How Ill Is Your IT Portfolio?

Measuring Criticality in IT Portfolios Using Epidemiology

Completed Research Paper

Florian Guggenmos
Project Group Business & Information Systems Engineering of the Fraunhofer FIT, University Bayreuth
Wittelsbacherring 10
95444 Bayreuth, Germany
florian.guggenmos@fit.fraunhofer.de

Peter Hofmann
Project Group Business & Information Systems Engineering of the Fraunhofer FIT, University Bayreuth
Wittelsbacherring 10
95444 Bayreuth, Germany
peter.hofmann@fit.fraunhofer.de

Gilbert Fridgen
FIM Research Center, University of Bayreuth
Project Group Business & Information Systems Engineering of the Fraunhofer FIT
Wittelsbacherring 10
95444 Bayreuth, Germany
gilbert.fridgen@fim-rc.de

Abstract

IT project portfolios, consisting of IT projects, also interact with the entire IT landscape. In case of a failure of only one element, existing dependencies can lead to a cascade failure, which can cause high losses. Despite the present effects of systemic risk, research into IT portfolio management lacks suitable methods to quantitatively assess systemic risk. We follow the design science research paradigm to develop and evaluate our 'on track' or 'in difficulty' (TD) method by applying the SI model, representing a recognized network diffusion model in epidemiology, in an IT portfolio context. We evaluate our method using a real-world dataset. We introduce a criticality measure for diffusion models in IT portfolios and compare the TD method's results and the alpha centrality to human judgment as a benchmark. From our evaluation, we conclude that the TD method outperforms alpha centrality and is a suitable risk measure in IT portfolio management.

Keywords: systemic risk, cascade failure, IT project portfolio, portfolio management, epidemiology, design science

Introduction

IT projects play key roles in companies, and not only for the evaluation and implementation of new technologies. However, a large number of complex IT projects confront decision-makers with significant challenges (Neumeier and Wolf 2017). The Standish Group's *Chaos Report* (2015) emphasized the frequency of IT project errors and the importance of project management. According to this study, 71% of all IT projects have been challenging or have failed. Also, Flyvbjerg and Budzier (2011) stated that about 16% of all IT projects exceed their budget by 200%. A Radar Group survey (2012) concluded that the excessive complexity resulting from dependencies between different projects in an IT project portfolio (ITP) is one reason for these budget overruns. In the real world, we cannot consider ITPs as isolated, since they are embedded in the organization's IT landscape. Thus, an ITP also interacts with different IT infrastructures such as legacy systems, IT services, or applications.

One approach to consider these dependencies is to consider IT projects as elements of complex IT project networks (Beer et al. 2015; Guo et al. 2019; Neumeier et al. 2018; Radszuwill and Fridgen 2017; Wehrmann et al. 2006; Wolf 2015). IT projects in an ITP are interdependent in different ways: Among others, projects can use the same infrastructure, require limited resources, or depend on previous projects' results. Thus, the question arises how a delay or failure in one or more projects will affect other projects in the same ITP. These networks' interconnectedness induces systemic risk and can trigger cascade failures (Beer et al. 2015). Thus, the failure of one project can lead to additional failures in other projects. So, the failure of one project may collapse an entire ITP, resulting in substantial financial losses or even bankruptcy (Beer et al. 2015; Wolf 2015). This phenomenon of significant losses owing to a failed project is known as a "black swan" (Taleb 2007). In practice, experts usually analyze the risks of ITPs according to their personal experience. However, owing to a large number of projects and the resulting large number of dependencies, real-world ITPs are usually overly complex. Thus, experts can no longer consider all the aspects, and they may disregard aspects that may be decisive. It is only possible for persons to interpret ITPs in simplified cases. In practice, this requires decision-supporting algorithms and methods for applicabillity.

Previous research in IT portfolio management (ITPM) considers ITPs as complex networks. But ITPM mainly focuses on direct dependencies and thus neglects the effects of systemic risk owing to indirect dependencies. Thus, popular portfolio risk measures, such as portfolio variance, which is based on pairwise dependencies, are unsuitable in the context of ITP. The same reasoning applies to popular centrality measures such as degree centrality, closeness centrality, or betweenness centrality (Wolf 2015). Researchers have only recently begun to also investigate indirect dependencies (Beer et al. 2015; Guo et al. 2019; Neumeier et al. 2018; Radszuwill and Fridgen 2017; Wolf 2015). Nonetheless, there are very few quantitative approaches in the literature to consider indirect dependencies in ITPs. For instance, Wolf (2015) suggested using alpha centrality, which considers many key aspects and therefore provides significantly better results than the abovementioned approaches. Although alpha centrality considers indirect dependencies, it does not consider how fast a failure spreads in a network. But propagation speed is a key factor in determining the criticality of single projects or the entire ITP. Further, to our best knowledge, all articles use artificial, self-generated data for ITPs to evaluate approaches. Thus, we conclude that there is a need to develop new approaches or to transfer existing approaches from other fields that also fulfil propagation speed as well as all other requirements for application in the context of ITPM, as stated by Wolf (2015), and to evaluate new approaches using real-world data. Popular approaches that quantify systemic risk in complex networks while considering propagation speed are so-called network diffusion models. Other research areas such as social networks (e.g., Watts and Dodds 2007), supply networks (e.g., Tang et al. 2016), the energy network (e.g., Crucitti et al. 2004), or epidemiology (e.g., Kermack and McKendrick 1927) already use network diffusion models to quantify cascade effects. For instance, Watts and Dodds (2007) investigated how quickly opinions spread in social networks. Also, in epidemiology, propagation speed is a key factor, since a disease may be even more critical if it spreads rapidly (Brockmann and Helbing 2013).

Addressing the identified research gap, we analyzed systemic risks in ITP via network diffusion models, using a real-world dataset that describes an ITP consisting of IT projects and their dependencies as well their interactions with the entire IT landscape. Since Guo et al. (2019) have already adopted a model from supply chain management, we will extend the research by transferring a model from epidemiology. We focus on epidemiology owing to its inherent characteristics in the context of risks and dependencies. Owing to the assumption of negative effects of dependencies in ITP, a large number of dependencies may cause high systemic risks. Analogously, in epidemiology, diseases spread faster in the case of a higher population density or a more networked population. In contrast, in the context of energy systems, most algorithms focus on the power grid, and dependencies mainly indicate power cables. Thus, such algorithms regard a network as less critical in the case of more power cables (edges), since others can compensate for the failure of another one. In epidemiology, multiple approaches simulate the spreads of diseases and epidemics in society. These algorithms predict among others the risks of different pathogens. Thus, if there is a viral outbreak, these algorithms help to draw conclusions about a disease's origin and to predict its further spread. Probably the best-known model for simulating the spread of disease is the so-called susceptible-infected model (SI model) of Kermack and McKendrick (1927). We investigate whether and how we can use the SI model in ITPM to predict the spread of systemic risks and thus to improve ITPM. Our research question is as follows:

What can we learn from transferring and applying the SI model from epidemiology to complex IT portfolios?

