ACTA INFOLOGICA, 2018; 2(1): 33-44
Research Article

ACTA INFOLOGICA
ISSN: 2602-3563

Validity Issues in Linked Data Driven IS Research

Ziya Nazım Perdahçı*, Mehmet Nafiz Aydın, Kenan Kafkas

ABSTRACT

This research adopts a complex system approach to linked data, which has a trace aspect and to examine validation issues in linked data driven IS research. Thereby a relevant question arises: What are the validity issues in the overall network analysis process applied on such linked data? This research argues that validity issues are vital to research in linked data and requires a complex system approach so that true value of linked data can be discerned and applicable to the real-world cases. Particular emphasis is placed on the validation issues in empirical research on linked data concerned with the educational system. This paper should be considered as a contribution to the efforts of those who are struggling with the validity issues in SNA. The intention of the work is to build a checklist that can be used to check the validity of the data, methods, and algorithms for transdisciplinary research teams who utilize theory of networks in general and SNA in particular in a particular domain, which is an educational system for the focus of this research. The findings may help the school administrators, instructors and student advisors in the decision making processes.

Keywords: Trace Data, Social Network Analysis, Network Science.

İlişkisel Verilere Dayalı Bilişim Sistemleri Araştırmalarında Geçerlik Konuları

ÖZ

Bu araştırmada ilişkisel veri odaklı Bilişim Sistemleri araştırmalarında görülen geçerlilik sorunlarını incelemek amacıyla, ilişkisel veri karmaşık sistem yaklaşımı benimsenmiştir. Bu durumda akla şu soru gelmektedir: Buna bezer ilişkisel veri üzerinde yapılan genel ağ analizi çalışmalarında geçerlilik sorunları nelerdir? Bu araştırmada ilişkisel veri araştırmalarında geçerlilik sorunlarının hayati derecede önemli olduğu ve ilişkisel verinin gerçek değerinin anlaşılması için karmaşık sistem yaklaşımının gerekli olduğu savunulmaktadır. Özellikle eğitim sistemleri ile ilgili deneysel araştırmalarda geçerlilik sorunları üzerinde durulmuştur. Bu çalışma Sosyal Ağ Analizinde geçerlilik sorunlarıyla karşılaşanların çabalarına katkı olarak düşünülmelidir. Disiplinlerarası çalışmalarında ağ teorisi özellikle de eğitim alanında Sosyal Ağ Analizi kullanan araştırma ekipleri için veri, metot ve algoritmaların geçerliliğini kontrol etmek amacıyla kullanılacak bir liste oluşturmak hedeflenmektedir. Elde edilen bulgular okul yöneticileri ve öğretmenlere karar verme süreçlerinde yardımcı olabilir.

Anahtar Kelimeler: İz Verisi, Sosyal Ağ Analizi, Ağ Bilimi.

Information of Author(s):

Ziya Nazım Perdahçı
ORCID: 0000-0002-1210-2448
nz.perdahci@msgsu.edu.tr
Mimar Sinan Fine Arts University

Mehmet Nafiz Aydın
ORCID: 0000-0002-3995-6566
mehmet.aydin@khas.edu.tr
Kadir Has University

Kenan Kafkas
ORCID: 0000-0002-1034-569X
kenankafkas@gmail.com
Kadir Has University

DOI: 10.30801/acin.356598

Submit Date: 21.11.2017
Accept Date: 22.02.2018
Publish Date: 26.06.2018

(*) Contact Author
Address: Mimar Sinan Fine Arts University, Department of Informatics, Bomonti, İstanbul, Turkey
Telephone Number: +90 212 246 00 11 Ext:6102

ACTA INFOLOGICA
ISSN: 2602-3563

1. INTRODUCTION

In the last decade in almost every field, data have become abundant, more accessible, and more diverse. For companies as well as academics, combining enterprise-wide data with open data to generate business intelligence brings up new opportunities and challenges (Behrendt et al., 2014). Recently, the term "linked data" is suggested to refer to bringing together all relevant digital data on the Internet for the sake of open data integration (Dong and Srivastava 2015). The current work extends the very idea of linked data from a typical integration context to trace data, which reveals the business context and complex relations of the things and their interactions. This aspect is crucial to make use of both enterprise and open data where the notion of "linked" emphasizes what and how data are derived from business context. In this regard, this research adopts a complex system approach to linked data, which has a trace aspect and to examine validation issues in linked data driven IS research.

