Different networks for different purposes
A network science perspective on collaboration and communication in an engineering design project
Piccolo, Sebastiano A.; Lehmann, Sune; Maier, Anja M.

Published in:
Computers in Industry

Link to article, DOI:
10.1016/j.compind.2022.103745

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Piccolo, S. A., Lehmann, S., & Maier, A. M. (2022). Different networks for different purposes: A network science perspective on collaboration and communication in an engineering design project. Computers in Industry, 142, [103745]. https://doi.org/10.1016/j.compind.2022.103745

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Different networks for different purposes: A network science perspective on collaboration and communication in an engineering design project

Sebastiano A. Piccolo a,⁎, Sune Lehmann b, Anja M. Maier c,d

a University of Copenhagen, Denmark
b DTU Compute, Technical University of Denmark, Denmark
c DTU Management, Technical University of Denmark, Denmark
d Department of Design, Manufacturing and Engineering Management, University of Strathclyde, Scotland

A R T I C L E   I N F O

Article history:
Received 2 January 2022
Received in revised form 24 June 2022
Accepted 8 July 2022
Available online 21 July 2022

Keywords:
Collaboration
Communication
Network science
Data science
Engineering design

A B S T R A C T

During the design and implementation of engineering systems, as project members solve problems and exchange information with each other, a complex network of organisational interactions is established. The study of organisational networks can provide us with a deeper understanding of how organisations function and can help us devise strategies to improve collaboration and communication. However, it is widely believed that communication networks accurately capture the underlying collaboration amongst the members of an organisation. Here, through the use of multilayer and temporal network analysis methods, we analyse collaboration (through the project activity log) and communication (via emails) in a large engineering project and find that, on the contrary, collaboration and communication networks are only weakly correlated. More importantly, collaboration and communication networks show diametrically opposite purposes, underscored by crucial topological differences. While the email communication network is organised to facilitate the spread of information, the collaboration network is organised to suppress error propagation. We discuss the implications of our findings on how to design organisational networks and how to foster effective communication.

© 2022 The Author(s). Published by Elsevier B.V.
CC BY 4.0

1. Introduction

Communication among the members of an organisation is crucial to establish relationships, solve problems, and reach the organisation's goals (Fulk and Boyd, 1991; Maier et al., 2008). Relationships and communication among the members of an organisation can be represented and analysed by means of network analysis; that is, by studying networks of individuals (Borgatti et al., 2009; Monge et al., 2003). In order to generate trustworthy insights through network analysis, the organisational structure should be faithfully mapped into a network. Indeed, if the network under analysis fails to capture important relations among individuals, the results and findings can potentially be misleading. The traditional way to map and measure communication networks is through surveys and self-reported measures (Borgatti et al., 2009). For instance, people are asked to report with whom they communicate or collaborate and to rate the frequency of communication on a Likert scale. However, self-reported measures are error-prone (Borgatti et al., 2009) and can be affected by both recency and salience biases (Eagle et al., 2009). A recency bias is one where memories are biased toward recent events, while a salience bias is one where memories are biased toward more vivid events (Eagle et al., 2009). These biases, along with the error-proneness of surveys and self-reported measures, might result in the omission of crucial interactions among individuals, with the consequence of skewing the final results of the analysis.

In order to avoid these problems, scholars have moved towards the use of other data sources to map communication networks. It is well known that in organisations people create informal networks in order to accomplish their tasks efficiently (Krackhardt and Hanson, 1993; Cross et al., 2002). As such, researchers have analysed email exchanges as a proxy for organisational communication and collaboration (Guimera et al., 2006; Tyler et al., 2005; Abuja and Carley, 1999). This stream of research includes detecting communities of practice through clustering procedures (community detection) (Tyler et al., 2005; Guimera et al., 2006), relating task ‘routineness’ and network structure to performance (Abuja and Carley, 1999), inferring social status of individuals (Hu and Liu, 2012), or studying communication and coordination, e.g. among software developers

⁎ Corresponding author.
E-mail address: sebastiano.piccolo@jur.ku.dk (S.A. Piccolo).

https://doi.org/10.1016/j.compind.2022.103745
0166-3615/© 2022 The Author(s). Published by Elsevier B.V.
CC BY 4.0
(Bird et al., 2006). These studies provide evidence that the email network captures certain aspects of the underlying organisation. Organisational communication networks in Engineering Design are of paramount importance (Eppinger and Browning, 2012). For instance, an approach to designing systems, termed mirroring hypothesis (Colfer and Baldwin, 2016), suggest that the product architecture, understood as a network of technical components, should mirror the structure of the organisational communication network, as such a mirror is supposed to be beneficial in terms of fewer defects in the product and faster project completion time (see a review of these approaches in Section 2.1). Furthermore, researchers have developed a number of approaches to define how to collocate people in order to improve organisational collaboration. These approaches apply a clustering procedure to an organisational communication network and collocate people who appear in the same cluster (see a review of these approaches in Section 2.2). One implicit assumption on which the aforementioned approaches rely is that communication networks accurately capture the underlying collaboration amongst the members of an organisation. However, such an assumption has not yet been tested. Here, by analysing collaboration and communication in a large engineering project, we undermine this assumption: collaboration and communication networks are only weakly correlated and exhibit striking topological differences.

We use archival data, from a large-scale engineering project focused on designing a renewable energy power plant (see the data description in Section 3), to analyse and characterise similarities and differences between collaboration and communication networks amongst the project members. Our work is based on two longitudinal datasets (20 months), with daily resolution. Our data consists in 1) the activity log that describes the collaboration between the project members through the activities (tasks) that they have performed throughout the project and 2) the emails that these project members exchanged throughout the project. This data allows us to study how collaboration and communication evolved throughout the project.

We provide empirical evidence that interactions in organisations have multiple modalities and evolve over time. The different modalities, expressed here as network layers, do not necessarily provide much information about each other. In our case, the collaboration and the email layers evolve differently because, as we show, they serve different purposes. As a result, their topology aligns with their purposes.

The layer of collaboration revolves around the technical tasks that need to be performed to complete the project. This layer exhibits characteristics and topological structures that are useful to slow down the propagation of possible errors. The layer of email communication is consistent with the behaviour of a coordination network (Kapucu, 2006). In fact, we find that email communication happens mostly between people who do not collaborate directly but have to coordinate with others and share information in order to progress with their tasks. This layer exhibits topological structures that are useful to facilitate the spread of information.

We found that ~ 71% of collaboration dyads exhibit low email communication frequency and that the total length of dyadic collaboration correlates only weakly with the total number of emails exchanged. This invalidates the belief that dyads that exchange a high (low) number of emails, or with high (low) frequency, are necessarily strong (weak) ties. Indeed, we show that a strong tie in one layer can be a weak tie in another.

In sum, organisational networks are not monolithic and cannot be understood as single static entities. Organisational networks are intrinsically multilayered and evolve over time. These layers are weakly correlated and provide limited information about each other; that is, most of the connections that are present in one layer are not present in the other. More importantly, and perhaps surprisingly, each layer serves a different purpose through specific network properties and structures. In particular, we find that the two purposes are diametrically opposite.

The paper is organised as follows. In Section 2, we review the literature at the intersection between network science and engineering design. In Section 3 we describe the data we used, the methods to extract collaboration and communication networks from the data, and the methods of analysis. In section 4, we present our analyses and results. In Section 5, we connect our findings to the extant literature discussing their implications for research and practice with respect to: 1) the link between organisational and product structure; 2) the connection between communication and organisational structure; 3) network theory and complex systems; and 4) the emerging research field of Data Science for Engineering Design (Chiarello et al., 2021). Finally, in Section 6, we conclude the paper by summarising our main findings.

2. Background

The use of Network Science in Engineering Design has a long tradition, which dates back to 1981, when Donald Steward used a graph to represent task dependencies and cycles, with the goal to find a task schedule that would reduce iterations and, therefore, the total completion time (Steward, 1981). This method is known as Design Structure Matrix (DSM) and uses an adjacency matrix to visualise the structure of a system. The DSM has been used, typically, in three domains (Eppinger and Browning, 2012; Browning, 2015; Yassine, 2019): 1) processes, as networks of tasks; 2) products, as networks of components; and 3) organisations, as networks of people. Approaches that combine more than one domain take the name of multiple domain matrices (MDM) (Eppinger and Browning, 2012; Lindemann et al., 2008).

Since Steward’s seminal paper, the DSM literature has proposed a number of prescriptive approaches to schedule projects in order to optimise their duration (Cho and Eppinger, 2005; Meier et al., 2007; Yang et al., 2012), divide product architectures in modules (Schaeffer, 2007; Borjeson and Hölttä-Otto, 2014; Deng et al., 2012), and determine groups of interrelated people (Jaber et al., 2015; Khoriaty et al., 2018; Yang et al., 2019).

Descriptive approaches focus on quantifying the degree of modularity of a DSM (Gershenson et al., 2004; Jung and Simpson, 2017), understanding and predicting change propagation (Clarkson et al., 2004; Giffin et al., 2009), performing systematic comparisons between different architectures (Piccolo et al., 2020), and understanding the role of people (e.g. information leaders, coordinators, gatekeepers, etc) in a communication network (Batallas and Yassine, 2006).

