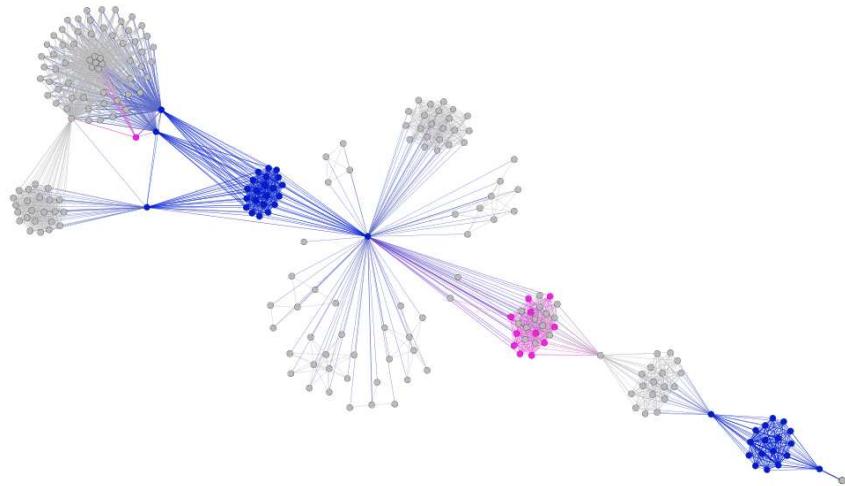


# Social network analysis as a method in the data journalistic toolkit (master thesis)



Tilburg university

Communication and information studies  
(Data Journalism)

Adriana Homolová

469544

**Supervisors:** H.A.J. (Hille) van der Kaa  
M.M. (Matje) van de Camp MA

**Second reader:** prof. dr. A. (Aske) Plaat

---

## Table of contents

Preface .....	4
Abstract .....	5
1. Introduction .....	6
Problem statement .....	8
Scientific relevance .....	8
Social relevance.....	8
Contributions .....	9
Structure of the thesis .....	9
2. Theoretical framework.....	10
2.1 Surfing on the third wave .....	10
2.2 Journalism .....	11
2.3 Data journalism .....	12
2.4 Social network analysis.....	13
2.4.1 Basic principles.....	14
Nodes and edges .....	14
Network .....	15
Social network graph.....	18
2.6 Corrupt behavior.....	19
Corruption.....	19
Collusion.....	20
Conflict of interest .....	20
2.7 Indicators of corrupt behavior within public procurement.	21
2.9 Corrupt behavior in Slovak public procurement .....	22
2.9 Social Network Analysis of corruption .....	24
2.10 Research questions .....	25
3. Methodology .....	26
3.1 Research methods.....	26
3.2 Variables.....	26
3.3 Study design .....	27
4. Data .....	30
4.1 Collection and content.....	30
Transparency data set .....	30
Chamber of commerce data set.....	34

---

4.2 Pre-processing .....	35
4.3 Network .....	35
5. Experiment and results .....	37
5.1 Companies network.....	37
Introduction.....	37
Results.....	37
Conclusion .....	42
5.2 Individuals network.....	44
Introduction.....	44
Results.....	44
Conclusion .....	49
6. Discussion .....	50
7. Conclusion and recommendations .....	52
Bibliography .....	54

---

## Preface

Nearly a year ago, I have started this research out of sheer curiosity. Diving deep into the unknown, I have found joy as well as hardships in the process. Writing this thesis has taught me a lot. Not only about the subject, but also about research as such and eventually about myself as well. What has started out as curiosity has eventually grown into hard work. Nevertheless, hard work bores fruit, and I am now very excited that I can successfully close this metaphorical chapter despite of all the obstacles and struggles that have crossed my path during the year.

I would hereby like to express gratitude to my two supervisors. Hille van der Kaa for feedback and lots of support throughout the difficult process. Matje van de Camp for inspiration as well as valuable consultations on the topic of social network analysis and extensive feedback as my writing progressed. I would also like to thank my second reader Aske Plaat for useful insights and reassurance in the relevance of my research. Furthermore to Matej Kurian and Zuzana Dančíková from Transparency International Slovensko for helping me to familiarize myself with procurement data and the public procurement in Slovak healthcare in general. Much appreciated was also the willingness of Michala Habala who repeatedly and promptly provided me with data sets from the Slovak Chamber of Commerce. And finally, I would like to thank Michal Fábry for his guidance in all those times I got lost between zeros and ones.

