Bottleneck detection of information flow in Finnish tram track infrastructure using graph theory

Tuomas Lievonen

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 22.7.2025

Supervisor

Asst. Prof. Philine Schiewe

Advisor

M.Sc. Mikail Kesenci

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the master's thesis

Author Tuomas Lievonen

Title Bottleneck detection of information flow in Finnish tram track infrastructure using graph theory

Degree programme Master's Programme in Mathematics and Operations Research

Major Systems and Operations Research **Code of major** SCI3055

Supervisor Asst. Prof. Philine Schiewe

Advisor M.Sc. Mikail Kesenci

Date 22.7.2025 **Number of pages** 58 **Language** English

Abstract

In this thesis, we are studying the information flow within a tram track infrastructure project using graph theory. The thesis is divided into three distinct sections. First, we conduct a literature review, where we explore different means of information flow within infrastructure and construction sectors. The second section introduces and defines graph theory concepts we are using later in the thesis. Lastly, we test the theoretical graph theory concepts on a real world case study.

In literature review, we discuss various methods, methodologies and tools that are nowadays used for handling information flow in construction. In addition to general information, we are also doing a deeper comparison between three European countries, Germany, Poland and Finland. We find that real time project management and knowledge management methods are widely used. To facilitate these methods, digital information sharing is becoming the go-to for large construction projects.

In graph theory section, we define basics of graph theory along with three different computational concepts: Maximum flow, minimum cost flow and betweenness centrality. These are the main tools we are using to study the information flow during our case study.

The case study involves information flow of a real life example of a Finnish tram track infrastructure project. We are able to create multiple graphs representing the example through interviews and compute graph theory computations on the graphs.

As a result, we find that maximum flow is an effective analysis method to find bottlenecks within a communication graph. Minimum cost flow gives us a theoretical minimum for sending all the necessary information through the graph. Finally, betweenness centrality gives a ranking on the importance of specific nodes to the information flow of the whole graph.

Even though the results show that the analysis methods used were in fact viable to study the information flow within a project, multiple issues emerged from the practical use of the model. There is a need to further clarify units for the edge capacities and node demands. Furthermore, the model functions only as a retrospective analysis and cannot directly be used beforehand to study the effects of information flow. Finally, the obtained results should not be trusted blindly but used more of as indicative approximations due to the dynamic nature of real world information flow.

Keywords Information flow, Bottlenecks, Tram track infrastructure

Aalto University
School of Science

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Diplomityön tiivistelmä

Tekijä Tuomas Lievonen

Työn nimi Pullonkaulojen havainnointi tiedonvirtauksessa Suomen raitiotieinfrastruktuurissa verkkoteoriaa hyödyntäen

Koulutusohjelma Master's Programme in Mathematics and Operations Research

Pääaine Systems and Operations Research **Pääaineen koodi** SCI3055

Työn valvoja Apulaisprofessori Philine Schiewe

Työn ohjaaja Di Mikail Kesenci

Päivämäärä 22.7.2025 **Sivumäärä** 58 **Kieli** Englanti

Tiivistelmä

Tässä työssä tutkimme tiedonvirtausta raitiotieinfrastruktuurissa verkkoteoriaa hyödyntäen. Työ on jaettu kolmeen erilliseen osaan. Ensimmäisenä teemme kirjallisuuskatsauksen, jossa tarkastelemme eri tapoja tiedonvirtauksen hallintaan infra- ja rakennusalalla. Seuraavaksi esittelemme ja määrittelemme työssä käytettäviä verkkoteorian konsepteja. Lopuksi teemme käytännön esimerkin, jossa tutkimme teoreettisia verkkoteorian konsepteja tosielämän tapauksessa.

Kirjallisuuskatsauksessa keskustelemme eri metodeista, metodologioista sekä työkaluista, joita käytetään nykypäivänä tiedonvirtauksen hallintaan rakennusalalla. Perustietojen lisäksi syvennymme tarkemmin vertaamaan kolmen Eurooppalaisen valtion toimintatapoja keskenään, Saksan, Puolan ja Suomen. Havaitsemme, että reaaliaikainen projektin hallinta sekä tiedonhallinta ovat usein käytettyjä metodeja. Hyödyntääkseen näitä metodeja tehokkaasti digitaalinen tiedonjako on nousemassa ensisijaiseksi tavaksi suurissa rakennusprojekteissa.

Verkkoteoria osuudessa määrittelemme verkkoteorian perusteet sekä kolme laskennallista konseptia. Nämä ovat maksimi virtaus, minimi kustannus virtaus sekä keskinäisyys. Käytämme pääsääntöisesti näitä työkaluja käytännön esimerkissä.

Käytännön esimerkissä tutkimme suomalaisen raitiotiehankkeen tiedonvirtausta. Luomme tapauksesta verkkoja haastattelujen avulla kuvastamaan tapauksen tiedonvirtausta ja suoritamme verkoille verkkoteoriaan pohjautuvia laskentoja.

Tuloksena havaitsemme, että maksimi virtaus on tehokas analysointi menetelmä pullonkaulojen etsimiseen kommunikointiverkossa. Minimi kustannus virtaus antaa teoreettisen minimin kustannuksille, joilla saadaan kaikki tarvittava tieto kuljetettua verkon läpi. Viimeisenä keskinäisyys antaa järjestyksen verkon solmujen tärkeydestä koko verkon tiedonvirtaukselle.

Vaikka tulokset näyttävätkin analyysimetodien olevan tomivia ratkaisuja projektin tiedonvirtauksen analysoimiseen, useita ongelmia löytyi mallin käytännön toteutuksesta. Yksikköjen määritelmiä tulee täsmentää erityisesti linkkien kapasiteeteille sekä solmujen tarpeille. Lisäksi malli toimii vain takautuvana analyysinä eikä sitä voida käyttää suoraan ennustavasti tutkimaan tiedonvirtausta. Lopulta saatuja tuloksia kannattaa hyödyntää vain suuntaa-antavina arvioina tosielämän tiedonvirtauksen dynaamisesta luonteesta johtuen.

Avainsanat Tiedonvirtaus, pullonkaulat, raitiotieinfrastruktuuri

Contents

1 Introduction

Trams are a popular way of public transportation in many large cities due to their reliability, capacity and speed. In Finland, trams have been operating since 1890. First, using horses to pull the trams along tracks and later from early 1900s using electricity (Suomen Raitiotieseura, 2024). Currently there are trams in Helsinki, Espoo and Tampere with ongoing projects of extending the tracks in Helsinki and Tampere. Future projects have also been announced at least in Vantaa and possible tracks are also in development in Turku and Oulu (Turun raitiotie, 2025; Herneoja et al., 2023).

There is a clear desire for building tram tracks as they have been shown to have larger capacities than some other forms of public transportation such as buses and they are more sustainable. Trams also, do not limit traffic as much as they are rarely driving directly among other cars. These projects are large and often have multiple companies participating with different expertise. The stages of building a tram track consists of the ordering, planning, budgeting, scheduling, procurement, production, quality control and handover. These stages can be divided into the planning stage and production stages with some overlapping. Thus, efficient information flow between all relevant parties from planners to contractors and back needs to be ensured. This is not necessarily the case, however, as communication breakdowns have been observed from ground level all the way to upper management. This can cause delays, monetary losses and at worst serious accidents on work sites.

During tram track construction, cities often also improve the overall services and comforts around the tracks at the same time. This tempts construction companies to build new apartment buildings near the tracks which accelerates urbanization as the cities grow and modernize themselves. This makes tram track infrastructure projects one of the more complex construction projects requiring constant and efficient information flow between different sectors.

The need for efficient information flow can be divided into four key concepts. It reduces failure, leads to innovations, positively influences stakeholders and leads to better decision making. All four of these concepts can be managed and improved by using different information flow methods, methodologies and tools.

To lead a large scale project efficiently, real time project management and knowledge management are two vital methods. These methods are widely used in improving information flow and keeping everybody in a project on track. While both project management and knowledge management have been used throughout history, they have become more and more necessary in complex construction projects to improve information flow.

Some methodologies that can be customised for information flow management purposes are information and communication technology and data-analysis. These are commonly used for ensuring high quality management and function as general means to share and analyse information.

Tools are the most concrete practices that are used to facilitate the methodologies and methods. These can vary from singular analysis tools to other tangible ways of improving information flow and are often practical and adaptable to wide variety of scenarios.

Advancements in technology and more specifically digitisation have drastically changed the overall process and handling of information flow. More and more companies are developing and improving their digital platforms to make information easily accessible for all relevant parties and it has shown promise in efficiency in workplaces.

Since the mid 20th century, technological advancements have been made quite consistently and until the last 20 or so years digitalisation especially has experienced huge leaps of progress. Cloud based solutions are used daily for sharing and accessing information easily. They also allow for real time communication and analysis tools which can be shared with relevant parties. Further improvements and development, for example, with AI will make information sharing even easier, simpler and quicker.

We are taking a closer look into the work process, and more specifically, information flow in European tram track infrastructure and construction industry comparing them to Finnish counterparts. These are larger than the Finnish tram track industry and should help us find ways of improving information flow. Construction industry in general is also fairly similar to tram track infrastructure regarding their usage of information and the paths the information follows.

To further study the information flow in a Finnish tram track industry, we are developing a model to detect the process and paths information follows from a source to a sink. This will be done by applying graph theory and we are especially interested in locating bottlenecks within the information flow. This will help us to try to develop methods to combat these bottlenecks and ensure more efficient flow of information. Graph theory can also be used to detect which node is the most vital in the chain of information to see whether there is a need to have a backup for a specific node.

Graph theory is chosen as it is a great mathematical methodology to follow the interactions between different people or groups of people within a system. We can easily track which nodes are connected to each other and how well information can flow through the system. We are using different algorithms and computations from graph theory, such as maximum flow, minimum cost flow and betweenness centrality to measure the efficiency of the process. The maximum flow along with minimum cut is used to find the bottleneck edges of the information flow. Minimum cost flow is used to find the theoretical minimum for the complete information sharing. Finally, betweenness centrality is used to find the importance of nodes.

In addition to introducing theoretical models and algorithms, we are doing a practical example applying graph theory to follow the information flow in Finnish tram track industry to try to detect bottlenecks and hopefully uncover possible solutions. This method will be then applied for a test scenario, where we follow information flow in a change of plans in a Finnish construction company.

Our main focus is to locate bottlenecks in information flow on our example, but we are also interested in finding the most critical nodes and the most optimal costs in information sharing. The critical nodes will help us determine who cannot be ill or on vacation for information to flow through as fast as possible while the optimal cost gives us a theoretical minimum for the costs of the information flow for the specific case.

2 Information flow in European construction

Information flow is a very broad term. We will focus on information flow in the European construction industry from the viewpoint of construction companies. We study the information flow in construction through methods, methodologies and tools. Methods are general means of studying information flow. They can be very broad and even abstract. Methodologies are more focused than methods, which often have some theoretical foundations. Finally, tools are the instruments or ways we use to facilitate information sharing in practice.

2.1 General

The communication problem in construction is a well documented event as construction sites usually have a lot of people working at the same time doing vastly different things. The basic working process in construction and infrastructure projects consists of planning, procurement, construction, monitoring and handover. These are the standard steps needed to successfully complete a complex project. With large projects each of these subdivisions may have multiple different organizations and contractors working on them. In order for the project to be completed on time and on budget, there needs to be open and efficient communication between all relevant parties.

In addition, the information also has to flow properly in between the different production processes which can consist of the design, cost, scheduling, quality and regulatory information. There are many different tools to manage information within a project. Nowadays these are mostly in digital form, which in theory provides a quick way of transporting information as well as keeping the gained information stored for future needs.

The need for efficient communication between different production processes helps keep the project on budget and on time. For example, planning and procurement need to have efficient communication between them so that procurement knows what is needed for the construction process and can then purchase them in advance. Another key motivation for efficient communication in construction is sustainability. Creating and building sustainable solutions that last are important in general to help the world as well as keep customers and stakeholders satisfied. Sustainability is also an important factor for a company's reputation (Pietrzak and Pietrzak, 2022). Most importantly, though, efficient communication reduces risks and improves safety. Construction industry is on the more dangerous side of professions due to many demanding jobs often done in difficult circumstances with changing variables. To improve safety, almost all construction companies already have dedicated units to monitor safety on sites, but with efficient information flow, safety can still be improved upon (Zhang et al., 2013).

Time is another big concern in large construction projects. Even though extensive planning is done for both budgets and time usage for projects, there is always a desire to finish the work earlier than scheduled and below budget. This is where commonly used methods in construction are used to help with information flow. These can be managed, automated and followed by utilizing different softwares and digital tools

(Memon et al., 2014).

The problem of communication in construction has been studied before. The key aspects on why good communication in construction is vital can be divided into four key points. It reduces failure, leads to innovations, positively influences participating stakeholders and lastly leads to better decision making (Hoezen et al., 2006). In order to satisfy all four key points, many different methods, methodologies and tools have been developed for improving information flow.

Construction projects have become increasingly larger and therefore more complex, but luckily information sharing has had huge developments. In complex projects, there are professionals in multiple different fields that need to have a common understanding of the end product. This is where proper information sharing and information flow management come into play. Information flow can be thought of as an interaction, where new information arrives and together with already known information a decision is made and current information is updated. Information that is ignored and not used becomes waste. This process relies heavily on the previously mentioned methods, methodologies and tools and also introduces a concept known as integration which is the act of using available information and knowing how to share it with teams working on the project (Phelps, 2012).

2.1.1 Methods

The first method we are looking into is one of the most common industry trends, real time project management. That is managing a project using information obtained and processed in real time. This can be achieved and monitored using different methodologies and digital tools.

The goal in real time project management is to be able to follow the progress of projects at the time they happen and make decisions based on data and knowledge at hand. Since projects are complicated and often consist of many different subprojects, it can become fairly difficult to stay on top of things. This is where multiple information sharing methodologies and tools come into play. There has been rapid development going on and technological advancements help keep information stored and easily available for analysis and further use. Collaborative tools are favoured and used in real time project management. These tools are nowadays often cloud based and therefore they are easily accessible by selected people at any time (Levitt, 2011).

Another actively used method along with project management is knowledge management. The key in knowledge management is to create value from known knowledge. In knowledge management, information is already processed and internalized and focuses on utilizing knowledge in the best possible way by using known expertise to create innovations and to form best practices (Dave and Koskela, 2009).

Similarly, when we are employing known data and processes from past projects, we can use the term knowledge infrastructure. Its goal is to help manage current projects based on experience from past projects. Knowledge infrastructure can be viewed as a subcategory to knowledge management, where instead of using knowledge obtained from working on the current project it is only using knowledge from past experiences by utilizing for example ways of working that have previously yielded results (Jalote, 2003).

The third and final method that is becoming more widely used is lean thinking, which affects all key concepts in communication. Its key idea is to reduce waste to deliver value to customers effectively by achieving steady and predictable flow of information. Waste can be anything from loitering or procrastinating to doing too much at faulty quality (LCI). The basic principles of lean are to focus on long-term results, create better conditions for working, encourage everyone to improve their practices constantly and lastly create connections that ensure constant improvement at organizational level (Ojalehto, 2023).

One idea in lean construction methods to improve information flow is to create platforms that hold all necessary information. These platforms would bridge the communication between different stages of projects such as planning and production. Using the platforms for information flow would inherently improve the total process as information would be readily available when needed. This makes lean act as more of a submethod to project management (Dave et al., 2014).

Each of the methods have an effect on the four key points of vital communication in construction. Good project management reduces failure and mutually improves decision making while additionally positively influencing stakeholders. Knowledge management, where we actively share and use known knowledge, helps in decision making and could lead to further innovations. Finally, lean thinking helps reduce failure, make better decisions and lead to innovations due to actively focusing on and encouraging people to create better workflow and thus improving the overall success of the project.

2.1.2 Methodologies

One methodology often used in the modern construction industry is information and communication technology (ICT). It is a platform that allows information and knowledge to be shared among collaborators working on a project and provides collaborative knowledge management solutions for large construction projects. Technological tools have been developed and implemented in making information and knowledge sharing as efficient as possible. Knowledge management method benefits from ICT due to simplifying the sharing of gained knowledge. While the digital tools provide good grounds to store and share knowledge, the most important aspects recognized by companies in sharing knowledge is still basic human interaction (Dave and Koskela, 2009).

Another similar methodology often used in the modern construction industry is information management. In short, it is actively collecting, preserving and manipulating acquired data. If it is done digitally, we can use the term information technology (IT). Information management is a broad term that consists of everything from obtaining information, having access to information and storing the information. The information in information management is often some sort of data that can then be analysed and reported (Gyampoh-Vidogah et al., 2003). It is heavily tied to project management as it helps project managers share and obtain information systematically and formally (Froese, 2005).

Trust in information flow is another aspect needed for efficient production process that is closely related to knowledge management. Trust is especially important with projects with multiple organizations such as consortiums or alliances. Trust can be divided into two forms, initial trust and gradually formed trust (Droege et al., 2003). Trust is key when organizations can not be sure how the other party will use shared information. This can naturally hinder information flow. Many factors affect the trust of an organization, such as reputation and field of work. These factors are taken into account when information is shared. In consortiums or alliances, trust is often earned more easily as the goal is common for all parties and information must be shared more openly to achieve the goal. Trust is heavily tied to knowledge management method as we need to obtain trust to openly share knowledge among collaborators to enhance the end product.

Data analysis is a commonly used methodology with many different tools for specific needs. The basics of data analysis is collecting some raw data and using different tools to analyse the data. Some notable applications of data analysis in the information flow include, for example, optimization of decision making, identifying communication patterns for possible bottlenecks and predictive analytics. Analysing all of these can help make decisions that enhance information flow (Durugbo et al., 2013).

The problem with data analysis can be the quality of the data, however there are multiple different tools, one can use when analysing data. The benefits from data analysis is that it is extremely common, meaning there are people, who know what they are doing. Other benefits include the variety of use cases such as, optimization, predictive analysis and overall progress analysis. With optimization we can develop better ways of doing things with given resources. Predictive analysis can be used to make better decisions based on previous results. Finally, overall progress tracking can be used to make quick decisions that help elevate the project and keep workers on track of the progression. Thus data analysis is closely related to knowledge management and project management methods.

Graph theory is the last methodology we discuss when it comes to information flow. It is quite closely related to the aforementioned data analysis due to its mathematical base and applicability. Graph theory is focused on the relationship of nodes and edges that often contain some additional data. It is a powerful methodology to study connections between nodes that can represent, for example people, from an information flow standpoint. Graph theory can be used to optimize the flow within a graph and additionally can function as a visual representation of a specific structure. It can be regarded as a subcategory to knowledge management as the base idea with graph theory is to analyse data and turn it into knowledge.