To answer this question, we developed a method, following the design science research (DSR) paradigm (Gregor and Hevner 2013; Hevner 2007; March and Smith 1995). We conducted our research according to the DSR cycle (relevance, design, and rigor) introduced by Hevner (2007), and structured our article according to the DSR publication schema of Gregor and Hevner (2013). We contribute to the literature by developing and evaluating a method called the *on track or in difficulty method* (the TD method), which applies the SI cascade model in an ITPM context to simulate the damage caused by the failure of individual IT projects (we focus on entire IT projects, neglecting single project tasks) or the dependent elements of the IT landscape. Besides offering nascent design theory (Gregor and Hevner 2013), an instantiation of our method using real-world ITP data allowed us to rigorously evaluate our method. To evaluate the method, we compared the method's results to the results of human judgment as well as alpha centrality, which is a suitable systemic risk measure in ITPM (Wolf 2015). Thus, we contribute to the discourse on cascading effects in ITP in a way that is relevant to both research and practice.

Theoretical foundations and literature review

A real-world ITP may consist of many IT projects and different dependency types between them. However, an ITP also depends on and affects other technical components such as legacy systems, databases, and applications within the firm's IT infrastructure. The academic literature uses graph theory to represent ITPs (c.f. Beer et al. 2015; Wolf 2015). Different graphs represent different aspects of such ITPs. A broad representation in practice are the so-called Gantt charts, which mainly represent temporal (mainly inter-temporal) dependencies of individual work packages or subprojects. However, project management uses this representation to organize different tasks within a project (e.g., milestone planning) rather than to manage systemic risks. In the literature, there are, among others, approaches that deal with the modeling of ITP via Petri nets and the description of risk spreads (Fridgen et al. 2015; Gilbert Fridgen, Christian Stepanek, Thomas Wolf 2012). In contrast, Beer et al. (2015), Ellinas et al. (2015), Guo et al. (2019), Radszuwill and Fridgen (2017), and Wolf (2015) considered an ITP as a complex network consisting of nodes (projects) and edges (project dependencies). A *complex network* describes a specific network type (represented by graphs) that is neither random nor regular (Strogatz 2001). Instead, these networks are characterized by a unique distribution of edges. Two of the most represented complex networks in the literature are scale-free networks (Barabási and Albert 1999; Barabási and Bonabeau 2003) and the so-called small world networks (Watts and Strogatz 1998). Since the distribution of the edges follows a power law, few nodes have very many edges (hubs), and many other nodes only have very few edges. By comparison, nodes in random networks have a very balanced number of edges. Depending on the application context, these edges can represent different dependencies (e.g., acquaintances in social networks or power lines in a power grid).

In a real-world ITP, there are several dependency types. While some articles in the literature focused on certain dependency types (c.f. Lee and Kim 2001; Santhanam and Kyparisis 1996; Tillquist et al. 2002; Zuluaga et al. 2007), others presented a framework of different dependencies (c.f. Wehrmann et al. 2006; Zimmermann 2008). The literature often distinguishes between intra-temporal and inter-temporal dependencies. Thus, intra-temporal dependencies describe dependencies within one step of time (both projects run simultaneously), and inter-temporal dependencies describe dependencies between different time steps (one project ended before the other starts). Beer et al. (2015), for instance, provided a more detailed subdivision of intra-temporal and inter-temporal dependencies.

Owing to these dependencies, a failure in a single project (e.g., project 1) can also affect project 2, which for instance depends on project 1's results. A *cascade failure* describes the phenomenon that this failure not only affects project 2, but also project 3. Project 3 does not directly depend on project 1, but it does depend on project 2. So project 3 does indirectly depend on project 1. The literature describes these as *transitive dependencies* (Beer et al. 2015; Wolf 2015) and the spread of the cascade failure as systemic risk (Battiston et al. 2012; Haldane and May 2011; Helbing 2013; Neumeier et al. 2018).

In the case of cascading failure, a failure in only one element of the network can cause a domino effect, owing to interconnectedness (Buldyrev et al. 2010; Crucitti et al. 2004; Guo et al. 2019; Motter and Lai 2002; Wang and Rong 2009). This is a well-known problem in various fields. Especially in the context of the financial sector, *systemic risk* became very popular during and after the financial crisis since 2007. The bankruptcy of Lehman Brothers caused a cascade of bankruptcies of other financial institutions across the world. The research into cascade failure has focused on critical infrastructures, such as power grids (Crucitti et al. 2004; Motter and Lai 2002; Wang and Rong 2009), opinion dynamics in social networks (Castellano et al. 2009; Galam 2002; Holley and Ligget 1975; Kempe et al. 2003; Sznjad-Weron and Sznajd 2000; Watts and Dodds 2007), and epidemiology (Brockmann and Helbing 2013; Kermack and McKendrick 1927; Pastor-Satorras and Vespignani 2001). But there are also generic cascade models for complex networks and for interdependent networks (networks of networks) (Buldyrev et al. 2010; Gao et al. 2011; Gao et al. 2012). Research into project management or ITPM has not fully covered systemic risk's effects.

During the past few years, research in the ITPM field has increased significantly. For instance, Beer et al. (2015) introduced an approach to quantify project benefits and risk considering transitive dependencies. Further, Wolf (2015) provided an overview over different centrality measures, concluding that alpha centrality is a suitable risk measure not only in social networks, but also in ITPM. Ellinas et al. (2015), Ellinas et al. (2016), Ellinas (2018), and Ellinas (2019) investigated general aspects of systemic risk. Guo et al. (2019) introduced a cascade failure model to simulate the spread of failures between different tasks within a single project. This cascade failure algorithm is based on Goh et al.'s (2001) model, which describes load distribution in general scale-free networks (Barabási and Albert 1999). While all these articles contribute to the study of systemic risk in the context of ITPs, all these approaches and their evaluation assume self-generated data. Further, while they all considered different aspects of systemic risk (e.g., indirect dependencies), they neglected propagation speed, among others. In epidemiology, for instance Brockmann and Helbing (2013) noted that this is a significant factor in analyzing cascade effects. To demonstrate this, Brockmann and Helbing (2013) illustrated the geographical spread of diseases as a circular wave using so-called effective distance, which re-interprets the geographical distance between two cities and indicates how fast a disease could spread owing to the existing flight connections between them.

Research approach

Design process

To rigorously develop and evaluate our method, we followed the DSR paradigm (Gregor and Hevner 2013; Hevner 2007; March and Smith 1995). As an artifact in DSR, methods describe how to perform goal-directed activities (March and Smith 1995). For method construction, we relied on the DSR cycles (relevance, design, and rigor) introduced by Hevner (2007). Having discussed problem relevance in the introduction section, we will now describe how we constructed our method. In the design process, we began with Kermack and McKendrick's (1927) SI model instead of starting from scratch. In the next iterations of the design process, we relied on the academic literature as justificatory knowledge (Gregor and Hevner 2013). We provide an overview over the existing knowledge in academic literature in the theoretical foundations and literature review section. For this purpose, we specified the following search query as a basis for a structured literature search: *(IT project OR project portfolio OR project management) AND (systemic risk OR cascade failure OR complex networks)*. We then conducted a keyword-based search (title, abstract, and keywords) in ScienceDirect as well as another keyword-based search (title and abstract) in the AIS Electronic Library. Further, we limited our search to research articles by including academic journals and academic conferences. This procedure resulted in 1,212 papers from ScienceDirect and 193 from the AIS Electronic Library. After eliminating duplicates, we identified 96 papers as potentially relevant for our study based on their titles and abstracts. After examining these papers in detail, including analyzing the full text, we eliminated 79 papers and finally considered 17 as our core literature. We then identified further relevant literature using forward and backward search of citations in the set of primary sources, as recommended by Webster and Watson (2002), considering these papers in our literature section. This approach allowed us to develop a method that is grounded on existing knowledge and that supports project management.