This master thesis is the conclusion of my Master study of communication and information sciences at the Tilburg University.

---

## Abstract

Corruption in public spending poses a big problem in many countries. It results in ineffective spending of public resources and has a negative impact on mechanisms of the free market as well as on the country's economy. One of possible signs of corruption in public spending is low competitiveness. An example can be found in the Slovak procurement market. The median of companies that participate on the competition for public contracts in Slovakia is not only far under the EU average, but also the lowest in comparison to other EU countries. However, corruption in public spending and procurement still remains difficult to expose. Tools that can help with uncovering corrupt practices in public procurement are therefore much needed for supervising institutions, such as news agencies. A possible improvement in the monitoring of public spending is the usage of social network analysis (SNA). SNA has been used in social sciences for decades. However, it is not often used in journalistic research. The aim of this research is therefore to explore to what extent can social network analysis be used to analyze public procurement networks in journalistic research. A case study on freely available Slovak public procurement data shows manifestations of the so-called “red flags of corruption” in the public procurement network. Companies that acquire above average amounts of contracts become apparent after visualization of contract flows between public procurement market participants. Furthermore, information flows established by shared high management between companies also become visible. Visualization of the professional network of individual directors discloses conflict of interest situations as well as locates the most important individuals for the communication within the network. These results suggest that SNA can serve as a useful tool in the monitoring of government spending and even uncover possible corruption patterns. Therefore, social network analysis can become a welcome addition to the data journalistic toolkit.

*Keywords:* social network analysis, social network visualization, data journalism, public procurement, Slovakia

---

## 1. Introduction

As theories on news media suggest, journalists play an important role in democratic societies (McCombs, 2003; Lippmann, 1920; Dahlgren, 1995; Middleton, 2009). Not only do the news media play a role in providing citizens with information that lead to the forming of public opinion, but they also serve as “watchers” by monitoring the activities of public institutions. Along with the information explosion that emerged after the worldwide implementation of the Internet, the journalists have also obtained new tools to achieve these goals. *Data journalism* is an emerging form of journalism that creates stories based on data analysis and visualization. The term covers a wide range of disciplines that use data to reveal stories. Within these disciplines, data analysis tools help journalists to understand the complex issues within the data (Gray et al., 2011). In addition to the tools available for analyzing data, there has been an increasing amount of data as well. One of the main sources of these data are public administrations which, in order to increase transparency, publish more and more of their data freely on the Internet. This development gives journalists the chance to monitor public authorities closer. In this respect, government purchasing of goods and services (public procurement) is an important point of interest. Reports by non-governmental organizations show, that this process tends to be sensitive to corrupt behavior such as bribery and cartel forming (OECD, 2007; OECD, 2010). Corrupt behavior in public procurement leads to inefficient spending of public funds and has an impact on the economy of a country (Lindskog et al., 2010).

One of the methods of the data journalistic tool kit is Social Network Analysis (SNA). The method has originated from the social sciences in the 1950s and is used to examine relationships between actors in a social network. SNA has not been used frequently by journalists so far (Stray, 2013). The goal of this thesis is to examine whether SNA can be used in journalistic research in order to examine country’s public procurement network, and especially if SNA is a promising method in this regard.

The inspiration for the thesis has three main sources. The first two are two projects called Muckety<sup>1</sup> and TheyRule<sup>2</sup>, which aim to track the “connections of the rich, famous & influential”. While TheyRule focuses on how big companies form interlocks by sharing members of their

