind of issue, the authors suggest an active research preparation phase with the special care of the study's timeline. Reliability in the scope of the research conducted should consider activities aimed at obtaining a complete set of respondents, adequate to the discrepancies defined in the study. The network system allows full automation of this phase with the scripts prepared to load lists of employees with the assigned attributes such as years of experience, department and team memberships, age, etc.

In the Network system, which uses Netwisor_WebSurvey to collect data electronically, there is a mechanism for their encryption with the 3DES algorithm, and

the encrypted data is placed on a central "cloud" server. Ensuring data confidentiality should also be included in the first information addressed to respondents to dispel the doubts arising in connection with this type of research that may minimize the effectiveness of the data collection process. This stage of the ONA methodology is a specific method of collecting data on relationships based on lists of respondents, their attributes, levels of values, goals, measures of processes, and the functions themselves. The mere assignment of people involved in implementing strategies is included in the scope of respondents' attributes.

Language and communication: Another critical issue is to make sure that the language of the analyzed organization corresponds to the language used in the survey questions. If there is a discrepancy, adjust the language or, if possible, conduct meetings with staff to clarify the meaning of the language and linguistic constructs used in the survey. The language can be adjusted through meetings with a working group of respondents and agreeing with them on understanding the linguistic meanings used in the questions. Where possible, a workshop with all respondents to clarify the terms used in the survey is a way to ensure a higher level of understanding of the questions.

Figure 3. Example of the introduction message in the Netwisor_WebSurvey system. Authors' own work.

UE Uniwersytet Ekonomiczny
we Wrocławiu

Language: English

Respondent: **Piotr Andrzejak**
(piotr.andrzejak@ue.wroc.pl)

Welcome to ONA survey

Welcome to the ONA (Organizational Networks Analysis) study. The starting point is to build an image of your Team and market based on the individual opinions of TopTeam Participants

In reply, please indicate the name and surname of the person you consider to be the most important from your point of view in the context of the question asked. You can rank the importance of the indicated people by the numbers 1, 2, 3. To indicate the equal importance of the listed people, assign them the same importance. By default, there is space for up to three people, but if necessary, you can add additional lines with the "+" button. [Note - If necessary, further groups of questions relate to individual opinions on various topics.]

By pressing the "Start" button below you will go to the questions. It is important for us to know how you see dependencies on the TopTeam market.

Source: Uniwersytet Ekonomiczny we Wrocławiu.

To ensure the high quality of the research and thus the highest possible response rate, it is assumed that the study process begins with adequately prepared communication from the organization side. The messages about the research are sent by e-mail or other means of communication used in the organization. The proposed communication

should fulfill two primary tasks: firstly, to explain the reason and purpose of the conducted research, and secondly, to make the investigation credible in the eyes of respondents by sending a message about it from top management involved in activities related to the identified discrepancies accepted for analysis. An example of a letter introducing the study, propagated with the Netwisor_WebSurvey system, is presented in Figure 2.

Messages about research to be conducted must also specify the time frame of the data collection survey. It is assumed that the time necessary to perform this task should be minimized following the assumption of the network analysis, which says about changing dependencies. As a result, constructed logical networks that form an inherent organizations' nervous system are adequate only in the adopted moment or period.

4. Conclusions

Analysis of the available tools for conducting SNA in organizations, altogether with a literature review, additionally deepened with the practical experience gathered through the years of active participation in the number of studies performed in the different organizations resulted in the identification of other challenges process of data collection. Nevertheless, authors are aware that this is not a complete list, and new challenges might eventually appear. Issues identified covers a wide range of activities in the initial phase of conducting ONA. From regulatory constraints such as data confidentiality in the age of GDPR, through proper boundaries identification of the study to perform, to soft and even linguistic aspects of effective communication provided to the research participants to achieve a high response rate so crucial for the validity of the entire ONA.

In the next step, we described a set of remedies potentially valuable for overcome difficulties that block wider adoption of the SNA method. We addressed issues that are especially important in the age of COVID-19, such as the almost lost or at least hindered possibility to collect network data based on observations. Partially presented in this paper, the original Netwisor system allows to conduct of ONA in an utterly remote way. It is a proposal of a platform that not only integrates all phases of ONA but essentially, through its built mechanisms, it allows to collect as complete and correct data as possible, with the minimal effort needed from the administrator of the study.