Evaluation strategy

Besides creating an artifact, it is crucial to demonstrate its inherent utility (Hevner 2007; March and Smith 1995). To evaluate the method's effectiveness, we developed an instantiation of our method to calculate the method's results using real-world data (step 1) and compared them to human judgment (step 2) as well as the results of alpha centrality (step 3), an established systemic risk measure in the context of ITPs (Wolf 2015). Since we used a simplified real-world ITP (an excerpt from the entire ITP), consisting of only 10 elements (eight IT projects and two IT landscape components), we assumed that the experts deliver the 'correct' solution and regard human judgment as the benchmark for the TD method and alpha centrality. To our best knowledge, this is the first time that alpha centrality has been applied and evaluated on real-world data. To compare the rankings of steps 1, 2, and 3, we used Kendall's τ (Kendall and Stuart 1945) to apply a non-parametric rank correlation test. Following this evaluation strategy, we evaluated whether the developed method delivers comparable results to alpha centrality and human judgment. To collect human judgment, we first presented a graph of a real-world ITP (Figure 4) to several probands and asked them to create a descending ranking of the criticalities (with the most critical element first). Second, we asked the probands to explain how they came to their conclusions and which criteria they used to create their rankings. As probands, we selected nine experts from our institutional network with project management experience. Instead of setting a time limit, we asked them to take as much time as they needed to properly assess the ITP's elements.

Artifact description

Modeling IT portfolios as complex networks

Similar to Beer et al. (2015), we regard ITP as complex networks. We represent projects as nodes and dependencies as edges. Since dependencies can be both unidirectional and bidirectional, the displayed edges are directed. Also, we modeled the dependency strengths as weights of the directed edges. We limited the intensity to the interval $[0; 1]$. An intensity of zero indicates a non-existent dependency (no edge). The closer the value is to one, the stronger the dependency is.

Using epidemiology models in the portfolio management context

Compared to Guo et al. (2019), we wanted to improve ITPM using a cascade algorithm that is typically used in epidemiology. Kermack and McKendrick (1927) introduced three basic models to simulate the spread of diseases in a population: the SI cascade model, the SIS cascade model, and the SIR cascade model. These models provide the basis for many modern approaches in this context. All three models are based on the same considerations, but differ in the considered states.

The SI cascade model considers two states: susceptible (S) and infected (I). A person can only reach these two states in sequence. For instance, this means that a person who is currently healthy is susceptible to illness (state S). This person can become ill (state I) owing to a spontaneous mutation or external influences. Epidemiology describes this person as patient zero and indicates the outbreak of a disease. It is also possible that a susceptible person comes into contact with a person who is already ill. In this case, the disease is transmitted through personal contact (i.e. acquaintances). Persons in state I can only infect susceptible persons if they come into contact with them. For the SI model, Kermack and McKendrick (1927) assumed that an infected person cannot become susceptible again. The so-called infection rate (β) indicates the possibility of transitioning from state S to state I (S → I). The infection rate is constant over time and for all people, and depends on the specific disease. The cascading process only ends when there are no more susceptible persons left, which implies that everyone is ill. Figure 1 provides a schematic representation of the cascading effect in epidemiology for four exemplary time steps t based on the SI cascade model, with susceptible persons in white and infected persons in grey.

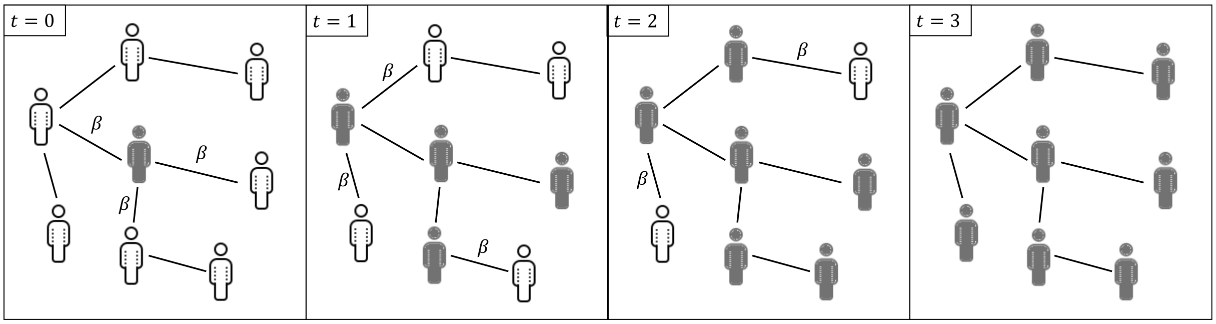

Figure 1. The spread of an illness based on the SI model

In contrast to the SI cascade model, the SIS cascade model provides the chance that infected persons become healthy again. Thus, the SIS cascade model's results will not determine. Further, the SIR cascade model considers an additional state. The state *recovered* (R) indicates that a person has recovered from state I and is now immune to re-infection. In addition, there are further models based on the approaches of Kermack and McKendrick (1927). These models include the SEIR cascade model and the SEIS cascade model (Aron and Schwartz 1984), which considers an exposed incubation time (state E), as well as the SWIR cascade model (c.f. Lee et al. 2017), which considers an additional state, weakness (state W).

In the context of an ITP, we assume that a project is either on track or in difficulty. Consequently SI model and SIS model remain. We decided to transfer the SI cascade model instead of the SIS cascade model, for two main reasons. First, the consideration of the possibility to transfer from state I back to state S would lead to a dynamic network without clear results, since the results depend on the observed time period. Since the use of epidemiology in the context of ITPM is new in research, we decided to keep the model as simple as possible and therefore to consider determining algorithms. Second, in our opinion, the consideration of the SIS cascade model extends the SI model. So, we regard the investigation of the SIS cascade model as a possible extension of our work that is subject to further research. We regard elements of ITPs as *on track* (state T) or *in difficulty* (state D). Further, we regard persons as projects and personal contact as a dependency. The infection rate indicates these dependencies' intensities.

We concluded that the basic SI cascade model does not consider all aspects of our ITP, requiring further adaptations. First, we needed the TD model to consider directed dependencies. In the context of epidemiology, it is evident that a person cannot come into contact with another person, but not vice versa. However, in the context of ITP, it is possible that project A depends on project B without B depending on A. Thus, we re-interpreted the SI cascade model's dependencies as directed dependencies. We therefore replaced undirected edges with directed ones (arcs). Further, dependencies' intensities in a real-world ITP are not constant for all dependencies. To consider this in the TD method, we also re-interpreted the infection rate β. In our TD method, the value of β is based on a dependency's intensity. Further, in contrast to alpha centrality, which is based on the adjacency matrix A, the TD method can more easily consider multiple dependencies between two elements. Analogous to the SI cascade model or alpha centrality, the TD model does not consider the propagation speed. Thus, we had to first define a key figure (criticality measure) that interprets the TD model's results such that propagation speed is considered.