The methodologies are all functioning in making the information flow as clear and efficient as possible by improving the four key concepts. Functioning ICT solutions provide knowledge available to all relevant parties leading to better decision making and possible innovations. Trust between collaborators influences stakeholders positively and could lead to innovations when collaborators openly share ideas between each other. Data analysis and graph theory help reduce failure by actively analysing current processes and therefore also help make better decisions.

2.1.3 Tools

Lastly, we will discuss some commonly used tools which help facilitate information sharing. Building information model (BIM) is an information management tool that is used for visually sharing and storing information. Most track building projects are not just building the tracks and overhead lines in their place, but also include other urban city construction such as sidewalks, cable pulling, traffic signs and lights, light posts, renovating sewer networks and planting trees. BIM is a helpful tool to visualize the plans of the different aspects of the project as well as visualizing the different constructed sections. For the end BIM product that contains possible changes from the plans, the term as-built model is often used (Nguyen and Adhikari, 2023).

Digital twin is a close relative to BIM. Digital twin is a digital representation of what is being built that is updated constantly. The difference between a digital twin and a BIM is that digital twins operate in real time and can be used for simulation and optimization purposes whereas BIM are static models used for design and construction. Digital twin is created by taking measurements of points and lines from the construction site that are then transformed into digital form and modelled to visually represent the measurement. This is a great method to store information and have it accessible later if needed while also enabling real time analyses. Another positive is that in digital form one can often review underground elements as well with a suitable visual software and diagnose possible problems that might have happened (Nguyen and Adhikari, 2023).

Artificial intelligence or AI is the newest tool that could have multiple different applications, though we have not yet figured how we could get full use out of it. In general, the benefit of AI is that it automates many processes based of history. By giving AI past data it can make helpful and meaningful decisions itself. As AI can be used to automate many processes, it can also help with improving information flow. AI can, for example, be used in construction briefings to transcribe and analyse the ongoing conversation. This can then be used to give feedback and help workers to internalize the given information easier and return back to it if necessary (Pettinger and Nelson, 2025).

A more specific application of AI known as generative AI (GenAI) has shown promise in decreasing risks in construction. Generative AI as the name suggests is focusing on generating new content based on previous data. Whether it is used for technological, operational or integration aspects of risk management, it is vital to understand the perception GenAI provides and create models that can analyse and communicate possible risks fully with the help of GenAI. In construction, GenAI integration can be used, for example, for BIM information handling to accelerate the modelling processes and thus reducing risks. Due to the general interest among companies and researchers and its rapid development, GenAI models are likely to become widely used in future project management purposes. (Mohamed et al., 2025)

GenAI also comes with some challenges. Multiple issues have been documented that include hallucinations, generalizability and accuracy of the results. These issues rise from the limited training the AI has due to available data and costs of training. These issues can and probably will be lessened with time and further development of GenAI (Ghimire et al., 2023).

Another issue with AI and GenAI is that it produces ethical challenges. This includes data security as well as human employment and are more difficult to solve. In addition, GenAI requires human insight to check the work done by the AI to make sure the results it gives are in fact accurate and proper. AI results in construction should never be trusted blindly as AI may miss some information leading to safety issues (Ghimire et al., 2023).

Our final tools we are introducing are about graph theory. For graph theory, there are multiple different tools to choose from. Throughout the rest of the thesis we are using maximum flow, minimum cost flow and betweenness centrality, however there are also other computation tools and centrality measures that could be used. Maximum flow can be used to determine the amount of information a network can handle. Closely related to the maximum flow is minimum cut, which can be used to determine the edges that act as bottlenecks for the graph. Minimum cost flow problem on a network is finding an optimized flow from source to sink that satisfies all necessary supply and demand while keeping the costs to a minimum. Betweenness centrality measures the importance of nodes by computing the ratio of shortest paths to shortest paths through a studied node. This gives each node a value showing how often the shortest paths from other nodes go through the node thereby determining the importance of that node to the graph (Jungnickel and Jungnickel, 2005; Needham and Hodler, 2019).

Finally, we are going to relate the tools into how they affect the key concepts of information flow. Both BIM and digital twin are already heavily used in infrastructure construction processes to visualize the plans and to simulate the efficiency of the plans. They can also be used to track processes in real time which is especially effective in construction and leads to better decision making and reduces failure. Additionally, an as-built model has great influence for stakeholders as the model holds all information used during construction in one place for easy access for future follow ups. AI can affect all key points based on the specific use case. This makes it difficult to separate some of the key points where it would have little to no effect. Lastly, data analysis and graph theory tools can be used to greatly improve decision making as well as reduce failure.

2.2 Country specific information flow

Now that we have discovered and discussed some general methods, methodologies and tools used to handle information flow in Europe, we will focus more on three specific countries in Europe to find which methodologies and tools they use and how. One of the countries is Finland, but we want to compare the information flow with two other countries. In central Europe, construction and tram track industry and infrastructure are generally larger than in Finland. We will focus more on the most prominent countries with large tram track infrastructure. In this case, we have chosen

Germany and Poland. The assumption is that these countries have mastered the tram track infrastructure due to their large infrastructure networks and we can learn valuable methods of communication and information flow from them. It should be noted that with many infrastructure projects, there is often more than just building roads or in this case tram tracks. Usually, these projects include general urban landscape developments as well, since it is optimal to try to inconvenience the general public as little as possible while doing as much as you can at once. This further makes it important to have good communication between all smaller construction projects to create a cohesive final product to the city. Furthermore, even though tram track infrastructure is its own special case of construction, there are not too many differences that would make general construction sources unreliable to adapt into tram infrastructure.

2.2.1 Germany

Germany is one of the largest and most advanced countries in Europe with large infrastructure. This indicates that they are doing something right to achieve their large public transport system and more specifically the tram track infrastructure. Germany has built tram tracks to increase public transport capacity and comfort. Trams also function overall as more sustainable and cost-efficient compared to some other types of transport (European Investment Bank, 2023).

In Germany many construction projects operate on a Design-bid-build (DBB) model, where contractors are selected in a two-step process including qualifications, experience and finally price (Michaela Boneva, 2022). Afterwards, the selected companies often form consortiums, where the companies form a temporary joint partnership to finish a project (Railway Technology, 2017). Consortiums are often more free form collaborations that aim to bring more value into a project. In order to achieve a working consortium, information must be shared with the participating organizations.

Planning processes for infrastructure projects have followed a similar guideline in Germany for years often running over their allotted budget and time. The issues come from long approval processes and complexity of the projects with often changing plans. Lacking knowledge management has been regarded as one possibility for the delays of infrastructure projects in Germany. Improving the success of infrastructure projects can be achieved by improving project management using modern techniques during the first stages of a project (Sözüer and Spang, 2012).

Accomplishing a functioning working flow for a project, information sharing must be fast and efficient. Real time project management in a German rail project has previously been handled by developing information and communication technologies (ICT) by implementing logistics, operator and technical requirements into a new ICT platform. This can later on be used to develop the digital twin for the project as well (Deutsche Bahn, 2021).

With ICT solutions, the information flow can generally be improved and simplified. ICT solutions are, however, not necessarily the easiest way to improve communication and they have shown issues regarding information flow improvement due to their complexity. To benefit from ICT solutions, it requires extensive knowledge of the tools used. Additionally, it has been theorized that improving information flow is more of a social than a technical problem (Butzin and Rehfeld, 2013).

An important digital tool in sharing information in German construction is Building information modelling. BIM has been used in complex building construction in Germany, which would indicate its usefulness also in tram track infrastructure construction. It is used as the basis for interface optimization throughout all phases of projects from planning to production. Additionally, to improve information flow, regular communication about construction timelines and content with the stakeholders is required. Naturally, all stakeholders need to have sufficient knowledge in BIM adaptation to effectively use it to share information with the other parties (Hartmann et al., 2023).

The information flow in German rail construction projects relies more and more on modern digital tools that allow for real time project management and data driven solutions. The projects are often done in consortiums of multiple contractors due to their size and complexity, which further entails need for efficient information flow. Germany uses and develops ICT solutions and BIM for improving information flow. Even with the promising results that digital tools bring, they are not a magical solution for improving communication and requires sufficient input from project managers and workers alike.

2.2.2 Poland

Next, we will take a closer look into the information flow in construction in Poland. New tram lines were completed in Poland recently where there were multiple organizations working in a consortium to achieve the multi-year project (Polimex Mostostal, 2024). Historically Poland has used consortiums in tram track construction (Ferrovial newsroom, 2018). The reason Poland is investing in tram tracks is to try to encourage people to use public transport rather than cars for safety reasons and because trams are more sustainable (European Commission, 2019). In addition, the trams have better capacity than traditional buses during rush hours (European Commission, 2020).

In Poland, a study of information flow in construction projects was conducted in 2020 using surveys from a total of 160 construction companies. The study found that good communication contributes highly to project's success and is perceived by experts as the most vital aspect in achieving desired results when carrying out the construction projects. Also, the lack of effective communication directly leads to delays and additional costs. In order, to guarantee efficient information flow, it was suggested that further research is directed into developing tools for project managers to actively and optimally control the information flow to reach common goals. It was also found that obtaining good results was more due to making sure that information reaches critical people on time rather than just having enough data overall (Kania et al., 2020).

IT tools have been shown to improve information management in Poland providing faster and more accurate information and more transparency in communication. This increases productivity and reduces costs of the project in the long run. Additionally, applying IT systems is becoming the norm in the field increasing the desirability and value of the company while also gaining a competitive edge to other construction companies (Romańczuk et al., 2023).

While Poland still often opts for the traditional planning and production of projects, they, similar to Germany, are starting to adopt BIM as a tool for information flow in construction. There is a clear need for BIM implementation as it is becoming the standard around the world in the developed countries. The issue with adopting BIM solutions are often due to the high costs and lack of experience, however Poland is planning to introduce standards for public projects that should initiate wider use of BIM in construction projects in the future (Zima et al., 2020).

Even though digital tools have had huge developments and have been shown to improve information flow, construction companies in Poland have been hesitant to apply IT solutions due to their initial costs and lack of experience. Some companies actively implement new solutions, but some still use the old and trusted yet slow techniques of information sharing; spreadsheets and paper (Kapliński, 2009).

Taking these into account, the information flow in Polish infrastructure needs to follow strict regulations as the projects are often done in consortiums. This does not allow as free flowing information flow as some other forms of collaboration. Furthermore, the costs associated with implementing IT solutions can cause hesitancy among company leaders. Nevertheless, implementing these solutions have shown great promise in helping information flow and providing better results.

2.2.3 Finland

Lastly, we will take a look into Finnish construction and infrastructure firms and their information flow methods. Construction companies in Finland are generally larger than any tram track infrastructure company. Same principles from construction information flow can still be easily applied to the tram track industry as it is essentially a more focused example of construction.

At the moment, there are a few tram track infrastructure projects going on in Finland in Helsinki and in Tampere. All of these projects are done using alliance model of working which is similar to consortiums but generally more long lasting. An alliance consists of multiple different organizations working towards a common goal. The main idea is that in alliances all parties are equal in that all wins and losses are shared equally. In alliances, even though they consist of multiple different companies that are commonly competitors, the information must be shared openly. From the previously mentioned methodologies trust plays a role in information sharing. Trust, though, is ensured within the alliance model as mentioned everyone tries their best to create the best possible results as it guarantees better winnings for all parties (YIT).

Another key benefit from alliances is that it brings people from different backgrounds together working on a solution. As some construction companies participating in alliances are more focused towards planning, and some towards production. For example with many tram projects in Helsinki, bridge building over a sea needs to be done. This brings even more complexity to the projects and requires more specific expertise and more accurate information flow among planners and workers from different backgrounds.

Finland is also active with the lean construction in the tram track industry. The alliances have adopted lean construction in trying to reduce waste and provide value for stakeholders. Some tram track projects that have used lean in Finland are Jokeri Light Rail, Crown Bridges and Pirkkala-Linnainmaa. The benefits of lean are measured using different measurements and by organizing regular follow up workshops (NRC Group, 2024).

To improve information flow with digital tools, Finland has extensively adopted and developed the use of BIM as a design and simulation tool for complex structures. This can be seen as BIM design for Kruunuvuori Bridge in Helsinki was the winner of a Tekla BIM award in 2024. A final BIM product will also be separately modelled that functions as the as-built model and digital twin of the bridge (Tekla, 2024).

All in all, Finnish tram track industry is often operated using the alliance model of working. This brings more expertise into the projects, but also increases the demand for efficient information flow. To help facilitate the information flow, lean construction method has been thoroughly adopted and utilized in the projects. Technological advancements have also been made on the ground of building information modelling to store and share as detailed information as possible and additionally simulate different conditions for the operating tram.

2.3 Information flow development

To develop communication practices, there should be someone in charge of the project who is able to communicate between all parties. They need to have a very large capacity to keep and understand all received and necessary information. This is where project managers and real time information management tools come into picture.

The use of digital tools is also beneficial as they provide the best possible way of real-time project management to date. The use of more advanced tools like artificial intelligence can also already be utilized to a degree. It is going through the biggest advancements and being developed more than ever and is the hottest new trend on the market. AI can be used to automate many operations currently operated on by manpower. It is AIs biggest advantage, however it has been criticized for making certain jobs obsolete.

Information sharing is already being done by using cloud based tools while trying to limit the use of previously well versed solutions such as e-mails. E-mails are a great tool to be used when contacting multiple people to share general information but it is quite slow compared to cloud based information sharing platforms. There is no guarantee that e-mails are answered and information may already change by the time the e-mail is sent. Therefore, real time communication is useful for sharing plans and production progress. E-mails do have a place, though, in giving a summary but for active information sharing it is far too slow.

2.4 Conclusions

There are multiple different methods, methodologies and tools that can be used to manage information flow in construction. It is not feasible to assume that all possible methods and methodologies could be taken into account in a project, but rather try to take the key points from many different useful methods. It may also be more harmful than beneficial to try to force too many methods of information sharing with people as overflow of information can be problematic.

It is also important to distinguish the differences between methods, methodologies and tools used in information flow. Methods are the broadest category and can be fairly flexible. Methodologies are more structured and theoretical in studying the information flow and how it can be managed whereas tools are the actual practical specific ways of sharing information.

Information flow management from the viewpoint of tram infrastructure is already often done utilizing BIM modeling and digital twins. AI is also becoming more widely used due to its rapid development.

Graph theory can be used to measure the efficiency of information flow, using network flow computations such as maximum flow, minimum cost flow and betweenness centrality. This is what we will be studying further by utilizing theory and testing it on a case study.

Finally, there are many different studies being done regarding information flow in construction. Construction industry, however, is very large with many different subindustries such as tram track infrastructure from which there are not many studies done. Since tram track infrastructure is not that different from other forms of construction such as building construction, the results obtained from literature can mostly be applied to the tram track infrastructure case as is.

There are no drastic differences occurring in different European countries and it usually comes down to the model of working and not necessarily to the act of information flow. Alliance models are great in bringing people from different background together to work on complex projects further emphasizing need for efficient and active communication. Mostly it boils down to the size of the projects. The key in information flow is cloud based solutions that can be operated on in real time. These methods ensure the goals of reducing total costs, constructing sustainably and keeping the workers safe by automating many processes. The main issues come from educating people on the use of digital devices and data integration problems. These can make the initial investment slightly larger, however the end results are often better. In the end, the benefits outweigh the possible problems in using modern methods, methodologies and tools.

The methods, methodologies and tools are not only restricted to help with information sharing, but also provide value improving the projects in other ways. For example, lean thinking is used to increase productivity and reduce waste and thus it is not entirely about improving information flow. Additionally, data analysis is more about analysing existing data to create better decisions. All of the methods, methodologies and tools can however still be used in some ways to improve information flow, which in turn increases productivity and overall success of the project.

3 Graph theory

Graph theory is a field of mathematics that studies graphs. A graph is a set of nodes which can be connected to each other with edges. Graph theory has variety of applications such as finding the most cost effective path between a set of nodes, finding the optimal location for nodes or finding the most important nodes in a graph.

We are using graph theory to study the behaviour of information flow within an organization to search for possible bottlenecks, important nodes and least costly paths for information to travel.

3.1 Notation

We will define the key aspects of graph theory we are going to use in the thesis. The following notation is used throughout the thesis.

Definition 3.1. (Verstraëte, 2020) A graph or a digraph $G(V, E)$ is a pair (V, E), where V is the set of nodes in a graph and E is a set of pairs of elements of V called edges. A node set $V = \{u_1, u_2, ..., u_n\}$ is a set that contains all the nodes of the graph. An edge set $E = \{e_1, e_2, ..., e_m\}$ contains all edges of the graph that connect two nodes to each other. Each edge is connected to two nodes, where the starting node for the edge is the initial node and the end node for the edge is the terminal node. An edge can be written as $e = (u, v)$, where u is the initial node of the edge and v is the terminal node of the edge.

Lower case letters u, v and w refer to an arbitrary node on the graph and s and t refer to the source and sink respectively.

Definition 3.2. (Cormen et al., 2022) A flow network is a weighted digraph $G(V, E)$, where each edge (u, v) has a non-negative weight $a(u, v) \geq 0$ and a non-negative capacity $c(u, v) \geq 0$. In addition, there are at least two nodes, a source s and a sink t.

A flow function is a function $f : V \times V \to \mathbb{R}$ that satisfies two feasibility conditions. The capacity constraint and the conservation of flow.

The value of flow on a network out of the source s is defined as: $|f| = \sum_{v \in V} f(s, v) - \sum_{v \in V} f(v, s)$.

Capacity constraint is defined as: $\forall\ (u, v) \in E$, the inequality $0 \leq f(u, v) \leq c(u, v)$ must hold.

Conservation of flow is defined as: $\forall\ u \in V \backslash \{s, t\}$, the equality $\sum_{(u,v) \in E} f(u, v) - \sum_{(v,u) \in E} f(v, u) = 0$ must hold. When $(u, v) \notin E,\ f(u, v) = 0$.

A flow is a unit that is sent from a source s to a sink t. $f(u, v)$ denotes the flow in the graph between nodes u and v. An $s - t$ flow is the total flow sent from source s to sink t. A flow network may contain multiple sources s and sinks t. These can be reduced to just one source and one sink by adding artificial nodes known as a supersource and a supersink. A supersource is a source that is connected to each individual source with a weight of 0 and a capacity of ∞. A supersink is similar but connected to all the individual sinks.

Definition 3.3. (Boykov and Veksler, 2006) A cut in a graph is a partitioning of nodes into two disjoint subsets. A cut-set is the set of edges, where the endpoints of an edge are in different subsets. An $s - t$ cut is a cut, where the source s and sink t are in different subsets. The capacity of an $s - t$ cut is the sum of capacities of edges in the cut-set.

Definition 3.4. (Zhang and Chartrand, 2006) A path of n steps on a graph from node u_0 to node u_n is an alternating sequence of nodes and edges, where each node apart from the start or end node of the path is a terminal node for the previous edge and an initial node for the following edge with no repeated nodes or edges apart from the possible starting and ending node. A path can be written in the form $P = (u_0, u_1, ..., u_{n-1}, u_n)$, which shows the nodes in order of visit.