The trace data present in Information Systems has certain characteristics. Among other types there is a data type called event-based data, which is often times enabled by conventional transaction information systems. The events mentioned here are usually records of various interactions between at least two entities. The recorded data turns up to have a linked structure and as a result of this linked structure a complex system emerges. To examine these complex systems, it is necessary to apply network science. Thereby a relevant question arises: What are the validity issues in the overall network analysis process applied on such linked data? The value and importance of the digital trace validation becomes immediately clear, taking into consideration the current studies (Jungherr 2015).

Howison et al. (2011) articulate the validity issues in network analysis of digital trace data and propose a number of issues that researchers should take into account. Addressing validity issues is vital to research in linked data and requires a complex systems approach so that true value of linked data can be discerned and applicable to the real-world cases. To better articulate the validation issues in a real-world context we utilize empirical research on linked data in a school information system as a case.

Although almost all the metadata in network studies in education comes from school management information systems, the crucial linked data is obtained generally by means of face to face surveys. Therefore, Social Network Analysis (SNA) in education does not completely rely on trace data. However, most of the methods and an important portion of the data are common in both papers. For this reason, validity issues match for both. This paper should be considered as a contribution to the efforts of those who are struggling with the validity issues in SNA. The intention of the work is to build a checklist that can be used to check the validity of the data, methods, and algorithms for transdisciplinary research teams who utilize theory of networks in general and SNA in particular in a particular domain, which is an educational system for the focus of this research.

Educators quickly adapted this situation and began to involve data more often in their decision making processes. Data Driven Decision Making (DDDM) concept is introduced in education (Marsh et al., 2006). Many advanced software emerged to meet the needs of educators. Learning Management Systems and School Management Information Systems became widely used in schools by both governments and private enterprises. Management Information Systems (MIS) are being used by schools to support a range of administrative activities including attendance monitoring, assessment records, reporting, financial management, and resource and staff allocation (O'Brien, 1998). As a result, a huge amount of data is collected and stored in relational database systems. Now, from governments to private sector, the education administrators and instructors rely on the information that is obtained by analysis of the data.

Although individuals play a key role in education, they are not isolated entities. Therefore, interactions among individuals also provide valuable information. This type of linked data requires a specific analysis, namely Social Network Analysis. Statistical data analysis in classical sense lacks the ability to capture the essence of complex systems that emerge from intricate human relationships. With SNA algorithms in a school environment for instance, key players in the network can be found or correlations between certain attributes of students can be calculated. The data necessary for this type of analysis may be the friendship ties among students or collaboration ties among groups. Additionally, SNA requires the data that is available in the traditional School Information Systems. For instance, a typical analysis would include the study of the friendship relations among

ACTA INFOLOGICA
ISSN: 2602-3563

students, involving calculation of the correlations between success rates and friendship preferences. This requires combining the relationship data with metadata, referred to as attributes such as gender or test scores. The findings may help the school administrators, instructors and student advisors in the decision-making processes.

2. METHOD AND BACKGROUND

The method of this study involves three steps: First, a model explaining general validity issues linked data driven IS research (Howison et al., model). Second, a specific research case (SNA in education) examplifying and elaborating the issues that might be encountered. Third, a framework on which the issues are explained in detail. Similar to the research design adopted in (Howison et al., 2011), we frame our study of validity issues respect to the decision matters researchers face with in network analysis of linked data driven IS research.