Measuring criticality

Finally, we needed the TD method, which applies the TD model, to deliver a criticality score for each element in the ITP. The criticality score indicates the extent to which a failure of an element would affect the rest of the ITP. In epidemiology, this criticality describes how many people become ill if a particular person becomes ill initially. However, in the context of ITPs, we do not interpret an element's criticality solely based on the ultimate components in difficulty. Since criticality also depends on the ability to react, it is essential to incorporate how quickly the failure of the initial element in difficulty spreads to other elements. Thus, spread depends on the ITP structure. We therefore considered elements of ITP that require shorter reaction times as more critical than elements that slowly infect other elements. To consider this, we built our criticality risk measure (CM_i) for element i as follows: First, we measured for each time step t how many elements changed from state T to state D (see equation 1).

$$\Delta elements_{i,t}^{D} = elements_{i,t}^{D} - elements_{i,t-1}^{D} \tag{1}$$

We can sum up all $\Delta elements_{i,t}^{D}$ to consider how many elements get into difficulty owing to the initial failure. However, in case we only sum up $\Delta elements_{i,t}^{D}$, we would neglect how many elements get into difficulty in which time step t. Thus, we would not consider how quickly a failure spreads. Figure 2 illustrates that the number of elements in difficulty would be an unsuitable criticality measure. For instance, for $t = 15$, an initial failure of element A (the black triangle) and element B (the white triangle) both caused a cascade failure and finally infected 10 elements. Thus, we would have to rank both elements as equally critical. Looking at the course of the two curves, we see that element A infects elements quicker than element B. Since we consider elements that require shorter reaction times as more critical than elements that slowly infect other elements, we based our criticality measure on geometric series and weighted each $\Delta elements_{i,t}^{D}$ based on t (see equation 2).

$$CM_i = 1 + \sum_{t=1}^{n} \frac{\Delta elements_{i,t}^{D}}{t^{\gamma}} \tag{2}$$

The weighting of the new elements in state D by parameter t ensured that the criticality measures rated the slower propagation of infection as less critical than faster propagation of infection. Thus, our criticality measure considers propagation speed. The additional parameter γ is a further adjusting parameter. With this parameter, the user can determine how strongly the criticality measure considers the propagation speed. We limited the parameter γ to the interval of $[0;\infty)$. For the lower bound ($\gamma = 0$), the criticality measure does not weight and therefore neglects the propagation speed. By increasing γ, the user increases propagation speed's influence.

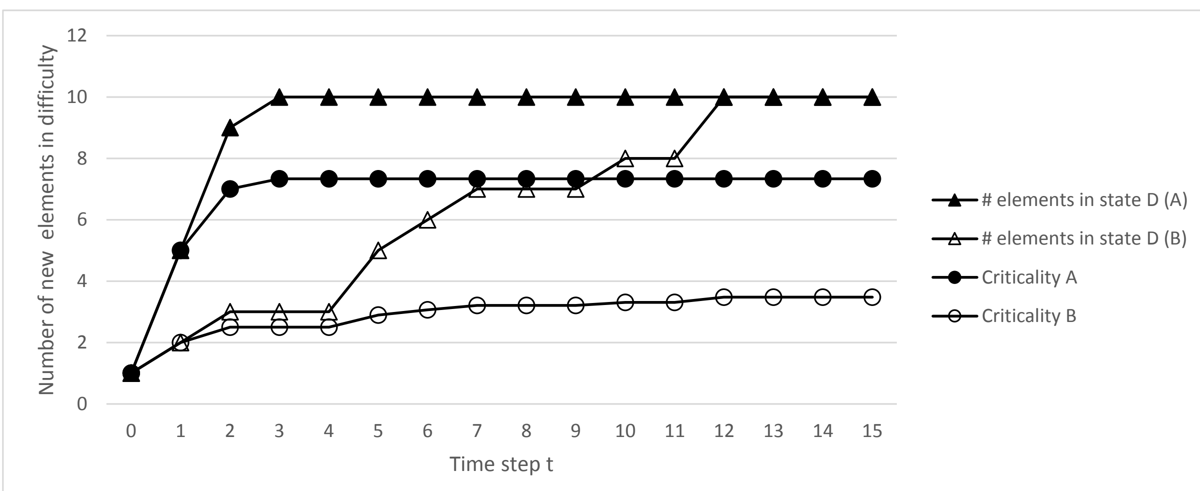

Figure 2. Exemplary application of the criticality measure

Figure 2 illustrates the criticality measure (based on equation 2) using black circles for element A and white circles for element B, compared to an unweighted version (black and white triangle), which corresponds to the total number of elements in state D resulting from state D of element A (# elements in state D (A)) and element B (# elements in state D (B)). We concluded that the weighted criticality measure ranked A as more critical than B.

Evaluation

Introducing the dataset

We used a real-world dataset representing an ITP and its dependencies to the IT landscape. By preparing the dataset, we deleted incomplete data rows. Finally, the prepared dataset contained 23 IT projects representing the ITP, four applications, five legacy systems, and one customer service representing the IT landscape. The dataset distinguishes between unidirectional and bidirectional dependencies, and contains statements about the expected effects of the dependency, and rates the expected effects of the dependency as low, medium, and high. Unfortunately, the dataset contains no project parameters such as costs or expected revenues. Table 1 shows an exemplary excerpt from the dataset and how we interpreted it in the context of complex networks. Moreover, we interpreted the data field expected effect as the dependency's intensity. To model the graph, we interpreted low as 0.25, medium as 0.5, and high as 0.75. Further, the dataset partially contained more than one dependency between two elements. This can be justified by the fact that the dataset is aggregated, for instance, at the project level. In this case, individual partial aspects of one project cause dependencies of different intensities on another project.

Element A	Element B	Direction	Expected effect	Interpretation
Project 1	Project 3	bidirectional	medium	$P1 \overset{0.5}{\longleftrightarrow} P3$
Project 8	Project 1	unidirectional	high	$P8 \overset{0.75}{\longrightarrow} P1$
Application 3	Project 12	unidirectional	high	$A3 \overset{0.75}{\longrightarrow} P12$

Table 1. Exemplary excerpt from the real-world dataset

Applying the TD method to the dataset

We implemented the TD model using Python. We used the library NDlib, which provides several functions to describe, simulate, and study network diffusion processes (Rossetti et al. 2018), including Kermack and McKendrick's (1927) SI cascade model. We used the code as the basis for the implementation of the TD method, which builds on the SI model. The TD method expects a total of three input parameters. The first input parameter G, is the ITP and the IT landscape modeled as a complex network (graph). We formalized this parameter using the NetworkX Python library (Hagberg et al. 2008). The graph G consists of n elements (nodes) and m dependencies (arcs). The second input parameter D, is an array of elements that indicates initial elements in difficulty (i.e. the origin of the cascade effect). This parameter's data type is also a class of the NetworkX library and contains one or more nodes of the graph. The third input parameter n, is an integer indicating the maximum time steps to run during one simulation. The result of this simulation is an array that indicates, for each element and time step t, the elements' states.

