A Process Framework for Big Data Research: Social Network Analysis Using Design Science

Denis Dennehy, School of Management, Swansea University, Swansea, Wales.

Samrat Gupta, Indian Institute of Management, Ahmedabad, Gujarat, India.

John Oredo, Department of Library and Information Science, University of Nairobi, Kenya.

Abstract

Managing and harnessing big data is increasingly being reported as an approach to generate business value, optimize decision-making, and achieve competitive advantage. There is strong evidence that research on big data has gained significant attention from both the academic community and analytics community. To date, research has largely focused on the technical aspects of big data and its applications in specific contexts, but with limited attention given to the underlying process. Yet, it is well accepted that understanding the processes required to leverage big data is a critical factor to realize the claimed benefits of big data. We address this knowledge deficit by designing a process framework to guide novice users to effectively apply social network analysis and improve the outputs of big data research projects. The framework is the artifact that emerged after applying the principles of design science research. The artifact was validated by a social network analysis of credit networks in India.

1. Introduction

There has been a contagious enthusiasm by academics and practitioners surrounding the notion of 'big data' and how it will revolutionize decision-making (Modgil et al., 2021; Choi et al., 2018; Fossa Wamba et al., 2017). To facilitate data-driven decision making, organizations need to invest in big data initiatives (Grover et al., 2018) to develop efficient and effective processes that will translate big data into meaningful insights (Davenport, 2018). At the same time, concerns are being raised that investing in big data initiatives does not necessarily lead to more effective decision making (Hirschheim, 2021; Ghasemaghaei et al., 2018; Dennehy et al., 2021). Further, decision-making "*is not just an act of decision-making between a given set of parameters, but it is also about the continuous act of shaping and designing of organizations and their stakeholders' experiences*" (Avital *et al*., 2009, p. 154).

Big data can be characterized in terms of seven Vs, comprising of volume, variety, velocity, veracity, value, variability, and visualization, which present various challenges in data management (Mikalef et al., 2017; Seddon and Currie 2017; Gandomi and Haider, 2015). There are two key processes for extracting insights from big data: data management and big data analytics (BDA) (Gandomi and Haider, 2015). Data management refers to the processes and technologies required to collect, store, and prepare data for analysis, while BDA refers to the entire process of managing, processing, and analyzing the data characteristics (e.g., Vs) to create actionable insights to deliver sustained business value, measure performance and achieve competitive advantage (Fosso Wamba et al., 2015; Watson, 2014). BDA can be categorized into three types (e.g., descriptive analytics, predictive analytics, and prescriptive analytics), which have implications for the technologies and architectures used for BDA (Watson, 2014). Developing 'big data analytics capabilities' is an emerging technological capability to effectively deploy an organizations' data, technology and talent through firm-wide processes, roles and structures (Mikalef *et al.*, 2019).

To date, research has largely focused on the technical aspects of big data (Mikalef et al., 2017) and its application in specific contexts (e.g., marketing, healthcare, smart cities, supply chains), but with limited attention given to the underlying process (Grover and Kar, 2017). This knowledge deficit is concerning as it is critically important to understand the processes required to leverage big data and create business value through data-driven decisions (Mikalef et al., 2017). Further, processes are important because they enable organizations to standardize employee work activities, enhance their process execution, as well as benefit from process standardization (Rosemann and vom Brocke, 2015; Schaefer et al., 2013). This is particularly important for novice users (e.g., students, graduates) who may have limited knowledge or expertise about an organisational process and therefor require support in their process execution (Morano et al., 2020).

The context of this study is the credit networks banks and organizations in India. We apply social network analysis whereby social phenomena are represented and studied by data on overlapping dyads as the units of observation (Brandes et al., 2013). Social network analyzes consists of a series of mathematical techniques that, using network and graph theories, can be used to understand the structure and the dynamics of complex networks (Pallavicini et al., 2017). A complex network is a system for which it is difficult to reduce the number of parameters without losing its essential global functional properties (Costa et al., 2007).

Numerous tools have been developed to fulfil the task of analyzing and describing complex social networks (Kim and Hashtak, 2018; Valeri and Baggio, 2021).

Social network analysis has been used to in a range of contexts including disaster management (Kim and Hashtak, 2018), tourism management (Valeri and Baggio, 2020), conspiracy thoeries abouve Covid-19 and 5G (Ahmed et al., 2020), disease ecology (Albery et al., 2021), migration and transnationalism (Bilecen et al., 2018), and online collaborative learning (Saqr et al., 2018). Despite the large body of literature addressing the topic of social network analysis, there is a noticeable absence of process frameworks that can guide novice researchers and practitioners.

We address this gap in knowledge by proposing a design-based process framework to guide novice and experienced researchers and practitioners in the use of big data. We ground our framework on two streams of literature, namely social network analysis and design science research (DSR). In our research project we follow the design science research (DSR) approach and address the following research aim:

To design a process framework for the effective application of social network analysis in big data research projects.

Design science research (DSR) is a problem-solving paradigm that seeks to 'design and evaluate' innovative artifacts (e.g., concepts, models, methods, and instantiations) with the desire to improve an environment, by introducing the artifact and associated processes for creating it (Holmström et al., 2009; March and Smith, 1995; Hevner et al., 2004). While several process models have been proposed for DSR projects (e.g., Nunamaker et 1991; Walls et al., 1992; Hevner, 2007; Kuchler and Vaishnavi, 2008) we adopt the model proposed by Peffers et al., (2007) as it is the mostly widely cited DSR model (vom Brocke et al., 2020) and although it is presented in a nominally sequential order, it is iterative in practice (Peffers et al., 2007).

DSR is about understanding and improving the search among potential components to construct an artifact that is intended to solve a real-world problem (Baskerville, 2008). Essentially, DSR addresses 'wicked problems', or using Simon's (1973) terminology, 'ill-structured' (Brooks Jr, 1987; Rittel and Webber, 1974), which are "decision situations where decision-makers may not know or agree on the goals of the decision, and even if the goals are known, the means by which these goals are achieved are not known and requisite solution designs to solve the problem may not even exist" (Holmström et al., 2009, p. 67).

The remainder of this chapter is structured as follows. First, a synthesis of key literature related to social network analysis and the principles and structure of complex networks is presented. Next, justification for adopting a design science research methodology is provided. Then, a rich context of the financial credit networks of Indian banks and organizations is provided. Followed by a discussion about the proposed process framework and implications for research and practice. The chapter ends with a conclusion.