The growth of data sources produced on online interactive platforms have drawn significant attention from IS researchers, but the validity issue in SNA in IS research context has remained an open issue (Whelan et al., 2016). Scholars with exception of (Howison et al., 2011) have addressed this issue within their own research contexts. For instance, (Nia et al., 2010). have examined the validity of Network Analysis in open source projects. For our research purpose, we need a model that should achieve theoretical cohesion (full chain of reasoning across all the phases in SNA), and provides researchers with meta-level issue analysis (by raising abstraction level to overcome limitations of case specific results). In line with these reasons, we adopted the model proposed by (Howison et al., 2011). The model (Figure 1), proposed by (Howison et al., 2011) is composed of six elements connected by five links raising ten issues. These links are the transition areas between steps starting from Information System to the research construct. First, when working with a digital trace data particular attention is to be paid to the information system producing that data. The misuse of the system can cause misinterpretations of the collected data. Another issue that should be considered in this phase is a reliability of the data generated. In the next step, the complication of converting digital trace data into nodes and links should be solved. In order to do that, the researcher should make one of the most crucial decisions, which is determining the type and intensity of the links, as well as deciding on the missing links. Creating a network where the order of the events matter can raise a problem of temporal aggregation. In another step, while using a network to obtain some measures, a researcher should address network tool effects and temporal mismatch. There is a large selection of software tools available for social network analysis (SNA), thus choosing proper software for an analysis is an important step, since these tools can help researchers with avoiding errors as well as at the same time can threaten to validity in their use. The temporal mismatch issue can be addressed by deciding the period of time over which measures derived from that network will be measured. The last issues a researcher should consider emerge when aligning a measure and a construct are data completeness and inference and inappropriate importation.

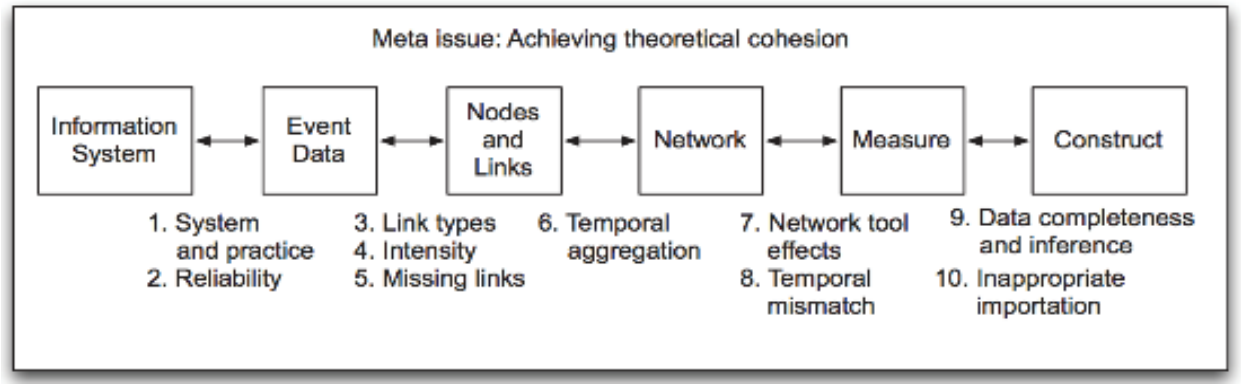

Figure 1. Howison et al. (2011) introduce five links in the chain of reasoning and corresponding validity issues to achieve theoretical cohesion.

Howison et al. (2011) state that "In practice, the process of achieving alignment between a theoretical context and the chain of reasoning underlying valid measurement is an iterative one, most likely involving multiple

ACTA INFOLOGICA
ISSN: 2602-3563

adjustments and decisions and revisiting these to achieve a cohesive logic." Thus, applying the model (Figure 1) has been an iterative process, which brought out three phases.

2.1. Social Network Analysis in Education

From the Network Science perspective, a network consists of two types of simple components, nodes and links. Nodes may represent an individual in a social network or an enzyme in a cell. The connections between nodes are called the links. They may represent kinship between individuals in social network or chemical interaction between enzymes in a cell. The term graph refers to mathematical representation of a network. It is analogous to a wiring diagram. The terms network, node and link are mostly used for referring real world complex systems whereas the terms graph, vertices and edges are used when referring to a mathematical representation of the real-world systems. These are only subtle differences and these terms are often used interchangeably (Barabási and Pósfai, 2016).

A node can have more than one link. Total number of links of a node is called its degree. If links in a network have distinct direction from one node to another, this type of network is called a directed network. In an undirected network links do not have directions. There are two types of degrees in directed networks; in-degree which is the number of links pointing towards a node, out-degree that is the number of nodes pointing out from a node

Social Network Analysis requires node and link data and in education networks, this corresponds to nodes being students or teachers and links being the relationships among nodes. The metadata about the nodes are called node attributes, which can be any data that range from the name or address of the student to test achievement scores or the name of the course taken. The link data is type of data, which defines a relationship between two nodes for instance, if two students study together or take the same course, the two nodes representing the students are linked in the graph. These data are mostly available in School Management Information Systems; however, additional data can be gathered by surveys. After collection, the data is prepared for analysis. The next stage is modelling. Deciding the types of nodes and the links is called modelling. Figure 2. shows a simple model of a network where nodes are students and links are drawn if the two students are best friends.