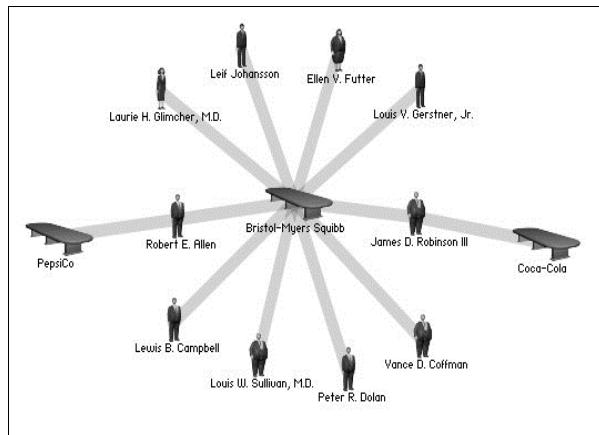
---

<sup>1</sup> <http://news.muckety.com/>

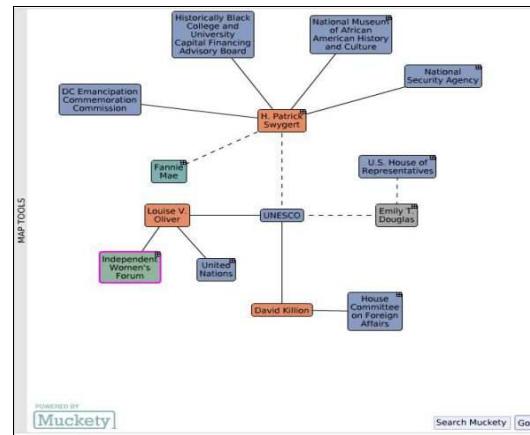
<sup>2</sup> <http://theyrule.net/>

executive boards, the scope of Muckety is broader, including for example also actors and politicians.

*TheyRule*



*Muckety*



The third source of inspiration is a report on the situation of Slovak public procurement market. According to this report, that has been prepared for the European Commission, the competitiveness on the Slovak public procurement market is the smallest in the EU. The median number of bids placed in Slovak tenders is only 1, while the median in the EU is 4 (Strand et al., 2011). This lack of competitiveness raises questions about the health of the Slovak public procurement market. A recent change in legislation concerning Slovak public procurement makes it mandatory to put all procurement contracts online (Act on Public Procurement 2010). The information published by the government is processed into structured data sets by Transparency International Slovakia. The resulting structured data provides an excellent opportunity for journalists and the general public to apply various methods of analysis to monitor flow of contracts and thereby the spending of tax money.

It should be noted, that it appears that Slovakia is unique in this respect, since we did not encounter such legislation anywhere else. Even though tenders where the value exceed certain thresholds<sup>3</sup> are published in the European Tenders Electronic Daily (TED) database,, we are not aware of an overview of contracts that have been procured at lower costs in any other country.

<sup>3</sup> Only includes big projects, the threshold value is individual per type of services that are being procured

---

We are also not aware of any previous research addressing the use of SNA to analyze the public procurement networks. Our research questions will therefore be based on theoretical works: on the theory of public procurement, on work on the red flags of corruption within public procurement (OECD, 2007; OECD, 2010), and on SNA theory (Borgatti et al. 2009; Wasserman & Faust, 1994).

## **Problem statement**

In this thesis, we apply SNA to a network of public institutions and the companies that deliver services to them. The problem statement of this research is:

*“To what extent can social network analysis be used to analyze public procurement networks in journalistic research?”*

This question gives rise to four associated research questions, namely:

- (1) *Do measures of the position of an actor within a network point out individuals that are important to the information flow in the public procurement network?*
- (2) *Does the social network graph expose conflict of interest situations?*
- (3) *Does the social network graph expose interlocking directorates?*
- (4) *Does the social network graph display differences in the magnitude of the flow of contracts between governmental institutions and private companies?*
- (5) *Does visualization of the social network graph in time reveal possibly corrupt patterns in awarding contracts between governmental institutions and private companies?*

We will answer these four research questions by applying SNA on so-called *red flags* of corruption that may indicate corrupt behavior in the public procurement network.

## **Scientific relevance**

We are not aware of any prior research on analyzing public procurement using SNA. Therefore, this thesis represents an innovative approach.

## **Social relevance**

---

In the journalistic field, the outcomes can be used to analyze the mutual relationships and connections of the actors instead of only the attributes of the actors themselves. Analyzing relationships between actors brings an additional depth to various analysis of data that form a network. Utilized by journalists, SNA can be applied in investigating not only public procurement networks, but also various journalistic researches where the relationships between actors of a story matter. For example looking for the most important actors within the top management of private and public institutions in a country<sup>4</sup> or mapping influential users on Twitter<sup>5</sup>. With more data becoming available, future research can expand the approach used in this thesis on other data in order to create models appropriate for examining human or nonhuman networks with SNA.

## **Contributions**

The main contribution of this thesis is that we create a methodology to use SNA to investigate corruption in public procurement.