In general, sequencing/scheduling is the recurring approach for task networks and clustering is the recurring approach for product and organisational networks (Browning, 2001; 2015). A full review of these approaches is beyond the scope of the present paper and we refer the reader to specialised literature (Eppinger and Browning, 2012; Browning, 2001; 2015; Yassine, 2019).

In the following, we review the literature at the intersection between Network Science and Engineering Design. In Section 2.1, we review the link between organisational structure and product architecture with focus on the mirroring hypothesis; that is, the idea that the product architecture should mirror the organisational communication structures. In this paper, we challenge this hypothesis and its rationale.

In Section 2.2, we review the connection between communication and organisational structure with focus on approaches that aim at improving organisational collaboration. In this paper, we challenge a common assumption behind these approaches; that is, that communication accurately captures the collaboration amongst the members of an organisation.
In Section 2.3, we review a stream of literature that builds on Network Science and complex systems, which relates network properties to performance of engineering projects. This stream of literature provides us with a number of network metrics that we use, in this paper, to quantify and characterise the differences between communication and collaboration. Finally, in Section 2.4, we review the use of Network Science in the emerging research field of Data Science for Engineering Design. These studies consider the engineering project from a socio-technical perspective and use multiple data sources, analysing social and technical networks together to reach a higher understanding of the project at hand.

2.1. The link between organisational structure and product architecture

In 1968, Melvin Conway stated that organisations which design systems are constrained to produce designs which are copies of the communication structures of these organisations (Conway, 1968). This sociological observation became known with the name of Conway’s law. In principle, Conway’s law did not imply, per se, any beneficial or detrimental effect on the design or system produced. In Engineering Design, Conway’s law is known as mirroring hypothesis (Coller and Baldwin, 2010) and has been interpreted as a positive fact. The mirroring hypothesis proposes that an alignment between the product (or the process) architecture and the organisational structure is beneficial in terms of fewer defects exhibited by the product (Sosa et al., 2004; Gokpinar et al., 2010), build success probability in software development (Kwan et al., 2011), faster completion time (Cataldo et al., 2008), and smaller propagation cost (MacCormack et al., 2012). The rationale behind this hypothesis is that a mirror between the structure of the technical system to be designed and the structure of the organisation would make it easier to manage its complexity, as people who collaborate should communicate more frequently (Coller and Baldwin, 2016). Many scholars have addressed the mirroring hypothesis. Sosa et al. (2004) aligned a product network and a team network, in a situation in which there was a one-by-one mapping between components and teams. They found that misalignment was more likely to happen between product modules than within product modules. In the data that they analysed, these unmatched interfaces between different modules exhibit the most significant problems. Gokpinar et al. (2010) performed a similar mapping between product and organisational networks, and computed a measure termed coordination deficit for each component; they correlated it with warranty claims as a measure of quality and found that components associated with higher coordination deficit were more likely to receive warranty claims. They noted that their measure was correlated with the degree of a component and alternative specifications of coordination deficit were not statistically significant. Cataldo et al. (2008) computed a similar measure, termed socio-technical congruence, which does not need a one-by-one mapping between teams and components. They found a positive effect of socio-technical congruence on the completion time. Kwan et al. (2011) used the same measure, extending it to consider also the weights of interactions, but found conflicting findings. There were some benefits, but only for certain build types and in some cases higher socio-technical congruence was associated with lower build success probability. Finally, they also observed that a large proportion of zero-congruence builds were successful, and that socio-technical gaps in successful builds were larger than gaps in failed builds. Similar findings were also found in a recent study (Chaudhari et al., 2022), where the authors noted that an excessive level of mirroring can be detrimental.

In short, the mirroring hypothesis is an appealing design hypothesis, which posits that aligning the product (or process) architecture to the organisational structure (with respect to the underlying communication network) is beneficial to both the product, in terms of fewer defects, and its implementation process in terms of faster completion time. The rationale of the hypothesis is that such an alignment makes it easier to manage the technical complexity of the system. Scholars have attempted to test the mirroring hypothesis, but results are conflicting. In this paper, using an approach similar to the one used by Cataldo et al. (2008), we test this hypothesis as formulated by Conway, by evaluating whether the collaboration network aligns with the communication network. Furthermore, by characterising the difference between collaboration and communication we evaluate the rationale behind the mirroring hypothesis.

2.2. Communication and organisational structure

Effective organisational communication is fundamental in order to establish trustworthy relationships amongst colleagues, solve problems, exchanging information in a timely manner and, ultimately, reach the organisation’s goals (Fulk and Boyd, 1991; Maier et al., 2008). Failure to establish effective communication, and in particular the presence of double agendas between project members can have strong negative effects on the project and its completion (Bjørn et al., 2014; Piccolo et al., 2018c). Given the paramount importance of organisational communication, researchers put a lot of effort into analysing organisational communication networks in order to improve communication and collaboration between the members of an organisation (Eppinger and Browning, 2012). Descriptive approaches focus on detecting communities of practice by detecting clusters in organisational networks (Tyler et al., 2005; Guimera et al., 2006), detecting gatekeepers, information leaders, coordinators, etc. (Batallas and Yassine, 2006), or inferring the social status of individuals (Hu and Liu, 2012). Prescriptive approaches focus on collocating people who communicate frequently in order to improve communication and collaboration in organisations (Allen, 1977; Eppinger and Browning, 2012). The usual way to determine how to collocate people is to discover clusters in a communication network and to move in the same office or building the people who figure in the same cluster (Jaber et al., 2015; Khoriyat et al., 2018; Yang et al., 2019). These approaches try to exploit the way people are connected in the attempt to facilitate the exchange of information within cluster while minimizing the cost of information passing between clusters. Scholars have considered different linkages between people to achieve this purpose: the relationships between people due to their assignment to decisions (Yang et al., 2015), the frequency of communication as captured by the number of deliverables exchanged (Khoriaty et al., 2018), or multiple attributes together (Yang et al., 2019).

In sum, the analysis of organisational networks can offer deep insights about the way people communicate and organise during engineering projects. However, the goodness of the insights depends on whether the organisational network under analysis accurately captures the desired information and contains all the relevant links between any two people. In particular, prescriptive approaches to collocation rest on the implicit assumption that communication accurately captures the collaboration amongst the members of an organisation. In this paper, by analysing the similarity and correlation between communication (via emails) and collaboration (through tasks performed together) networks we test this assumption.

2.3. Network theory and complex systems

Now, we turn our attention to a stream of literature that relates network properties to performance of engineering projects. Smith and Eppinger (1997), considering task networks and estimates of task completion time, developed a model to predict the total project duration and the total work time for each task (including the rework
time). They performed an eigenvector analysis to show that the rate of convergence of the project is intrinsically linked to the eigenvectors of the task network; in particular, to the eigenvectors associated to the largest eigenvalues. Similar result applies in the context of epidemics, where the largest eigenvalue of the adjacency matrix is related to the epidemic threshold (Wang et al., 2003).

Braha and Bar-Yam (2007), building on statistical mechanics of networks (Watts and Strogatz, 1998; Barabási and Albert, 1999; Albert and Barabási, 2002), focused on commonalities across a number of product or process networks. They found that these networks share structural properties such as 1) right-skewed degree distributions, which implies that most of the links are attached to a few nodes, called hubs; 2) low network density, which means that the number of connections is much smaller than the total number of possible connections; 3) high clustering coefficient and small average path length (compared to a random graph), which imply the small world property (Watts and Strogatz, 1998). These three properties are common to many networks and influence their behaviour (Albert and Barabási, 2002; Easley and Kleinberg, 2010; Newman, 2010), suggesting the existence of universal properties and behaviour of complex systems (Thurner et al., 2018) Furthermore, Braha and Bar-Yam (2007) studied an error propagation process on the networks and found the important role of hub tasks: in their simulations, interventions targeted at hub tasks are the most effective to improve the convergence rate of the project.

Piccolo et al. (2020) used simulated annealing to generate 2100 small world networks with right-skewed degree distribution and varying degrees of modularity. They showed that modularity suppresses error/change propagation but decreases topological robustness (network robustness to node removal). They found that the inverse of the largest eigenvalue of the network increase as modularity increases. As such, modularity increases robustness to error propagation. However, the authors noticed that this increase in robustness is likely to be due to negative assortativity. Additionally, they showed that the inverse of the first non-zero eigenvalue of the network Laplacian increases as modularity increases. This means that in modular architectures error propagation processes are slower.

In an empirical analysis, Parraguez et al. (2019) used a temporal network analysis to evaluate how the modularity of a design process changes over time. They contrasted the observed and expected patterns of modularity and showed that drops of process modularity were associated with periods of integrative activities and rework. Finally, they performed a statistical analysis to understand how selected network properties affect process modularity. They found that modularity is associated with decentralised, efficient, sparse, and more resilient networks.

Piccolo et al. (2018a) departed from single domain networks and used a bipartite network to consider the allocation of people to tasks of a real design process. They found that the network is sparse, exhibits right-skewed degree distributions, and the small world property. They evaluated topological robustness to node removal and have shown that the network was more vulnerable to the removal of highly connected people than to the removal of highly connected tasks, thus evaluating a form of bus factor. Through a set of simulations they have shown that topological robustness is determined by the degree distributions. Finally, they simulated cascades on the network to illustrate the importance of people who are connected to many tasks (termed generalists). These generalists have more power in suppressing cascades than other members.