In Figure 1 is a graph that we will use to demonstrate the different graph properties and computations we are using in our analysis. The graph has $|V| = 10 =: n$ nodes and $|E| = 15 =: m$ edges. We can represent this graph as $G(\{s, 1, 2, 3, 4, 5, 6, t_1, t_2, t\}$, $\{s1, s3, 12, 1t_1, 24, 35, 36, 3t_2, 45, 4t_1, 54, 56, 6t_2, t_1t, t_2t\})$. Out of these nodes, the graph has one source s and two sinks, t_1 and t_2 and a supersink t. It is not uncommon to have information start from one source that is needed by multiple end users or have information come from many sources and end up into one sink or a combination of the two. A supersource and a supersink are used when we want to study the properties of the whole graph. The weights or costs of each edge and the capacities of each edge are depicted next to an edge, where the first number denotes the weight and the second number denotes the capacity. All edges leaving the supersource to actual sources and all edges from actual sinks to the supersink will be assigned a weight of 0 and a capacity of ∞. This will ensure that the added supersource or supersink does not affect the computations made on the graph. It should be noted that when using the standard flow network computations as in this thesis, using a supersource or a supersink only works if there is only one type of commodity or flow. Otherwise, if the graph contains multiple commodities it is a multicommodity flow problem and the use of supersources and supersinks, as well as, computation algorithms become more complex (Hall et al., 2007).

One possible path P for the graph in Figure 1 from source s to supersink t is, for example, $P = (s, 3, 6, t_2, t)$.

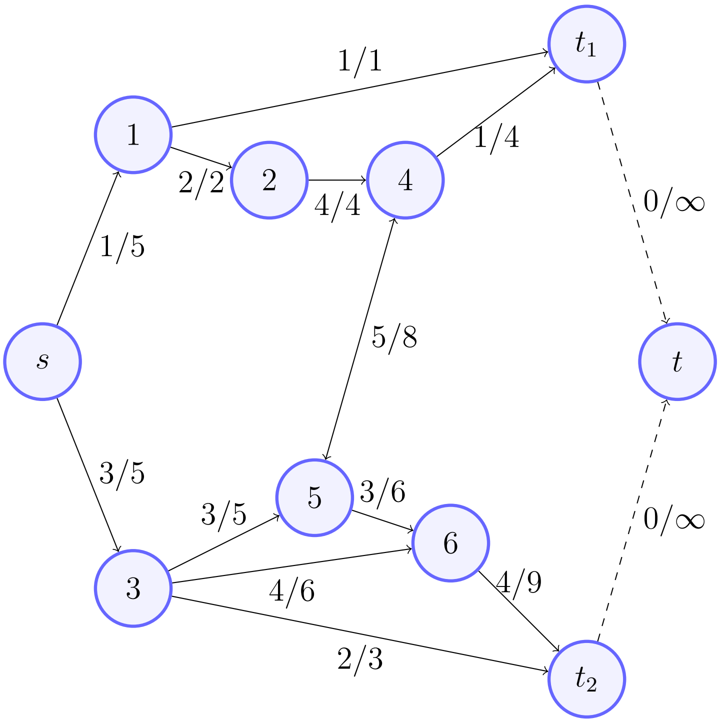

Figure 1: Graph used for examples of different graph measures, where nodes t_1 and t_2 are the actual sinks and t is a super sink.

3.2 Shortest paths

Shortest paths in graphs measure the shortest path from start node to target node regardless of capacity if such exists. The shortest path works for both unweighted and positively weighted graphs. Special cases of shortest paths include single source shortest path (SSSP) and all pairs shortest path (APSP). In SSSP we want to find the shortest path from one root node to all other nodes. In APSP we are want to find the shortest path between all possible pairs of nodes. Shortest paths between all pairs of nodes are needed to compute the betweenness centrality we use to determine the importance of nodes.

The definition of shortest path from one node u to another v is as follows:

Definition 3.5. (Brossard and Morrow, 2010) Let $G(V, E)$ be a directed or undirected graph, as in Definition 3.1, with edge weights $a : E \to \mathbb{R}_+$. Assume that nodes u and v are connected by some path. We want to find the path from u to v for which path $P = (u_0, u_1, ..., u_n)$, where $u_0 = u$ and $u_n = v$, minimizes the sum $C = \sum_{i=0}^{n-1} a(u_i, u_{i+1})$.

Both the SSSP and APSP follow from the Definition 3.5 by computing the shortest path for required pairs of nodes. The SSSP can be solved, for example, with Dijkstra's algorithm or Bellman-Ford algorithm (Schrijver, 2012) and the APSP can be solved, for example, with Floyd-Warshall algorithm or simply by running Dijkstra's algorithm multiple times (Chan, 2007).

In Figure 2, we have computed the shortest path from source s to sinks t_1 and t_2 separately and with unit weights and capacities. The capacity of an edge does not affect the result of the shortest path. It is quite straightforward to see that for both problems the shortest path is of length two.

All the numerical computations are done using Python packages networkx, matpltolib.pyplot, pandas and numpy.

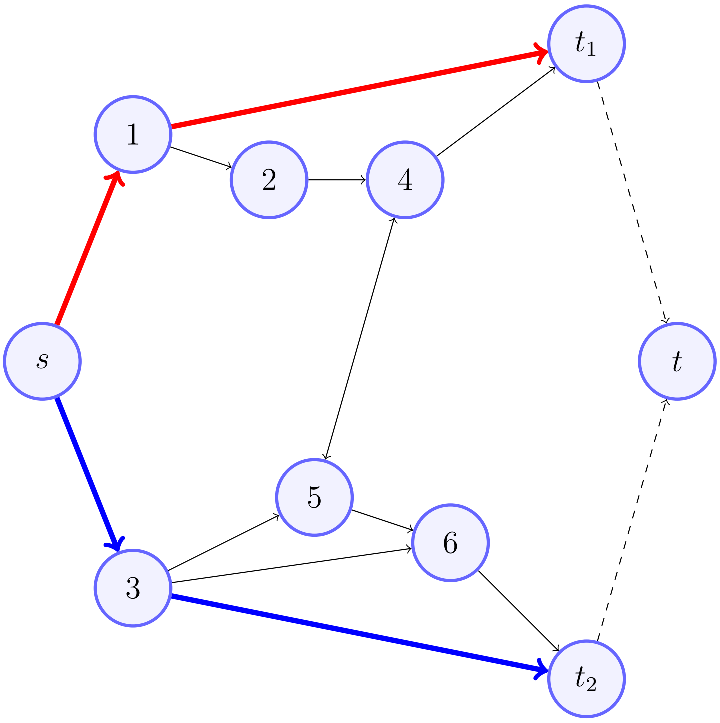

Figure 2: Graph in which the shortest path from source s to sink t_1 is in red and from source s to sink t_2 is in blue.

3.3 Maximum flow

Maximum flow in a graph corresponds to the maximum amount of information that can flow through the graph from a source s to a sink t in a given time. The value of flow was defined in Definition 3.2. Our goal is to maximize this flow (Cormen et al., 2022).

To find a maximum flow in a graph, there are two constraints that need to be satisfied. These are the capacity constraint and the conservation of flow. These were defined in Definition 3.2 and Definition 3.2, respectively. The capacity constraint restricts the flow on an edge so that it does not exceed the capacity of the edge. The conservation of flow states that apart from source and sink nodes the flow flowing into a node must equal the flow out of the node.

The maximum flow problem can thus be expressed as a LP problem (Cormen et al., 2022):

$$
\begin{aligned}
\max \quad & \sum_{v \in V} f(s,v) - \sum_{v \in V} f(v,s) \\
\text{s.t.} \quad & \qquad\qquad f(u,v) \leq c(u,v), \forall\ (u,v) \in E, \\
& \sum_{(u,v)\in E} f(u,v) - \sum_{(v,u)\in E} f(v,u) = 0, \qquad \forall\ u \in V \backslash \{s,t\}, \\
& \qquad\qquad f(u,v) \geq 0, \qquad \forall\ (u,v) \in E
\end{aligned}
\tag{1}
$$

In Figure 3 we have computed the maximum flow for the graph from source s to supersink t. The maximum flow value from s to t is 8.

Maximum flow can be computed, for example, with the Ford-Fulkerson algorithm or Edmonds-Karp algorithm. Ford-Fulkerson runs in $O(E|f|)$ time and the Edmons-Karp runs in $O(VE^2)$ time (Cormen et al., 2022).

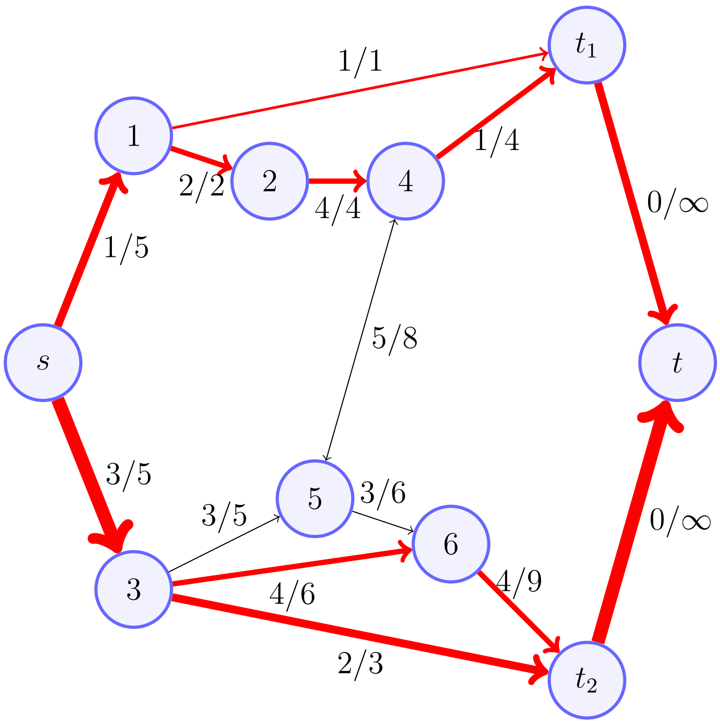

Figure 3: Graph that shows the maximum flow from source s to supersink t in red. The width of the edge shows the amount of flow flowing through the edge.

Maximum flows can be also computed by finding the minimum cut of the graph, where a cut is defined as in Definition 3.3. This is a well-known theorem in graph theory known as the max-flow min-cut theorem and it is written more formally in Theorem 3.6.

Theorem 3.6. *(Korte et al., 2011) In a network, the maximum value of an $s-t$ flow is equal to the minimum capacity of an $s-t$ cut.*

The minimum $s-t$ cut gives as a result two partitions, where the source s and the sink t are in different partitions and the edges connecting the partitions are the minimum cut edges. These edges can be thought of as the bottleneck edges, where by decreasing the capacity of any of the bottleneck edges, the maximum flow of the system decreases. Note that increasing the capacity of any of these edges will increase the maximum flow of the system provided that there is only one minimum cut in the first place. The sum of capacities of the edges in a minimum cut set equals the maximum flow of the system.

3.4 Minimum cost flow

Minimum cost flow problem (MCFP) is a problem, where we try to minimize the costs of flow. In order to fully define the problem, we need to introduce a demand for the network. Demand is the amount of flow we want to pass through the network from sources to sinks.

Definition 3.7. (Ahuja et al., 1993) Demand d_u of the system is defined as:

$$\begin{aligned} d_u > 0, &\text{for source nodes} \\ d_u < 0, &\text{for sink nodes} \\ d_u = 0, &\text{for intermediate nodes} \end{aligned}$$

The demand of the network is different, whether the nodes are sources, sinks or neither (intermediate nodes). Source nodes produce flow into the network and sink nodes consume flow from the network. Intermediate nodes are nodes through which flow passes to reach the sink from the source. For the MCFP to be well defined, the sum of the total demand of the network must be equal to zero.

$$\sum_{u \in V} d_u = 0 \tag{2}$$

Minimum cost flow is thus finding a flow in a flow network that has the least cost associated with it while supplying all the demand. The LP formulation of minimum cost flow in a graph is defined as (Ahuja et al., 1993):

$$\begin{aligned} \min \quad & \sum_{(u,v) \in E} a(u,v) f(u,v) \\ \text{s.t.} \quad & f(u,v) \leq c(u,v), \forall\ (u,v) \in E, \\ & \sum_{(u,v) \in E} f(u,v) - \sum_{(v,u) \in E} f(v,u) = d_u, \quad \forall\ (u,v) \in E, \\ & f(u,v) \geq 0, \quad \forall\ (u,v) \in E \end{aligned} \tag{3}$$

There are again a few constraints that need to be satisfied. These are similar to the constraints in the maximum flow problem, as the capacity constraint is the same as in Definition 3.2 and the conservation of flow is the same as in Definition 3.2 for intermediate nodes. In addition, the conservation of flow implicitly contains the feasibility condition for the total demand of the network as defined in Equation (2).

This simply states that the total demand of the network must be zero. In other words, the total flow leaving the sources must equal the total flow entering the sinks

(Ahuja et al., 1993). Otherwise there would be excess supply or excess demand in the network and the problem would be infeasible.

We thus want to minimize the sum of edge costs the flow passes through such that the capacity constraint and conservation of flow is satisfied for the whole network. In addition, the flow has to be larger or equal to zero for all edges $(u, v) \in E$.

Note that the shortest path problem is a special case of MCFP, where a flow unit of 1 is sent from source s to sink t and the edges have infinite capacity.

Given the graph shown in Figure 1, the minimum cost flow of the graph is depicted in Figure 4, where the supply from source is 8 units and each sink has a demand of 4 units. All red edges are used while black edges are not used and the width of an edge shows the amount of flow flowing through the specific edge. The minimum cost of flow for the example graph is 56.

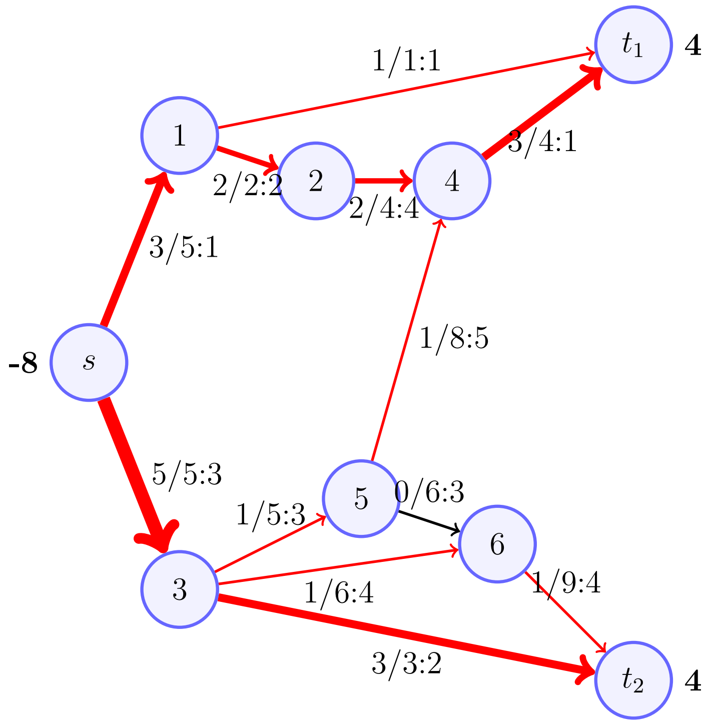

Figure 4: Example of minimum cost flow with supply of 8 and demand of 4 to both sinks t_1 and t_2. The minimum cost flow value is 56. The red edges have flow flowing through them and the width of the edge shows the amount of flow. The edge labels are of the form $f(u, v)/c(u, v) : a$.

3.5 Betweenness centrality

Betweenness centrality is a graph theory concept which measures the importance of a node in a graph. It gives each node in the graph a numeric value based on how often the node appears on the shortest path between other nodes indicating the importance of that node to the overall graph. The larger the value is, the higher the importance of that node is.

Betweenness centrality can be defined as follows (White and Borgatti, 1994):

Definition 3.8. Given a directed graph $G(V, E)$, let σ_{uw} be the number of shortest paths between nodes u and w and let $\sigma_{uw}(v)$ be the number of shortest paths between nodes u and w that pass through the node v at some point in the path. Then the betweenness centrality for node v is:

$$C_B(v) = \sum_{u \neq v \neq w} \frac{\sigma_{uw}(v)}{\sigma_{uw}} \tag{4}$$

In betweenness centrality, we compute the proportion that node v is along the shortest path between two other nodes and sum over all possible pairs of nodes. In weighted graphs, the shortest paths for betweenness centrality are the paths that minimize edge weights.

This method is costly to compute betweenness centrality as we need to compute the shortest path between all pairs of nodes and also find the amount of shortest paths that have an intermediary node v. The graphs in this thesis, however, are not going to be large enough to warrant a faster algorithm. If needed, one such algorithm is detailed by Ulrik Brandes in 2001 (Brandes, 2001).

In Figure 5 is shown the betweenness centrality measures for each node. The larger the node is, the larger the betweenness centrality measure is. This shows that nodes 4 and 5 have the largest measures indicating that they are the most important nodes in the graph. This makes sense considering that both of these nodes have to be in the shortest path between pairs of nodes from upper and lower part of the graph, for example, shortest path from node 2 to node 6 is $P = (2, 4, 5, 6)$.

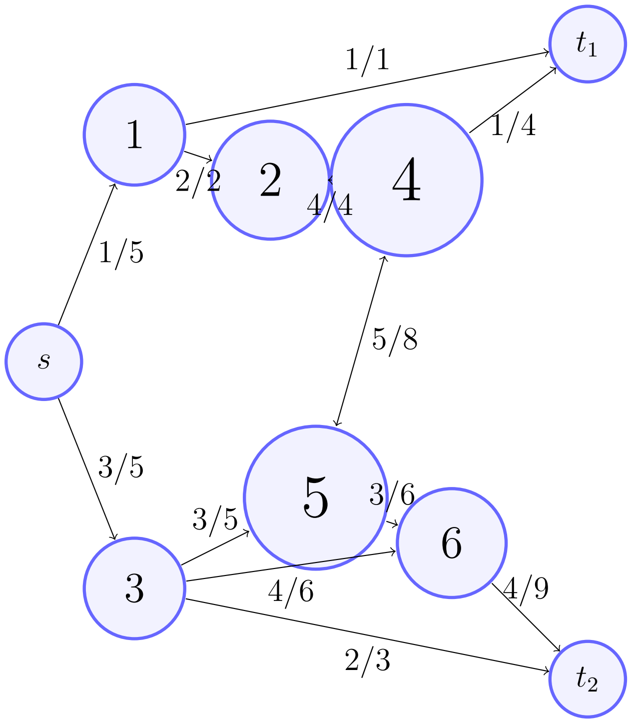

Figure 5: Graph showing the betweenness centrality measures for each node. The size of the node shows the importance of the node to the system.