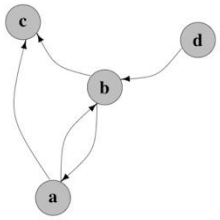

Figure 2. A simple directed network model.

In other words, if student "a" claims that student "b" is his or her close friend, then an edge with an arrow pointing from node "a" to node "b" connects them. Looking at this toy model closely reveals the fact that student "a" accepts student "c" as a close friend, student "c" however, does not perceive him or her as close friend. In these kinds of networks, links have directions which makes the network directed network. In the organization science literature, such a directed network model for friendship has not been examined extensively (Smirnov and Thurner, 2017).

ACTA INFOLOGICA
ISSN: 2602-3563

3. FINDINGS FOR PROPOSED FRAMEWORK

3.1. Framing Validity Issues in Linked Data in Network Science

In this study, we present a framework (see Figure 3) that illustrates general structure of a typical scientific research involving linked data along with digital trace data. Our intention is to address the validation issues that may arise during the research process. The framework contains flow of data through steps of research up to the scientific output. One can see that the framework contains a number of feedback loops to enhance scientific rigor and validation. In essence, it is an operational means to support research inquiry by incorporating relevant theoretical accounts. Noticably, linked data is considered from a complex system point of view, which allows us to bring theory of network and information system research together. In doing so, the scientific output may include descriptive, predictive, and prescriptive IS which in turn enhance researchers to provide feedback to information systems.

Real World System

Information systems collect data from the system, which is in fact the Real World. However, interactions between them is not one way. That is to say, the system and IS constantly interact and shape each other. Customers interacting with sales representatives, employees interacting with each other or the environment, students working mutually on a project, are examples of such activities in a real-world system.

Information System

Information Systems keep track of these activities and a huge amount of data cumulate over time. This event-based trace data has a linked nature. The interactions leave trace of events, which can later be collected. Since events take place among entities of the system, the entities can be linked to each other and this phenomenon can be represented as a network.

Linked Data

Two different types of data are obtained from two different systems. Offline enterprise data is gathered from the IS and depending on the case, ground truth data or open data is gathered from the real-world system. For instance, ground truth may correspond to various observed relationships among students in school environment. On the other hand, open data may correspond to other social interactions gathered out of school premises.

The resulting network is a complex system containing substantial number of interacting components. To examine such systems, network science approach is required.

Network Science and IS Research

The necessary steps to analyze these networks is explained in SNA process section. This section covers the network analysis particularly the Social Network Analysis in order to limit the scope of the paper.

The Scientific Output

The scientific output of the entire process is the description of the system, predictions towards the future and prescriptions of the IS problems. Finally, these outputs are sent to the IS as a feedback.

During the transitions of each step of the process, researchers should check the validity of the data, methods and findings. The validity issues section addresses the possible validity issues and explains them in detail.