For our simulation, we defined the input parameters as follows: For parameter G, we formalized our real-world dataset and transformed it into a weighted digraph. Since we simulated the cascade effect in case of an initial failure of only one element, we parameterized D with only one element (e.g., $D = ['Project\ 1']$). Finally, we set $t_{max} = 100$. Preliminary calculations indicated that this value is more than sufficient for the used portfolio size in order to impede a premature termination of the algorithm. In each case, the algorithm stagnated for $t < 15$. To obtain statistically significant results for all projects, we performed the simulation of these cascade effects 1,000 times. The simulation outputs a $n \times t_{max}$ matrix. This matrix contains the state of each element in each time step. Based on the output, we calculated the number of elements in state D for every time step. Finally, we combined the output of all 1,000 runs into one matrix using the average. Then, we changed the initial element in difficulty (e.g., $D = ['Project\ 2']$) and started the algorithm again. We carried on until we had calculated the cascade effect for each element (i.e. n times).

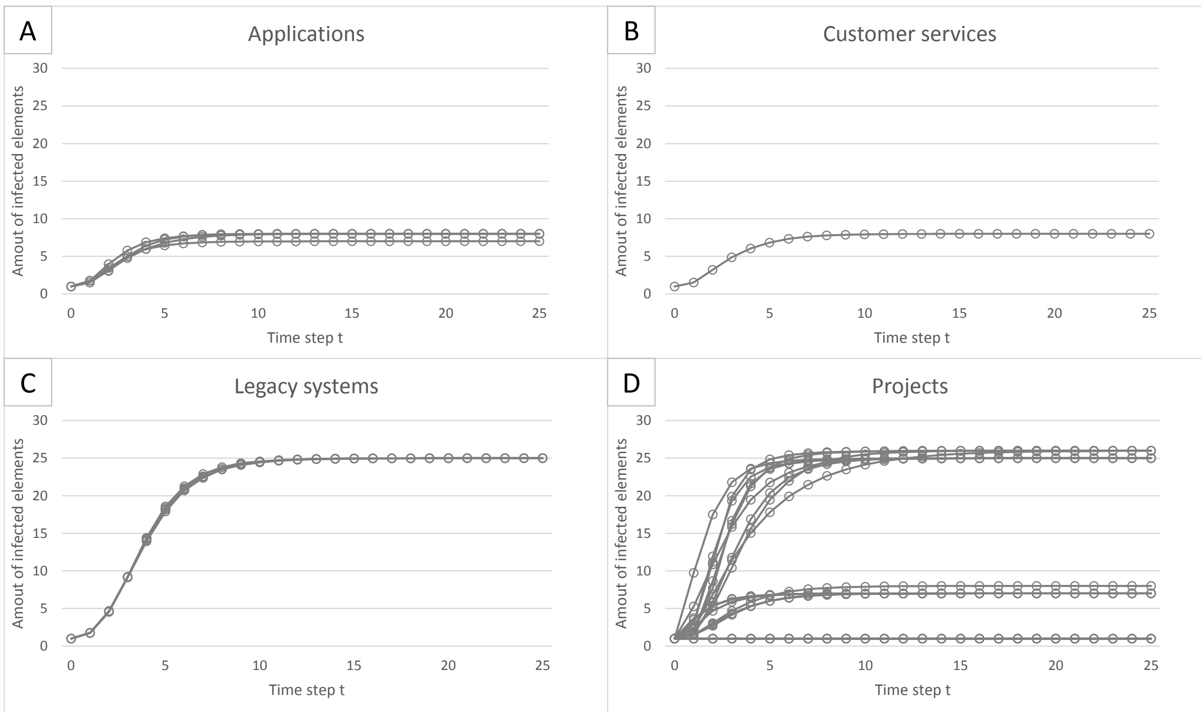

Figure 3. Applying the TD method to the real-world dataset

To illustrade possible results of an output of the TD model, we plotted graphs for different element types in the entire real-world ITP and the IT landscape in Figure 3. Graph A illustrates that an initial failure of an application causes lower systemic risk because it only affects 8 respectively 9 out of 33 elements. For graph B, we concluded that an initial failure in the only considered customer service only affected 8 out of 33 elements. Thus, its criticality is similar to the applications. Graph C illustrates that the initial failure of any legacy system affects 25 out of 33 elements. Graph C therefore confirms existing knowledge that any failure in a legacy system may cause significant damage to the ITP. Finally, graph D illustrates that the ITP contained several projects with low criticality, several with medium criticality, and several that are highly critical. Projects with no other elements depending on them are less critical. Therefore, the failure of one of these projects would not initiate a cascade effect. From the fact that we can see several medium and several highly critical projects, we concluded that the real-world ITP contains two main clusters. These clusters describe projects that strongly depend on one another but only less affect elements that are not part of the cluster. In ITPM, we know such constellations as programs.

Evaluating the method's effectiveness

We will now examine whether and how exactly the TD method's results represent the real criticality of the individual elements of an ITP. Since we considered it too complex for a person to sensibly rank a portfolio with 33 elements, we decided for our experimental study to select only a section from the real-world ITP. Figure 4 plots our reduced real-world ITP (containing eight IT projects), which interacts with two elements (one application and one legacy system) from the IT landscape, which we used for the evaluation.

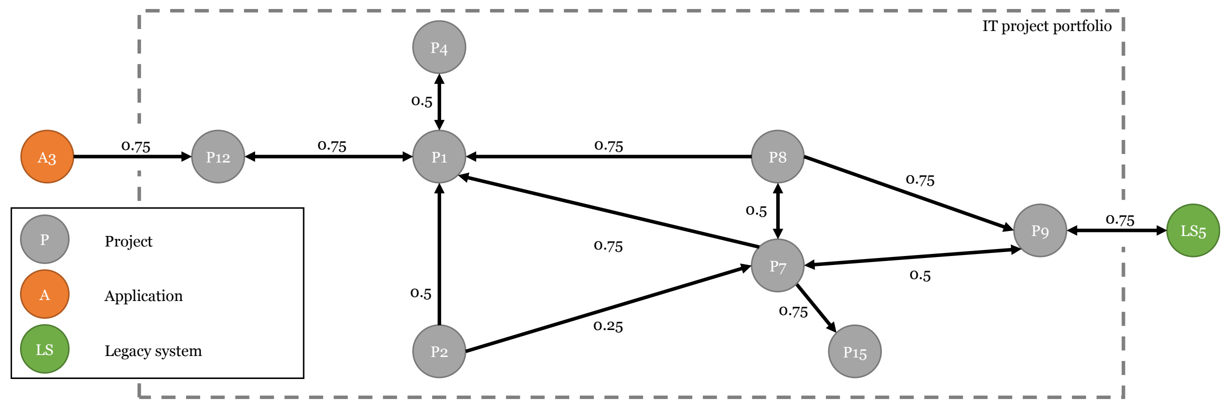

Figure 4. Network for evaluation

First, we computed the TD method and ranked all elements according to their criticality based on our criticality measure (equation 2). The choice of γ determines the ranking resulting from the TD method. So, to choose a suitable value for γ, we first performed a sensitivity analysis (see Figure 5). As illustrated in Figure 5, a lower value for γ (e.g., 0.1) led to a ranking that mostly neglects propagation speed and regards P2 as most critical. The resulting ranking will change at every point their two curves intersect. On the basis of this analysis, we decided on $\gamma = 1.0$ for this evaluation. We assumed $\gamma = 1.0$, since there is no intersection point of the nearby curves and we considered both aspects of criticality (propagation speed and total number of affected elements). We determined that the TD method rated P7 as the most critical (rank 1), followed by P8 (rank 2) and P2 (rank 3), and P4 (rank 9) and P15 (rank 10) as the least critical. As we can see in Figure 4, no other elements depended on P15 (i.e. no outgoing arcs). Thus, an initial failure in one element does not cause a cascading effect – only the element itself failed. Table 2 presents the ranking of criticalities for the TD method, illustrated by observing projects 2, 8, and 9.