4 Case study

Based on our model defined previously, we want to test it by studying a practical real world example. We will use the model to follow the information flow regarding a change of plans in a Finnish tram track industry project. This case has multiple people and two different organizations associated with it, which should make it a good test for our model.

4.1 Research material and methods

For the practical study of information flow, we are using graph theory. The goal is to follow all units of information through the graph from the source to the sinks and see how many different paths the information actually goes through, what are the best possible routes and how important different people are in the graph. All of these can be used to determine the efficiency of the information flow. To do these analyses for the resulting graph we are employing our previously defined graph theory concepts of maximum flow, minimum cost flow and betweenness centrality.

We implement the information into a graph as a flow network, where each node is a person the information goes through and each edge is directed to show the information flow direction. Each edge has some capacity as people cannot possibly communicate infinite amount of information to each other. In addition, the edges are weighted to simulate the amount of cost information flow has when flowing from one node to the next.

We should note that graph theory on its own does not take into account the changes that may happen in the information flow through nodes. This is not taken into account in the graph computations on this thesis due to their complexity as it would require either a method that simulates the effects of deteriorating information or the use of dynamic graphs. For more about dynamic graphs, see for example (Hoppe, 1995; Fonoberova, 2010).

To create a graph to test our model, interviews from various people associated with the project at hand are conducted. These interviews are used to obtain information about the connections people have with each other as well as the data during the information sharing. The data in this case is the weight and capacity of the edge and the total demand of the system. The demand needs to be taken into account for the computation of a minimum cost flow.

To find bottlenecks in the graph, we are computing minimal cut sets. We also conduct different graph computations to find other interesting aspects of the information flow, such as the most important node and the minimum cost needed to push all of the information through the system. The maximum flow is used to find the limit of the information flow in the graph. The minimum cost flow is used to measure the costs of sending the required amount of information through the entire graph. This flow may be smaller than the maximum flow of the system. Finally, betweenness centrality is used to study the effect of individuals in the graph. All of these methods are also applied to graphs, where individual nodes are removed to study the efficiency of information flow in a limited graph. Graphs where one node is removed at a time is used to simulate a person being unavailable for some reason, for example, due to vacation or illness. We can then study how well the information flows with this restriction.

4.2 Bottlenecks

One of our biggest goals is to find bottlenecks in information flow. A bottleneck in traditional terms, is an integral part of a network, where flow gets blocked and accumulates backlog. That is that there is more flow going into an opening than there is leaving out. A minimum cut can be used to find these bottlenecks. A minimum $s-t$ cut in essence gives two subsets containing all the nodes of the graph, where the start of the flow, source s and the the end of the flow, sink t are in different subsets. The edges that connect these two subsets are the tightest constraints that the system can have. If any one of the edges would be removed, the flow of the network would decrease and therefore the edge is a bottleneck for the system. On the other hand, increasing the capacity of any of these edges should be done first when wanting to increase the overall flow in the system, although this does not guarantee increase in maximum flow. Increasing the capacity of an edge in a minimum cut set increases the maximum flow only if the minimum cut is unique.

4.3 Weights and capacities

A flow network has capacity and often weight associated with each edge in the graph as well as demand for the source and sink nodes. In order to take into account the time or other possible costs it takes for the information to traverse an edge, we add weights and capacities into our graph model. The weights are non-negative $a(u, v) \geq 0$ and capacities are non-negative $c(u, v) \geq 0$.

We use the edge weights to measure the time it takes for the information to travel from one node to the next and capacities to limit the amount of information that can flow from one node to the next. Demand on the other hand is the actual amount of information we want to send through the system. The difficulty in the practical example comes from assigning the weights, capacities and demands for the nodes and edges.

In order to assign the weights into our model, we do not take into account every possible way a piece of information can travel, but rather focus on the two most prominently used and easy to track information sharing methods. One is written e-mails, documents and messages and the other is arranged meetings. Now to create a graph that resembles real life as closely as possible, we need to find a reasonable way to represent these two information sharing methods using the same units, since messages are easy to measure in quantity, but meetings are easy to measure in amount of time. Therefore we try to approximate the amount of time a message takes to write, send and receive to change the quantity units into time.

Before we combine the weights into a single graph, however, we first create two separate weighted graphs, where one of the graphs is weighted with only messages and the other is weighted with meeting times. We can then compare these two graphs and their computations to see what differences they have and if they seem reasonable.

If we measure the combined weights as time, we can approximate the amount of time a message takes to be internalized. Typically we can say that a message has benefits in that you can always return to it to check what information it contains, but meetings are better in internalizing the new information and provides an easier way of asking questions.

4.4 Interviews

In order to test and use our model in a real world scenario, we need to create a graph from scratch and assign edge weights, capacities and demands. We conduct interviews of people working in a Finnish construction project and use the information gained to build the graph.

There were in total 8 interviews with people from the construction project. The goal of the interviews was to gain a thorough understanding of the problem and to find out how it was handled with the emphasis on the information flow within different groups. Another goal was to gain approximate data to be used as weights, capacities and supply/demands on the graphs.

The interviews were conducted by sending each interviewee the summary of the case and the interview questions beforehand. The most relevant interview questions are in Appendix A. During the interview they were asked to estimate the amount of messages and meetings they participated in and to clarify who shared the information and who received it. Using the information received from the interviews we create a graph that approximates the information flow through the process. It should be noted that the data received from the interviews is not exact but it serves as a starting point to test the use of the mathematical model described previously.

There were some challenges during the interview process. In addition to the data only being approximate, it was not always easy to understand the direction of the flow and the timeline. These issues will impact the resulting graph. Furthermore, estimating the capacity for edges proved really challenging for interviewees and therefore, if used in further studies the definition should be improved.

Another observation from the interview process that shows that the obtained data is only approximate is that there were cases where two or more people were interviewed that had information exchange with each other. In these cases for example, when one person A shares that they have sent x number of messages to another person B and received y amount from them, person B in this scenario should reply with the same or at least very similar amount of messages to person A. That is person B should have received x number of messages from person A and sent y number of messages to them. This was not the case and the answers were on occasion very different. In these cases, data was used from the person that seemed more confident in their answers or had actually done a history check on the messages sent and received or meetings they participated in.

We should note that since interviewees were also asked to estimate the total demand of the system, various answers were once again received. This is most likely due to the unit for the demand as it was not clearly defined due to its somewhat abstract nature and therefore the demand may alter slightly for different people. For this reason, we will not use the demand received from interviews in our graph computations, but rather use the maximum demand possible obtained from the maximum flow computation.

5 Results

After the interviews were conducted, it was time to use the data obtained from the interviews to create a graph showing the relations between people and organizations. Since there were two distinct organizations in the project, it is beneficial to visually separate the different communication groups into two groups where one group represents one side of the project and the other is side two of the project. Both these groups have communication between themselves, but also inter-group communication with each other.

Two graphs are initially created with both having different weights according to the interviews. The first graph has weights that represent the amount of messages sent from one node to another. The second graph has weights corresponding to the amount of time in meetings. These graphs both have the same nodes but the edges are different. Afterwards, we combine these two graphs into one graph that has combined weights of the two graphs to represent the full picture of information flow in the case study.

The information in the graphs begin from a source s that is connected to the owners of the two organizations. There are a total of three sinks and a supersink when needed. Sink t_1 is the sink for group one and sink t_2 is the sink for group two. These sinks represent the information that is being stored and saved by both organizations individually. The third sink and the most important sink is the sink t. This sink is the sink for the total project with both organizations contributing. In this case the sink t represents the finished change of plans for the construction project.

The total demand we use for the minimum cost flow computation is the maximum flow value of each graph. This tests the maximum efficiency of the graph, but in reality the demand could be lower than the maximum flow and obtained from the interviews. By using the maximum flow as the demand, the further study of removing nodes from the system should also give more interesting results as the system is now operating on full capacity.

After we have made the two graphs with weights corresponding to different means of information flow and the graph that has combined weights, we can compute the maximum flow, minimum cost flow and betweenness centrality for these graphs and additionally the minimum $s - t$ cuts for the combined graph. Since the two original graphs represent only half of the full picture of the real world information flow as in reality information is shared using both methods, the computational results obtained are used more as indicative results compared to the computations done with the graph with combined weights. We will go into more detail with the computations and the combination process used to obtain the graph with combined weights in the next sections.

All the data for the computations has been randomized to protect any sensitive data from the interviews. This includes the weights, capacities and the division of demands of the source and sinks. The division of the demands into the different sinks is arbitrary, but done so that the sink t that represents the entire project has the largest value, as it is the most important individual sink.

5.1 Message number as weights

Our first graph contains the amount of messages as weights shown in Figure 6. The gray nodes are the source and sinks respectively. The blue nodes represent group one of the project and the red nodes represent the group two of the project. The dashed edges from sinks t_1, t_2, t to supersink T are to signify that the superink is not added for the betweenness centrality computations as it would alter the results due to adding new shortest paths for each of the sinks. The supersink is added only for the maximum flow computation from source s to supersink T.

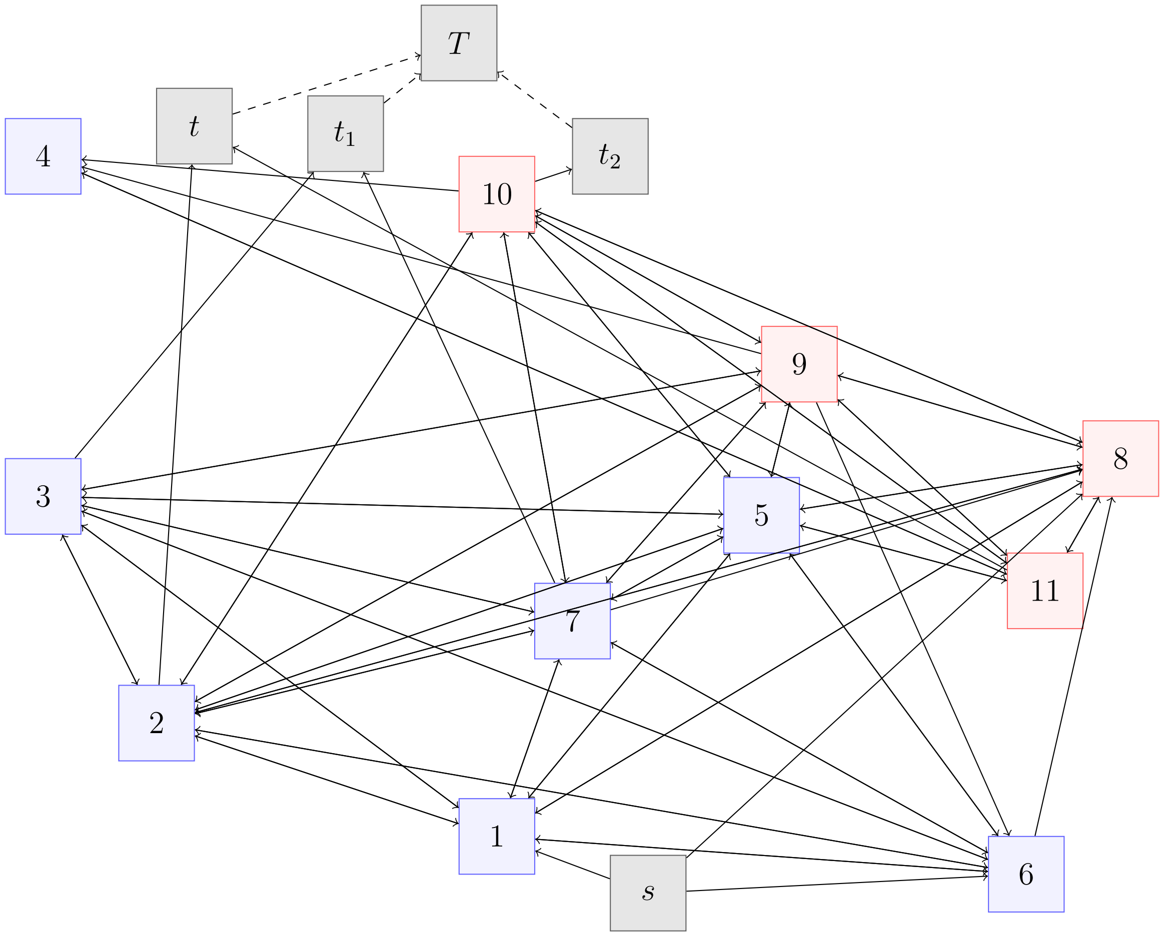

Figure 6: Resulting graph from the case study interviews with the amount of messages sent as weights. Grey nodes represent the source and sinks, blue nodes are group 1 personnel and red nodes are group 2 personnel. A supersink T is connected with dashed edges to signify that it is not required in all computations.

In Figure 7, the maximum flow from the source to each of the sinks is shown. From source s to sink t_1 in red, from s to t_2 in blue, from s to t in green and from s to supersink T in orange. The maximum flow has been computed for each scenario separately and added into one figure. The width of the edge represents the amount of flow flowing through that edge. The maximum flow for message graph from source to different sinks is from $s \to t_1$ is 6, from $s \to t_2$ is 3, from $s \to t$ is 8 and from $s \to T$ is 15.

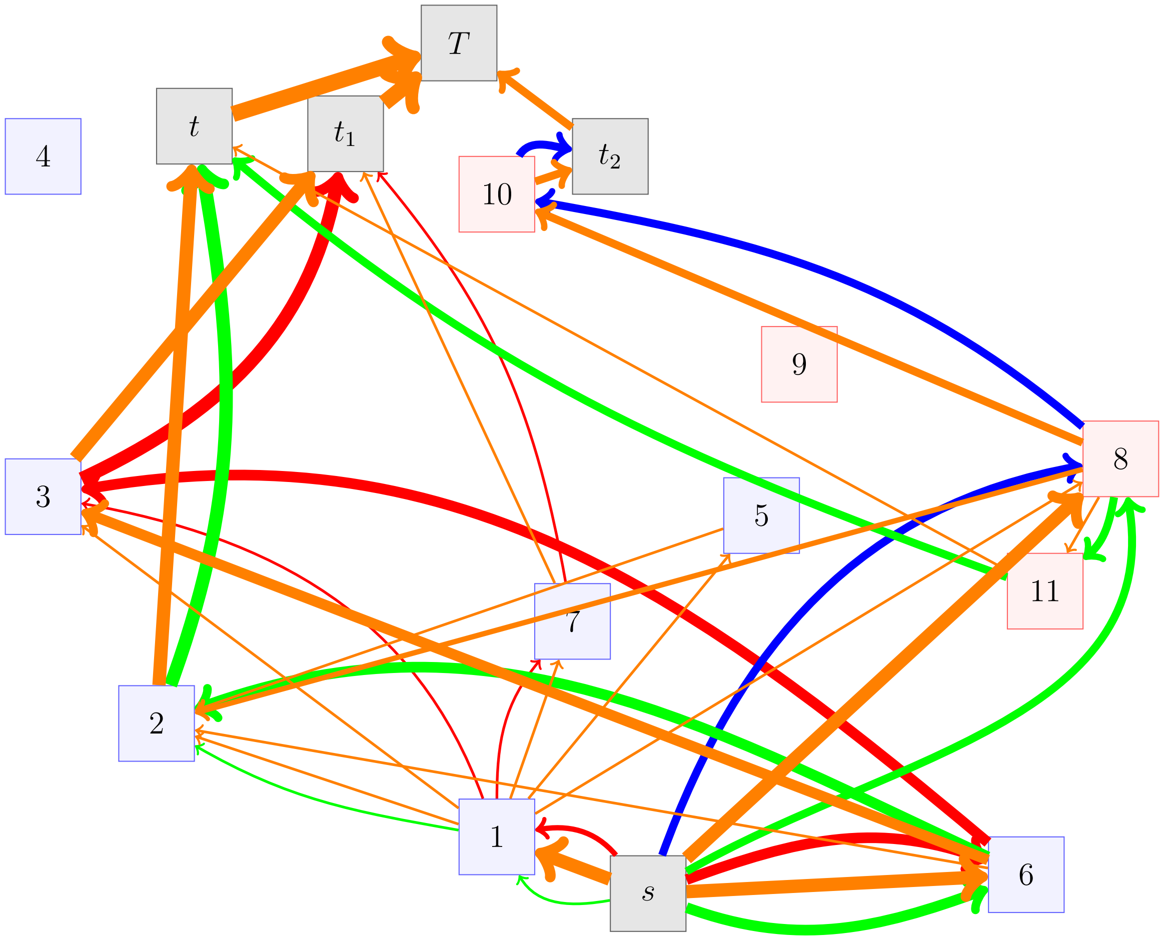

Figure 7: Maximum flow of the graph with amount of messages as weights from s to t_1 in red, from s to t_2 in blue, from s to t in green and from s to supersink T in orange. All unused edges have been removed.

We now know that the maximum amount of information we can send through the whole graph is 15. Since we want to use the maximum flow possible to test the limitations of the graph, we send 15 units out from source s and divide the demand between the three sinks accordingly in line with the interviews. The demand for the source and sink nodes for the graph with message number as weights are $s = -15$, $t_1 = 5$, $t_2 = 3$ and $t = 7$. We can now compute the minimum cost flow for the graph.

The minimum cost flow is shown in Figure 8. The computations are done using the supply and demands as described previously. The red edges show the flow of the system with the width indicating the amount of flow. The black edges are all the edges that are in the original graph but are not used in transporting flow through the system. The minimum cost flow value for the system is 323. This result can now be interpreted as that it requires a minimum of 323 messages to satisfy all the demand for the whole system.