Statistical analyses support the theoretical propositions reviewed above. Sosa et al. (2011), analysing software products, found that architectures with more right-skewed degree distributions are more likely to exhibit a lower number of defects than architectures with less right-skewed degree distributions. They also found that architectures with positive assortativity are more likely to exhibit higher number of defects than architectures with lower assortativity. Sosa et al. (2013) also found that modular components exhibit fewer defects than integrative components and that components involved in cycles exhibit a higher number of defects than components not involved in cycles. The number of defects increases with the length of the cycle and the degree of the component in the cycle. Similarly, Piccolo et al. (2019) found that design documents associated to modular tasks receive fewer iterations than documents associated with integrative tasks. Furthermore, the number of iterations increases with the number of dependencies in input and output of the task to which the design document is associated.

In this section, we have reviewed in great detail a number of studies that relate structural properties of networks to dynamical processes such as error propagation or information diffusion. As such, in the context of Engineering Design, this stream of literature relates product architecture, process architecture, or organisational structure to performance of projects. The insights that we have reviewed here are not specific to a single network, but generalise to any network (this is the primary goal of statistical mechanics of networks). In particular, hubs, network density, clustering coefficient, average path length, assortativity, and modularity are recurring network properties that influence robustness to error propagation and efficiency of information diffusion. In this paper, we build on this stream of literature and use the aforementioned properties to quantify and characterise the structural differences between collaboration and communication, in order to understand their different purposes.

2.4. Data science and engineering design

The emerging field of Data Science for Engineering Design has been defined by Chiarello et al. (2021) as a field of studies that uses computational tools to extract knowledge from data in order to address Engineering Design challenges and provide actionable insights. A recent stream of literature has introduced Network Science in the field of Data Science for Engineering design (Bird et al., 2009; Piccolo et al., 2018b; 2019; Piccolo, 2019). Some of the studies reviewed before use Network Science to analyse engineering projects and provide actionable insights, and could in principle figure here as well. Instead, here we focus on studies that consider the engineering design project from a socio-technical point of view and, rather than compressing multiple networks in simple indicators, as done by the studies reviewed in Section 2.1, put together all the relevant information provided by technical and social networks to better understand the project at hand and derive actionable insights.

These studies provide evidence that a socio-technical perspective, which combines the information provided by more networks, has more predictive and explanatory power than the information provided by a single network in isolation. Bird et al. (2009) studied the failure proneness of components in Windows Vista and across six releases of the Eclipse development environment. They computed a number of network metrics on both dependency and contribution networks and trained a logistic regression classifier to predict failure proneness. They showed that the algorithm exhibited the best performance when the algorithm was trained on features from both networks or when the measures were computed on a multilayer socio-technical network. Piccolo et al. (2019) studied the number of revisions of design document as a function of the network of interactions between teams and the network of task dependencies. They also considered product modularity and the number of stakeholders connected to a design document. They found that iterations increase when the number of stakeholders increases, when external suppliers are involved, with the number of input and output dependencies of the task to which a document is connected, and with the centrality of the team that produces or reviews the design document. They have shown that a socio-
We advance the state of the art in many ways: and the number of interfaces of the task. connected to the task, the number of people allocated to the task, and the number of interfaces of the task.

The present paper finds its place within this stream of literature. We advance the state of the art in many ways: 1) we analyse collaboration and combination networks from a temporal and multilayer network perspective; 2) we provide empirical evidence that organisational networks have multiple modalities, which evolve over time; 3) we show that these modalities are weakly correlated and not descriptive of each other; 4) we find that one of these modalities has a structure that helps to suppress error propagation, while the other has a structure that facilitates information flow, which implies that these modalities have diametrically opposite purposes. This last fact alone provides compelling reasons to rethink the mirroring hypothesis and other prescriptive approaches.

3. Data and methods

3.1. Data description

The data we analyse in this paper refers to a large-scale engineering project (€160,000,000+) of designing and implementing a renewable energy power plant, conducted by a multi-project Scandinavian company (Parraguez et al., 2015; 2016). To build the collaboration and the email communication networks we use the following datasets from the aforementioned project.

1. An activity log, which records the activities performed by the company's personnel throughout the duration of the project. As such, this activity log describes the relationships between people and activities (task). Each activity is identified by a unique code that is assigned by the software that the company uses to plan its projects and to break down the work into tasks. Similarly, each person is identified by a unique username. The activity log contains triples of the form <Person, Activity, Date>.

2. The complete email exchange between all the people involved in the project including employees, suppliers, external consultants, and project partners. The complete email archive contains 54,000+ emails. From the complete email archive we can extract relationships of the form <Sender, Receiver, Date>.

In these datasets, people are uniquely identified with the username with which they access the company's IT systems. Each person is associated with one username, and the username is the same for the email address. As such, the comparison between the two datasets is facilitated. There are 80 employees in total, who were responsible for planning the work, designing the biomass power plant, managing contractors and external partners, liaising with suppliers, and supervising the on-site work of construction of the power plant. It's worth noting that the number of people working on the construction site, in the implementation stage, was much higher as it involved external partners, contractors, and other workers.

3.2. Extracting networks from the data

The collaboration network is built from the activity log by projecting the bipartite network person-activities onto the people network. As a result of this projection, two people are linked if they collaborate on common tasks. Mathematically, let $B \in \{0, 1\}^{n \times m}$ be a matrix representing the way in which $n$ people are connected to $m$ activities, the collaboration network that links people through activities is obtained with the following projection: $A = B \times B^T$, where $B^T$ is the transpose of $B$. The binarised matrix $A$ is an adjacency matrix (see Appendix A) which describes the collaboration between the 80 employees. The email communication network is obtained by extracting the email addresses in the fields "To" and "CC".

The presence of the timestamp in our data allows us to build weighted and temporal representations of these networks. With weighted networks we can represent the strength of the connections between any two people in both networks. For the collaboration network, we define the strength of the interaction between two people as the duration of their collaboration in number of days. Similarly, for the email network, we can represent the strength of each connection by counting the number of emails exchanged between any two people.

To build temporal networks, instead of aggregating all the connections in a single network, we aggregate connections in temporal snapshots. Therefore, a temporal network is modelled as a series of static networks, one for each time step. Fixed a time step $t$, a temporal network is an ordered sequence of adjacency matrices $(A_{t1}, A_{t2}, \ldots, A_{tn})$.

3.3. Strategy of analysis

In order to evaluate both the mirroring hypothesis and the assumption that communication accurately captures collaboration, we need to evaluate how similar the email communication network and the collaboration network are. We do this by using both Pearson correlation and Jackard similarity coefficients (see Appendix A) on the adjacency matrices of communication and collaboration networks. We compute Pearson correlation and Jackard similarity on:

- the time aggregated binary networks: to evaluate the extent with which a link between any two people in one network is also present in the other network.
- the time aggregated weighted networks: to evaluate the extent with which weak (strong) ties in one network are also weak (strong) ties in the other network.
- the temporal networks: to evaluate whether the similarity between the two networks increases (decreases) during specific stages of the project.

Subsequently, to evaluate how the two networks change over time and to understand their topology, we evaluate the following structural properties: number of nodes, density, clustering coefficient, assortativity, average connectivity, number of connected components, average path length, average degree, and shell index (mathematical definitions and descriptions of these measures are found in Appendix A). Building on the results reviewed in Section 2.3, we can interpret the values of these topological measures to understand the behaviour of the two networks with respect to error propagation and information diffusion.

We complement the temporal analysis by simulating an SI spreading process (see Appendix A) on the two networks using the finest temporal resolution available in the data (daily snapshots). Through such an analysis, we are able to evaluate which network exhibits the highest spreading power. In particular, we are interested in evaluating the number of people that can be reached and the time with which the maximum diffusion can be achieved. By evaluating together the similarity between the two networks, their topological differences, and their spreading power we can understand the purpose of each network, which, in turn, will allow us to reason about the rationale of the mirroring hypothesis.

We conclude our analysis by characterising the email communication network by evaluating how the emailing frequency changes with respect to the collaboration period, for any two people. Specifically, we compute the emailing rate for the period preceding the collaboration, during the collaboration, and after the
collaboration. We do this to understand if people increase the emailing frequency during their period of collaboration or if the most of emailing communication happens outside the collaboration period. This analysis will let us understand if people use the emails as a tool to communicate during collaboration or if they use emails as a coordination tool.

4. Analysis and results

In this section, we perform the analyses of the data following the strategy that we have described previously. We describe our analyses, methodological choices, and how to interpret the measures that we use in detail in each of the following subsections. Overall, our results clearly show that the collaboration network and the email communication network are fundamentally different and serve two different purposes. As such, in our case, communication does not capture collaboration.

In Section 4.1, we compute the correlation between the time aggregated networks. For both unweighted and weighted networks we find that this correlation is low. We observe that in the collaboration network people are clustered around the technical tasks to be performed, while in the communication network the connections between people transcend the boundaries of the tasks.