From the perspective of data collection, organizational network analysis is still a long and complicated process that requires researchers' due diligence, appropriate preparation, and awareness of details that are not adequately addressed could result in lower quality of outcomes. The process is demanding from the very beginning - defining research boundaries till the last steps - constructing networks helpful from the different analysis perspective and failing or being late at any moment within risk to cause the entire research invalid or outdated. The authors hope that this paper would

minimize the burden of efficient network analysis in the organizations and help obtain a higher quality of conducted research.

References:

- Bird, C., Gourley, A., Devanbu, P., Gertz, M., Swaminathan, A. 2006. Mining Email Social Networks. In: *Proceedings of the 2006 International Workshop on Mining Software Repositories*, 137-143. ACM.
- Borgatti, S.P., Brass, D.J., Halgin, D.S. 2014. Social Network Research: Confusions, Criticisms, and Controversies. In: *Contemporary Perspectives on Organizational Social Networks*. Emerald Group Publishing Limited.
- Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G. 2009. Network Analysis in the Social Sciences. *Science*, 323(5916), 892-895.
- Cross, R., Ehrlich, K., Dawson, R., Helferich, J. 2008. Managing Collaboration: Improving Team Effectiveness through a Network Perspective. *California Management Review*, 50(4), 74-98.
- Cross, R.L., Parker, A. 2004. *The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations*. Harvard Business Press.
- Eagle, N., Macy, M., Claxton, R. 2010. Network Diversity and Economic Development. *Science*, 328(5981), 1029-1031.
- Ferrin, D.L., Dirks, K.T., Shah, P.P. 2006. Direct and Indirect Effects of Third-Party Relationships on Interpersonal Trust. *Journal of Applied Psychology*, 91(4), 870.
- Green, P. 2018. The Role of Organisational Network Analysis in People Analytics 2018. Retrieved from: <https://www.linkedin.com/pulse/role-organisational-network-analysis-people-analytics-david-green/>.
- Haythornthwaite, C., Wellman, B., Mantel, M. 1995. Work Relationships and Media Use: A Social Network Analysis. *Group Decision and Negotiation*, 4(3), 193-211.
- Huisman, M. 2009. Imputation of Missing Network Data: Some Simple Procedures. *Journal of Social Structure*, 10(1), 1-29.
- Johnson, R., Kovács, B., Vicsek, A. 2012. A Comparison of Email Networks and Off-Line Social Networks: A Study of a Medium-Sized Bank. *Social Networks*, 34(4), 462-469.
- Kawa, A. 2013. Analiza Sieci Przedsiębiorstw z Wykorzystaniem Metody SNA. *Przedsiębiorczość i Zarządzanie (Analysis of the Network of Enterprises with the Use of the SNA Method. Entrepreneurship and Management)*, 14(13), 77-87.
- Koji, A., Kanehira, N., Olguin, D., Waber, B.N., Kim, T., Mohan, A., Gloor, P., Laubacher, R., Oster, D., Pentland, A.S. 2008. Sensible Organizations: Changing Our Businesses and Work Styles through Sensor Data. *Journal of Information Processing*, 16, 1-12.
- Kotsios, A., Magnani, M., Vega, D., Rossi, L., Shklovski, I. 2019. An Analysis of the Consequences of the General Data Protection Regulation on Social Network Research. *ACM Transactions on Social Computing*, 2(3), 1-22.
- Kutik, B. 2019. Time to Care About ONA! HRExecutive.Com 2019. Retrieved from: <https://hrexecutive.com/time-to-care-about-ona/>.
- Monge, P.R., Noshir, S. 2003. *Theories of Communication Networks*. Oxford University Press, USA.
- Quintane, E., Kleinbaum, A.M. 2011. Matter over Mind? E-Mail Data and the Measurement of Social Networks. *Connections*, 31(1), 22-46.

- Sue, H., Fuller, A., Johnston, B. 2009. Chasing Shadows: Defining Network Boundaries in Qualitative Social Network Analysis. *Qualitative Research*, 9(5), 645-661.
- Visier. 2018. The Age of People Analytics Research Report 2018. Retrieved from: <https://hello.visier.com/age-of-people-analytics-research-report/>.
- Wasserman, S., Faust, K. 1994. *Social Network Analysis: Methods and Applications*, 8. Cambridge university press.
- Wawrzynek, Ł. 2019. Wykorzystanie Zarządczej Analizy Sieci Społecznych w Doskonaleniu Procesów Biznesowych (The use of Management Analysis of Social Networks in the Improvement of Business Processes), 300. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
- Zbieg, A., Zak, B., Jankowski, J., Michalski, R., Ciuberek, S. 2012. Studying Diffusion of Viral Content at Dyadic Level. In: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 1259-1265. IEEE.