ACTA INFOLOGICA
ISSN: 2602-3563

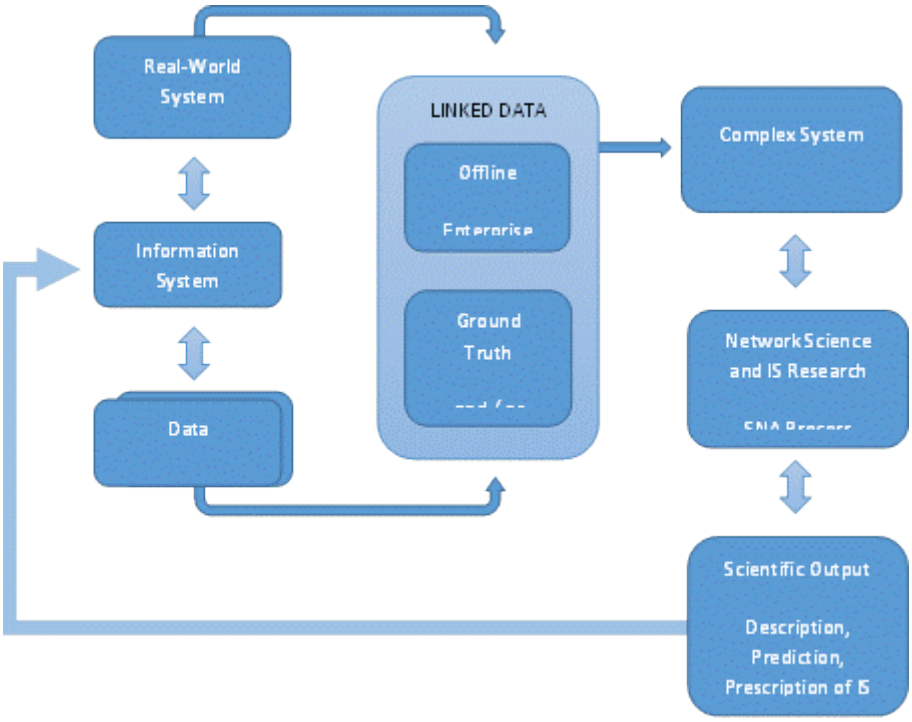

Figure 3. Framework for Validation Issues in Linked Data Driven IS Research

3.2. Network Analysis of Linked Data

The SNA Process

Four basic steps are taken into consideration when approaching a social network research problem from Network Science perspective: Data preparation, network modelling, network analysis and interpretation (Aydın and Perdahçı, 2014). Researchers follow these steps when looking for answers to their research problems. The process has an iterative structure. The data is obtained from various sources depending on the necessities of the social network that is subject of the study. In the literature, self-reported friendship data collection is one of the well-known techniques (Labun et al., 2016). These data are then prepared to suit the requirements of the analysis and visualization tools so that they can be appropriately imported into working environments.

Following the data preparation, the modelling step takes place where the nodes and the links of the network are decided. After this step, analysis begins which is mainly visualizing the constructed network and applying suitable SNA algorithms. The findings that are obtained in this step are scrutinized and the process moves to the next step. In the interpretation step, if the findings do not satisfy the research objectives, the process enters in a loop where the previous steps are repeated by going back to modelling step. The network model is adjusted for new requirements and the process moves to next iteration. For instance, the first iteration would be a network model in which nodes are students and undirected links are reciprocated friendships among students. Should the analysis fail to capture the true nature of the friendship relations a second iteration would involve remodeling the friendship relations as a directed network (one student sees the other student as a friend, but the other student does not.)

Then the SNA process continues, and the findings are reported. In some situations, researchers may decide that there is a problem with the data they collected in the beginning. For instance, they might decide to add a node attribute. In that case, process moves back to the beginning and the entire process is repeated with the changes made.

ACTA INFOLOGICA
ISSN: 2602-3563

Modelling

The choices made at the modelling step are critical to representing a complex system as a graph. For instance, individuals who are regularly interacting with each other can be connected to create a professional network. On the other hand, the relationships among individuals who are calling and mailing each other define an acquaintance network. The analysis of first type of networks for instance, can play role in a company's success in terms of management. The analysis of second type of networks can play a key role in marketing products or services (Aydın and Perdahçı, 2014). Likewise, when analyzing social networks in education modelling step can be critical.

There are several options as there are several types of interactions in a school environment. These interactions can be acquaintance, close friendship among students or teaching relationship between teachers and students. Additionally, sharing activities such as projects, classes, team sports and various clubs can create interactions. The relationship should be determined by choosing the best option, which fits the research objectives. Furthermore, multiple relationships can be selected as links. In that case, the network becomes multi layered. To illustrate a multi layered model, let the close friendship relations among a set of students constitute a friendship network, while linking the same set of students according to their shared projects, make up a collaboration network. The combination of these two networks in one network in which some of the links represent friendship and others represent collaboration, results in a multi layered or, in other words, multiplex network. The actors in a network can also vary in terms of node selection such as, students, instructors, administrators. However, selecting nodes is usually simpler with respect to link selection.