2. Review of Social Network Theory

A social network is a collection of actors (nodes) that include people and organizations linked by a collection of social relations (Laumann et al., 1978). It is widely employed in the social sciences, behavioral sciences, political science, economics, organizational science and industrial engineering (Garton et al., 1999). The fundamental components of a social network study are the actor (node) and the connection (link) The nodes can be individuals, corporates or groups and other social units and the nodes are linked to each other by ties (Wasserman and Faust, 1994). As a geographical map describes the landscape, networks offer a tantalizing tool to model the complex systems existing in real world. Network theoretic modelling and visualization helps in managing and apprehending the enormity of complex systems. Networks help in understanding the basic patterns of interactions within the components and thus aid in understanding the complexity in real-world systems (Boccaletti et al., 2006). For instance, banking systems of several countries including India, Peru, Italy, Mexico, and US have captured accurate repositories of their interbank network for systemic risk analysis (Bargigli et al., 2015; Cuba et al., 2021; Gupta and Kumar 2021; Soramaki et al., 2007). In these networks (generally referred as *complex networks*), the connection patterns between the nodes are neither purely regular nor random – they are complex (Fortunato, 2010; Soramaki et al., 2007). The goal of modelling and analyzing complex networks is to reproduce the observed collective behavior in the real world by simplifying the rules of interaction between the components constrained in the network.

2.1 Properties and structure of complex networks

Complex networks and their implications on dynamical processes forms a broad area of study. Some of the most important research areas in complex networks are related to models of networks, structural properties of networks, module discovery in networks, motif discovery in networks, link prediction in networks and visual representation of networks. The properties that characterize the structural aspects of complex networks are presence of giant component, small world effect, scale-freeness, high clustering coefficient and presence of modular structure which are discussed next.

Presence of Giant Component: The real world complex networks either contain a giant component or they are fully connected (Barabási, 2014). Giant component containing a finite fraction of all the nodes emerges if the average degree represented by $< d >$ is greater than 1 (Barabási, 2014). However, all the nodes of a network are absorbed by the giant component if average degree $< d >$ is greater than $ln|N|$ where $|N|$ is the number of nodes in a network (Barabási, 2014). Though many real world networks such as the internet and power grid do not satisfy the criteria of being fully connected (Barabási, 2014; Pagani and Aiello, 2013), the social network of humans in the world with a population of around 7.5 billion satisfies the criteria of being fully connected as average degree $< d > \approx 1{,}000$ is greater than $ln(7.5 \times 10^9) \approx 22.73$ (Barabási, 2014).

Presence of Small World Effect: Complex networks are characterized by the small world phenomenon implying that any two randomly selected nodes in a network are connected within short distances or hops (Watts and Strogatz, 1998). In practical terms, small world effect has been manifested as "six degrees of separation" meaning that between any two individuals even on the opposite side of the globe there exists a path of at most six acquaintances (Travers and Milgram, 1969). Considering a network with average degree $< d >$ and $|N|$ number of nodes, small world effect can be explained by the following calculation. Any node in this network is on an average connected to:

$< d >$ nodes within 1 hop.

$< d >^2$ nodes within 2 hops.

$< d >^3$ nodes within 3 hops.

….

$< d >^h$ nodes within h hops.

Precisely, the expected number of nodes up to distance h from starting node can be formulated as:

$$E(h) \approx 1+ < d > + < d >^2 + < d >^3 + \cdots + < d >^h = \frac{< d >^{h+1} - 1}{< d > -1} \approx < d >^h \qquad (2.1)$$

Assuming that the maximum number of hops or diameter of the network is h_{max} and given that the total number of nodes in the network is $|N|$, it can be mathematically expressed as:

$$E(h_{max}) \approx |N| \qquad (2.2)$$

$$< d >^{h_{max}} \approx |N| \qquad (2.3)$$

$$h_{max} \approx \frac{ln\ |N|}{ln\ < d >} \qquad (2.4)$$

Thus, equation 2.4 represents the mathematical formulation of small world effect and also offers a good approximation for average path length $< h >$ between any two nodes in complex network (Barabási, 2014). Since $ln\ |N| \ll |N|$, the dependence of diameter (h_{max}) or average path length ($< h >$) on $ln\ |N|$ implies that distances in real networks are much smaller than size of system. Moreover, the denominator $ln\ < d >$ implies that denser the network, smaller the average distance between the nodes of network.

Scale Free Property of Complex Networks: The degree distribution of nodes in random networks is Poisson such that degree of each node is typically given by $d \approx < d >$ (Barabási and Albert, 1999). See Figure 1(a) that illustrates the poisson degree distribution in random networks. On the contrary, complex networks have a statistically significant probability of each node having a much higher degree than the average degree $< d >$ (Barabási and Albert, 1999). Therefore, complex networks are free of a characteristic scale and called scale-free networks (*ibid*). The degree distribution of nodes in these networks is given by power-law:

$$P\ (d) \approx d^{-\gamma} \qquad (2.5)$$

where the value of γ approximately lies between 2 and 3 which implies that there are few hubs in a network that are highly connected and dominate the topology of the network (Dorogovtsev and Mendes, 2002). See Figure 1(b) that illustrates the power law degree distribution in complex networks. Power law degree distribution has been observed in several real-world networks such as internet, world-wide-web, international trade networks and citation networks

(Rosvall, 2006). In a non-network context power law has been observed in the rank of word frequencies, size of cities and distribution of incomes (Zipf, 1949).

Clustering Coefficient: Measures the extent to which nodes in a network tend to cluster together (Boccaletti et al., 2006). Intuitively, clustering coefficient represents the probability that two connections of a person relate to each other in a social network. It has two versions: first based on global aspect of clustering in the network and second based on local (node-wise) indication of clustering (Boccaletti et al., 2006).

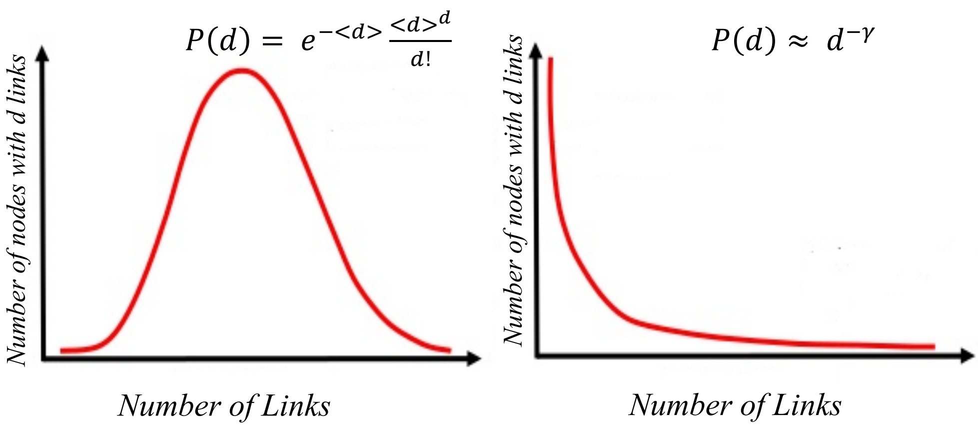

Figure 1 (a) Poisson Degree Distribution in Random Networks (b) Power Law Degree Distribution in Complex Networks

The global clustering coefficient is the measure of probability that two adjacent neighbors of a node are also adjacent to each another (Newman, 2003). This relation leads to formation of triangles within the network. Thus, more the number of triangles within a network, higher the clustering coefficient. Mathematically, it can be represented as:

$$GC = 3\frac{N_\Delta}{N_\wedge} \tag{2.6}$$

where N_Δ denotes the number of triangles wherein each of the three nodes is connected to remaining two nodes. $N_\wedge$ denotes the number of connected triplets wherein atleast one node is connected to other two. The multiplication factor of three indicates that each triangle forms three connected triplets and value of GC lies between 0 to 1 (Newman, 2003).