Second, we presented our simplified graph (Figure 4) to nine probands from our institutional network with project management experience and asked them for their assessments of the individual elements' criticalities. We asked them in detail to create a descending ordinally scaled order of the elements in the simplified real-word ITP and IT landscape, based on their personal experience. Finally, we aggregated the individual probands' results into an overall order using the median for each element. The aggregated human judgments stated P7 as most critical, followed by P8. Analogous to the TD method, they also deemed P15 as least critical. To quantify the agreement of our probands' rankings, we used Kendall's coefficient of concordance, also known as Kendall's W (Kendall and Smith 1939). This coefficient measures the extent of agreement between rankings over n characteristics (elements) and m judges (probands) based on ordinal scaled data. The value of W ranged in the interval $[0; 1]$. Here, a value of $W = 1$ indicates a perfect match of the probands (all probands gave the same order). A value of $W = 0$ indicates that there was no correlation between the rankings. We used this approach to test the null hypothesis (H_0): *There is no agreement among the judges*. Further, we set the significance level to $\alpha = 0.001$. The analysis of the individual probands' rankings resulted in $W = 0.8243$ and a p-value of 0.0000. Since $0.0000 < 0.001$, we rejected H_0 and concluded that all probands agreed very well and delivered very similar rankings. Since we used a very simple graph for the evaluation, we regard the probands' results as the 'correct' order and used it as the benchmark for the TD method and alpha centrality.

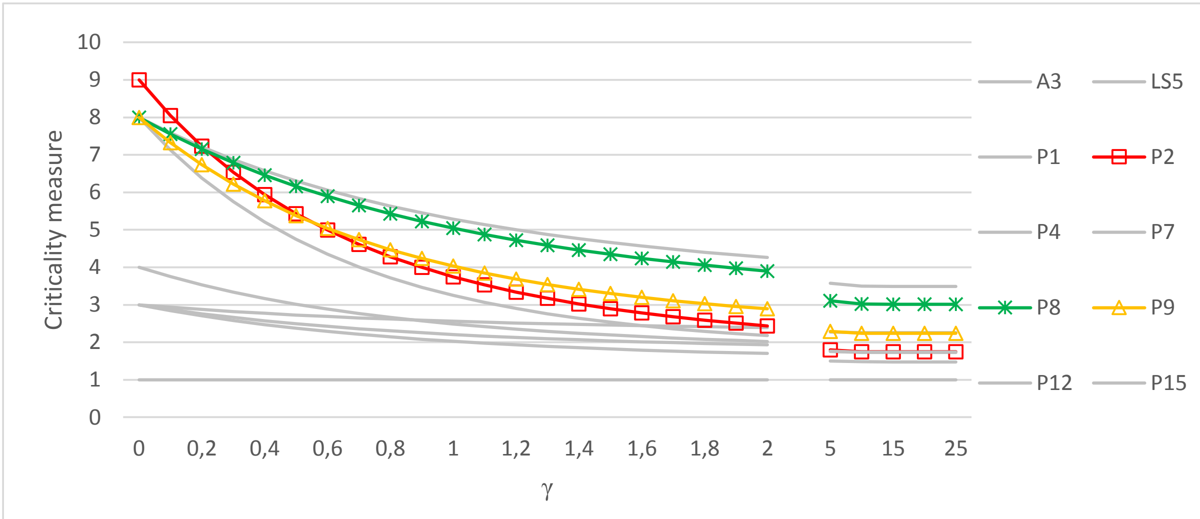

Figure 5. Sensitivity analysis for the parameter γ

Third, we compared the TD method's results to alpha centrality, a proven approach for measuring systemic risk in ITP (Wolf 2015). To run alpha centrality, we transferred the graph to a directed and non-binary adjacency matrix. Following Wolf (2015), we chose a high α and set $\alpha = 0.94$ ($\alpha_{max} = 1/\lambda_{max} \approx 0.9475$) and the vector e to $e = (1, 1, \dots, 1)^T$. Alpha centrality ranked P8 as the most critical, followed by P7 and P9, and P15 as the least critical.

Rank	TD method	Human judgment (median)	Alpha centrality
1	P7 (5.29)	P7 (1.00)	P8 (488.96)
2	P8 (5.04)	P8 (2.00)	P7 (425.35)
3	P9 (4.03)	P1 (3.50)	P9 (400.86)
4	P2 (3.75)	P9 (4.00)	LS5 (283.60)
5	LS5 (3.25)	P2 (5.00)	P2 (104.58)
6	P1 (2.56)	P12 (6.00)	P1 (7.71)
7	A3 (2.49)	LS5 (7.00)	P12 (6.44)
8	P12 (2.21)	A3 (8.00)	A3 (5.54)
9	P4 (2.03)	P4 (9.00)	P4 (4.62)
10	P15 (1.00)	P15 (10.00)	P15 (1.00)

Table 2. The results of the evaluation

We calculated Kendall's τ for the pairwise correlation to benchmark the results of the TD method and alpha centrality to the results of human judgment (see Table 3). We concluded that all results were significant for at least $\alpha \leq 0.005$. Thus, we rejected H_0 (*There is no correlation between two ordinal scaled rankings*) for any cases. Overall, this implies that the TD method is, besides alpha centrality, also a suitable approach to consider systemic risk in the context of assessing criticality in ITPM. In our case, the TD method provided even better results than alpha centrality based on the comparison of Kendall's τ ($0.7778 > 0.7332$).

	TD method	Alpha centrality	Human judgment
TD method	1.0000**	0.8667**	0.7778*
Alpha centrality	0.8667**	1.0000**	0.7332*
Human judgment	0.7778*	0.7332*	1.0000**
Notes: * $p \leq 0.005$; ** $p \leq 0.001$.			

Table 3. Pairwise correlation using Kendall's τ

Discussion

The evaluation, especially the results of Kendall's τ, indicated that the TD method delivers almost identical results compared to recognized methods. The calculation of Kendall's τ proved that the TD method ($\tau_{TD;H} = 0.7778$) correlates best with the assessments of human judgment, which we consider to be the benchmark. However, there was also a significant agreement between the results of alpha centrality and human judgment ($\tau_{AC;H} = 0.7332$), confirming Wolf's results (2015). In both cases, we rejected H_0. However, there are slight differences between the methods.