In Section 4.2, we perform the temporal network analysis. We compute the correlation between collaboration and communication networks over time, as the project evolves, extending the previous findings. The results confirm that the correlation between the networks is low and remains low throughout the whole project. By evaluating and analysing a number of topological properties, we find that the topology of the two networks changes over time. This analysis suggests the existence, in the collaboration network, of structural properties which suppress error propagation processes, while the email network exhibits properties that facilitate diffusion processes. As such, we show that collaboration and communication network have different structures and purposes.

In Section 4.3, we further test this intuition through simulating and studying a susceptible-infected (SI) epidemic process on the two temporal networks. We show that in the email network the epidemic process reaches more people and is always faster than it is in the collaboration network. Furthermore, we show that the spreading power of the email network is almost optimal, if we compare it to the spreading power of a temporal network that combines all the connections between people transcends the boundaries of the tasks.

Finally, in Section 4.4, we consider the emailing rate, in the case of each collaboration dyad and in three different stages: before, during, and after the collaboration period. We find that the emailing frequency tends to be generally low and constant throughout the project for most people. Only 15 % of dyads exhibit an increase of email communication during the collaboration period. As such, we find that the email network acts as a coordination network.

4.1. Collaboration and communication networks are weakly correlated

In the collaboration network, we connect two people if they collaborate on common tasks. Thus, an edge in the collaboration network represents a relation between two individuals that has a certain duration and persistence over time. We argue that the collaboration network has a technical and planned nature which constrains the possible connections between people. In fact, people are allocated to tasks for specific reasons, which depend on the technical requirements of the tasks and on the way tasks are connected by technical dependencies, as the output of a task can be the input of another task. Therefore, the collaboration network reflects the technical and planned nature of the biomass power plant that needs to be designed and implemented. However, we stress here that our data describes the implementation of the project by recording which people worked on which task. As such, here we analyse the collaboration as implemented during the project. In contrast, an edge in the email network represents a contact between two individuals. This contact does not necessarily have persistence over time and can have an unplanned (emergent) nature. Previous work has shown that the same set of people can form diverse networks and can behave differently in each of these networks (Mones et al., 2017; Agrest et al., 2015). As a result, multilayer networks are not generally structurally reducible by aggregation into a single layer.

In this paper, we deal with a network which is driven by the technical aspects of the biomass power plant (the collaboration network), and with a network which is emergent and driven by the need to communicate with other project members (the communication network). As such, we are in the position to analyse this project from a socio-technical perspective, addressing the mirroring hypothesis and the common assumption that communication correctly captures collaboration.

A visual inspection of the networks’ adjacency matrices (Fig. 1) provides visual confirmation that the networks exhibit different structures: While in the collaboration network people are clustered around the tasks they perform, the email communication network has a denser and more homogeneous structure. To highlight this fact visually, we detected cliques on the collaboration network and arranged rows and columns of all the matrices in decreasing order of clique size. As such, a given person is associated to the same row (column) in each adjacency matrix. If we consider the link-weights of the networks (Fig. 1 bottom panel), that is the duration (in days) of the collaboration and the total number of emails exchanged between any two people, the clustered and more localised nature of the collaboration network – as opposed to the more global and homogeneous nature of the email network – becomes more clear. This striking difference between the two networks underlines the importance of the socio-technical perspective to study engineering projects.

In addition, the visual comparison in Fig. 1 also suggests that it is more likely that an edge in the collaboration network appears also in the email network than vice versa. To confirm these insights from visual inspection, we computed the structural similarity between the networks using the Jaccard similarity and the correlation coefficient between the networks (weighted and non-weighted) evaluating the statistical significance using the quadratic assignment procedure (QAP) test, which is a permutation test for matrices (see Appendix A). Fig. 2 shows the results of the permutation tests. The correlation between the non-weighted networks is positive ($\rho = 0.35$) and statistically significant ($p < 0.0001$). The correlation is smaller if we consider the weights ($\rho = 0.14$) and still significant ($p = 0.001$). Finally, the degree of structural similarity between the networks, where we only consider the proportion of edges in common, is small ($J = 0.26, p < 0.0001$). Despite the statistical significance, the correlation between the networks is weak; the correlation coefficient is small in both weighted and non-weighted cases. This means that a large part of the email communication occurs between people who are not linked in the collaboration network, such as people from different departments or people working on different components.

Additionally, we computed the conditional probabilities of finding an edge in one network given that the edge appears in the other network. With C being the collaboration network and E being the email network, we computed the following: $P(C|E) = 0.32$ and $P(E|C) = 0.61$. In words, if two people are linked in the collaboration network, the probability of observing at least one email between them is 0.61. In contrast, if two people are linked by an email exchange, the probability that these two people are also linked by a collaboration on a task is only 0.32. This means that the email network is primarily used for communication that involves people that are not collaborating on the same tasks. This is reasonable as people
who collaborate on the same tasks might prefer more direct communication, such as face-to-face communication.

4.2. Collaboration and communication networks have different structures and purposes

In Section 4.1, we argued that the two networks, when considered in their time-aggregated form, are fundamentally different: in the collaboration network people are clustered around the tasks they perform, while the email network shows a more distributed and global structure. In this section, we quantify and describe the structural differences between the two networks and their co-evolution over time. This is important since we know from the literature that, when it comes to temporal networks, the study of the time-aggregated network is not necessarily descriptive of the network dynamics (Holme and Saramäki, 2012). Furthermore, as the two networks represent relationships between the same people in the context of a large engineering project, it is interesting to evaluate the structural differences between the networks, in order to understand specifically how collaboration and communication are related, in the project under investigation. Here, we are not interested in the detailed flux of the email network, which is not matched by the slowly evolving collaboration network; therefore, we consider the network evolution in monthly snapshots.

We analyse how the correlation between the two networks and a number of their topological properties change over time; the results are shown in Fig. 3. The measures and their purposes are explained in Table 1. More details and formal definitions can be found in

![Fig. 1. Visual representation using adjacency matrices of the collaboration network, the email communication network, and their aggregation (sum). Matrices are arranged in a descending order of clique size. Cliques are computed on the collaboration network and within each clique there is no particular order. The order is maintained for the other matrices as well so that the i-th row in both the collaboration and email networks refers to the same person. The top panel shows the binary adjacency matrices and the bottom panel shows the weights. Edges in the collaboration network are weighted by the duration of the collaboration in number of days; edges in the communication networks are weighted by the number of emails exchanged. The collaboration network shows a clustered structure which revolves around the tasks performed, while the email communication network has a different structure.](image1)

![Fig. 2. Statistical significance of network similarity. The vertical red line indicates the test statistics and the black histogram the null hypothesis distribution computed through 10,000 simulations. A: Pearson’s correlation and QAP test between the non-weighted collaboration network and the non-weighted email communication network ($\rho = 0.35$, $p < 0.0001$). B: Pearson’s correlation and QAP test between the weighted collaboration network and the weighted email communication network ($\rho = 0.14$, $p = 0.001$). C: Jaccard’s similarity and QAP test between the collaboration network and the email network ($J = 0.26$, $p < 0.0001$).](image2)
Appendix A and in the following when we analyse the values of these measures. The topological measures that we have selected allows us to understand the robustness of a network against node removal, the resilience against error propagation processes, the efficiency of a network in routing information, the presence of hubs that can act as powerful spreaders, the presence of synergy between these hubs, and the modularity of these networks.

Before moving to the analysis of the topological differences between the two networks, let us consider the relatively low persistence of both networks, which shows that their topology changes considerably over time (Fig. 3j). The persistence is the Jaccard similarity coefficient between two consecutive snapshots and can be understood as the percentage of links in common between two consecutive snapshots. We observe that the persistence of the email network (red line) is higher than the persistence of the collaboration network (blue line). The lower persistence of the collaboration network is in line with the project progression where the people connected to tasks that have been completed can either become inactive or can connect to different people as they are now working on other tasks.

We now consider the correlation of the two networks over time by computing the Pearson correlation coefficient $\rho(C_t, E_t)$ and the Jaccard similarity coefficient $J(C_t, E_t)$ for each temporal snapshot $t$. We observe that both correlation and similarity over time between the two networks are low, confirming that the two networks are different and indicating that they also evolve differently (Fig. 3k and l). This results also extends the insights from the previous section: collaboration and communication networks are weakly correlated and, when considering the temporal dimension, this correlation is even lower.

As the networks appear to be quite different and change considerably over time, we are motivated to characterise the evolution of their topology. Now, we compute the topological measures reported in Table 1. Each measure has its own purpose, as reported in Table 1; however, when these measures are interpreted together we get a much more complete and nuanced view of the networks. We do this in the following.