Unlike GC which is dependent on the global properties of the network, local clustering coefficient LC_i of a node n_i is given by the ratio of links existing between the nodes in the adjacent neighborhood of n_i divided by the number of possible links between them (Watts and Strogatz, 1998). Mathematically, this notion can be formulated as:

$$LC_i = 2\frac{\left|\{l_{jk}: n_j, n_k \in N_i, l_{jk} \in L\}\right|}{d_i(d_i - 1)} \tag{2.7}$$

where $N(n_i)$ denotes the neighborhood subset of node n_i and d_i is the degree of node n_i. The local clustering coefficient LC is then determined by taking the average of local clustering coefficients of all the nodes in a network, as given by following formula:

$$LC = \frac{1}{|N|}\sum_{i=1}^{|N|} LC_i \tag{2.8}$$

Though, there are several other measures to reveal the structure of real-world networks, a network can be considered as complex if number links present in the network is much less than the total possible number of links within it. The average node degree is greater than one and power law exponent is greater than two (Barabási, 2014). Moreover, the clustering coefficient of the complex network should be much higher than that of the corresponding random network and average path length should be reasonably close to that of the corresponding random network (Albert and Barabási, 2002).

Modular Structure: A high clustering coefficient gives an indication about the network topology and presence of clusters of nodes in a network. These clusters of nodes have been referred to as cohesive subgroups, modules, complexes, depending on the context and research discipline. In today's digital world wherein, networks are increasingly being mapped, modules can be viewed as groups of humans, places, banks, photos, events, web pages or any other real world entity. In unipartite social networks clustering of nodes has been studied theoretically as homophily, one of the underlying tenets of social network theory (McPherson, Smith-Lovin, and Cook, 2001). Homophily is the tendency of individuals to mingle with other individuals of their own kind. This tendency may be induced by preference such as gender and ethnicity or by constraints such as organization and educational standards.

"Modularity" is a network level measure to determine the degree of homophily or goodness of modular structure in a complex network (Newman, 2006). Mathematically, modularity can be expressed by the following formula (Clauset, Newman, and Moore, 2004):

$$Q = \frac{1}{2|L|}\sum_{vw}\left[A_{vw} - \frac{d_v d_w}{2|L|}\right]\delta(C_v, C_w) \tag{2.9}$$

where $|L|$ is the total number of edges in a network, d_v is the degree of the node n_v and d_w is the degree of node n_w. A_{vw} is the adjacency matrix in which $\delta(C_v, C_w) = 1$ if n_v and n_w are in same module and 0 otherwise.

2.2.2 Modules in unipartite and bipartite networks

Modelling real-world complex networks wherein data is interweaved in the form of nodes and links is one of the main research goals of big data analytics (Chang 2018; Hu and Zhang, 2017). Graph theory, one of the most cited theories in business and information systems offers conceptual guidance to model and analyze the interactions in complex networks (Houy et al., 2016). A network consisting of a set of nodes and a set of links that join pairs of nodes is said to be unipartite or one-mode network. On the other hand, when a network consists of two different sets of nodes and a set of links where each link joins nodes in different sets, the network is referred to as bipartite or two-mode network (Gupta and Kumar, 2016; Huang and Gao, 2014).

Previously, researchers have made persistent efforts to investigate and infer modular patterns in complex networks. In the context of social networks modules have been studied theoretically as homophily, one of the underlying tenets of social network theory (McPherson et al., 2001). Homophily is the tendency of individuals to mingle with other individuals of their own kind. This tendency may be induced by preference such as gender and ethnicity or by constraints such as private or public organization (Borgatti and Foster, 2003). For example, two banks may belong to the same group if they belong to same (private or public) sector. In bipartite networks, the intuition behind modules can be developed by considering a set of nodes as banks in a banking system and another set of nodes as the firms where a bank-firm link exists if the firm has borrowed from a bank. Two banks are similar in terms of their credit relationships if they have provided loans to same firms (Gupta and Kumar 2021). Similarly in an event-participant bipartite network, individuals who participate in similar events are more likely to be associated with each other (Davis et al., 2009). Thus, common neighborhoods on one side of the bipartite network reflects the nodes belonging to a same cohesive subgroup on the other side and vice-versa. Previously, identification of modules in bipartite networks have been used for various applications such as mapping ontologies (Fonseca, 2003) and analyzing users and content in social media (Grujic et al., 2009). Moreover, investigation of cohesive subgroups in complex networks has multifarious applications such as modelling of contagion (Agarwal et al., 2012), marketing and product development (Landherr et al., 2010).

3. Methodology

3.1 Background to DSR

DSR is rooted in the seminal literature of Herbert Simon's semi 'The Sciences of the Artificial' (Simon, 1969). Interest in DSR has been growing across disciplines, notably engineering, computer science, and information systems (Baskerville, 2008). DSR is a 'paradigm' (Iivari, 2007) grounded in 'discovery-through-design' (Baskerville, 2008). DSR is "a lens or set of synthetic and analytical techniques and perspectives (complementing positivist, interpretive, and critical perspective) for performing research (Vaishnavi and Kuechler, 2004, p.1). DSR is about understanding and improving the search among potential components to construct an 'artifact' that is intended to solve a 'real world' problem (Baskerville, 2008). In this context, an artifact is broadly defined as constructs (e.g., the conceptual vocabulary and symbols of a domain), models (e.g., propositions or statements expressing relationships between constructs), methods (e.g., algorithms or a set of steps used to perform a task: how-to knowledge), and instantiations (e.g., the operationalisation of constructs, models, and methods) (Vaishnavi and Kuechler, 2004; March and Smith, 1995; Hevner et al., 2004).

In contrast to design practice (routine design), a 'knowledge using activity' (e.g., the application of existing knowledge to organisational problems), DSR is a 'knowledge producing activity' that addresses important unsolved problems in unique or innovative ways or solved problems in more effective ways (March and Smith,1995; Hevner et al., 2004). Although the iterations between design (development) and evaluate (experiment) is a significant difference between DSR and the theory-driven 'behavioural science' (Kuechler and Vaishnavi, 2008), both approaches share a common environment (e.g., people, organisations, and technology) (Silver et al., 1995). A paradigm difference between design science and behavioral science is that the former is 'problem understanding' while the latter in 'problem understanding' (Niehaves and Stahl, 2006). As mentioned previously, we adopt the DSR model proposed by Peffers et al., (2007) which is explained in the next section.