## **Structure of the thesis**

In Chapter 2 the theoretical framework is introduced. We begin within the context of recent developments in information expansion and the role journalists have in the information society. In Section 2.2 we explain the effects that information expansion had on journalism. Section 2.3 introduces the concept of data journalism. Section 2.4 is devoted to the SNA and the metrics it uses. In Section 2.5 we briefly explain the public procurement process and its position within the country economy. In Section 2.6 we familiarize ourselves with some types of corrupt behavior and in Section 2.7 we put them in context of public procurement and red flags of corruption. In Section 2.8 we focus on some of the journalistic research on Slovak public procurement. We also briefly address research on corrupted networks that has already been conducted in the past in Section 2.9. In Section 2.10 we specify our research questions. In Chapter 3 we introduce the methodology we use to answer our research questions. Chapter 4 addresses the data sources we use for the analysis. In Chapter 5 we discuss the results of the analysis.

---

<sup>4</sup> <http://top200.volkskrant.nl/>

<sup>5</sup> <http://econsultancy.com/uk/blog/63682-twitter-network-analysis-identifying-influencers-and-innovators>



---

## 2. Theoretical framework

### 2.1 Surfing on the third wave

The field of data journalism would not have come into existence without the information and technology boom of the past few decades. In the following paragraphs, we will describe the idea of the information society and information expansion that made this shift in journalism possible. We will also explain the relationship between information and journalism. This Section is based on the book *Theories of the Information Society* (2006) by Frank Webster.

The present society, sometimes called *information society*, is represented by information and communication technologies, where information is central, shared, everywhere and anytime. This information boom has influenced many parts of society, including the mass media, form of work, or education systems. The first claims of the technology having the power of bringing a systematic social change appeared already in the late 1970s. And shortly after, the idea of a “Third wave”<sup>6</sup> as the metaphor for information revolution was introduced by the futurist Alvin Toffler:

“A new civilization is emerging in our lives [...]. This new civilization brings with it new family styles; changed ways of working, loving, and living; a new economy; new political conflicts; and beyond all this an altered consciousness as well.“ (Toffler, 1980)

Even if Toffler’s books were in general regarded as “intellectually slight, derivative, analytically inept and naïve on almost every count“ (Webster, 2006), nearly a decade after it was published, the invention of the World Wide Web paved the way for the final expansion of last big medium during the mid 1990s, the Internet. Nowadays, information technology plays an integral part in contemporary society. The information boom has manifested itself on many levels, whether we talk about pasta recipes, gardening or quantum physics, accessing necessary knowledge has never been easier. However, having easy access to information does not necessarily imply being informed. There is a great diversity (and oftentimes contradiction) coming from the information we encounter. This diversity includes misinformation and

---

<sup>6</sup> the First being the agricultural revolution and the Second the industrial revolution

---

unscientific claims, but also differing opinions, traditions or practices that develop in various cultural contexts. While the quantity of the information is abundant, the quality of information does vary a lot. The dimensions of information quality also include for example its accuracy, believability, consistency and reliability (Yang et al. 2002). Highly accurate, believable, consistent and reliable information therefore can be labeled *high quality information*.

One of the institutions we rely on in our daily lives to filter out the lower quality information and to guide our focus to important matters is the news media. Here, the central role of information mediator is occupied by the journalist. In the next Section, we discuss the development of journalism as a vital part of democracy, as well as in what way it has developed to serve the needs of the information society.

## 2.2 Journalism

Journalists and the news media not only play a big role in the dissemination of information, but also influence the public understanding and interpretation of the provided news and therefore shape public opinion (McCombs, 2003). As Lippmann (1920) put it: “The only feeling that anyone can have about an event he does not experience is the feeling aroused by his mental image of that event.“. One can therefore argue, that the media have a responsibility to provide their audience with high quality information. This idea was addressed in the *social responsibility theory*, which was first formulated in 1959. It appeared in the influential book *Four theories of the Press* and emphasizes “the obligation to provide [the audience with] trustworthy and relevant news and information” (Siebert et al. 1959, as cited in Middleton, 2009). By shaping the opinion of the citizens, news media are “intimately linked” to the dynamics of democracy (Dahlgren, 1995). However, this is not the only role news media play in a democratic society. Another one is the so-called *watchdog role*. It refers to the protection of public interests by keeping an eye on affairs within political and civil institutions. The role includes policing these institutions and alerting citizens to abuses of power, such as corruption (Breit & Harrison, 2004). Journalists in a traditional sense were mostly supposed to be passive, naive and innocent - reporting on events but not creating news themselves (Meyer, 1991). However, in order for news media to be able to execute both of these roles, journalists need to be more than just mere observers. “The world has become so complicated, the growth of available information so explosive, that the journalist needs to be a filter as well as a transmitter, an