First, let us observe that the number of people involved in the project (Fig. 3a) increases over time in both networks; furthermore, the number of active people is always higher in the email network. The email communication network is also constantly denser than the collaboration network and its density grows over time while the density of the collaboration network is relatively static (Fig. 3b). This pattern indicates that the communication network becomes more connected and integrated over time. In fact, the average connectivity (Fig. 3e), which measures the extent to which redundant paths in the network exist and it is, therefore, a measure of robustness against node removal, is in line with this pattern. Additionally, the number of connected components (Fig. 3f) is consistently one for the email network and greater than one for the collaboration network. As such,
Table 1
Measures computed on the temporal networks with their explanation. The capital letter in bold locates the measure's subplot in Fig. 3. The first nine measures are topological network measures useful to understand the differences between collaboration and communication. The last three measures are employed to understand the persistence, and the correlation between collaboration and communication over time. More details and mathematical definitions are presented in Appendix A.

| Measure         | Explanation                                      |
|-----------------|--------------------------------------------------|
| Nodes           | Number of active nodes in a given snapshot       |
| Density         | Number of edges out of all possible edges. Measures how connected the network is. |
| Clustering coeff. | The extent with which the network is organised in cliques.       |
| Assortativity   | The extent with which hubs are connected to other hubs. A measure of resilience. |
| Average Connectivity | Measures redundant paths in the network and robustness against node removal. |
| Connected Components | Number of disconnected components. A coarse measure of network separation. |
| Average Path Length | The average distance between any two people in the network. A measure of efficiency. |
| Average Degree  | The average number of links per node. A measure of spreading power. |
| Shell Index     | The size of the largest core in the network. A measure of spreading power. |
| Persistence     | Measures the extent with which the network does not change (persist) over time. |
| $c(C_r, E_r)$   | Person correlation, at time $r$, between the two networks. |
| $J(C_r, E_r)$   | Jaccard similarity, at time $r$, between the two networks. |

compared to the collaboration network, the communication network is more cohesive, integrated, and less subject to problems arising from people being unavailable. The robustness against people unavailability is an important and desirable feature in engineering projects (Piccolo et al., 2018a). In fact, the bus factor – i.e. the number of people that need to be removed from a project (as if they were hit by a bus) in order for the project to stall – is an important indicator of project robustness (Piccolo et al., 2018a).

The clustering coefficient (Fig. 3c) increases over time for both networks although it increases more rapidly in the case of the collaboration network. The increase of the clustering coefficient is in line with the project progression and with the integration stage, where the work carried out by the different teams is integrated together. The average path length (Fig. 3g) is always smaller for the email network; however, we observe that for the collaboration network the average path length decreases over time. High clustering coefficient and small average path length are associated with shorter diffusion time, while positive assortativity is associated with longer diffusion time. As such, considering also that high clustering coefficient and small average path length are also associated with synchronisation being easier to achieve (Comellas and Gago, 2007; Sorrentino et al., 2006), the communication network is expected to be faster than the collaboration network at diffusing information.

In sum, the communication network is relatively dense, cliquish but integrative, with many redundant paths that make it robust against node removal, efficient with respect to information diffusion, and with the hubs positioned as information gatekeepers, a position that allows them to efficiently spread information. The collaboration network is relatively sparse, cliquish and modular with clearly defined clusters, with few redundant paths that make it less robust against node removal but more resilient against error propagation. The hubs in the collaboration network are connected with each other within their respective clusters, with the effect of slowing down error diffusion processes.

4.3. The different purposes: information diffusion and error suppression

Here, we complement the previous observations by simulating information spreading under different initial conditions on the two networks to test which one is faster. To accomplish this, we simulate an SI (Subjected-Infected) spreading process respecting the time evolution of the networks (see A). For our simulations, we assume that a person, for each time unit, can spread a signal only to the direct neighbours with a certain probability ($\beta$). As the collaboration network has a daily resolution, we consider daily snapshots for both networks, in order to make the results more comparable. We study the speed of spreading and the coverage, in terms of percentage of nodes reached, by varying the infection probability $\beta \in [0,1]$. In 15 points evenly spaced on a logarithmic scale, and the number of initial seeds $s \in [5,10,15]$ randomly chosen. We repeat the simulations, for each couple of parameters ($s, \beta$), 50 times to estimate the statistical uncertainty due to randomness. We simulate the SI on the collaboration network, the email network, and a third network that aggregates the links in the previous two networks.

We found that regardless the choice of $s$ and $\beta$, the email network has always higher coverage (Fig. 4) and is always faster at spreading the information (Fig. 5). Higher values of $\beta$ are associated with higher coverage and higher speed. Higher values of $s$ are associated with higher coverage and lower variance (i.e. the effect of randomness is lower), but not with the speed. Furthermore, the spreading
properties of the email network are basically indistinguishable from the spreading properties of the network that combines the edges of both email and collaboration networks. This shows that the topology of the email network dominates the spreading dynamics.

We find that the epidemic threshold is lower for the email network. This means that in the collaboration network the infection probability needs to be higher, than in the communication network, for the SI process to infect the same number of nodes (Fig. 4). Even when we set the infection probability to 1 the SI process needs more time to infect the whole collaboration network than it needs to infect the whole the communication network.

These simulations confirm that the collaboration network, compared to the email network, has topological properties that are useful to slow the error propagation processes: higher epidemic threshold, more pronounced modular structure, higher assortativity, and lower speed of propagation. These properties are useful to limit the error propagation and to have more time for interventions before a catastrophic diffusion. The email network, in contrast, has topological properties that are useful to spread the information faster and to reach more people.

The results in this section, along with the results in Section 4.2, reveal that collaboration and communication networks have diametrically opposite purposes.

4.4. The interplay between collaboration and communication: emails as a means of coordination

In this section, we aim to understand what the interplay between collaboration and coordination is. Specifically, we want to understand how the frequency of email communication changes throughout the project, for each dyad, with respect to the collaboration period. We computed, for each dyad, the daily emailing rate
before the collaboration ($\lambda_B$), during ($\lambda_D$), and after ($\lambda_A$). These rates are shown in Fig. 6 as 1) univariate distributions in the form of boxplots; and 2) pairwise bivariate distributions using kernel density estimations with the standard Gaussian kernel. The boxplots are shown on a logarithmic scale because of the skewed distributions given by the presence of many values close to zero. The visualisations show that the overall rate of communication tend to increase during the periods of collaboration. In addition, the three lambdas are positively correlated. The couples ($\lambda_B$, $\lambda_D$) and ($\lambda_D$, $\lambda_A$) exhibit higher correlation than the couple ($\lambda_B$, $\lambda_A$) as shown by the diverging joint distribution.

To complement the overall analysis shown in Fig. 6, we performed a statistical analysis of each dyad where we compare the three rates of email communication to understand whether they are statistically different. Specifically, we compare $\lambda_D$ with $\lambda_B$ and $\lambda_A$ with $\lambda_D$. This analysis will let us enumerate patterns of email communication that we show in Table 2. We observe that the majority of dyads (1st row) exhibit, regardless of the period, sporadic email communication throughout the whole project. Forty dyads (2nd row) show higher communication in the periods before and during collaboration, 39 dyads show higher communication during and after the collaboration (7th and 8th row), and 32 dyads show the higher rate of communication after the period of collaboration (3rd and 9th row). Only 10 dyads show higher communication rate before the collaboration (4th, 5th, and 6th row). Finally, only 62 dyads exhibit a statistically lower emailing rate after the collaboration (2nd, 5th, and 8th row).

Therefore, ~71% of the dyads show very low email communication rate which amounts to ~3 emails per month and only ~15% of the dyads lowers their email communication rate after the collaboration period. This is sustained throughout the project, meaning that people

Table 2
Patterns of dyadic communication with test of significance between rates. Each row describes one communication pattern. Each row reports the number of dyads, average emailing rates before, during and after collaboration ($\lambda_B$, $\lambda_D$, and $\lambda_A$) and the significance test of the pattern. The table highlights that most of the dyads show low emailing rate that is constant throughout the whole project.

| #dyads | $\bar{\lambda}_B$ | $\bar{\lambda}_D$ | $\bar{\lambda}_A$ | Significance test | Explanation |
|--------|-----------------|-----------------|-----------------|-----------------|-------------|
| 1      | 291             | 0.04            | 0.07            | $\lambda_B = \lambda_D = \lambda_A$ | Dyads that communicate with similar frequency throughout the project. |
| 2      | 40              | 0.27            | 0.49            | $\lambda_B = \lambda_D > \lambda_A$ | Dyads that decrease the amount of communication after the collaboration. |
| 3      | 28              | 0.03            | 0.11            | $\lambda_B = \lambda_D < \lambda_A$ | Dyads that increase the amount of communication after the collaboration. |
| 4      | 6               | 0.26            | 0.09            | $\lambda_B > \lambda_D = \lambda_A$ | Dyads that communicate with higher frequency before the collaboration. |
| 5      | 1               | 1.00            | 0.05            | $\lambda_B > \lambda_D > \lambda_A$ | Dyads that decrease the amount of communication as the project progresses. |
| 6      | 3               | 0.40            | 0.01            | $\lambda_B > \lambda_D < \lambda_A$ | Dyads that communicate with lower frequency during collaboration. |
| 7      | 18              | 0.03            | 0.20            | $\lambda_B < \lambda_D = \lambda_A$ | Dyads that communicate with lower frequency before the collaboration. |
| 8      | 21              | 0.19            | 1.31            | $\lambda_B < \lambda_D > \lambda_A$ | Dyads that communicate with higher frequency during the collaboration. |
| 9      | 4               | 0.04            | 0.23            | $\lambda_B < \lambda_D < \lambda_A$ | Dyads that gradually increase the communication frequency. |

Fig. 6. Emailing rate before ($\lambda_B$), during ($\lambda_D$), and after ($\lambda_A$) the collaboration period for each dyad. The y-axis in the box-plot is on a logarithmic scale. The notch in the box-plots represents an approximate 95% confidence interval around the median. The emailing rate, overall, tends to increase during the collaboration period. The bi-variate kernel density estimation (KDE) plots show that most of the dyads communicate with low frequency. The bi-variate KDE plots also show that the emailing rate is higher during the collaboration period than outside.
who collaborate to an engineering project might prefer other means of communication, such as face-to-face communication, documents, and sketches. As such, the email network is not a good proxy to capture the collaboration network and, therefore, it is not descriptive of the patterns of collaboration in the project studied here.