3.2 Process model adopted in this DSR project

We adopt the six-step process model (see Table 1) proposed by Peffers et al., (2007) as it is the mostly widely cited model (vom Brocke et al., 2020) and although it is presented in a nominally sequential order, it is iterative in practice. In addition, there are four possible entry points for research, namely, (i) problem-centered approach (i.e., if the research idea resulted from observation of the problem or from suggested future research in a paper from a prior project), (ii) objective-centered approach (i.e., by-product of consulting experiences whereby client expectations were not met), (iii) design and development-centered approach (i.e., existence of an artifact that has not yet been formally thought through as a solution for the explicit problem domain in which it will be used), and (iv) observing a solution (i.e., observing a practical solution that worked and the researchers working backwards to apply rigor to the process retroactively). The entry point for this research is the design and development-centered approach.

Table 1. A six-step process for design science research

#	Step	Description (Peffers et al., 2007)
1	*Problem identification and motivation*	Define the specific research problem and justify the value of a solution. Justifying the value of a solution is important as it (i) motivates the researcher and the audience of the research to pursue the solution and to accept the results and (ii) helps to understand the reasoning associated with the researcher's understanding of the problem.
2	*Define the objectives for a solution*	Infer the objectives of a solution from the problem definition and knowledge of what is feasible. The objectives can be quantitative (e.g., terms in which a desirable solution would be better than existing ones) or qualitative (e.g., a description of how a new artifact is expected to support solutions to problems not hitherto addressed).
3	*Design and development*	Create the actual artifact by determining its functionality and architecture. In DSR, an artifact can include constructs, models, methods, or instantiations.
4	*Demonstration*	Demonstrate the utility of the artifact to solve the problem. This could involve its use in experimentation, simulation, a case study, proof, or other ap- propriate activity.
5	*Evaluation*	Observe and measure how well the artifact supports a solution to the problem. At the end of this activity the researchers can decide whether to iterate back to step 3 to try to improve the effectiveness of the artifact or to continue to communication and leave further improvement to

		subsequent projects. The nature of the research venue may dictate whether such iteration is feasible or not.
6	*Communication*	Communicate the problem and its importance, the artifact, its utility and novelty, the rigor of its design, and its effectiveness to researchers and other relevant audiences (e.g., practitioners).

In this context of this study, we describe how each step as per the Peffer's et al., (2004) model aligns the theoretical elements of social network analysis.

Problem identification and motivation: We address the problem of how to discover patterns of interaction in a social network based on big data. According to Polites and Watson (2009), common objectives of a social network analysis include:

- Information flow analysis – to determine the direction and strength of information flows through the network, such as information that is passed from one actor to other actors within the network.
- Evaluation of actor prominence - determines the most influential actors within a network.
- Hierarchical clustering – used to identify cliques whose members are fully or almost fully connected such as groups of actors that highly communicate with each other.
- Block modeling and – aims at discovering the key links between different subgroups in the network such as actors that serve as information brokers across groups or subgroups.
- Calculation of structural equivalence measures – aims at discovering network members with similar characteristics such as actors that correlate thus can be considered alternatives for each other.

Design and development: The design of a solution involves creating an artifact. According to March (1995), constructs, models, methods, and instantiations are considered as the main artifactual types. Constructs refers to the "language" developed to capture the problem and its conceptual solution. Models use this language to represent problems and solutions. Methods describe processes which provide guidance on how to solve problems. Instantiations are problem specific aggregates of constructs, models, and methods. At this stage, an artifacts desired functionality and its architecture is determined as a prelude to the creation of the artifact. The resources required for moving from the objectives of a solution to the design of a solution includes the knowledge of theory that links the objectives to the solution (Peffers et al., 2006). In social network analysis, the design of a solution is governed by social network theory. The key network concepts that organize research on network effects are centrality, cohesion, and structural equivalence (Liu et., 2017).

Demonstration: During the demonstration phase, the effectiveness of the artifact to solve one or more instances of the problem is illustrated through experimentation, simulation, proof of concept or through other accepted means. The illustration of the efficacy of a solution can also be in the form of a case study using a prototype (Fisher, 2007; Geerts and Wang, 2007). The knowledge base required at the demonstration stage is that of how to use the artifact to solve the identified instance of the problem (Fisher, 2007).

Evaluation: The evaluation phase measures how well the artifact supports a solution to the problem. It involves comparing the observed results from the use of the artifact during demonstration to the objectives of the solution. During the evaluation phase, the utility, quality and efficacy of a design artifact should be demonstrated by executing evaluation metrics (Arnott and Pervan, 2012). The metrics are useful in establishing the performance of the new artifact. Metrics for evaluation may be based on the artifact's functionality, quantitative performance measures, satisfaction surveys, clients' feedback, and simulations. In social network analysis, evaluation metrics include contingency heuristics.

Communication: This phase entails communicating the problem and its importance, the artifact, its utility, novelty, and the rigor of its design. The communication is aimed at showing the effectiveness of the artifact to researchers as well as technology-oriented and management-oriented audiences. Effective communication of design artifacts requires the knowledge of the disciplinary culture. In this phase, we underscore how the artifact developed in this study can be applied by novice users (e.g., students, practitioners) to identify cohesiveness amongst actors in a social network, for example, credit relationships amongst banks and firms.

4. Case Study: Credit Networks amongst Indian Companies and Banks

In financial systems, interactions arising due to credit relationships could potentially lead to systemic risk (Gupta and Kumar, 2021). These systems can be modelled as bipartite networks consisting of two heterogeneous interacting agents (nodes) connected by credit relationships (links), as in bank-firm credit network. The analysis of real credit networks reveals that these networks have the characteristics of complex social networks i.e., high clustering coefficient, power-law degree distribution, and modular structure (De Masi et al., 2011). Once a bankruptcy occurs at a particular node in the network, it may promulgate wider in the network, leading to systemic consequences (Gupta and Kumar, 2021). For instance, several business houses in the hospitality, aviation, energy sector and fitness center such as Virgin Atlantic, Gold’s Gym, Avianca, CMX Cinemas and Apex Parks went bankrupt during the COVID-19 pandemic following a forecast of 35 % increase in global insolvency index by Euler Hermes (a credit insurance company) during the June’ 2020- June’2022 period[1]. Similarly, the 2008 global financial crisis leading to the insolvency of investment banks such as Lehmann Brothers, Merrill Lynch, and Bear Stearns exposed the entwined nature of financial systems. Another insolvency proceeding was initiated in 2020 owing to Reliance Capital’s default of INR 1417 crore to Yes Bank, a private sector bank in India. This resulted in several other Indian banks such as State Bank of India, HDFC bank, ICICI bank, Axis bank, and Kotak Mahindra bank investing several crores in the bank while acquiring stakes in the bank. The effect of Yes Bank’s collapse had a contagion effect across the country with stock market indices falling sharply and growth in credit rates. Several similar bankruptcies have occurred in the past in different parts of the world including impairment of Japan’s banking system in 1992 and Greek bank in 2010 and the effect of these bankruptcies has imbued throughout the respective country or even to other countries. Due to these disastrous incidents, policymakers and governments have put in immense effort to unravel the hidden risks in complex financial systems. These credit relationships between banks and firms garner interest margin as profit for banks and fuel business growth of firms. The multiple borrowing relationships hedge the companies against liquidation risk. On the other side, multiple lending relationships insure banks against a firm’s risk of failure. However, the propensity to form multiple or single relationships varies with internal and external conditions (De Masi et al., 2011). On the flip side, insolvencies whether be in banks or firms reduce the risk appetite of lenders and lead to an increase in lending rates. As the insolvencies can have contagious effect, detection of modules is an effective decision support system for credit risk assessment in a financial system. This case study uses the data from annual reports of 20 heavily indebted companies in India to map the credit bipartite network of these companies and their bankers. Subsequently, modules of banks are identified using the concepts discussed in review of social network analysis theory.