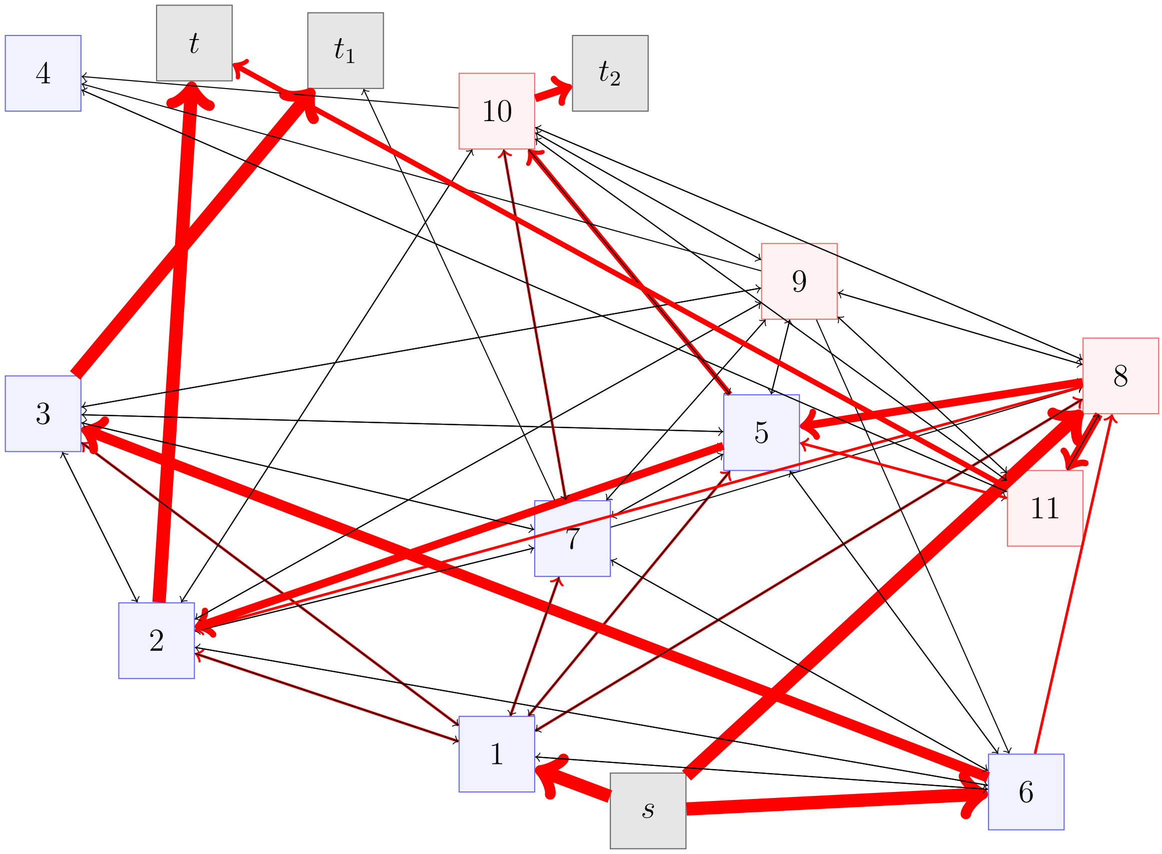

Figure 8: Minimum cost flow with message number as weights with demands equal to $s = -15$, $t_1 = 5$, $t_2 = 3$ and $t = 7$. The minimum cost that satisfies the demand of the system is 323.

The last computation we are using to analyse the graph is betweenness centrality. This is used to see which people in the graph are the most vital. Betweenness centrality for the message graph is shown in Figure 9, where the larger the node size is the more important the node is. It can be seen that node 5 seems to be the most important node in the case of sharing information with messages. Nodes 8 and 10 are the second and third most important nodes and they both come from a different group than node 5.

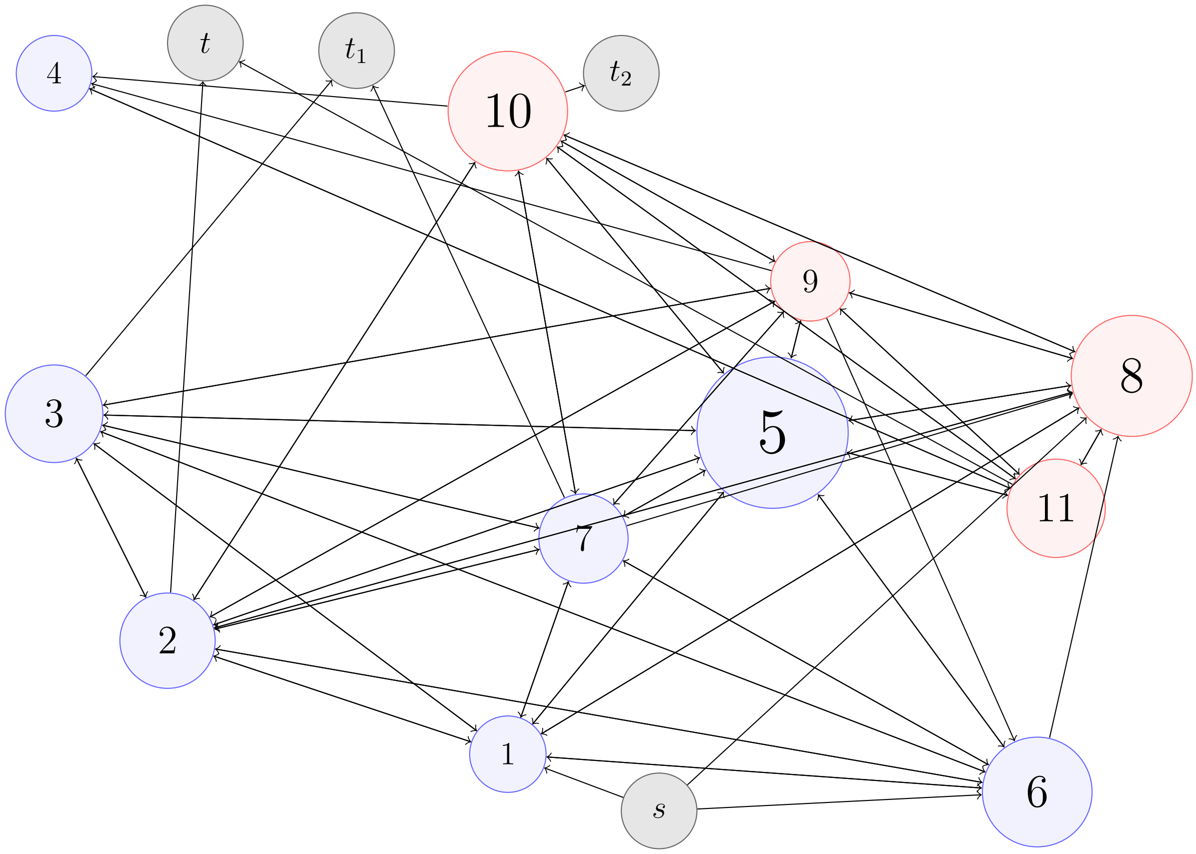

Figure 9: Betweenness centrality of nodes with the amount of messages as weights.

As a result when analysing the graph with message number as weights, the maximum amount of information we can send through the graph is 15. In order to send these 15 units of information through, it takes a minimum of 323 messages. Lastly, nodes 5 and 10 are the most central nodes in sharing information with messages. These results appear to be somewhat contradictory with each other when looking at the figures. Node 5 is the most important node based on the betweenness centrality computation, however with the maximum flow, node 5 is barely used at all in any of the computations. This most likely comes from the fact that betweenness centrality uses edge weights in computation whereas with the maximum flow computation, only edge capacity is used. The minimum cost flow, on the other hand, shares some similarities with both maximum flow and betweenness centrality due to needing both weights and capacities in its computation.

5.2 Meeting time as weights

Next, we have the graph with meeting times as weights. The graph has the same nodes as the message graph and is shown in Figure 10. The edges can be different and the weights are now measured in hours.

First, we can compare the message graph and the meeting graph visually. It is

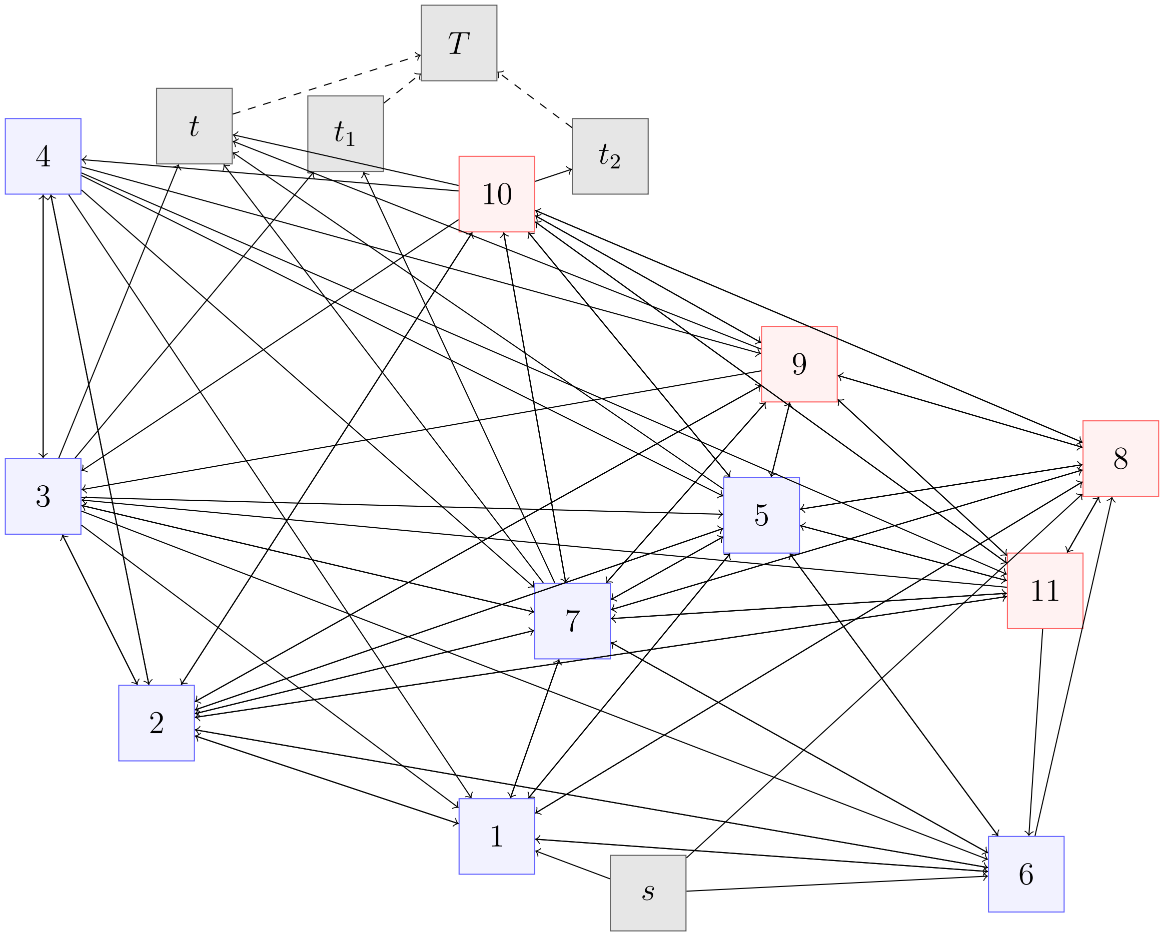

Figure 10: Resulting graph from the case study interviews with meeting time as weights. Grey nodes represent the source and sinks, blue nodes are group 1 personnel and red nodes are group 2 personnel. A supersink T is connected with dashed edges to signify that it is not required in all computations.

clear that the meeting graph is slightly denser than the message graph. This indicates that there were more people participating in meetings than sending messages.

We again start our computational analysis of the graph by computing the maximum flow. The results of the maximum flow are shown in Figure 11. The maximum flow values from source to sinks are as follows: From $s \rightarrow t_1$ is 6, from $s \rightarrow t_2$ is 3, from $s \rightarrow t$ is 15 and from $s \rightarrow T$ is 15. These are very similar to the message graph apart from the maximum flow from s to t. Comparing the two graphs we can see that in order to satisfy this larger maximum flow, the flow flows through many more edges than previously. This is most likely due to the increase in the amount of edges and thus the capacity. Yet, this increase does not increase the total flow the whole system. This indicates that even though in a single case the maximum flow increases, the same edge is used to carry flow into the other sinks as well limiting the amount of total maximum flow from s to T.

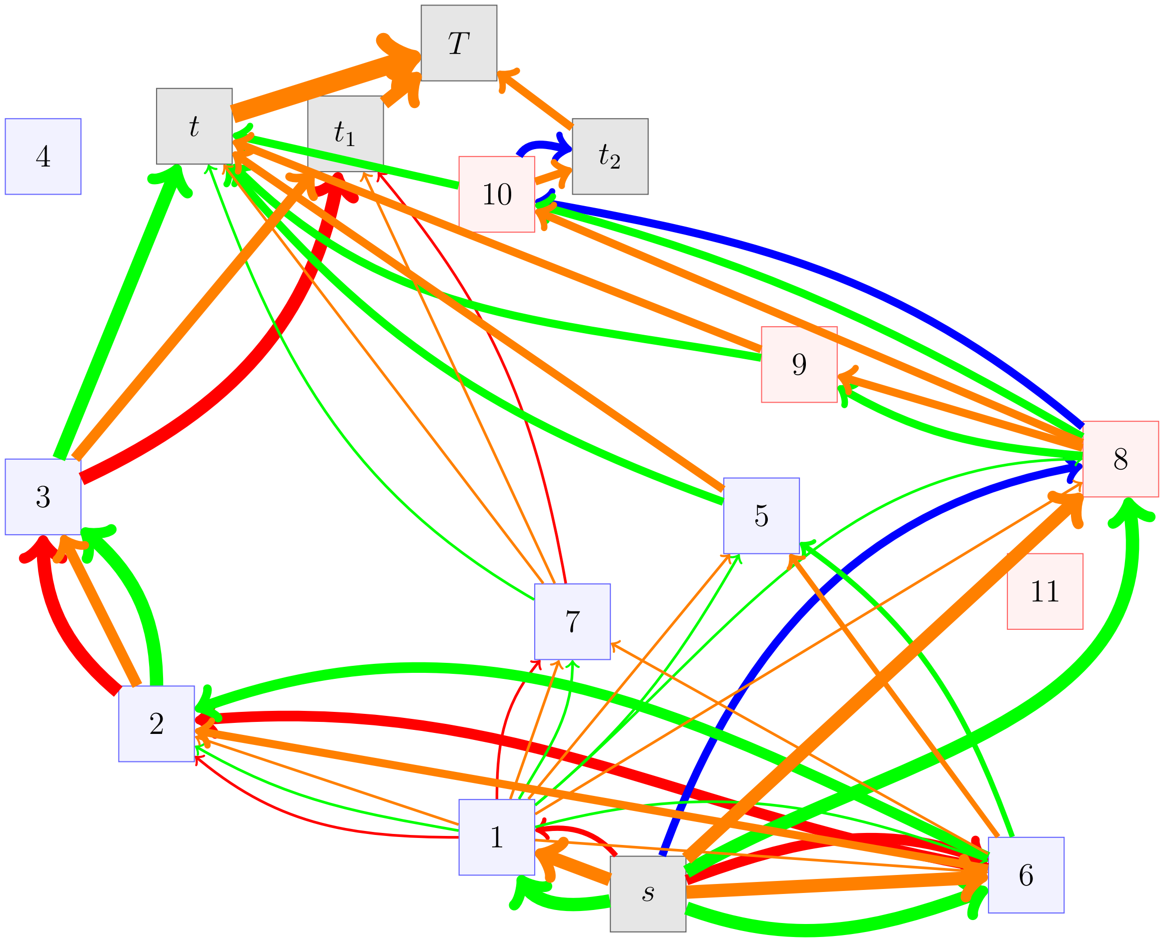

Figure 11: Maximum flow of the graph with meeting time as weights from s to t_1 in red, from s to t_2 in blue, from s to t in green and from s to supersink T in orange. All unused edges have been removed.

Next we compute the minimum cost flow using the maximum flow value from s to T as the total demand. We divide the demand between the three individual sinks as follows: $s = -15$, $t_1 = 3$, $t_2 = 2$ and $t = 10$. Before computing the minimum cost flow for meeting graph using Python Networkx, we need to be multiply the edge weights by a factor of 100 to convert the edge weights into integers as the Networkx function may give false results when using floating point numbers. The result is then divided by the same factor of 100 to obtain the actual result. The minimum cost flow for meeting graph is shown in Figure 12. The minimum cost flow cost for the graph is 238.47. Thus, it takes at least 238.47 hours to satisfy all the demand with meetings.

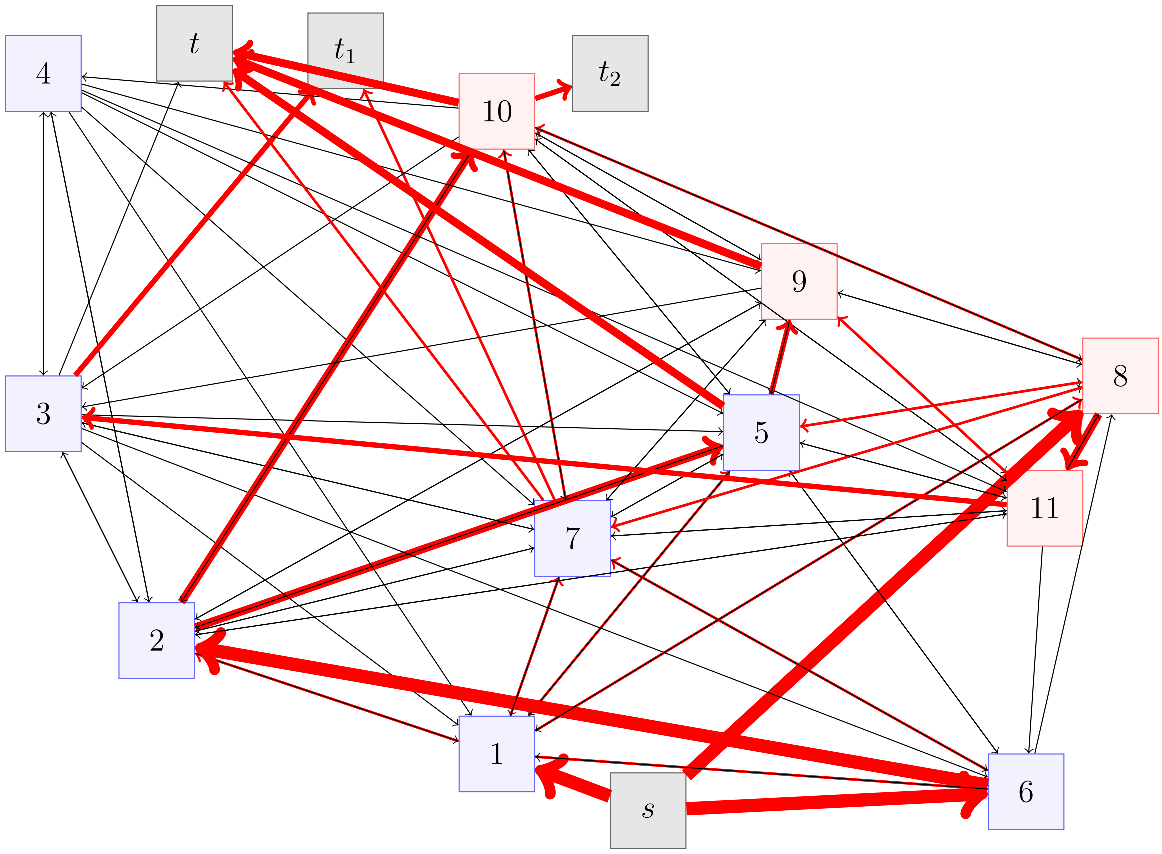

Figure 12: Minimum cost flow with meeting times as weights with demands equal to $s = -15$, $t_1 = 3$, $t_2 = 2$ and $t = 10$. The minimum cost that satisfies the demand of the system is 238.47.