Table 1. Possible inputs and outputs of SNA in school networks depending on the link type

	Friendship Network	**Collaboration Network**	**Course Affiliation Network**
Input	Who is friends with whom	Who studies with whom	Who takes same courses with whom
	Test achievement scores	Content of the project	Course subject
	Gender	Degree of contribution	Social clubs
	Class	Project grade	Team sports
	Social Background	Study major	Success rate of the activities
Used Metrics	Clustering	Clustering	Clustering
	Assortativity	Assortativity	Assortativity
	Modularity Opt.	Modularity Opt.	Modularity Opt.
	Centrality measures	Centrality measures	Centrality measures
Outcome	Correlations based on attributes	Correlations based on the specific project	Better groups, teams etc.
	Class Compositions	Modularity Opt.	Homopohily
	Broker nodes	Sustainability of Interconnectedness	Correlations based on the specific programs

Network Analysis Findings

In this step, the modeled network is visualized as a sociogram (Moreno et al., 1932; Freeman, 2004) by utilizing graph visualization tools. A sociogram is a map of social relations in which individuals are depicted as circles and related individuals are linked with each by lines. Here, the students are represented as nodes and the nodes are linked according to chosen model in the previous stage. This graph representation gives visual insight into the social network, which shows how the students are connected with each other. After visualization, basic

ISSN: 2602-3563

analysis takes place where basic structural properties of the network are calculated. These findings for instance, show the density of the network (i.e. how densely they are connected to each other). And, how much the students are clustered (how closely they are grouped). Further analysis finds the communities in the network or how students with different attributes mix with each other. Additionally, the students who form a link between communities can be found with SNA algorithms.

Although in an education setting there are several relationships that can be subject of study, the current study focuses on three types of networks: Friendship networks, collaboration networks, and affiliation networks. While the friendship networks focus on the friendship relations among students, collaboration networks focus on the shared activities of the students such as working on the same project. Attending the same course or being in the same social club generally represent the links of an affiliation network. These network types nececitate different inputs and the anlysis of these networks offers different outputs. Table 1. Shows some of the possible inputs for a research that utilizes SNA as a method and the possible outputs of that analysis.

4. DISCUSSIONS OF VALIDITY ISSUES FOR LINKED DATA IN A SCHOOL INFORMATION SYSTEM

In the following, we discuss our findings on validity issues encountered in linked data driven Information Systems research.

4.1. Reliability Issues from System Generated Data

SNA in school networks obtain node attributes such as student grades, name of the taken courses etc. from information systems. These are not system generated data and since such data is heavily checked by the administrators, teachers, students and parents, they can be considered reliable. For instance, if the age, gender or grade of a student is entered incorrectly into the system, the students instantly demand a correction. Therefore, in a school network the validity of the data gathered from Information System should not be an issue for statistical conclusions derived from this data.

4.2. Aligning Digital Data and Nodes & Links

A researcher, at the beginning of a study, has to decide how to use the data to build a network. The first desicion is to determine which entities in the data will constitute nodes and links. There are many options for this decision step, for instance, in an e-commerce sales data, products might be the nodes as well as the sellers or the buyers. The desicion should align the data and the network according to the research objective both contextually and theoreticaly. “In Social Network Analysis, however (emphasis on the word Social), nodes are almost always people, although at different levels of analysis they might be individuals, groups, or organizations” (Howison et al., 2011). Similarly, in education settings nodes are almost people namely, students, instructors or administrators. There may be instances where education tools or classrooms make up the nodes of the network along with students and instructors. Since these are all solid entities in education environments, they should not raise significant validity issues.

4.3. Choosing Multiple or Single Link Types

Choosing link types in modelling step of SNA is a validity issue that concerns the construct of the network. The number of link types in a network should align with the research interests and requirements. In a school, social network there are several relationships that can be studied: affiliation (course or lab attendance), collaboration (study or project groups), friendship (best friends or acquaintance). A network can be modelled using single link type or multiple types (multiplex network). An analysis of a single link type network may leave out the effects of other relationships among the nodes. On the other hand, a multiplex network analysis may lack the ability to detect the certain interactions taking place in the network. In that sense, number of link types should align with the research objectives.