This section has shown a difference in purpose between the two networks: the email network, as opposed to the collaboration network, is used by people who do not collaborate directly but need to exchange information. Furthermore, among the dyads that appear in both networks the email communication rate is low. As such, the email network exhibits the role of a coordination network.

5. Discussion and implications

In this section, we discuss and connect our results with extant literature, providing applicable insights for practitioners and scholars, and pointers for future research. We start by considering the link between organisational structure and product architecture with particular focus on the mirroring hypothesis, also known as Conway’s law (Section 5.1). We proceed to address the issues of communication improvement and people co-location, highlighting the importance of facilitating communication between people who do not collaborate on common tasks but have the need to exchange information (Section 5.2). We, then, position and discuss our findings within the Network Science literature considering the difficulties of mapping organisational networks as they are intrinsically multilayered, partially planned and partially emergent (Section 5.3). Finally, we conclude by positioning our work within the emerging research field of Data Science for Engineering Design (Chiarello et al., 2021). We discuss how our findings provide actionable insights for the stages of conceptual and embodiment design, and for the management of engineering projects. We conclude that Network Science is an important methodological and theoretical addition to the emerging research field of Data Science for Engineering Design (Section 5.4). We summarise the main implications of our paper in Table 3, considering researchers and practitioners.

5.1. The link between organisational structure and product architecture

In 1968, Melvin Conway stated that organisations which design systems are constrained to produce designs which are copies of the communication structures of these organisations (Conway, 1968). Accentuated differently, the topology (i.e. the network structure) of the product designed would resemble the topology of the communication network between the people or the teams involved. This sociological observation, which became known as Conway’s law, has been restated in the Engineering Design community with the name of mirroring hypothesis (Colfer and Baldwin, 2016). Conway’s law has often been interpreted as a desirable property which enables the development of higher quality products, with lower number of iterations and lower error propagation during the development process (Colfer and Baldwin, 2016) and has even become a theoretical underpinning of the micro-service architecture (Wolff, 2016).

Our results about the overall positive and statistically significant correlation between the email communication network and the collaboration network (Section 4.1), could be interpreted as weak evidence in favour of Conway’s law. However, this mirror fades away as we zoom in into the nature and purposes of the two networks. We discovered crucial differences between the email communication network and the collaboration network, showing that they have opposite purposes. The collaboration network exhibits properties that are useful to suppress error propagation; these properties would be detrimental in a communication network as they would create barriers to the diffusion of information. Conversely, the topological structures exhibited by the email communication network would be detrimental in a collaboration network, as they would make the process inherently more iterative and error prone.

As such, our results cast a doubt about the desirability of the mirror between product and organisational structures. After all, if organisations are truly bound to design products that mirror their communication structures, a dysfunctional communication structure will probably produce a dysfunctional product or will generate an error-prone design process. Our results are in line with other recent empirical works that have shown how an excessive mirror between product and organisation can be detrimental (Gavras and Kostakis, 2021; Chaudhari et al., 2022). In light of the above, a promising venue for future research is to understand when the mirror between product and organisation is desirable and how much the product should mirror the organisation.

To practitioners, we recommend to design products using the solid principles of modular design: sparse architecture with low density and highly cohesive but loosely coupled modules. These properties, as highlighted by our results, are useful to suppress the error propagation. Additionally, our results highlight the importance

Table 3

| Theme | Researchers | Practitioners |
|-------|-------------|---------------|
| The link between organisational structure and product architecture | 1. Rethinking the mirroring hypothesis and its rationale. 2. Researching how to design product architectures and organisational structures to serve different purposes. | 1. Designing modular products avoiding direct dependencies between hub components. 2. Collaboration should be organised and supported in order to enable fast response to changes and in order to suppress error propagation. |
| Communication and organisational structure | 1. Communication is, at best, weakly correlated with collaboration. 2. Organisational structures and communication networks evolve over time. 3. Researching methods to promote shared understanding and to remove communication barriers. | 1. Mapping the collaboration network is of paramount importance in order to effectively collocate people. 2. Implementing tools to facilitate communication and coordination, particularly across teams and departments. |
| Network theory and complex systems | 1. Organisations have multiple modalities and evolve over time. As such, they are best studied as multilayer temporal networks. 2. Researching the best way to support multiple modalities of communication in complex systems. 3. Researching how to facilitate the formation of informal networks with desirable properties. | 1. Mapping organisational networks and understanding how collaboration and communication evolve over time. 2. Re-arranging organisational networks according to the need and the problem to solve. 3. Implementing tools and culture to support people in finding experts in the organisation. |
| Data Science and Engineering Design | 1. Network Science offers important tools to understand engineering projects from a socio-technical perspective. 2. Network Science is an important addition in the Data Science toolbox for Engineering Design. | 1. Network Science is useful to study past projects, and the way they unfolded, in order to derive actionable insights for future projects. 2. Network Science is useful to plan the design of product, processes, and organisations. |
of assortativity (i.e. the degree correlation). Negative assortativity is associated with higher resistance to change propagation (Piccolo et al., 2020), a desirable property in the product structure. Therefore, as suggested in Piccolo et al. (2020), we recommend to add negative assortativity to modular design; that is, avoiding direct connections between system components with high number of dependencies.

In the following section, we discuss the implications of our finding on how to design organisational networks and how to foster effective communication.

5.2. Communication and organisational structure

Collocating people who communicate frequently has been considered a strategy to improve communication and collaboration in organisations (Allen, 1977). The usual way to determine how to collocate people is to discover clusters in a communication network and to move in the same office or building the people who figure in the same cluster (Jaber et al., 2015; Khoriayt et al., 2018; Yang et al., 2019; Eppinger and Browning, 2012). These approaches try to exploit the way people are connected in the attempt to facilitate the exchange of information within cluster while minimizing the cost of information passing between clusters. One common assumption behind this type of approaches is that the communication network correctly captures the collaboration between the project members. Our results undermine this assumption.

First, we show that the same people are connected in substantially different ways depending on the network we consider; as such, the discovered clusters can be fundamentally different, depending on the network considered. Second, we find that people who exchange many emails are those people who do not collaborate directly. This finding is in line with previous research about the use of virtual means of communication as opposed to face-to-face (Walsh and Maloney, 2007). Third, we argue that the two networks studied here have different purposes. These results strongly limit the scope of the aforementioned approaches of communication improvement through clustering of organisational networks. We posit that correctly capturing the information flow amongst the project members is of paramount importance to any attempt of improving collaboration and coordination.

Social interactions in organisations have multiple modalities, which evolve over time. Considering only one of the many modalities might hide important interactions, which can in turn frustrate attempts to improve the organisational structure. At the same time, as we show that these multiple modalities of interactions serve different purposes, it does not appear wise to confine them together. Clustering as many modalities as possible in an attempt to collocate people might yield sub-optimal solutions that can have a detrimental effect of organisational communication. This is, for instance, the case of ‘open’ work-spaces, that have been shown to decrease collaboration and productivity (Brennan et al., 2002; Bernstein and Turban, 2018). Finally, it is worth mentioning that collocation does not necessarily improve collaboration or productivity (Yang and Jin, 2008; Bjørn et al., 2014; Morrison-Smith and Ruiz, 2020; Brown et al., 2020) and that familiarity with the tasks and familiarity with team members is more relevant to team performance than geographical distance or co-location (Espinosa et al., 2007).

Our findings show the existence of two kinds of communication across the project members: the communication between people who work on the same tasks and the communication between people who work on different tasks but have a need to coordinate and exchange information. This suggest that practitioners can use the aforementioned clustering and co-location approaches to organise new teams for new product development projects, in order to co-locate people who are expected to benefit from face-to-face communication, because they are connected to the same tasks. In parallel, practitioners and project managers should strive to facilitate communication and coordination, in particular with people that work on different tasks and different departments. This is needed to create a common ground and to avoid double agendas (Bjørn et al., 2014; Piccolo et al., 2018c).

Future research can focus on developing approaches and methods to promote shared understanding, common ground, and remove barriers to communication. With regards to the structure of a communication network, we argue that a good communication network should facilitate the fast flow of information and the reachability of individuals. Future research can also focus on other network properties, than those shown here, that correlate with higher spread of information and higher reachability. Finally, future research should understand what can facilitate the formation of networks with these desirable properties.