We can now compare the costs of sending information through both the message and meeting graphs. Both have the same maximum flow value, so we could say that in total it took 323 messages and 238.47 hours to send the same amount of information through using different means. This is not, however, necessarily true as the division of the demands between the three sinks in both graphs are different from each other. This, in fact, makes the statement most likely false since the flow from source s to different sinks takes different paths depending on the demand required. In addition, some units of information could be transferred with less cost through messages whereas others through meetings.

Our final analysis tool again is betweenness centrality. The results for betweenness centrality are shown in Figure 13. This time nodes 2, 5 and 7 seem to be the most important nodes in the graph which are all from the same group. The only similarity between the message and meeting graphs seems to be the importance of node 5. That is that in both graphs a large proportion of shortest weighted paths between all pairs of nodes go through the node 5.

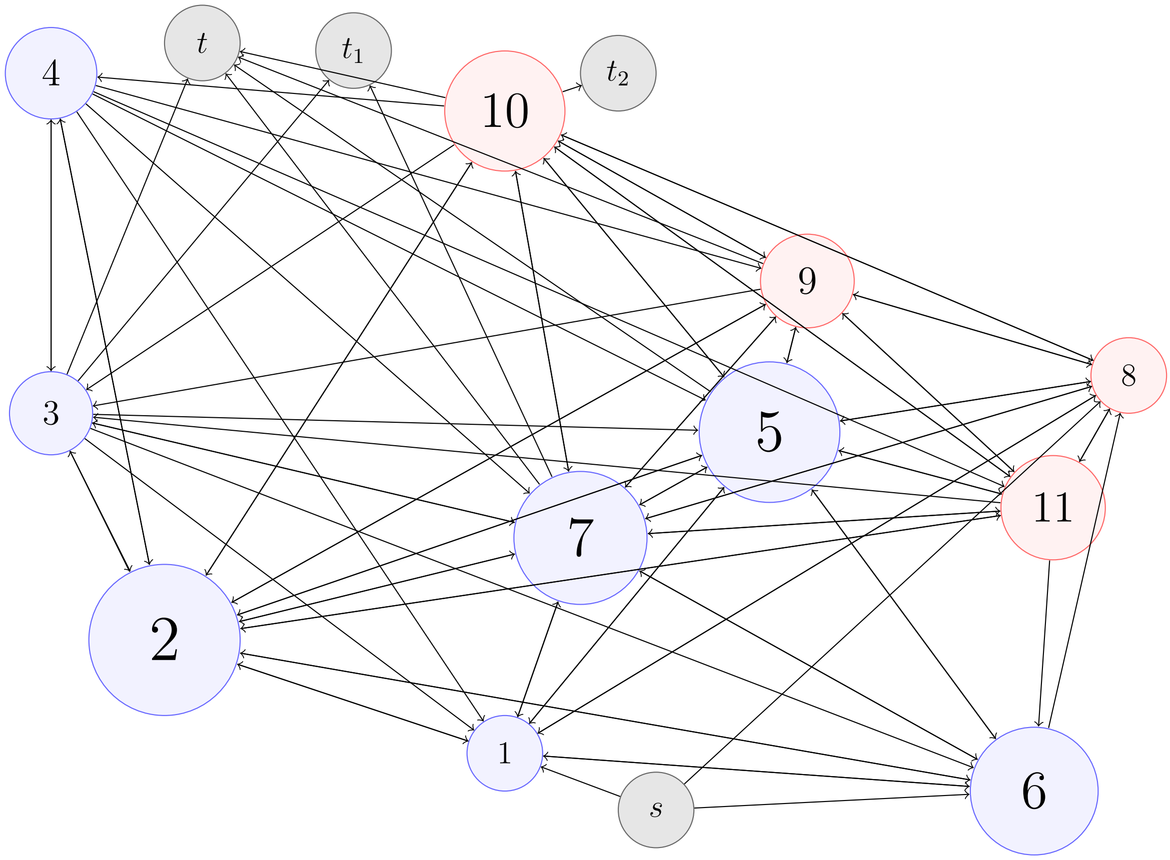

Figure 13: Betweenness centrality of nodes with meeting time as weights.

As a summary for the graph with meeting times as weights, the maximum amount of information the graph can handle is 15 which is equal to the previous message graph. The minimum cost for sending 15 units of information through the graph with demand divided between the sinks is 238.47 hours. Betweenness centrality gives similar results to the message graph as node 5 shares its position as the most important node for information flow. Similarly to the message graph, it appears that the minimum cost flow results follow the betweenness centrality results and maximum flow results quite well as a lot of flow flows through node 5 indicating its cheap central position while also sending flow through edges with larger capacity.

5.3 Combined weights

Now that we have created and computed the values for two separate graphs with different units for weights, we combine the graphs based on the answers received from the interviews. The interviewees were asked to estimate how often they have to read one message and how long does it take to read one message. These answers gave us an average of 2 times per message and 13.9 minutes reading time. That is that one message takes on average 27.8 minutes to internalize. We will round this number to 30 minutes or half an hour when we combine the weights. Thus the combined weights will be in hours.

In addition, the interviewees were asked to estimate how many messages would equal one meeting. From this, we received an average that the amount of information shared in one meeting is equal to the same amount of information shared with 17.7 messages. This information could be used to change the units of weights into amount of meetings rather than time.

It must be noted that the sample size for obtaining the necessary values to estimate the amount of messages that equal one meeting is very small. Also, the answers received from the interviews were very largely different with the standard deviation for the amount of times a message is read being 1.08, the reading time per message 9.43 minutes and the amount of messages that equal one meeting 10. Some interviewees also claimed that it is not possible to cover the same information obtained in a meeting using only messages. This is the main reason we are not using this information to combine the weights.

To create the combined graph, we need to combine both the capacities and weights as well as determine a total demand for the system. To combine the capacities, the sum of the edge capacities between the two original graphs is used. This follows directly from the interviews as the interviewees were asked to estimate their capacities separately for both methods of sharing information. This is thus a natural way of combining the individual capacities as we now have two possible means to share information.

The demand for the combined graph is assigned as the sum of the demands of the two graphs. That is, for the combined graph the demands are $s = -30$, $t_1 = 8$, $t_2 = 5$ and $t = 17$. Combining the demands this way should work as the message graph and meeting graph already on their own are capable of delivering 15 units of information and the combined graph has edge capacities equal to the sum of the capacities of the two graphs. Thus, the combined graph is able to deliver at least twice the demand of one graph.

Combining the weights is the most complicated. In essence, the combined weight is the weight of the meeting graph plus half the weight of the message graph.

As a summary the combination process of the graphs in more detail was done as follows: The nodes in all graphs are the same. The edges were combined by first checking if the edge exists in only one of the graphs. If it does, the edge is added as is with the same weight for the meeting graph and for the message graph the weight is divided by two due to the results from the interviews that one message equals 30 minutes or half an hour. If the edge exists in both message and meeting graph, the edge is added and the weight is the weight of the meeting graph plus half the weight of the message graph. The capacity is the sum of the edge capacities and the demand for source and sink nodes are the sum of the demands of the two graphs.

Now the combined graph should represent the actual real world information flow, where information was shared both in meetings and through messages. The resulting graph with the weights, capacities and demands combined as described previously is shown in Figure 14. We can now compute the maximum flow, minimum cost flow and betweenness centrality for the combined graph.

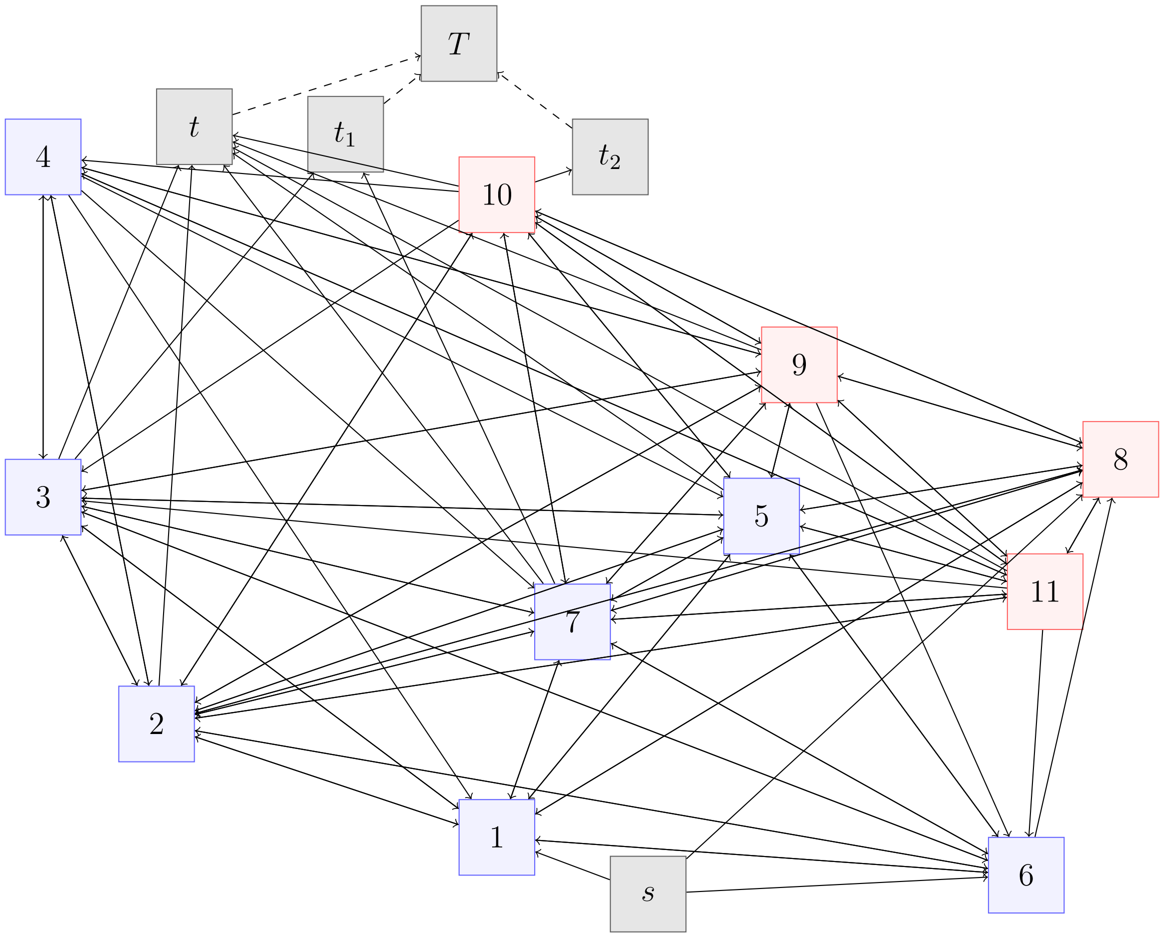

Figure 14: Resulting graph from the case study interviews with combined weights. Grey nodes represent the source and sinks, blue nodes are group 1 personnel and red nodes are group 2 personnel. A supersink T is connected with dashed edges to signify that it is not required in all computations.

5.3.1 Maximum flow

The maximum flow of the combined graph is shown in Figure 15 with the edge width indicating the amount of flow flowing through an edge and the colors as in the previous maximum flow graphs. The maximum flow value in the combined graph is from $s \to t_1$ is 12, from $s \to t_2$ is 6, from $s \to t$ is 23 and finally from $s \to T$ is 30.

The maximum flow from source to sinks in all cases is equal to the sum of the maximum flows from source to sinks in the message and meeting graph. This could be larger though if the combined edges that were only part of one of the original graphs could be used to bring value in increasing the maximum flow. It becomes more clear why this does not happen in this case when we compute the minimum cuts and locate the bottlenecks of the combined graph.

The minimum cut can be used to find bottlenecks in the information flow. These cuts can be found using Theorem 3.6 which indicates that minimum capacity of an

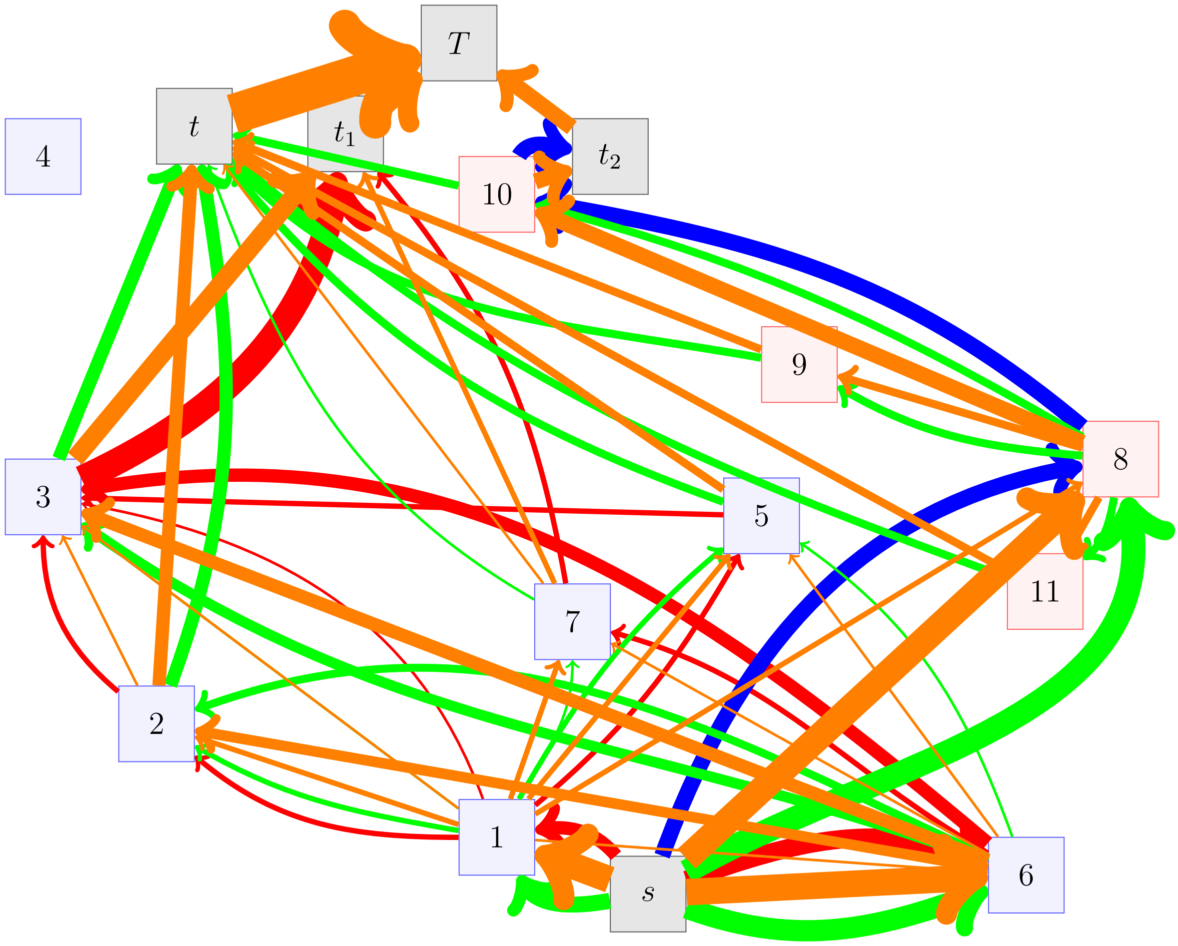

Figure 15: Maximum flow of the graph with combined weights from s to t_1 in red, from s to t_2 in blue, from s to t in green and from s to supersink T in orange. All unused edges have been removed.

$s - t$ cut is equal to the maximum flow of the system. If any of these edges were to be removed, the system would not be able to carry the same amount of flow. Conversely, to increase the capabilities of the system, capacities should be increased in these edges first and foremost.

In the combined graph the partitions of minimum cuts between source s and sinks t_1, t_2 and t always have the sink in one partition and all the other nodes in the other. This indicates that bottleneck edges for these are always the ones that are connected to the sink. It makes sense for the most part since there are fewer edges connected to the sinks than there are possible paths before the sink. Thus, we cannot have larger maximum flow than the sum of the two graphs for any of these cases even if the added edge would increase the capacity in the middle of the graph, since the system already operates on full capacity for the edges connected only to the sinks.

The minimum $s - T$ cut is the opposite to the other three cuts and it is illustrated

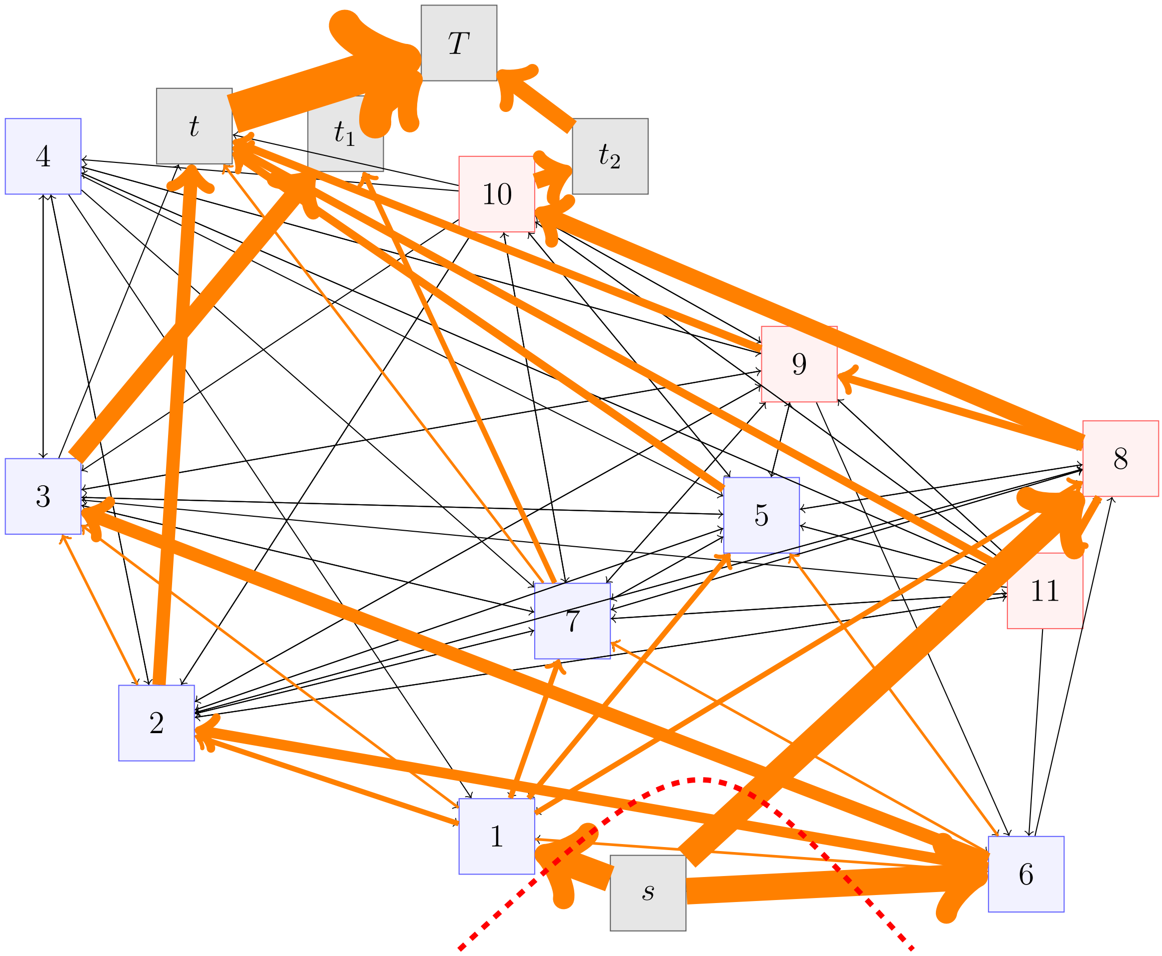

Figure 16: Minimum $s - T$ cut of the graph with combined weights.

in Figure 16. In the figure, we see that the minimum cut separates the graph into two partitions, where one partition only holds the source node s and the other partition holds all the other nodes including the supersink T. Thus the bottleneck edges of the total system are the three edges that leave the source s. This is also not a particularly surprising result as there are only three edges leaving the source s whereas there are multiple edges connecting other nodes to each other and we now also have more edges connected to the sink. In essence, only having three edges leave the source s while the rest of the graph is very dense makes it probable that the bottleneck edges of the system are the edges leaving the source. If the capacities of the edges leaving the source would be increased, the minimum cut would most likely shift.