4.1 Data Collection

[1] https://www.firstpost.com/india/insolvency-cases-have-gone-up-substantially-in-covid-hit-corporate-world-but-india-can-heave-a-sigh-of-relief-10176331.html

The data on the credit relationships between companies and banks in India is not available in an organized form at a single source. Therefore, we collected the data manually from annual reports of Indian companies for the financial year 2020– 21. We first shortlisted the high debt non-financial companies from moneycontrol.com which is a more than two decades old financial portal in India. This shortlisting process resulted in 20 companies. The industry type, total debt, and number of bankers for each company are shown in Table 2.

Table 2: Details of companies used in creation of credit network

	Company	**Industry Type**	**Debt**	**No. of Bankers**
1	Reliance	Conglomerate	1,97,403.00	17
2	NTPC	Power – Generation & Distribution	1,63,799.35	17
3	Power Grid Corp	Power – Generation & Distribution	1,45,415.99	8
4	ONGC	Oil Drilling and Exploration	77,065.12	1
5	IOC	Refineries	72,740.20	2
6	JSW Steel	Steel – Large	49,215.00	9
7	Indiabulls Housing	Housing Finance	34,136.17	28
8	HPCL	Refineries	33,003.40	9
9	Adani Ports	Infrastructure - General	31,570.43	18
10	BPCL	Refineries	31,314.82	9
11	NHPC	Power – Generation & Distribution	28,947.90	22
12	SAIL	Steel – Large	27,176.05	21
13	Jindal Steel	Steel	24,099.53	23
14	Alok Industries	Textiles	22,770.20	3
15	CESC	Power – Generation & Distribution	11,332.38	22
16	Future Retail	Retail	5,360.11	14
17	IRB Infra	Infrastructure - General	5,213.51	19
18	NLC India	Power - Generation & Distribution	13,365.62	5
19	Oil India	Oil Drilling and Exploration	15,398.83	4
20	Can Fin Homes	Housing Finance	5,552.62	1

As companies have borrowed from multiple banks, there are 56 banks in the dataset. We create a bipartite network such that each bank is linked to the companies to which it loaned money (see Table 3). This resulted in 20 nodes of companies on one side of bipartite network connected to 56 banks on the other side.

Table 3: Description of credit network

Characteristics	**Value**
Number of companies	20
Number of banks	56
Number of links	252
Average degree of companies	12.6
Average degree of banks	4.5

4.2 Analysis and Results

We applied the bipartite clustering approach to cluster the set of bank nodes in the bipartite credit network. The experimental results reveal six modules consisting of 4,9,13,6,14, and 10 banks as shown in Table 4. It is interesting to observe that the non-Indian origin banks fall into module 1 and module 6 except for Karnataka bank and IDFC First bank.

Table 4: Modules of banks in credit network

Module 1	**Module 2**	**Module 3**	**Module 4**	**Module 5**	**Module 6**
ANZ Bank	Bank of India	Canara Bank	Axis Bank	Bank of Baroda	Bank of America
Catholic Syrian Bank	Union Bank of India	Axis Finance Bank	SBI Bank	Punjab National Bank	Barclays Bank
Karnataka Bank	Standard Chartered Bank	Aditya Birla	HDFC Bank	India Overseas Bank	DZ Bank
Shinhan Bank	ICICI Bank	IDFC Bank	IndusInd Bank	Central Bank of India	Germany Export-import bank
	BNP Paribas	Jammu & Kashmir Bank	Federal Bank	IDBI Bank	Hamburg Commercial Bank
	CitiBank	Bank of Maharashtra	PFCL Bank	Indian Bank	IDFC First
	Deutsche	EXIM		DBS Bank	Mizuho Bank
	Credit Agricole	Punjab & Sind Bank		Cooperative Bank	MUFG Bank
	Hong Kong and Shanghai	Kotak Mahindra Bank		UCO Bank	JP Morgan
		AU Small Finance		IIFCL	Sumitomo Mitsui Bank
		RBL Bank		Yes Bank	
		United Overseas Bank		IFCI	
		SBI Life Insurance		Union Bank	
				India Infra	

Most of the private sector banks are identified as module 2. Modules 3,4, and 5 are mainly composed of public sector banks. This modular organization clearly indicates the extent of interdependency among Indian banks arising due to lending to companies. Module 3 and 5 with 13 and 14 banks each are the most critical modules for the Indian banking system and if a firm defaults to any of the banks in these clusters, adverse effects will spread to 13 banks (module 3) or 14 banks (module 5).

5. Discussion, Implications and Limitations

From the outset, the aim of this DSR project was "to design a process framework for the effective application of social network analysis to improve the outputs of big data research projects". Drawing on contemporary literature, we frame contributions of this study. A design science contribution must be 'interesting' to the research community (Gregor and Hevner, 2013), as well as be valued and accepted by the research community through its publication (Viashnavi and Kuechler, 2004). By adopting a DSR approach, we ensure that this study produces an interesting framework that will be of value and accepted by both researchers and practitioners. Further, the utility (e.g., practical knowledge) of the proposed framework provides a '*proof of value added*' (Davis, 2005).

Gregor and Hevner (2013, p. 345) propose a knowledge contribution framework for DSR which consists of four quadrants, namely, *Improvement* (develop new solutions for known problems, and *Invention* (Invent new solutions for new problems), and *Routine Design* (Apply known solutions for known problems) which would rarely be accepted as a research contribution (Vaishnavi and Keuchler, 2004), and *Exaptation* **(**Extend known solutions to new problems (e.g., adopt solutions from other fields). For *improvement, invention, and exaptation* to be considered as a significant research contribution, "it must be judged as significant with respect to the current state of the knowledge in the research area and be considered interesting" (Vaishnavi and Keuchler, 2004, p. 17). We believe the contribution of the proposed framework falls under the 'invention' quadrant as it provides a new solution to a new problem.

The proposed artifact also presents a theoretical contribution, as its construction is a special case of predictive theory that provides a prescription which when acted upon, causes an artifact of a certain kind to come into being (Gregor, 2006). The proposed process framework is an artifact aimed at actualizing the activities involved in social network analysis. Further, in an artefactual contribution, originality and novelty refer to the introduction of a particular artefact (Ågerfalk and Karlsson, 2021). The aim of validating the process framework using social network data of banks and firms within the Indian context provides evidence of 'satisficeability' of the artefactual contribution (Simon, 1969). Since the entry point of this research was at the 'design and development' phase, a rich description of the design process leading to the creation of the artefact and its instantiation using a case study of the credit networks of banks and firms in India provided empirical contribution. An empirical contribution captures data, measurements, observations, or descriptions regarding the artifact (Ågerfalk and Karlsson, 2021).