5.3. Network theory and complex systems

In this work we have shown that the email communication network and the collaboration network, although weakly correlated, are different and not much descriptive of each other. This finding is in line with previous work that has shown that different layers of a multilayer network are not predictive of each other (Mones et al., 2017; Agrest et al., 2015). This implies that a strong tie (Krackhardt, 2003) or a weak tie weak (Granovetter, 1973) in one layer might not be a strong (weak) tie in another layer. As such, studying communication networks and their performance can be challenging, especially from an empirical point of view, as communication has multiple modalities and capturing a faithful and complete map of the communication amongst individuals is difficult.

In addition, theoretical models (Lazer and Friedman, 2007; Piccolo et al., 2020) have shown that different network structures support different goals. For instance, a distributed (modular) organisational network structure is associated with better solutions as teams can work in parallel, exchange their solutions, and iterate them further (Lazer and Friedman, 2007). However, such a distributed network will take more time than an integrated network to reach consensus (Lazer and Friedman, 2007; Piccolo et al., 2020). These theoretical propositions are also supported by some empirical findings (Pentland, 2012; Piccolo et al., 2018c; Parraguez et al., 2019). The empirical findings presented here add to this stream of literature, showing 1) that different network structures serve different purposes and 2) that communication in complex systems (such as engineering projects) intrinsically need to have multiple modalities, as complex systems typically address/face multiple problems. This raises an important question for future research: what is the best way to support such multiple modalities of communication in complex systems?

Another important aspect that we highlighted in this paper is that the email network is emergent and self-organising, showing the partially planned and partially emergent nature of engineering projects. As our results show, the engineering process analysed here evolves over time and the organisational structure, as captured by the collaboration and email layers, evolves over time as well. We find that the email network evolves in a way that overcomes certain difficulties that might have been present in the collaboration network, such as communication between distant people (i.e. connected to different tasks). As the email network is emergent in nature and the desirable structures that we found are emergent and evolutionary as well, a question raises spontaneously: Is it possible to facilitate such a process? The question is important, because evolutionary processes, self-organization, and emergence are common features of complex systems (Thurner et al., 2018; Newman, 2010; Easley and Kleinberg, 2010). As these processes are unplanned, decentralised, and path dependent it is important to ask if we can influence them, for instance to make them happen faster,
and if such an influence might have detrimental effects in the long run.

Finally, a related topic is that of navigability in networks; that is, the ability of efficiently routing information in the network when the actors do not have complete information of the network topology. In organisations, it is of crucial importance to efficiently reach who knows what, so that important knowledge and specific expertise can be reached in a timely manner (Balog et al., 2012; Tang et al., 2015). Many complex systems exhibit a network structure that promotes efficient navigability when nodes do not have global knowledge of the network in which they are in (Granovetter, 1973; Newman, 2010; Boguna et al., 2009). The email network we analysed here, compared to the collaboration network, offers higher navigability. Future research can study how to promote more navigable networks and which routing protocols offer the best navigability performance.

5.4. Data science and engineering design

Engineering Design is a process of problem solving that involves the joint effort of many people, who work on a variety of tasks, in order to reach a common goal such as the development of a product (Eppinger and Ulrich, 2015). Data Science is a multidisciplinary field, which uses computational tools to extract knowledge from data and derive applicable insights. It uses both exploratory and confirmatory analyses, iterating between induction and deduction, mirroring the scientific discovery process (Box, 1976; Guo, 2012; Piccolo, 2019). The use of Data Science in Engineering Design has a long history, and has been increasing in the past two decades (Chiarello et al., 2021).

Chiarello et al. (2021) have recently reviewed the literature at the intersection between Data Science and Engineering Design. They mapped the extant literature into design process phases and highlighted the most common computational approaches. Finally, they have detailed challenges for both Engineering Design and Data Science. Our work connects with this emerging stream of literature in many ways.

First, we complement the work of Chiarello et al. (2021) in reference to the Data Science methods used in Engineering Design, with Network Science. We do this in two ways: 1) we intersect Engineering Design and Network science with our literature review, and 2) we offer a case study of applications of Network Science through our analyses, which employ a broad array of methods using a temporal multi layer network perspective.

Second, with respect to the engineering design phases, as derived by Chiarello et al. (2021), our analyses address some fundamental challenges within the topic of project management, with particular emphasis on organisational communication. As discussed in Section 5.2, our analyses has highlighted structural network properties that are desirable for facilitating the information flow amongst the project members. Moreover, our analysis has allowed us to show the limitations of traditional approaches to improve communication and the overall difficulties of analysing communication networks, as they have multiple modalities. Finally, as discussed in Section 5.1, our findings have important implications on the mirroring hypothesis/Conway’s law. We offer a more nuanced view, where the product architecture does not have to be a perfect copy of the organisational structure, and that it can be advantageous to design the two for different purposes.

Third, our results on the collaboration network and its structures aimed at suppressing error propagation gave us important insights on some desirable properties of product architecture. Namely, modularity and negative assortativity. As such, our results provide actionable insights in the stages of Conceptual Design and Embodyment Design, with particular focus on the product architecture, adding to a long stream of literature on the topic.

As such, we believe that Network Science, with both its theory and methods, is an important addition to the Data Science toolbox for Engineering Design research and practice.

6. Conclusions

In this paper we have investigated collaboration and communication, in an large scale engineering project, through the lenses of Network Science. From archival data consisting in activity logs and email exchanges, we obtained network representations of collaboration and communication. By evaluating the overlap between the networks, studying multiple topological properties, and simulating information propagation we offer the following findings. First, we provide empirical evidence that organisational networks are multi-layered, evolve over time and, therefore, cannot be understood as monolithic static entities. Second, contrary to a widespread belief, collaboration and communication networks are weakly correlated and provide little information about each other. Third, collaboration and communication networks have different purposes, which are highlighted by specific topological differences. Specifically, we find that the collaboration network exhibits a topology that serves the purpose of suppressing error propagation, while the communication network exhibits a structure that serves the purpose of facilitating information propagation. Finally, we discussed the implications of our findings on 1) the link between organisational structure and product structure with respect to the mirroring hypothesis/Conway’s law; and 2) how to design organisational networks and how to foster effective communication. We also discussed our findings in the context of Network theory and Data Science for Engineering Design.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the case company for providing the data and the anonymous reviewers for their suggestions and comments.

CRediT authorship contribution statement

Sebastiano A. Piccolo: Conceptualization, Investigation, Methodology, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. Sune Lehmann: Conceptualization, Methodology, Supervision, Writing – review & editing. Anja M. Maier: Conceptualization, Methodology, Data curation, Supervision, Writing – review & editing.
Appendix A. Network analysis methods

Adjacency matrices

The architecture of a system (i.e. the way in which its elements are connected) can be modelled as a network. A network is a pair \( G = (V, E) \) where \( V \) is the set of nodes that represent the elements of the system and \( E \) is the set of edges that represent the connections between the elements of the system. Each edge connects two nodes. For each graph we can define its adjacency matrix \( A \) defined as \( A_{ij} = 1 \) if nodes \( i \) and \( j \) are connected and \( A_{ij} = 0 \) otherwise. A graph is undirected if \( A_{ij} = A_{ji} \), otherwise is directed. If the set of nodes \( V \) can be partitioned in two sets \( V_1 \) and \( V_2 \), such that nodes in \( V_1 \) are only linked to nodes in \( V_2 \) (and vice-versa), such a graph is bipartite.

Density

The density \( \delta \) of a given network \( G \) is defined as the ratio of the actual number of edges in \( E \) to the number of all the possible edges; that is, for undirected networks

\[
\delta = \frac{2 \times |E|}{|V| \times |V| - 1} \tag{A.1}
\]

Average degree

The average degree \( \langle k \rangle \) is defined as the sum of the edges incident on each node over the number of nodes. With \( k_i \) being the degree of node \( i \), the average degree is defined as

\[
\langle k \rangle = \frac{\sum k_i}{|V|} \tag{A.2}
\]

Average shortest path length

Let \( d(i, j) \), where \( i, j \in V \) denote the shortest distance between nodes \( i \) and \( j \), the average shortest path length \( l_c \) is defined as

\[
l_c = \frac{1}{|V| \times |V| - 1} \sum_{i \neq j} d(i, j) \tag{A.3}
\]

This measure indicates the average length of the shortest path between any two nodes and gives an idea of the efficiency of the network with respect to information flow.

Clustering coefficient

The clustering coefficient of a network captures the cliquishness of a network. It is defined as the number of closed triplets (cycles of length three), also called triangles, over the total number of triplets (paths of length two connecting three nodes) in the network. Let \( C_3 \) denote the number of triangles and \( L_2 \) the number of triplets in \( G \), the clustering coefficient \( C_c \) is

\[
C_c = \frac{C_3}{L_2} \tag{A.4}
\]

The clustering coefficient can also be computed at a node level with the same formula by centering the calculations on the node under analysis.