First, we assumed that all probands correctly analyzed the graph we presented to them. We did not limit the probands' considerations by time or other restrictions. The results exclusively depended on their personal experience and preferences, for instance about whether they prioritize the total number of finally affected elements or the propagation speed. Many probands told us that it was challenging to consider cyclical dependencies, because these quickly increase the considerations' complexity. Thus, the probands usually either ignored cyclical dependencies or considered them in a very simplified way. At this point, alpha centrality and the TD method also differ. While alpha centrality considers cycles and thus assigns a higher centrality to the elements that are part of the cycle, the TD method neglects cycles owing to its binary consideration of state T. We regard this as an advantage of alpha centrality.

Second, the TD method can easily consider multiple different dependencies between two elements. These multiple dependencies would have strongly challenged the probands. Thus, we did not use elements with multiple interdependencies for the evaluation. However, since these elements can occur in real ITPs, we will briefly discuss this case. Also, several dependencies between two elements can have different intensities. Multiple dependencies are not a problem when calculating the TD method, since it considers each edge individually and calculates whether or not the state passes over this edge to the second element. For persons, such considerations are very complex and may lead to the approximation of values (e.g., the mean value over all dependencies). Alpha centrality, as introduced by Bonacich and Lloyd (2001), cannot deal with such a fact, because the adjacency matrix can only consider one value between two elements. Here, we can also use approximate values for simplification. We regard this as an advantage of the TD method.

Third, a further difference between the three approaches lies in the results' interpretability. Since the results of human judgment are ordinally scaled, we can only interpret them in terms of the sequence. Alpha centrality also delivers ordinal scaled results. The TD model initially only provides an overview over which and how many elements were in state T at time steps t. To be able to interpret these results, we introduced the criticality measure (CM_i), which assigns each element a score depending on the number of the final affected elements and the propagation speed. We regard these values as cardinally scaled. Thus, we could both interpret the TD's criticality values in terms of the order, and we can make statements about how much more critical one element is compared to another. However, the quality of the cardinal scalability of the TD method partly depends on the choice of parameter α. We regard this as a significant advantage of the TD method.

Conclusion

Considering the high interconnectedness in a real-word ITP, clearly, modern ITPM must consider these dependencies for risk management, owing to emerging systemic risk. Although there are several approaches to quantifying systemic risk in other fields, most methods in research and practice have neglected these effects.

With this in mind, we used the design science approach of Hevner (Gregor and Hevner 2007) to transfer a method from another field to ITPM. We took Kermack and McKendrick's (1927) SI model from the field of epidemiology, re-interpreted it, and evaluated our so-called TD method using human judgment. Further, we compared the results of the TD method and alpha centrality, a suitable systemic risk measure in the context of ITPM (Wolf 2015), to human judgment. Based on Kendall's coefficient of concordance, introduced by Kendall and Smith (1939), we concluded that our probands agreed very well. Further, based on Kendall's τ, introduced by Kendall and Stuart (1945), we also conclude that the ranking of the TD method significantly agreed with human judgment and with alpha centrality. Thus, we regard the TD method as suitable to quantify systemic risk in the context of ITPM. In our case, the TD method even outperformed alpha centrality, based on Kendall's τ.

We have made three primary theoretical contributions. First, we have created a risk measure that indicates the criticality of real-world ITPs based on how many other elements depend on it and on how fast cascade failure will spread. Second, we have transferred the SI model from epidemiology to ITPM. Further, we re-interpreted the model so that it fits the properties of ITPs (the TD model). By introducing the TD method, we have extended the pool of available risk measures in ITP by a further approach that explicitly also considers systemic risk and is more flexible than other methods (e.g., alpha centrality). Third, to our best knowledge, we – for the first time – evaluated systemic risk measures (e.g., alpha centrality) in ITPM using real-world data. Thus, we have extended Wolf's (2015) results. We have also made managerial contributions. We have given project and portfolio managers a further decision tool to improve risk management in ITP.

Our paper also has limitations. First, we based the evaluation of the TD method on human judgment with only nine probands from our institutional network. To further expand our review, we suggest further concretizing the evaluation by designing it as a supervised laboratory experiment, so as to eliminate possible bias contained in our evaluation. In doing so, we would be able to evaluate the TD method's quality and to refine the choice of γ in our criticality measure. Second, the concretized evaluation (lab experiment) should be based on several graphs from different datasets and an increased number of probands. Third, further research should investigate the interpretability of the TD method's results. We regard our results as cardinally scaled. This must be investigated in detail. Fourth, we have merely re-interpreted the SI model (TD method). Future research should therefore focus on how to extend the TD method so as to consider further aspects of real-world ITPs. These include for instance the consideration of project-specific parameters (e.g., the expected benefit or the duration). Further, we propose examining which additional states should be considered in future research. Thus, if projects are in difficulty, they can finally heal and get on track again. Accordingly, we suggest further extending research to the field of dynamic network analysis.

References

Aron, J. L., and Schwartz, I. B. 1984. “Seasonality and period-doubling bifurcations in an epidemic model,” *Journal of Theoretical Biology* (110:4), pp. 665–679.

Barabási, A.-L., and Albert, R. 1999. “Emergence of Scaling in Random Networks,” *science* (286:5439), pp. 509–512.

Barabási, A.-L., and Bonabeau, E. 2003. “Scale-free networks,” *Scientific american* (288:5), pp. 60–69.

Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B., and Stiglitz, J. E. 2012. “Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk,” *Journal of economic dynamics and control* (36:8), pp. 1121–1141.

Beer, M., Wolf, T., and Zare Garizy, T. 2015. *Systemic Risk in IT Portfolios – An Integrated Quantification Approach*. presented at: 36th International Conference on Information Systems (ICIS), Fort Worth, USA.

Bonacich, P., and Lloyd, P. 2001. “Eigenvector-like measures of centrality for asymmetric relations,” *Social networks* (23:3), pp. 191–201.

Brockmann, D., and Helbing, D. 2013. “The hidden geometry of complex, network-driven contagion phenomena,” *science* (342:6164), pp. 1337–1342.

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. 2010. “Catastrophic cascade of failures in interdependent networks,” *Nature* (464:7291), p. 1025.

Castellano, C., Muñoz, M. A., and Pastor-Satorras, R. 2009. “Nonlinear q-voter model,” *Physical review. E, Statistical, nonlinear, and soft matter physics* (80:4 Pt 1), p. 41129.

Crucitti, P., Latora, V., and Marchiori, M. 2004. “Model for cascading failures in complex networks,” *Physical Review E* (69:4), p. 45104.

Ellinas, C. 2018. “Modelling indirect interactions during failure spreading in a project activity network,” *Scientific reports* (8:1), p. 4373.

Ellinas, C. 2019. “The domino effect: an empirical exposition of systemic risk across project networks,” *Production and Operations Management* (28:1), pp. 63–81.

Ellinas, C., Allan, N., Durugbo, C., and Johansson, A. 2015. “How Robust Is Your Project? From Local Failures to Global Catastrophes: A Complex Networks Approach to Project Systemic Risk,” *PloS one* (10:11), e0142469.

Ellinas, C., Allan, N., and Johansson, A. 2016. “Project systemic risk: Application examples of a network model,” *International Journal of Production Economics* (182), pp. 50–62.