Now that we have already determined the maximum flow of the combined graph and found some bottleneck edges in the system, we want to study how well the system handles flow if we were to remove one node at a time from the system along with all the edges connected to that node. We are interested in seeing will the maximum flow stay the same or change in the system under these circumstances. The findings of maximum flow value from source to different sinks with one node removed at a time are depicted in Table 1, where the maximum flow value for each source-sink pair is bolded and the minimum value is underlined.

	Maximum flow			
Removed node	$s \to t_1$	$s \to t_2$	$s \to t$	$s \to T$
-	**12**	**6**	**23**	**30**
1	**12**	**6**	20	20
2	**12**	**6**	<u>18</u>	29
3	<u>2</u>	**6**	<u>18</u>	26
4	**12**	**6**	**23**	**30**
5	**12**	**6**	20	29
6	**12**	**6**	19	<u>19</u>
7	10	**6**	22	29
8	**12**	**6**	19	<u>19</u>
9	**12**	**6**	20	**30**
10	**12**	<u>0</u>	20	**30**
11	**12**	**6**	20	**30**

Table 1: Maximum flows from source to different sinks of the combined graph with one node removed at a time. The largest maximum flow values in each case are in bold and the smallest maximum flow values are underlined.

We notice that the maximum flow from source s to supersink T is in most cases close to 30 as with the full graph. The largest drop in maximum flow value comes when node 1, 6 or node 8 is removed. If we look back at the maximum flow figure in Figure 15, we can see that there is a large amount of flow directed through these nodes as they are the three nodes that are directly connected to the source s and contain edges that act as the bottleneck edges. Therefore, if one of them is removed, all the excess flow would have to be directed through other nodes, but the capacity constraints do not allow this fully.

For the maximum flow between the source and the other original sinks, removing one of the nodes connected to the source does not change the amount at all for maximum flow values from s to t_1 and from s to t_2 as we can direct all of the remaining flow through the other two nodes. Studying the maximum flow from s to t we can see that the largest reduction in maximum flow occurs when nodes 2 or 3 are removed although removing one of the nodes 1, 6, or 8 also reduces the maximum flow a lot. Therefore, we can deduce that removing a node with a bottleneck edge always reduces the maximum flow as it should, however it does not necessarily give the largest reduction of flow.

5.3.2 Minimum cost flow

Next, we compute the minimum cost flow, where the demands for all source and sink nodes are summed together from the two original graphs. The results of minimum cost flow are visualized in Figure 17. The edge widths are naturally larger than in the previous two figures since we are sending twice as much information through the

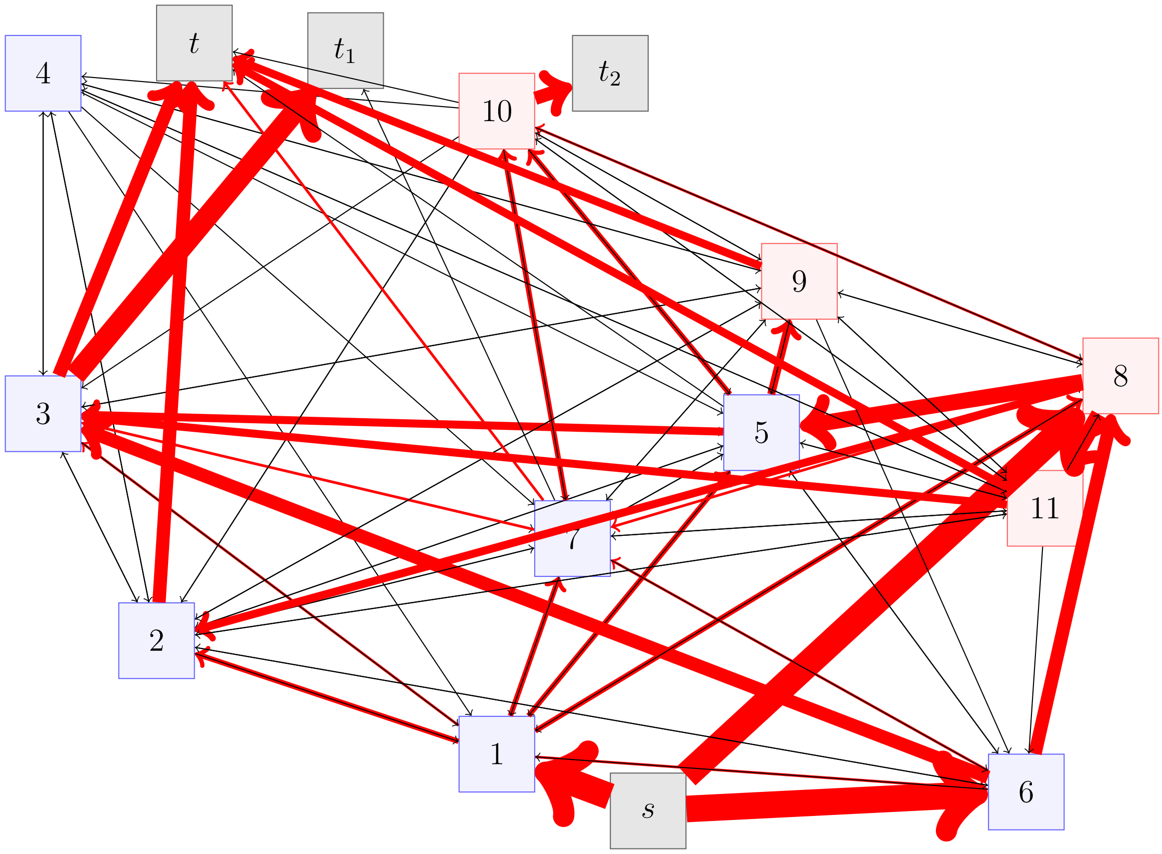

Figure 17: Minimum cost flow with combined weights with demands equal to $s = -30$, $t_1 = 8$, $t_2 = 5$ and $t = 17$. The minimum cost that satisfies the demand of the system is 697.8.

graph. The minimum cost flow cost for the graph is equal to 697.8. This can now be interpreted as in order to send all required information, it takes at least 697.8 hours.

By comparing this result to the two original graphs, we notice that the minimum cost flow cost is significantly larger than if we combine the minimum costs of flow of the two original graphs with similar logic as the weights. This comes up to $323/2+238.47 = 399.97 \approx 400$ hours. The reason for this comes from the combination process of the edge weights as traversing an edge that was in both the original graphs now has an increased weight. Even though the capacities are combined by simply summing the edge capacities of the original graphs, this does not mean that the flow would be directed through these edges as with the original graphs' minimum cost flows. In fact, for all units of flow through an edge that was in both original graphs, there is now more cost associated with it and edges in only one of the original graphs cannot possibly be used to send all information through the network thus increasing the total cost of the minimum cost flow.

If the combination process of the graphs would have been done by appending the edges of the two graphs as is meaning that there could actually be two edges going from one node to the next, the minimum cost flow would be at most equal to the combination of the total costs. The cost would most likely even reduce in this case since the edges that are only in one original graph could be used to reduce the cost of flow in the combined graph.

We continue the minimum cost flow analysis by doing the same analysis we did with the maximum flow. That is, we remove one node at a time from the graph and see how much the minimum cost increases or if the problem becomes infeasible. We use the same demands as with the full graph. These results are shown in Table 2. We see that depending on the removed node the minimum cost either does not increase substantially or in most cases the problem becomes entirely infeasible. This means that we are not able to satisfy the total demand of the system due to the capacity constraints which is not surprising as we are testing the graph with maximum possible demand. In this scenario, the people 1, 2, 3, 5, 6, 7, 8 nor 10 cannot be unavailable for the demand to be satisfied as there are no substitute nodes that could handle the extra flow previously handled by edges connected to these nodes.

Removed node	Minimum cost
-	697.8
1	∞
2	∞
3	∞
4	697.8
5	∞
6	∞
7	∞
8	∞
9	698.64
10	∞
11	721.1

Table 2: Minimum cost flow costs with different nodes removed. The demands are unchanged.

5.3.3 Betweenness centrality

Our last analysis method for the combined graph is going to be the betweenness centrality. The capacities do not affect this unlike with the maximum flow and minimum cost flow, respectively. However, the weights do still carry an effect on the betweenness centrality as the shortest paths in a weighted graph is the weighted shortest paths. The results of the betweenness centrality for the combined graph is shown in Figure 18. In this case, node 3 has the biggest role in information flow followed by both nodes 5 and 8. The node 5 does still carry a large effect as with both the individually weighted graphs, however node 3 is now by far the most important. This could be due to the combination process as there are slightly more unique edges in the two original graphs to and from node 3 than for other nodes. These edges are then added as is and have smaller weights than edges that were in both original

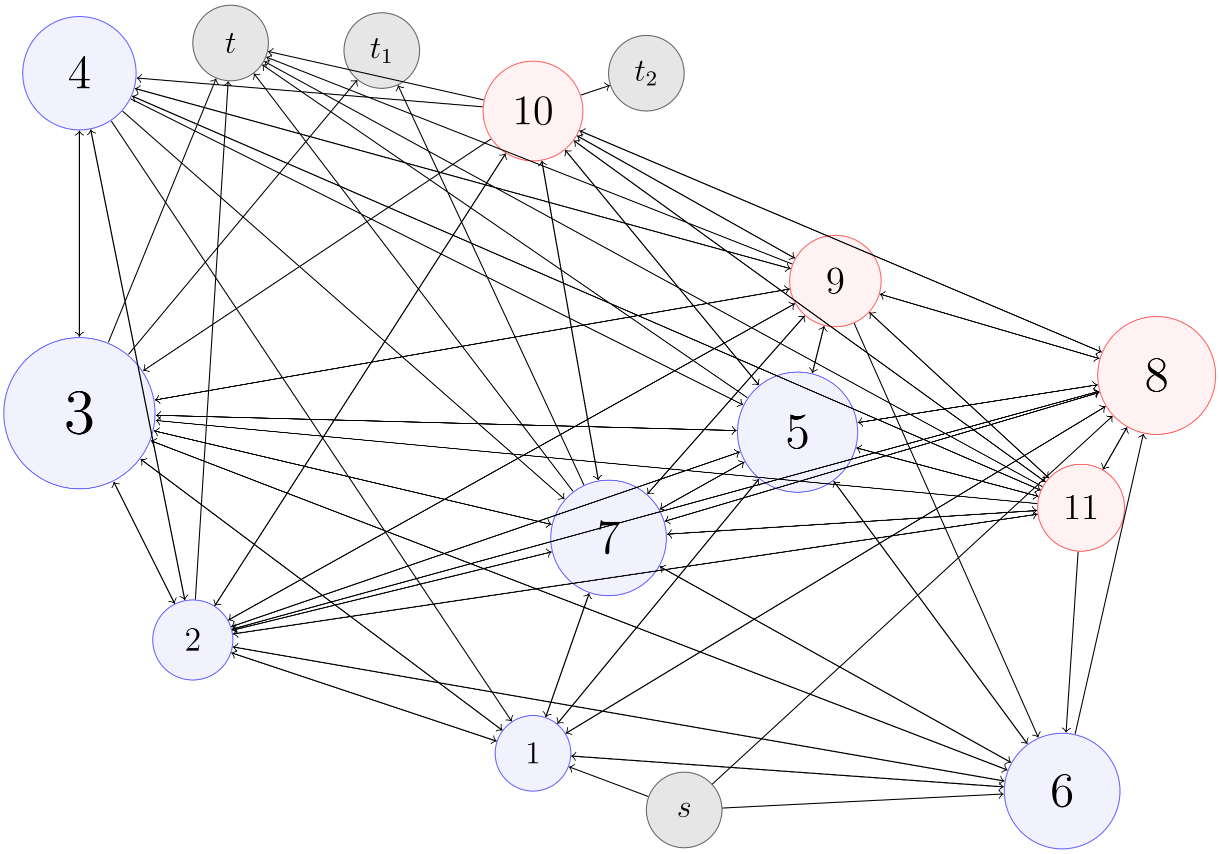

Figure 18: Betweenness centrality of nodes with the combined weights.

graphs resulting in shorter weighted paths near node 3. The betweenness centrality is measured, though, for all pairs shortest paths which indicates that the weights across all shortest paths have to be smaller through node 3 than through any other node.

Finally, we simulate the effects of people or nodes being unavailable and see how the betweenness centrality changes. The betweenness centrality values for all nodes with removal are shown in Table 3. The maximum values have been bolded to show the most important node in each case. As with the full graph, the betweenness centrality is maximal for node 3 in all cases apart from when node 3 is removed. This result further establishes node 3 as the most important node for information flow. Therefore it is more cost effective to use node 3 as an intermediary node for information flow between any other two nodes.

	Betweenness centrality										
Removed node	1	2	3	4	5	6	7	8	9	10	11
-	0	2	**34**	17	20	18	18	19	7	11	5
1	-	2	**29**	12	17	18	15	17	7	10	5
2	0	-	**28**	18	13	16	18	16	6	10	5
3	0	1	-	12	10	14	29	**30**	5	14	8
4	0	6	**27**	-	19	19	10	21	9	10	1
5	0	2	**20**	13	-	17	**20**	15	5	10	13
6	0	2	**26**	19	21	-	17	13	0	10	3
7	0	3	**33**	8	29	17	-	17	10	10	4
8	0	4	**41**	19	14	14	16	-	6	10	1
9	0	2	**22**	16	15	11	18	17	-	10	8
10	0	1	**33**	11	18	16	4	15	6	-	5
11	0	1	**36**	12	18	16	15	12	8	10	-

Table 3: Betweenness centrality measure for the graph with combined weights with different nodes removed from the graph at a time. The maximum value is in bold for each computation.

5.3.4 Conclusions

The combined graph is the one that should model the real life scenario the best as it is modelled with both messages and meetings as means to share information. It was created by combining the nodes and edges of the message and meeting graph accordingly. The resulting graph can carry double the amount of information to either of the original graphs with 30 units. The maximum flow could be more if the bottleneck edges of the combined graph would not be directly connected to the source or sinks as this could allow flow flowing through edges that were only in one of the original graphs used for the combination. The minimum cost flow is probably the most interesting result in the combined graph as the minimum cost is significantly higher than the individual costs of both original graphs combined. This happens due to the combination of weights as traversing the same edge in the combined graph than in the original graphs now has more weight associated with it bringing more costs. For betweenness centrality node 5 still remains important, however, it has been overshadowed by node 3. In the combined graph node 3 is the most important or central node for information flow between pairs of nodes. Again, the minimum cost flow results are somewhat similar to both the maximum flow and betweenness centrality results due to its goal of finding both small edge weights and large edge capacities. This time maximum flow and betweenness centrality also seem to share some similarities at least when it comes to node 3 and node 8 having large centrality measures and large amount of edges going in and out of these nodes in the maximum flow.

To improve total information flow within the network, there are a few possibilities depending on the situation that can be used. These will either increase the capacity of the system and therefore increase the maximum flow and possibly alter the bottleneck edges or decrease the costs of sending information through the graph. To increase capacity and thus the maximum flow of the system, one possibility would be to hire more people to aid the ones who are limiting the maximum flow. In theory these people would help share the load and increase the capacity of the edges if we assume that the two people handling the information would be able to divide it suitably. Another possibility to increase capacity would be to ensure that the people have access to all necessary information and educate them about the information to help them determine what is indeed important and what can be discarded.

Increasing the capacity and the maximum flow of the system is obviously great in that it allows more information to be shared through the network. Increasing the capacity, though, is only focused on the quantity of information and the costs can rise drastically with the increase of capacity. Additionally, increasing the maximum flow of the graph will eventually lead to a situation where it is pointless to increase the capacities of edges as there is never the need to send that much information through in the first place. Therefore, we should also try to decrease the costs of sharing information in order to improve the information flow. This could be argued to be the more important way to improve the overall network as if the capacity would not be high enough in the first place, the information sharing for a project would be doomed from the beginning.

There are at least two possible ways to decrease the costs of information flow through the graph. The first way would be to ensure the access and knowledge of information sharing tools that would limit the amount of time and effort it takes to share the information. The second possibility is to increase the amount of direct connections within the graph. If we have two people that need to have correspondence with each other it would most likely be beneficial to allow them direct communication among themselves rather than have intermediary nodes for the flow to flow through. This in itself does not necessarily guarantee cheaper costs though, if the new added communication edges would have large weights to begin with.

6 Discussion

In the literature review we discussed available articles and documents regarding how information flow can be and is being handled in European construction projects. We found that the emphasis lies in good project and knowledge management that are facilitated by different methodologies and tools. Digitalization has become the standard in storing and sharing information, and specifically in construction, building information models and digital twins are used to visually share information. Tram track projects and infrastructure construction projects in general are often very large consisting of multiple companies and people from various backgrounds. Thus, information flow needs to be efficient to keep the projects on time and on budget while keeping the workers safe.