We acknowledge two limitations of this study, which also offer directions for future research. First, the study is based on a single case of module discovery problem within an umbrella of social network analysis which by nature, limits generalizability (Yin, 2009). Second, the proposed framework has been developed based on the context of the social networks of credit networks of banks and firms in India, which is a highly regulated industry. Future research could test the applicability of framework in different contexts such as in non-regulated environments (e.g., tourism, education).

The proposed process framework (see Figure 2) for conducting social network analysis using big data provides a clear formal structure in a form of predefined activities such that the underlying processes are mapped to the domain application and knowledge outputs. In accordance with the principles of DSR process model (Dresch et al., 2015; Peffers et al., 2006, 2018; Gupta and Tiwari, 2021), the proposed framework first formally introduces the module discovery problem to impart familiarity with the importance and relevance of module discovery through real life examples. In the second step, an understanding of the various aspects of problem such as credit risk, insolvency, modularity in networks is provided. The third step deals with the explanation of the origin and scientific advances related to credit risk assessment and solution of module discovery problem. The fourth step puts forward multiple classes of problems associated with module discovery such as module discovery in unweighted and unipartite networks (Kumar, Gupta and Bhasker, 2017; Gupta and Deodhar, 2021), module discovery in weighted and unipartite networks (Gupta and Tiwari, 2021) module discovery in bipartite networks (Gupta and Kumar, 2021) are highlighted in this step. The fifth step brings forth the concepts related to binarization, transformation, and one mode projection for module discovery. The sixth step pertains to demonstration of the superior approach through its implementation and heuristic such as similarity involved herein. Once the working of solution is explained, in the seventh step, the concept of contingency heuristic (modularity) is discussed. Subsequently, the advantages of bipartite module discovery for credit risk assessment should be brought forward. Finally, how the similar approaches of bipartite modelling and module discovery could be used for other context and application domains should be highlighted.

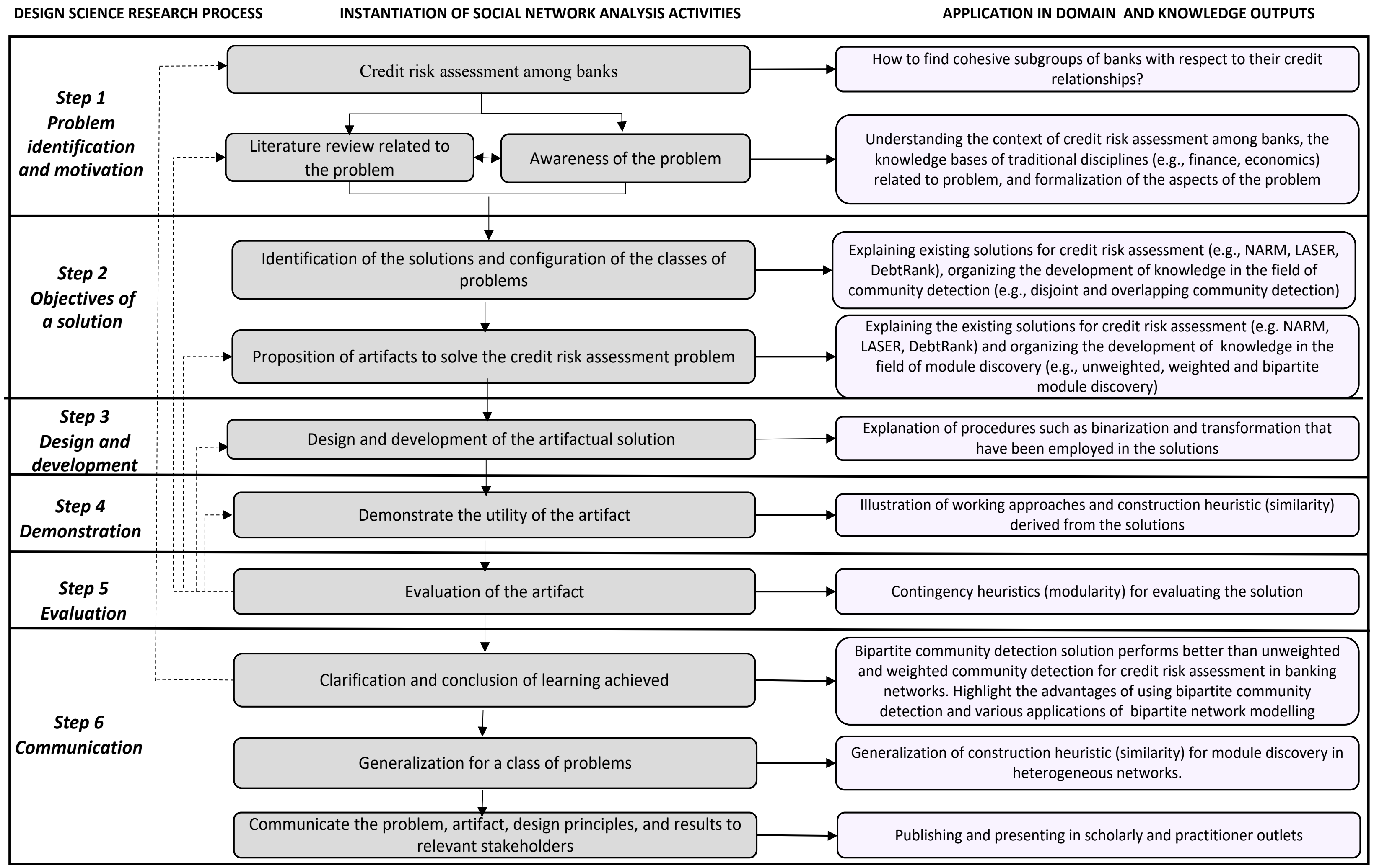

Figure 2 Process framework for conducting social network analysis using big data

6. Conclusion

This chapter provided a brief overview of the value of big data analytics and social network analysis that requires mathematical and computational techniques that can be used to understand the dynamics of complex real-world and artificial networks. The chapter also highlighted the importance of understanding the processes required to leverage big data to create business value through data-driven decisions. The chapter then describes a process framework that to guide novice users to effectively apply social network analysis and improve the outputs of big data research projects. This chapter therefore provides some interesting insights and opportunities for research and practice of social network analysis.

References

Ahmed, W., Vidal-Alaball, J., Downing, J. and Seguí, F.L. (2020). COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *Journal of Medical Internet Research*, *22*(5), p.e19458.

Ågerfalk, P. J., and Karlsson, F. (2021). Theoretical, Empirical, and Artefactual Contributions in Information Systems Research: Implications Implied*. In N. R. Hassan & L. P. Willcocks (Eds.), *Advancing Information Systems Theories: Rationale and Processes* (pp. 53–73). Cham: Springer International Publishing.