ACTA INFOLOGICA
ISSN: 2602-3563

4.4. Defining a link (Link Intensity)

Only existence of a link between two nodes is sufficient for some research problems however, in some cases link's intensity or strength is also a key factor. For example, in a student collaboration network, links can be established in a binary mode where a link has only two states: zero or one. In other words, a links indicates whether two students studied together or not. However, if the same students collaborated multiple times or they collaborated for longer periods, we might need a stronger link betweeen them so that the degree of their collaboration is represented as link intensity. Therefore, the links should have an attribute indicating the intensity of that relationship called weight of a link. Metrics such as centrality measures or community detection algorithms produce different results depending on the type of the links.

4.5. Defining a non-link (Missing Links)

In school social network analysis, often link data is obtained via face to face surveys. In a school environment, social networks consist of small number of nodes. Therefore, reaching to the students, instructors or administrators is not a challenge. This eliminates the validity issue from missing node aspect. However, this does not eliminate the non-link issue which means that knowing if a link exists between two nodes is important, in the same sense, knowing if a link does not exist is also important. In other words, lack of a link indicates lack of interaction between two nodes where in fact, this might be due to missing data, which can cause serious errors for example, in a friendship network, students are asked about their friends and the links of the network is established with that data. However, in the constructed network if two nodes are not linked this does not necessarily mean that they are not friends. If information flow in the network is needed to be measured, making sure that two students are exactly friends or not friends is critical. Betweenness centrality is a network measure that quantifies the brokerage position of a node. The nodes that have high betweenness centrality play a vital role in bridging communities in a network. For papers that examine this metric in school networks (Grunspan et al., 2014) absence of a node is as important as existence of it.

4.6. Temporal Aggregation

In dynamic analysis of a social network, some situations can cause validity problems. For example, an affiliation network considers two students as linked if they take the same course. However, if the students did not attend that course at the same time, this means that there has not been an interaction between those nodes in the network. This issue happens when links aggregate in time and can lead to faulty analysis results especially for the algorithms that depend on the path calculations. There are solutions to this problem and the techniques to deal with this issue are mentioned in (Howison et al., 2011).

4.7. Network Tool Effects

There are various tools to visualize the network map and calculate the network metrics (Csardi and Nepusz, 2006; Hagberg et al., 2008). Using these tools helps researcher to standardize the implementation of the algorithms. Otherwise, reimplementation of the algorithms by different researchers could cause other validity issues. Additionally, using common tools provide some shared ground for the work to be repeated by other researchers. "Nonetheless, the convenience these tools provide can also mask threats to validity in their use. First, programs use subtle variations of algorithms and slightly different names for the same algorithm, potentially leading to confusion and misinterpretation of results." (Howison et al., 2011). To overcome this problem, the same algorithms are calculated manually (pen-and-paper calculation) on a toy network and compared with the network tool's calculation results.

4.8. Temporal Mismatch

In most cases, social networks are analyzed in a static manner. In other words, obtained data is a snapshot of the network, which represents a short time interval. This can lead to validity issues where a network metric which changes over time as the network evolves may mislead the researchers (Huisman and Snijders, 2003; Leskovec et al., 2005). For example, betweenness centrality of a student may be high when the data is collected, then it may change due to the nature of social interactions. Moreover, not only measured values change but also the

ACTA INFOLOGICA
ISSN: 2602-3563

links in the network may change over time. For instance, a student may decide to partner with a different student in a collaboration network. Similarly, change of friends or even best friends is frequently seen among high school students. Researchers should make sure that the observed measurements span the entire time frame.

4.9. Questions of Data Completeness

Data about the relationships among actors of a network gathered in one environment no matter how complete may lack the information of interactions that happen out of that environment. For instance, researchers may study friendships among students and its correlations to numerous factors by collecting the relationship data via surveys or interviews. However, due to access or privacy issues, their information most likely will lack their social media interactions. This is another validity issue that is caused by incompleteness of the data. Depending on the subject of the study, incomplete link data will reduce the reliability of the SNA results.

4.10. Inappropriate Importation of Network Measure Interpretation.

Network measures and algorithms utilized in SNA are mostly transferred from other disciplines of science. For example, Pearson's correlation formula was originally used to describe a biological phenomenon (Pearson, 1896). It was later imported by several other disciplines such as economy, physics, chemistry etc. when it is needed to examine linear relationships between two quantitative variables. In network science, this metric is used to find out the mixing behaviors among nodes called assortativity (Newman, 2002). In specific case of social networks, this metric is used for instance, to measure mixing behavior among individuals (Bearman et al., 2004). The interpretation of the results might differ in different venues. Therefore, researchers should pay attention to the interpretations of their findings for avoiding validity issues that is caused by importation.