Next we introduced graph theory basics and a few computation algorithms that can be used to study the efficiency of graphs. We relate how information flow within a specific scenario can be studied with graph theory using the different computations.

The maximum flow is used to measure the maximum amount of information that can be passed through the graph. Additionally, related to the maximum flow, the minimum cut can be determined and used to find the bottleneck edges of the system. Minimum cost flow on the other hand is used to measure how much it costs to fully send a certain amount of information through the system. This varies with the demands associated with the sources and sinks. Finally, betweenness centrality is used to determine the importance of a node in the graph where the shortest paths are the shortest weighted paths or in other words, the lowest costs.

The maximum flow and minimum cut gave us the maximum amount of information and the bottleneck edges that limit the amount of information in the graph. Bottleneck edges are the edges that need to function for the maximum amount of flow to be sent through the graph. If we wanted to increase the total flow of the system, we should increase edge capacities first. If we were able to increase edge capacities for the bottleneck edges enough, eventually there comes a situation, where the total flow a network can handle increases to the point where maximum flow computations for the scenario would not give us any relevant information. This is due to the fact that the maximum flow and minimum cut always give an answer for a well defined system and we receive bottleneck edges as a result. In real life situations if the network can already carry more flow than we would ever try to send through, it becomes meaningless to find bottleneck edges. In these situations, the minimum cost flow computation would give us a better insight on the graph properties and we should focus more on reducing the costs of sending information through edges rather than increasing their capacities.

Minimum cost flow is used to find the theoretical minimum for sending a required amount of information through a graph. To compute a minimum cost flow, there needs to be demand for the source and sink nodes of the graph. In essence, the minimum cost flow computation results in a path for the flow that satisfies all the necessary demand with the least amount of cost. The minimum cost flow problem can be infeasible, if we cannot satisfy all the demand required. This happens when the capacities of the edges in the graph are too limiting for all the information we need. The strength of minimum cost flow comes when we explicitly know all the nodes and their demands along with all the edges and their weights and capacities. Then we can find out the theoretical minimum for the costs of information flow and make improving decisions.

Betweenness centrality is used to study the importance of each node in the graph by studying how often it lies somewhere in the shortest path between any other two nodes. To compute betweenness centrality, there is no need for capacities or demands making it simpler compared to maximum flow and minimum cost flow. All we need for betweenness centrality is the structure of the graph with weights if there is some cost associated with traversing an edge. Betweenness centrality gives a clear ranking on the importance of nodes in a graph and results from it can be used to make sure the important nodes or people have good conditions to continue operating.

The maximum flow, minimum cost flow and betweenness centrality all measure different aspects of a network. The maximum flow is focused on the largest possible amount of information and depends on capacity, while betweenness centrality uses edge weights and is used to find nodes that are along cheap paths between all pairs of nodes. Minimum cost flow computation depends on both the edge weights and capacities and focuses on cost efficiency with a given demand quantity. In certain situations this can be thought of as a combination of the other two, as we are interested in cheap paths similar to the betweenness centrality, while also using edges with large capacities as with the maximum flow. This combination, though, happens only when the network is well structured and has edge weights and capacities that are similar in size. Otherwise, if they are largely disproportionate, it could be possible to have a situation, where some nodes in a network are extremely important according to betweenness centrality but are hardly if ever visited in maximum flow or minimum cost flow due to the capacity constraint. The demand of the system also has to be very large if not the maximum demand as with smaller demands, the minimum cost flow does not have to worry about the edge capacities as much and can focus only on the cheapest paths. If the demand of the system is one, the minimum cost flow problem is the same as the shortest path problem.

In the case study, we studied information flow in a tram track construction project. We first made two separate graphs based on interviews with people from the project that had weights depicting the cost of information flow through messages and meetings respectively. To study the complete information flow in the project, these two graphs were combined into a single graph where each edge weight corresponds to the total costs of information sharing through the edge. Computations were then done on the graph to find out its efficiency. We found that bottleneck edges occur in the edges that are connected to the source or to the sinks depending on which maximum flow we are computing. We tested how the maximum flow changes if some edges would become unavailable and found that removing bottleneck edges always reduces the maximum flow, although it does not necessarily give the largest reduction.

The minimum cost flow computation in the case study was done to test the efficiency of the graph operating on maximum capacity. We found that the minimum cost flow is larger in the combined graph than if we were to combine the costs of the two original graphs similarly to the weights. This is due to the increased costs in the edges that were in both the message and meeting graphs that still have to be used to satisfy the demand of the system. Upon computing the minimum cost flow with removed nodes, there are only three nodes that could be removed with the system still being able to satisfy all necessary demand. This is another consequence of testing the minimum cost flow with the maximum flow. Whenever a node is removed that reduces the maximum flow to at least one of the sinks in our graph, the problem instantly and unsurprisingly becomes infeasible.

The betweenness centrality results in our case study show that node 5 is quite important in all graphs, however with the combined graph node 3 takes the title of the most central node. Node 3 also holds the position in all cases of removing nodes from the graph apart from the obvious case of it being removed. This indicates that we should make emphasis on keeping node 3 available at all times as it is the most useful node in the graph in most cases of information sharing.

Even though the methods used are efficient and based on well-known theory, the adaptation with real data can be quite challenging if it is not properly defined. Furthermore, the results obtained from using our model are based on historical data for it to work. Therefore, it cannot be used in advance to see where possible bottlenecks or other information sharing problems may occur as the data is different for every interaction in each case. The only exception to this would likely be the use of betweenness centrality as it could possibly be determined already in the middle of the project if we know the basic structure and have an estimation on the weights.

The graph we created also only measures the active part of the information sharing and does not take into account the delays that may happen in real life. We assume in our graph computations that once information is received by a node it is instantly sent forward, whereas in reality the information might be held for days before sending it forward. For example, the minimum cost flow computation in the graph might say that it is cheapest to send some amount of flow from a source to a sink through 5 different nodes as the sum of the edge weights connecting these nodes is the smallest even if there would be a path available that goes through less than 5 nodes. In reality if it would take a day for each person to send information forward it would in this case take 5 additional days for information to go through the cheapest path. This would actually likely make it more expensive than going through a more expensive but shorter path. This could be relatively easily fixed in our analysis by just adding weights to each edge corresponding to the delay it takes for information to be shared forward assuming the delay is constant for people. This delay, though, in reality likely has a dynamic nature based on the content and importance of the information making the fix more complex.

In reality there are also bound to be changes in the information throughout the process of sharing information. In order to take into account the changes that may happen during the information flow from one person to another, modifying the edge weights accordingly to try to limit the amount of information going through the undesired nodes is one possibility. Dynamic graphs are another option, however these graphs increase the difficulty of computation and computation time significantly. These are some methods that could be considered in further research.

There are also multiple other possible graph theory concepts that could be used in studying the information flow within a static network. If the network would have both multiple sources and sinks simultaneously, augmented network computations could be used. Additionally, there are also other graph theory tools that could be used in similar cases as in our case study. For centrality measures of nodes, there are many different measures other than the used betweenness centrality such as decay centrality which would take into account the possible decay of information in going through shortest paths in the graph (Bloch et al., 2023).

For further adaptation of the process, one of the biggest changes that would help the accuracy of the analysis is to gather the data during the actual project. During the interviews it was quite clear that most of the interviewees gave only estimates that varied drastically between different interviewees and sometimes failed to provide any numerical data. The fix could be to alter the questions and simplify them, and most importantly, define and give a clear definition of the units used in the case.

The combination process of the two graphs into one single graph that mimics the real life information sharing in the project was also probably not done optimally. We first assumed that information could be shared only with messages or meetings but later during the combination process we assumed that information flow through an edge in most cases requires both messages and meetings to be shared. This made the costs in the combined graph larger than they probably should be. For example, sharing one unit of information in the meeting graph through a random edge takes one hour, but in the combined graph sharing that same unit of information takes one hour plus some amount of messages. This inflation of weights in the combined graph makes the comparisons to the original graphs difficult as they do not convey information in the same scale.

In the end, while the graph theory model gives us results, they should not be trusted blindly and are more or less only approximate and indicative results. This is due to the difficulty of creating the graph and assigning weights and capacities and the overall dynamic nature of real life information flow.

References

Ravindra K Ahuja, Thomas L Magnanti, James B Orlin, et al. *Network flows: theory, algorithms, and applications*, volume 1. Prentice hall Englewood Cliffs, NJ, 1993.

Francis Bloch, Matthew O Jackson, and Pietro Tebaldi. Centrality measures in networks. *Social Choice and Welfare*, 61(2):413–453, 2023.

Yuri Boykov and Olga Veksler. Graph cuts in vision and graphics: Theories and applications. In *Handbook of mathematical models in computer vision*, pages 79–96. Springer, 2006.

Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.

Elliott Brossard and James Morrow. Graph theory: Network flow, 2010.

Anna Butzin and Dieter Rehfeld. The balance of change and continuity in the german construction sector's development path. *Zeitschrift für Wirtschaftsgeographie*, 57 (1-2):15–26, 2013.

Timothy M Chan. More algorithms for all-pairs shortest paths in weighted graphs. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 590–598, 2007.

Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2022.

Bhargav Dave and Lauri Koskela. Collaborative knowledge management—a construction case study. *Automation in construction*, 18(7):894–902, 2009.

Bhargav Dave, Silvain Kubler, Kary Främling, and Lauri Koskela. Addressing information flow in lean production management and control in construction. International Group for Lean Construction, 2014.

Deutsche Bahn. Logiktram: research for freight transport on light rail. *Web page*, 2021. Cited 30.05.2025. Available: https://db-engineering-consulting.com/en/updates/logiktram-research-for-freight-transport-on-light-rail.

Scott B Droege, Jonathan R Anderson, and Matthew Bowler. Trust and organizational information flow. *Journal of business and management*, 9(1):45–59, 2003.

Christopher Durugbo, Ashutosh Tiwari, and Jeffrey R Alcock. Modelling information flow for organisations: A review of approaches and future challenges. *International journal of information management*, 33(3):597–610, 2013.

European Commission. Tram upgrade enters second phase for greener, more accessible transport in szczecin, poland. *Web page*, 2019. Cited 30.05.2025. Available: https://ec.europa.eu/regional_policy/en/projects/Europe/tram-upgrade-enters-second-phase-for-greener-more-accessible-transport-in-szczecin-poland.

European Commission. New trams and upgrades to tram lines in bydgoszcz, poland. *Web page*, 2020. Cited 30.05.2025. Available: https://ec.europa.eu/regional_policy/en/projects/Europe/new-trams-and-upgrades-to-tram-lines-in-bydgoszcz-poland.

European Investment Bank. Rhein-neckar-tram 2023. *Web page*, 2023. Cited 30.05.2025. Available: https://www.eib.org/en/projects/pipelines/all/20220975.

Ferrovial newsroom. Budimex-strabag-zue consortium completes modernization of the "inner ring line" in warsaw. *Web page*, 2018. Cited 30.05.2025. Available: https://newsroom.ferrovial.com/en/news/budimex-strabag-zue-consortium-inner-ring-line/.

Maria Fonoberova. Algorithms for finding optimal flows in dynamic networks. *Handbook of power systems II*, pages 31–54, 2010.

Thomas M Froese. Information management for construction. 2005.

Prashnna Ghimire, Kyungki Kim, and Manoj Acharya. Generative ai in the construction industry: Opportunities & challenges. *arXiv preprint arXiv:2310.04427*, 2023.

Regina Gyampoh-Vidogah, Robert Moreton, and David Proverbs. Implementing information management in construction: establishing problems, concepts and practice. *Construction Innovation*, 3(3):157–173, 2003.

Alex Hall, Steffen Hippler, and Martin Skutella. Multicommodity flows over time: Efficient algorithms and complexity. *Theoretical computer science*, 379(3):387–404, 2007.

Sabine Hartmann, Dirk Gossmann, Suzan Kalmuk, and Katharina Klemt-Albert. Optimizing interfaces of construction processes by digitalization using the example of hospital construction in germany. *Buildings*, 13(6):1421, 2023.

Anne Herneoja, Anna-Sofia Hyvönen, and Kati Kiiskilä. Oulun seudun liikennejärjestelmäsuunnitelma 2040, 2023. Cited 08.07.2025. Available: https://www.ouka.fi/sites/default/files/attachments/OSLJS%20raportti_muok%2020.2.2024.pdf.

Mieke Hoezen, Isabelle Reymen, and Geert PMR Dewulf. The problem of communication in construction. In *International Conference on Adaptable Building Structures, ADAPTABLES 2006*, pages 12–14. Eindhoven University of Technology, 2006.

Bruce Edward Hoppe. *Efficient dynamic network flow algorithms*. Cornell University, 1995.

Pankaj Jalote. Knowledge infrastructure for project management. *Managing Software Engineering Knowledge*, pages 361–375, 2003.

Dieter Jungnickel and D Jungnickel. *Graphs, networks and algorithms*, volume 3. Springer, 2005.

Ewelina Kania, Elżbieta Radziszewska-Zielina, and Grzegorz Śladowski. Communication and information flow in polish construction projects. *Sustainability*, 12(21): 9182, 2020.

Oleg Kapliński. Information technology in the development of the polish construction industry. *Technological and economic development of economy*, (3):437–452, 2009.

Bernhard H Korte, Jens Vygen, B Korte, and J Vygen. *Combinatorial optimization*, volume 1. Springer, 2011.

LCI. Mitä on lean rakentaminen. Cited 16.06.2025. Available: https://lci.fi/lean-rakennusalalla/mita-on-lean-rakentaminen/.

Raymond E Levitt. Towards project management 2.0. *Engineering project organization journal*, 1(3):197–210, 2011.

Aftab Hameed Memon, Ismail Abdul Rahman, Ismaaini Ismail, and Noor Yasmin Zainun. Time management practices in large construction projects. In *IEEE Colloquium on Humanities, Science and Engineering*, pages 61–66, 2014.

Michaela Boneva. Germany. *Web page*, 2022. Cited 30.05.2025. Available: https://projectdelivery.enotrans.org/case-studies/germany/.

Mohamed Abdelwahab Hassan Mohamed, MKS Al-Mhdawi, Udechukwu Ojiako, Nicholas Dacre, Abroon Qazi, and Farzad Rahimian. Generative ai in construction risk management: a bibliometric analysis of the associated benefits and risks. *Urbanization, Sustainability and Society*, 2(1):196–228, 2025.

Mark Needham and Amy E Hodler. *Graph algorithms: practical examples in Apache Spark and Neo4j*. O'Reilly Media, 2019.

Tran Duong Nguyen and Sanjeev Adhikari. The role of bim in integrating digital twin in building construction: A literature review. *Sustainability*, 15(13):10462, 2023.

NRC Group. Lean merkitsee kaikkien resurssien optimointia. *Web page*, 2024. Cited 16.06.2025. Available: https://www.nrcgroup.fi/uutiset/lean-merkitsee-kaikkien-resurssien-optimointia.

Elina Ojalehto. Lean-menetelmien hyödyntäminen infrarakentamisessa. Opinnäytetyö, Ammattikorkeakoulu, 2023.

Charles B Pettinger and Barry Nelson. Daily planning conversations and ai: Keys for improving construction culture, engagement, planning, and safety. *American Journal of Industrial Medicine*, 68:S15–S22, 2025.

Andreas Floros Phelps. Managing information flow on complex projects. *Balfour Beatty Construction*, 2012.

Krystian Pietrzak and Oliwia Pietrzak. Tram system as a challenge for smart and sustainable urban public transport: Effects of applying bi-directional trams. *Energies*, 15(15):5685, 2022.

Polimex Mostostal. Construction of new tram lines in olsztyn officially completed. *Web page*, 2024. Cited 30.05.2025. Available: https://www.polimex-mostostal.pl/en/news/construction-new-tram-lines-olsztyn-officially-completed.

Railway Technology. Die kombilosung hybrid tram-train system, karlsruhe. *Web page*, 2017. Cited 30.05.2025. Available: https://www.railway-technology.com/projects/die-kombilosung-hybrid-tram-train-system-karlsruhe/.

Michał Romańczuk, Dominika Król-Smetak, Paweł Stępień, Katarzyna Kazojć, and Apoloniusz Kurylczyk. Determinants of management efficiency using it systems in the context of regionally dispersed construction enterprises in poland. 2023.

Alexander Schrijver. On the history of the shortest path problem. *Documenta Mathematica*, 17(1):155–167, 2012.

Meltem Sözüer and Konrad Spang. Challenges in the planning process of infrastructure projects in germany. In *Construction Research Congress 2012: Construction Challenges in a Flat World*, pages 2369–2378, 2012.

Suomen Raitiotieseura. Raitiotiet suomessa. *Web page*, 2024. Cited 18.09.2024. Available: https://www.raitio.org/suomen-raitiotiet-ja-raitiovaunut/suomen-raitiotiet/.

Tekla. Kruunuvuori bridge. *Web page*, 2024. Cited 25.06.2025. Available: https://www.tekla.com/bim-awards/kruunuvuori-bridge.

Turun raitiotie. Raitiotieltä turkuun sujuvaa arkea ja elinvoimaa. *Web page*, 2025. Cited 08.07.2025. Available: https://turunraitiotie.fi/.

Jacques Verstraëte. Introduction to graph theory, 2020.

Douglas R White and Stephen P Borgatti. Betweenness centrality measures for directed graphs. *Social networks*, 16(4):335–346, 1994.

YIT. Lean merkitsee kaikkien resurssien optimointia. *Web page*. Cited 16.06.2025. Available: https://www.yit.fi/infra/allianssimalli.

Ping Zhang and Gary Chartrand. Introduction to graph theory. *Tata McGraw-Hill*, 2:2–1, 2006.

Sijie Zhang, Jochen Teizer, Jin-Kook Lee, Charles M Eastman, and Manu Venugopal. Building information modeling (bim) and safety: Automatic safety checking of construction models and schedules. *Automation in construction*, 29:183–195, 2013.

Krzysztof Zima, Edyta Plebankiewicz, and Damian Wieczorek. A swot analysis of the use of bim technology in the polish construction industry. *Buildings*, 10(1):16, 2020.

A Interview questions

- What was your job description regarding the case?
- Was there a predefined change process that was followed in the case?
- Estimate to and from whom you sent or received messages, e-mails and or documents and how many.
- How many times did you read these messages to internalize the information?
- Estimate the amount of meetings regarding the case and the duration of these meetings.
- Who shared information and who received information in these meetings?
- Estimate how much information can you send or receive at once per method of sharing.
- Estimate how many different sets of information was needed to successfully perform the work.
- Estimate how many messages would be needed to equal the information gained in a meeting.
- Who were the most critical people in the process?
- Did you notice situations where the information did not flow as hoped?
- In your opinion, how could the information flow have been improved in the case?
- Which form of information sharing do you prefer, messages/documents or meetings?
- Does something come to mind about the case that was not discussed in this interview?