Arnott, D., and Pervan, G. (2012). Design Science in Decision Support Systems Research: An Assessment using the Hevner, March, Park, and Ram Guidelines. *Journal of the Association for Information Systems*, *13*(11). https://doi.org/10.17705/1jais.00315

Albery, G.F., Kirkpatrick, L., Firth, J.A. and Bansal, S. (2021). Unifying spatial and social network analysis in disease ecology. *Journal of Animal Ecology*, *90*(1), pp.45-61.

Albert, R., and Barabási, A.-L. (2002). Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1), 47.

Baskerville, R. (2008). What design science is not. *European Journal of Information Systems,* 17(5), 441–443. https://doi.org/10.1057/ejis.2008.45

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. (2006). Complex networks: Structure and dynamics. Physics Reports, 424(4–5), 175–308.

Barabási, A. L. (2014). Network science. Cambridge university press.

Barabási, A.-L., and Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509–512.

Bargigli, L., Di Iasio, G., Infante, L., Lillo, F., and Pierobon, F. (2015). The multiplex structure of interbank networks. *Quantitative Finance*, *15*(4), 673-691.

Bazzaz Abkenar, S., Haghi Kashani, M., Mahdipour, E., & Jameii, S. M. (2021). Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telematics and Informatics*, *57*, 101517. https://doi.org/10.1016/j.tele.2020.101517

Bella, E. di, Leporatti, L., and Maggino, F. (2018). Big Data and Social Indicators: Actual Trends and New Perspectives. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, *135*(3), 869–878. Retrieved from https://ideas.repec.org/a/spr/soinre/v135y2018i3d10.1007_s11205-016-1495-y.html

Bilecen, B., Gamper, M. and Lubbers, M.J. (2018). The missing link: Social network analysis in migration and transnationalism. *Social Networks*, *53*, pp.1-3.

Borgatti, S. P., and Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. Journal of Management, 29(6), 991–1013.

Chang, V. (2018). A proposed social network analysis platform for big data analytics. *Technological Forecasting and Social Change*, *130*, 57-68.

Choi, T.M., Wallace, S. W. and Wang, Y. (2018) "Big data analytics in operations management," *Production and Operations Management*, 27(10), 1868–1883.

Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. Physical Review E, 70(6), 066111

Cuba, W., Rodriguez-Martinez, A., Chavez, D. A., Caccioli, F. and Martinez-Jaramillo, S. (2021). A network characterization of the interbank exposures in Peru. *Latin American Journal of Central Banking*, *2*(3), 100035.

Davenport, T. H. (2018). From analytics to artificial intelligence. *Journal of Business Analytics*, *1*(2), 73–80. https://doi.org/10.1080/2573234X.2018.1543535

Davis, A., Gardner, B. B., and Gardner, M. R. (2009). Deep South: A social anthropological study of caste and class. University of South Carolina Press

Dennehy, D., Oredo, J., Spanaki, K., Despoudi, S., and Fitzgibbon, M. (2021). Supply chain resilience in mindful humanitarian aid organizations: The role of big data analytics. *International Journal of Operations & Production Management*, *41*(9), 1417–1441. https://doi.org/10.1108/IJOPM-12-2020-0871

Deng, Q., and Ji, S. (2018). A Review of Design Science Research in Information Systems: Concept, Process, Outcome, and Evaluation. *Pacific Asia Journal of the Association for Information Systems,* 10(1). https://doi.org/10.17705/1pais.10101

G. De Masi, Y. Fujiwara, M. Gallegati, B. Greenwald, J.E. Stiglitz, An analysis of the Japanese credit network, Evolutionary and Institutional Economics Review 7 (2011) 209–232. https://doi.org/10.14441/eier.7.209.

Dorogovtsev, S. N., and Mendes, J. F. (2002). Evolution of networks. Advances in physics, 51(4), 1079-1187

Dresch, A., Lacerda, D.P., & Antunes, J.A.V. (2015). Design science research. In Design science research (pp. 67–102). Springer.

Fonseca, Y. C. F. (2003). A bipartite graph co-clustering approach to ontology mapping. In Proceedings of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data. Colocated with the Second International Semantic Web Conference (ISWC-03).

Fortunato, S. (2010). Community detection in graphs. Physics Reports, 486(3), 75–174

Fossa Wamba, S., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., and Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., and Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics,* 165, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031

Fisher, I. E. (2007). A prototype system for temporal reconstruction of financial accounting standards. International Journal of Accounting Information Systems, 8(3), 139–164. https://doi.org/10.1016/j.accinf.2007.07.001

Gandomi, A., and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Garton, L., Haythornthwaite, C., and Wellman, B. (1999). Studying On-Line Social Networks. In Doing Internet Research: Critical Issues and Methods for Examining the Net (pp. 75–106). Thousand Oaks: SAGE Publications, Inc. https://doi.org/10.4135/9781452231471

Ghasemaghaei, M., Ebrahimi, S., and Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, 27(1), 101–113.

Geerts, G. L., and Wang, H. J. (2007). The Timeless Way of Building REA Enterprise Systems. *Journal of Emerging Technologies in Accounting*, 4(1), 161–182. https://doi.org/10.2308/jeta.2007.4.1.161

Gregor, S. (2006). The Nature of Theory in Information Systems. *Management Information Systems Quarterly*, *30*(3). Retrieved from https://aisel.aisnet.org/misq/vol3

Gregor, S., and Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2), 337–355. Retrieved from https://www.jstor.org/stable/43825912

Grover, V., Chiang, R.H., Liang, T.P. and Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), pp.388-423.

Grover, P., and Kar, A. K. (2017). Big data analytics: A review on theoretical contributions and tools used in literature. Global Journal of Flexible Systems Management, 18(3), 203–229.

Grujic, J., Mitrovic, M., and Tadic, B. (2009, July). Mixing patterns and communities on bipartite graphs on web-based social interactions. In 16th International Conference on Digital Signal Processing, 2009 (pp. 1-8). IEEE.

Gupta, S., & Deodhar, S. (2021). Understanding digitally enabled complex networks: a plural granulation based hybrid community detection approach. *Information Technology & People*, (ahead-of-print).

Gupta, S., & Kumar, P. (2016, August). Community detection in heterogenous networks using incremental seed expansion. In *2016 international conference on data science and engineering (ICDSE)* (pp. 1-5). IEEE.

Gupta, S., and Kumar, P. (2021). A constrained agglomerative clustering approach for unipartite and bipartite networks with application to credit networks. *Information Sciences*, *557*, 332-354.

Gupta, S., & Tiwari, A. A. (2021). A design-based pedagogical framework for developing computational thinking skills. Journal of Decision Systems, 1-18.

Hevner, A., March, S., Park, J., and Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28(1). Retrieved from https://aisel.aisnet.org/misq/vol28/iss1/6

Hevner, A., March, S., Park, J., & Ram, S. (2008). Design Science in Information Systems Research. *Management Information Systems Quarterly*, *28*(1). Retrieved from https://aisel.aisnet.org/misq/vol28/iss1/6

Hirschheim, R., 2021. The attack on understanding: How big data and theory have led us astray: A comment on Gary Smith's Data Mining Fool's Gold. *Journal of Information Technology*, 36(2), pp.176-183.