Degree assortativity coefficient

The degree assortativity coefficient is defined as the Pearson correlation coefficient of the degrees at either ends of an edge (Newman, 2002) and lies in the range \(-1 \leq r \leq 1\). It expresses the extent to which nodes with similar degrees are connected together in the network.

\[
r = \frac{\frac{1}{|E|} \sum_{(i,j) \in E} k_i k_j - \left( \frac{1}{|V|} \sum_{i \in V} k_i \right)^2 \left( \frac{1}{|V|} \sum_{j \in V} k_j \right)^2}{\left( \frac{1}{|E|} \sum_{(i,j) \in E} \frac{1}{2} (k_i^2 + k_j^2) - \left( \frac{1}{|V|} \sum_{i \in V} k_i \right)^2 \right)^{1/2} \left( \frac{1}{|E|} \sum_{(i,j) \in E} \frac{1}{2} (k_i + k_j) - \left( \frac{1}{|V|} \sum_{i \in V} k_i \right) \left( \frac{1}{|V|} \sum_{j \in V} k_j \right) \right)^{1/2}} \tag{A.5}
\]

Positive values of \( r \) indicate that nodes with high (low) degree tend to connect more to nodes with high (low) degree and vice versa. In this case the network is said to be assortative. Negative values of \( r \) indicate that nodes with high (low) degree tend to connect more to nodes with low (high) degree and vice versa. In this case the network is said to be disassortative.

Average connectivity

One measures of how well a network is connected is the connectivity, defined as the minimum number of nodes whose deletion results in a disconnected graph. Two nodes \( i \) and \( j \) in a network \( G \) are said to be \( k \)-connected if there are \( k \) or more pairwise internally disjoint paths between them. Being \( \kappa_G(i, j) \) the maximum value of \( k \) for which \( i \) and \( j \) are \( k \)-connected, the average connectivity of a graph is defined as:
The average connectivity gives the expected number of nodes that must fail in order to disconnect the network. Thus, this measure is related to the topological robustness of the network considered.

Coreness

A k-core is a maximal subgraph that contains nodes of degree k or higher. The coreness index, or shell-index, of a node i is the maximum integer k for which i belongs to a k-core, but not to a k + 1 core. Computing the coreness index is done by deleting recursively nodes with degree less than k, ∀k = 1, 2, ..., k until no node remains in the graph. The value k, that is the maximum coreness, is called degeneracy and represents a global property of the graph, indicating the existence of at least one core of at least k + 1 nodes with degree at least k. The degeneracy is also a measure of sparseness of the network.

The node coreness has been found to be a measure of spreading power (Kitsak et al., 2010; González-Bailón et al., 2011; Pei et al., 2014); that is, the number of nodes reachable from a given one. The nodes at the core of the network, because of their high density of linkages, are likely to have a synergistic mutually reinforcing effect, thus playing the role of opinion leaders (Pei et al., 2014).

Jaccard similarity coefficient and persistence

The Jaccard similarity coefficient measures the overlapping between two sets. Let X and Y be two sets, the Jaccard similarity coefficient is defined as follows

\[ J_{XY} = \frac{|X \cap Y|}{|X \cup Y|} \]  \hspace{1cm} (A.7)

In other words, the Jaccard similarity coefficient is the fraction between the number of elements common to both sets over the total number of unique elements in both sets. In this paper, we use the Jaccard coefficient to measure the structural similarity between the collaboration and the email networks; that is, to measure the extent with which an edge in one network is also present in the other network. We use this indicator also to measure the persistence of a temporal network between two consecutive snapshots.

Quadratic assignment procedure (QAP) test

Assessing the statistical significance of the correlation coefficient between two networks, requires some special attention because network data are auto-correlated as edges are not established randomly. Quadratic assignment procedure (QAP) is a permutation test based on random permutation of rows and columns of the adjacency matrix which lets us compute the statistical significance of a correlation coefficient between networks. Let X and Y be two adjacency matrices, we compute the correlation coefficient \( \rho = \text{corr}(X, Y) \). Subsequently, we generate a set of N permutation values \( \rho_1^*, \rho_2^*, ..., \rho_N^* \) by randomly permuting rows and columns of X and computing the correlation coefficient between the permuted matrix and Y. We can compute the p-value of the permutation test as:

\[ p_{\text{perm}} = \sum_{i=1}^{N} I(\rho_i^* \geq \rho) \frac{1}{N} \]  \hspace{1cm} (A.8)

where \( I(\cdot) \) is the indicator function. Note that the p-value described above is the right-tailed test; however, the conversion to the left- or two-tailed test is straightforward. Here, we described the QAP test for the correlation coefficient, but the same is possible for any other measure of correlation or similarity, such as the Jaccard coefficient.

SI spreading process

In order to understand how the network structure affects the way information flows in a network, we can simulate a dynamical process on the network and compare the properties of the simulated dynamical process across networks with different structures. The SI spreading process is the simplest epidemic process, where a node in the network can be in either the susceptible (S) or the infected (I) state; hence, the name SI. With probability \( \beta \) a node can transition from being susceptible to being infected. At the beginning of the simulation, all the nodes are in the susceptible state. Before the simulation, an initial number of seeds (infected nodes) \( s \) is chosen at random, and the probability of infection \( \beta \) is set. At each round of the simulation, a node i is chosen at random and its neighbourhood is considered. For each infected node in its neighbourhood, node i can be infected with probability \( \beta \). In the SI model, an infected node cannot recover and the simulation finishes when no more nodes can be infected. In Engineering Design, the study of dynamical processes on networks has been used to study the relations between process architecture and iterations (Smith and Eppinger, 1997; Braha and Bar-Yam, 2007), to study the role of people in the propagation of errors and how this relates to the way people are allocated to tasks (Piccolo et al., 2018a), or to study the multifaceted relation between modularity and robustness of complex systems (Piccolo et al., 2020).
Piccolo, S., Lehmann, S., Maier, A., 2020. Modularity, robustness, and change propagation: a multifaceted relation. In: Proceedings of the Design Society: DESIGN Conference, Cambridge University Press. pp. 2335–2344.

Piccolo, S.A., Lehmann, S., Maier, A.M., 2018a. Design process robustness: a bipartite network analysis reveals the central importance of people. Des. Sci. 4.

Piccolo, S.A., Maier, A.M., Lehmann, S., McMahon, C.A., 2019. Iterations as the result of social and technical factors: empirical evidence from a large-scale design project. Res. Eng. Des. 30, 251–270.

Piccolo, S.A., Trauer, J., Wilberg, J., Maier, A.M., et al., 2018b. Understanding task execution time in relation to the multilayer project structure: Empirical evidence. In: Proceedings of the DS 96: The 20th International DSM Conference, pp. 129–138.

Piccolo, S.A., Wilberg, J., Lindemann, U., Maier, A., et al., 2018c. Changes and sentiment: A longitudinal email analysis of a large design project. In: DS 92: Proceedings of the DESIGN 2018 15th International Design Conference, pp. 869–880.

Schaeffer, S.E., 2007. Graph clustering. Comput. Sci. Rev. 1, 27–64.

Smith, R.P., Eppinger, S.D., 1997. Identifying controlling features of engineering design iteration. Manag. Sci. 43, 276–293.

Sorrentino, F., DiBernardo, M., Cuellar, G.H., Boccaletti, S., 2006. Synchronization in weighted scale-free networks with degree-degree correlation. Phys. D: Nonlinear Phenom. 224, 123–129.

Sosa, M., Mihm, J., Browning, T., 2011. Degree distribution and quality in complex engineered systems. J. Mech. Des. 133.

Sosa, M.E., Eppinger, S.D., Rowses, C.M., 2004. The misalignment of product architecture and organizational structure in complex product development. Manag. Sci. 50, 1674–1689.

Sosa, M.E., Mihm, J., Browning, T.R., 2013. Linking cyclicality and product quality. Manuf. Serv. Oper. Manag. 15, 473–491.

Steward, D.V., 1981. The design structure system: a method for managing the design of complex systems. IEEE Trans. Eng. Manag. 71–74.

Tang, F., Ma, J., Thomas, E., 2015. Who knows what in npd teams: communication context, mode, and task contingencies. J. Prod. Innov. Manag. 32, 404–423.

Thurner, S., Hanel, R., Klimek, P., 2018. Introduction to the Theory of Complex Systems. Oxford University Press.

Tyler, J.R., Wilkinson, D.M., Huberman, B.A., 2005. E-mail as spectroscopy: automated discovery of community structure within organizations. Inf. Soc. 21, 143–153.

Walsh, J.P., Maloney, N.G., 2007. Collaboration structure, communication media, and problems in scientific work teams. J. Comput. Mediat. Commun. 12, 712–732.

Wang, Y., Chakrabarti, D., Wang, C., Faloutsos, C., 2003. Epidemic spreading in real networks: an eigenvalue viewpoint. In: 22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings. IEEE. pp. 25–34.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. Nature 393, 440–442.

Wolff, E., 2016. Microservices: Flexible Software Architecture. Addison-Wesley Professional.

Yang, M.C., Jin, Y., 2008. An examination of team effectiveness in distributed and colocated engineering teams. Int. J. Eng. Educ. 24, 400.

Yang, Q., Yang, N., Browning, T.R., Jiang, B., Yao, T., 2019. Clustering product development project organization from the perspective of social network analysis. IEEE Trans. Eng. Manag.

Yang, Q., Zhang, X., Yao, T., 2012. An overlapping-based process model for managing schedule and cost risk in product development. Concurr. Eng. 20, 3–17.

Yassine, A.A., 2019. Managing the development of complex product systems: an integrative literature review. IEEE Trans. Eng. Manag. 68, 1615–1636.