Flyvbjerg, B., and Budzier, A. 2011. “Why Your IT Project May Be Riskier Than You Think,” *Harvard Business Review* (89), pp. 23–25.

Fridgen, G., Stepanek, C., and Wolf, T. 2015. “Investigation of exogenous shocks in complex supply networks – a modular Petri Net approach,” *International Journal of Production Research* (53:5), pp. 1387–1408.

Galam, S. 2002. “Minority opinion spreading in random geometry,” *The European Physical Journal B* (25:4), pp. 403–406.

Gao, J., Buldyrev, S. V., Havlin, S., and Stanley, H. E. 2011. “Robustness of a network of networks,” *Physical review letters* (107:19), p. 195701.

Gao, J., Buldyrev, S. V., Stanley, H. E., and Havlin, S. 2012. “Networks formed from interdependent networks,” *Nature physics* (8:1), p. 40.

Gilbert Fridgen, Christian Stepanek, Thomas Wolf 2012. *Simulating Exogenous Shocks in Complex Supply Networks Using Modular Stochastic Petri Nets*. Proceeding of the 20th European Conference on Information System (ECIS), Barcelona, Spain.

Goh, K.-I., Kahng, B., and Kim, D. 2001. "Universal behavior of load distribution in scale-free networks," *Physical review letters* (87:27), p. 278701.

Gregor, S., and Hevner, A. R. 2013. "Positioning and presenting design science research for maximum impact," *MIS Quarterly* , pp. 337–355.

Guo, N., Guo, P., Dong, H., Zhao, J., and Han, Q. 2019. "Modeling and analysis of cascading failures in projects: A complex network approach," *Computers & Industrial Engineering* (127), pp. 1–7.

Hagberg, A., Swart, P., and Chult, D. S. 2008. "Exploring Network Structure, Dynamics, and Function using NetworkX," in *7th Python in Science Conference (SciPy2008)*, G. Varoquaux, T. Vaught and J. Millman (eds.), Pasadena, CA, USA. August 21, pp. 11–15.

Haldane, A. G., and May, R. M. 2011. "Systemic risk in banking ecosystems," *Nature* (469:7330), p. 351.

Helbing, D. 2013. "Globally networked risks and how to respond," *Nature* (497:7447), p. 51.

Hevner, A. R. 2007. "A three cycle view of design science research," *Scandinavian journal of information systems* (19:2), p. 4.

Holley, R. A., and Ligget, T. M. 1975. "Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model," *The Annals of Probability* (3:4), pp. 643–663.

Kempe, D., Kleinberg, J., and Tardos, É. 2003. "Maximizing the spread of influence through a social network," *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM* , p. 137.

Kendall, M. G., and Smith, B. B. 1939. "The problem of m rankings," *Annals of mathematical statistics* .

Kendall, M. G., and Stuart, A. 1945. "The advanced theory of statistics," .

Kermack, W. O., and McKendrick, A. G. 1927. "A Contribution to the Mathematical Theory of Epidemics," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* (115:772), pp. 700–721.

Lee, D., Choi, W., Kertész, J., and Kahng, B. 2017. "Universal mechanism for hybrid percolation transitions," *Scientific reports* (7:1), p. 5723.

Lee, J. W., and Kim, S. H. 2001. "An integrated approach for interdependent information system project selection," *International Journal of Project Management* (19:2), pp. 111–118.

March, S. T., and Smith, G. F. 1995. "Design and natural science research on information technology," *Decision Support Systems* (15:4), pp. 251–266.

Motter, A. E., and Lai, Y.-C. 2002. "Cascade-based attacks on complex networks," *Physical Review E* (66:6), p. 65102.

Neumeier, A., Radszuwill, S., and Garizy, T. Z. 2018. "Modeling project criticality in IT project portfolios," *International Journal of Project Management* (36:6), pp. 833–844.

Neumeier, A., and Wolf, T. 2017. "Getting a Grip on IT Project Complexity – Concluding to Underlying Causes," *38th International Conference on Information Systems (ICIS), Seoul, South Korea* .

Pastor-Satorras, R., and Vespignani, A. 2001. "Epidemic spreading in scale-free networks," *Physical review letters* (86:14), p. 3200.

Radar Group 2012. "The Impact of Data Silos in IT Planning (White Paper)," .

Radszuwill, S., and Fridgen, G. 2017. "Forging a Double-Edged Sword: Resource Synergies and Dependencies in Complex IT Project Portfolios," *38th International Conference on Information Systems (ICIS), Seoul, South Korea* .

Rossetti, G., Milli, L., Rinzivillo, S., Sîrbu, A., Pedreschi, D., and Giannotti, F. 2018. "NDlib: a python library to model and analyze diffusion processes over complex networks," *International Journal of Data Science and Analytics* (5:1), pp. 61–79.

Santhanam, R., and Kyparisis, G. J. 1996. “A decision model for interdependent information system project selection,” *European Journal of Operational Research* (89:2), pp. 380–399.

Standish Group 2015. *Chaos Report 2015*. https://www.standishgroup.com/sample_research_files/CHAOSReport2015-Final.pdf.

Strogatz, S. H. 2001. “Exploring complex networks,” *Nature* (410:6825), p. 268.

Sznjad-Weron, K., and Sznajd, J. 2000. “Opinion evolution in closed community,” *International Journal of Modern Physics C* (11:06), pp. 1157–1165.

Taleb, N. N. 2007. *The black swan: The impact of the highly improbable:* Random house.

Tang, L., Jing, K., He, J., and Stanley, H. E. 2016. “Complex interdependent supply chain networks: Cascading failure and robustness,” *Physica A: Statistical Mechanics and its Applications* (443), pp. 58–69.

Tillquist, J., King, J. L., and Woo, C. 2002. “A representational scheme for analyzing information technology and organizational dependency,” *MIS Quarterly* , pp. 91–118.

Wang, J.-W., and Rong, L.-L. 2009. “Cascade-based attack vulnerability on the US power grid,” *Safety science* (47:10), pp. 1332–1336.

Watts, D. J., and Dodds, P. S. 2007. “Influentials, networks, and public opinion formation,” *Journal of consumer research* (34:4), pp. 441–458.

Watts, D. J., and Strogatz, S. H. 1998. “Collective dynamics of ‘small-world’networks,” *Nature* (393:6684), p. 440.

Webster, J., and Watson, R. T. 2002. “Analyzing the past to prepare for the future: Writing a literature review,” *MIS Quarterly* (26:2), pp. xiii–xxiii.

Wehrmann, A., Heinrich, B., and Seifert, F. 2006. “Quantitatives IT-Portfoliomanagement: Risiken von IT-Investitionen wertorientiert steuern,” *WIRTSCHAFTSINFORMATIK* (48).

Wolf, T. 2015. *Assessing the Criticality of IT Projects in a Portfolio Context using Centrality Measures*. 12th International Conference on Wirtschaftsinformatik (WI), Osnabrück, Germany.

Zimmermann, S. 2008. “IT-Portfoliomanagement–Ein Konzept zur Bewertung und Gestaltung von IT,” *Informatik-Spektrum* (31:5), pp. 460–468.

Zuluaga, A., Sefair, J. A., and Medaglia, A. L. (eds.) 2007. *Model for the selection and scheduling of interdependent projects,* IEEE.