Holmström J, Ketokivi, M., and Hameri AP. (2009) Bridging practice and theory: a design science approach. Decision Sciences 40: 65-87.

Houy, C., Frank, J., Niesen, T., Fettke, P., and Loos, P. (2016). On the usage of theories in the field of Wirtschaftsinformatik: A quantitative literature analysis.

Huang, Y., and Gao, X. (2014). Clustering on heterogeneous networks. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4(3), 213–233.

Hu, J., and Zhang, Y. (2017). Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. *Scientometrics*, *112*(1), 91-109.

Kim, J. and Hastak, M. (2018). Social network analysis: Characteristics of online social networks after a disaster. *International Journal of Information Management*, 38(1), pp.86-96.

Kuechler, B. and Vaishnavi, V. (2008). On theory development in design science research: anatomy of a research project. *European Journal of Information Systems*, 17(5), pp.489-504.

Kumar, P., Gupta, S., & Bhasker, B. (2017). An upper approximation based community detection algorithm for complex networks. Decision Support Systems, 96, 103-118.

Landherr, A., Friedl, B., and Heidemann, J. (2010). A Critical Review of Centrality Measures in Social Networks. Business & Information Systems Engineering, 2(6), 371–385.

Laumann, E. O., Galaskiewicz, J., and Marsden, P. V. (1978). Community Structure as Interorganizational Linkages. Annual Review of Sociology, 4, 455–484. Retrieved from https://www.jstor.org/stable/2945978

Liu, W., Sidhu, A., Beacom, A., and Valente, T. (2017). Social Network Theory. In Rossler, P., Cynthia, H., and Liesbet, van Z. (Eds.), Social Networks and Macrosocial Change. Wiley. https://doi.org/10.1002/9781118783764.wbieme0092

March, S. T., and Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2.

Marshall, J., and McKay, P. (2005). A Review of Design Science in Information Systems. *ACIS 2005 Proceedings*. Retrieved from https://aisel.aisnet.org/acis2005/5

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 415–444

Mikalef, P., Pappas, I.O., Krogstie, J. and Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management*, 16(3), pp.547-578.

Modgil, S., Gupta, S., Sivarajah, U. and Bhushan, B. (2021). Big data-enabled large-scale group decision making for circular economy: An emerging market context. *Technological Forecasting and Social Change*, 166, p.120607.

Morana, S. and Maedche, A. (2020). Designing Process Guidance Systems the Case of IT Service Management. In Design Science Research. Cases (pp. 177-203). Springer, Cham.

Niehaves, B. and Bernd, C. Stahl. (2006). Criticality, epistemology and behaviour vs. Design – information systems research across different sets of paradigms. *ECIS 2006 Proceedings*. 166. https://aisel.aisnet.org/ecis2006/166

Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23), 8577–8582.

Newman, M. E. (2003). The structure and function of complex networks. SIAM Review, 45(2), 167–256

Nunamaker, J. F., Chen, M., and Purdin, T. D. M. (1990). Systems Development in Information Systems Research. *Journal of Management Information Systems,* 7(3), 89–106. https://doi.org/10.1080/07421222.1990.11517898

Rosvall, M. (2006). Information horizons in a complex world. Doctoral Dissertation, Department of Physics, Umeå University, Umeå.

Nunamaker, J. F., Chen, M., and Purdin, T. D. (1991). Systems development in information systems research. *Journal of Management Information Systems*, 7(3), 89–106.

Otte, E., and Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, *28*(6), 441–453. https://doi.org/10.1177/016555150202800601

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77.

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., and Bragge, J. (2006). *The Design Science Research Process: A Model for Producing and Presenting Information Systems Research*. Retrieved from https://jyx.jyu.fi/handle/123456789/63435

Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: Introduction to the special issue on exemplars and criteria for applicable design science research. European Journal of Information Systems, 27(2), 129–139. https://doi.org/10.1080/0960085X.2018.1458066

Polites, G., and Watson, R. (2009). Using Social Network Analysis to Analyze Relationships Among IS Journals. *Journal of the Association for Information Systems*, *10*(8). https://doi.org/10.17705/1jais.00206

Saqr, M., Fors, U., Tedre, M. and Nouri, J. (2018). How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *PloS one*, *13*(3), p.e0194777.

Seddon, J.J. and Currie, W.L. (2017). A model for unpacking big data analytics in high-frequency trading. Journal of Business Research, 70, pp.300-307.

Silva, T. C., de Souza, S. R. S., and Tabak, B. M. (2016). Network structure analysis of the Brazilian interbank market. *Emerging Markets Review*, *26*, 130-152.

Soramäki, K., Bech, M. L., Arnold, J., Glass, R. J., and Beyeler, W. E. (2007). The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, *379*(1), 317-333.

Simon, H. (1969). The Sciences of the Artificial. Cambridge, MA: MIT Press.

Simon, H. (1996). The Science of the Artificial (Third Edition). Cambridge, MA: MIT Press.

Travers, J., and Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 425–443.

Vaishnavi, V. and Kuechler, W. (2004/21). “Design Science Research in Information Systems” January 20, 2004 (updated in 2017 and 2019 by Vaishnavi, V. and Stacey, P.); last updated November 24, 2021. URL: http://www.desrist.org/design-research-in-information-systems/.

Valeri, M. and Baggio, R. (2020). Social network analysis: Organizational implications in tourism management. *International Journal of Organizational Analysis*.

Venable, J. R., Pries-Heje, J., and Baskerville, R. L. (2016). FEDS: A framework for evaluation in design science research. European Journal Information Systems, 25(1), 77–89.

Velichety, S., and Ram, S. (2021). Finding a Needle in the Haystack: Recommending Online Communities on Social Media Platforms Using Network and Design Science. *Journal of the Association for Information Systems*, *22*(5), 1285–1310. https://doi.org/10.17705/1jais.00694

vom Brocke, J., Hevner, A. and Maedche, A. (2020). Introduction to Design Science Research. In Design Science Research. Cases (pp. 1-13). Springer, Cham.

Walls, J. G., Widmeyer, G. R., and El Sawy, O. A. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, *3*(1), 36–59. https://doi.org/10.1287/isre.3.1.36

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of "small-world"networks. Nature, 393(6684), 440–442

Watson, H. (2014). Tutorial: Big Data Analytics: Concepts, Technologies, and Applications. *Communications of the Association for Information Systems*, 34(1). https://doi.org/10.17705/1CAIS.03462

Wagner, G., Prester, J., and Schryen, G. (2021). Exploring the Scientific Impact of Information Systems Design Science Research. *Communications of the Association for Information Systems*, *48*(1). https://doi.org/10.17705/1CAIS.04837

Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Winter, R. (2008). Design science research in Europe. *European Journal of Information Systems*, *17*(5), 470–475. https://doi.org/10.1057/ejis.2008.44

Yin, R. K. (2009). Case study research: Design and methods. Sage Publications, Inc.

Zipf. G.K. (1949). Human Behavior and the Principle of Least Effort: An Introduction to human Ecology. Ravenio Books