---

organizer and interpreter as well as one who gathers and delivers facts” (Meyer, 1991). As a result of the new demands on the news reporters, a new type of journalism started to develop in the 1960s. It involved applying research methods used in social sciences such as statistical analysis or survey methods on the journalistic practice. We use the term *journalistic research* to refer to the process of using these methods in journalism. The term *precision journalism* was coined in 1970 by Philip Meyer in his book of the same title. The new journalism aimed at being precise, objective and truthful. This shift armed journalists with tools that increased the efficiency of information gathering. However, while precision journalism was devoted more to the data gathering process, in the information society information is easily accessible. This change called for another approach to news making. The focus has therefore shifted to the methods of processing the information (Gray et al., 2011). At this point, journalism needed to change once again.

In the next Section, we introduce the term data journalism. We explain reasons for its existence, as well as introduce some tools and terms that data journalism operates with.

### **2.3 Data journalism**

Data journalism emerged as an answer to the worldwide phenomenon of increasing quantities of available information. It provides the storyteller with necessary tools to analyze information and to create stories from the emerging patterns. It stresses the usage of the computational methods available to journalists in order “to take in and process [the information] with the speed and volume that it comes to us.” (Gray et al., 2011).

*Data* refers to factual information. Whether accessible as a database, spreadsheet or a text file, data is nowadays acquired from the Internet. A structured machine-readable collection of data is called a *data set* and it most commonly takes the form of a table with various rows and columns. A data set can be produced and published for example by a statistical institute or a non-governmental organization (NGO), but it can also be created by collecting information from various sources such as websites or by creating quantitative or qualitative surveys. *Open data* are data sets that are released freely and are meant to be reused and shared without restrictions (OKF, 2012). In the last few years, governments around the world began to embrace open data as a tool to increase government transparency. As a result, many governments started to publish some of their own data sets online, as for example the United Kingdom, the United States, The

---

Netherlands, but also countries from the former Eastern Block. Open data have also been released by various NGOs and news sites such as The Guardian, who publishes the data sets they have used in the creation of their own news stories for their readers to explore. In this way, institutions provide access to the raw, unprocessed sources of their investigation. “Rather than the numbers belonging to the experts, they belong to all of us - and data journalism is part of that reclaiming of the facts.” (Rogers, 2012) Together with large amounts of information, tools that help clean, analyze and visualize the data have become freely available for the wide public to use. For example: Scraperwiki<sup>7</sup> allows non-programmers to use website scraping tools; Google refine<sup>8</sup> and Google fusion tables<sup>9</sup> help to clean and merge different sets of data; Open source applications like PSPP (the free alternative to SPSS) or the programming language R provide easy access to statistical analysis; Gephi<sup>10</sup> can analyze and visualize social networks; Datawrapper<sup>11</sup> visualizes interactive charts and Piktochart<sup>12</sup> provides tools to create infographics without the need to use vector graphics software (such as the Adobe Illustrator). In a 2012 article in The Guardian, Simon Rogers compared data journalism to the Do It Yourself essence of punk. The punk movement encouraged young people to engage with musical instruments with no previous musical education. One does not have to be a programmer or a developer to be a data journalist. The emphasis lies more on the ability to harness the power that these tools can provide (Rogers, 2012).

The wide availability of governmental raw data increases governmental transparency. Journalists are nowadays able to come closer to more and more of what is happening within the public and private institutions. This allows them to provide more accurate information to the public. In order to do so, they need to deploy various tools to analyze the data. While for example statistical analysis focuses on the attributes of the actors themselves, SNA provides tools for analysis of their mutual relationships. As we will see in the following Section on SNA, relationships are formed widely and by all kinds of interactions.