5. CONCLUSION

This paper addresses the validity issues that researchers face when conducting Social Network Analysis. Although its scope is SNA studies in general, education domain applications is used to exemplify the validity issues. Prior to addressing these issues, network concept is briefly explained. Furthermore, a conceptual model is presented which covers Network Science processes and how the linked data advances starting from the real-world system and IS to complex systems and finally analyzed to produce scientific output.

The validation issues mostly arise between the phase transitions. Data reliability at the beginning when deciding the nodes and links are not likely to cause serious validation problems since these entities are well defined in education networks. However, decisions about the link types, weights or even non-existence of a link are potentially critical validation checkpoints during the SNA process. Another type of validation issue arises due to temporal issues when deciding the analysis to be static or dynamic. The interpretation of the algorithm results should involve the effects of time over the network. Additionally, as in every scientific research, the utilized tools have validity issues as well as the measures. Researchers should be aware of the strengths and weaknesses of their tools and metrics.

The Network Science is relatively a new discipline. Therefore, researchers should be informed about the pitfalls throughout the processes. This paper does not claim to address all possible issues rather; it is intended to be used by researchers as a starting point to avoid these issues and as a validation checklist after their research. To that end, issues are collected and examined in detail furthermore; practical solutions are offered to facilitate the researchers in their network analysis efforts.

ACTA INFOLOGICA
ISSN: 2602-3563

REFERENCES

Aydın, M. N., & Perdahçı, N. Z. (2014). Ağ Bilimi Yaklaşımı Ve Çevrimiçi Etkileşimli Sağlık Platformunun Bir Örnek Olarak İncelenmesi: Informa Yönetim Bilişim Sistemleri Dergisi, 1(2), 60-80.

Barabási, A. L., & Pósfai, M. (2016). *Network science*. Cambridge university press.

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American journal of sociology*, *110*(1), 44-91

Behrendt, S., Richter, A., and Trier, M. (2014). Mixed methods analysis of enterprise social networks. *Computer Networks*, *75*, 560-577.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1-9.

Dong, X. L., & Srivastava, D. (2015). Big data integration. *Synthesis Lectures on Data Management*, *7*(1), 1-198.

Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*, *1*.

Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE-Life Sciences Education*, *13*(2), 167-178

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX* (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States)

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, *12*(12), 767.

Huisman, M., & Snijders, T. A. (2003). Statistical analysis of longitudinal network data with changing composition. *Sociological methods & research*, *32*(2), 253-287.

Jungherr, A. (2015). Analyzing political communication with digital trace data. *Cham, Switzerland: Springer*.

Labun, A., Wittek, R., & Steglich, C. (2016). The co-evolution of power and friendship networks in an organization. *Network Science*, *4*(3), 364-384.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005, August). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 177-187). ACM.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). Making sense of data-driven decision making in education.

Moreno, J. L., Whitin, E. S., & Jennings, H. H. (1932). Application of the group method to classification: National Committee on Prisons and Prison Labor. *Psychology Abstracts*, *6*, 2872.

Newman, M. E. (2002). Assortative mixing in networks. *Physical review letters*, *89*(20), 208701

Nia, R., Bird, C., Devanbu, P., & Filkov, V. (2010, May). Validity of network analyses in open source projects. In *Mining Software Repositories (MSR), 2010 7th IEEE Working Conference on* (pp. 201-209). IEEE.

O'Brien, J. A. (1998). *Management information systems: Managing information technology in the networked enterprise*. McGraw-Hill Professional.

ISSN: 2602-3563

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, *187*, 253-318

Smirnov, I., & Thurner, S. (2017). Formation of homophily in academic performance: Students change their friends rather than performance. *PloS one*, *12*(8), e0183473.

Whelan, E., Teigland, R., Vaast, E., & Butler, B. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization*, *26*(1-2), 1-12.

This work was presented at the 4th International Management Information Systems Conference and published in the conference abstract book.