---

<sup>7</sup> <https://scraperwiki.com/>

<sup>8</sup> <https://code.google.com/p/google-refine/>

<sup>9</sup> <http://www.google.com/drive/start/apps.html#fusiontables>

<sup>10</sup> <http://gephi.org/>

<sup>11</sup> <http://datawrapper.de/>

<sup>12</sup> <http://piktochart.com/>

---

## 2.4 Social network analysis

Social network analysis is a social science method that provides insight into the structure of networks. A network consists of certain actors and their mutual relationships. The actors within a network can represent basically anything or anyone that can form an arbitrary relationship with something or someone else, as for example people, computers, animals or even neurons. For instance, individual people are embedded into various types of relationships and therefore they form various networks. This ranges from the most basic family networks, through friendships, up to business-to-business professional networks. Networks as these form the basis of a functioning society. Social network analysis is a method that helps us to analyze these networks. According to Freeman (2011), it involves the following four properties:

- (1) it emphasizes the analysis of relationships between actors
- (2) it analyzes data that characterize these relationships
- (3) visual outcome plays an important role in revealing patterns in these relationships
- (4) it utilizes mathematical models to describe and explain these patterns

### 2.4.1 Basic principles

Social network analysis can produce two forms of output. One is a visual representation of the network, also called a *social network graph*. The graph can be static, meaning that it represents the network at a certain point in time. It can be also dynamic, changing according to the formation, strengthening and dissolution of relationships in time. It is also possible to calculate a set of quantitative metrics that uncover additional characteristics of the network (Johnson & Reitzel, 2011), as for example centrality measures. SNA research provides insights into social regularities, and more importantly into the patterns of information flow and human behavior. This Section introduces some of the key principles of SNA. It is based on the article by Borgatti et al. (2009), the influential book on SNA by Wasserman and Faust (1994) and an article on development within the SNA field by Faust (2011).

#### Nodes and edges

---

The basic element of a network is a relationship between two actors, or *nodes*, also called a *dyadic relationship*. Social network analysis is not only concerned with describing the attributes of individual actors, but also with the relationships among the actors. A node represents a certain actor, for example a person or an institutions involved in the network. An actor is connected to another actor with an *edge* (also called a *link* or a *vertex*), which represents a certain relationship (e.g. A gave a gift to B). Multiple actors and their relationships form a network. The relationship can be either *directed* (e.g. A gave money to B creates a directed relationship from A to B) or *undirected* (e.g. A went to school with B). The number of connections a certain actor has is called a *degree* (e.g. A has three friends yields a degree of three). Some connections can be measured by their frequencies, such as interactions (e.g. A sent three e-mails to B). Frequency of an interaction is used to measure relational strength, the so-called *weight* of the edge. Increased communication therefore leads to a stronger relation. Measurement of the number of relationships that also takes into account the frequency of the interactions is called a *weighted degree*.

The differences between types of actors (nodes) within a network also have an impact on the analysis of that network. While a network of someone's Facebook friends yields only one type of nodes, a network of Facebook pages and their fans would create a network with two types of actors: the first type would be the fan pages and the second type would be the people that 'liked' the pages. The types of actors that form a network are called *modes* and different *modality* requires different approaches to the network analysis. The computational techniques have to take into account that for example the Facebook fan page will probably have a significantly higher number of relationships (being "liked" by the lots of people) than the fans themselves.

## Network

There are different levels of quantitative analysis that one can perform on a network. For the purpose of this thesis however, we are going to be concerned only with the positions of a node within a network. We focus on the *centrality* of a node. Centrality is a measure of the importance of nodes within the network. There are three computational methods on how to compute the centrality according to the actor's position within the network. It often requires the

---

calculation of *shortest paths*. The shortest path is the minimal amount of “hops” one node has to do to reach another one (Figure 1).

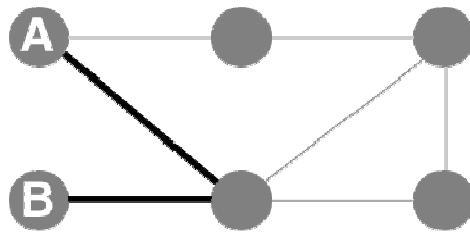


Figure 1. The shortest path from A to B

*Closeness centrality* refers to the closeness of one actor to the rest of the actors in the network through the shortest paths it forms. The less number of “hops” an actor has to make to reach the rest of the other actors in the network, the higher its closeness centrality (Figure 2).

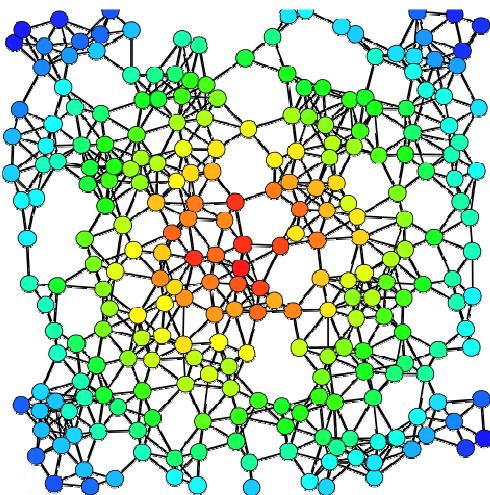
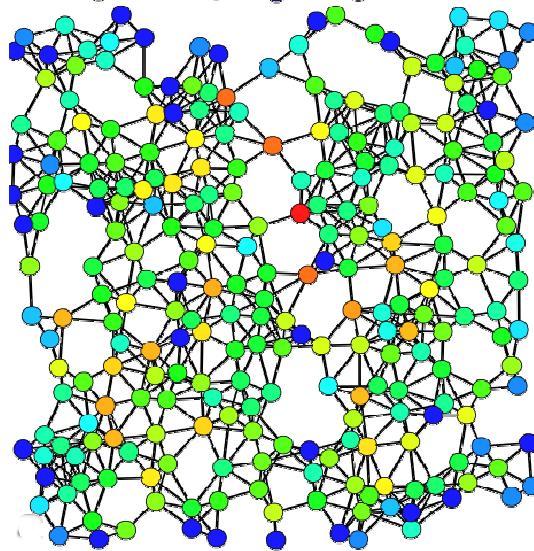


Figure 2. Red nodes have the highest closeness centrality, blue nodes have the lowest closeness centrality. Reprinted from Wikipedia, the free encyclopedia by Claudio Rocchini. Retrieved from [https://en.wikipedia.org/wiki/File:Centralit\\_y.svg](https://en.wikipedia.org/wiki/File:Centralit_y.svg)

*Betweenness centrality* measures how often an actor is on the shortest paths between all pairs of nodes within a network. These actors oftentimes serve as hubs for information exchange, as

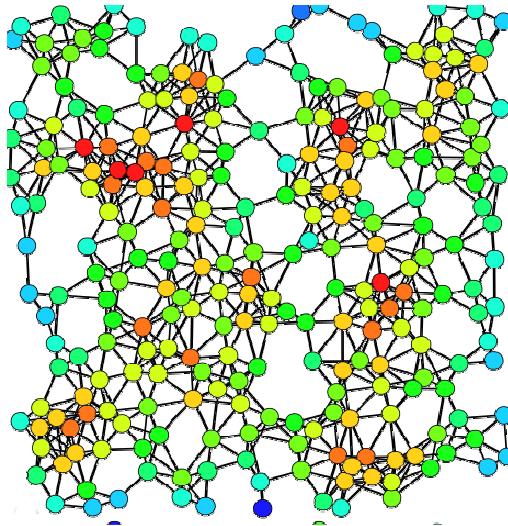
---

information from one actor has to pass through them to get to another one. Therefore they can control for example the communication within the network (Figure 3).



*Figure 3.* Red nodes have the highest betweenness centrality, blue nodes have the lowest betweenness centrality. Reprinted from Wikipedia, the free encyclopedia by Claudio Rocchini. Retrieved from <https://en.wikipedia.org/wiki/File:Centrality.svg>

And finally, the simplest measure is *degree centrality*. It refers to the number of relationships (degrees) an actor forms with other actors within the network (Figure 4).



*Figure 4.* Red nodes have the highest degree centrality, blue nodes have the lowest degree centrality. Reprinted from Wikipedia, the free encyclopedia by Claudio Rocchini. Retrieved from  
<https://en.wikipedia.org/wiki/File:Centrality.svg>

### Social network graph

Descriptive information about the actors can be reflected in the visual representation to help the visual analysis of the network. Nodes and edges can have different colors and sizes (or thickness in case of the edges) that correspond to their attributes. For example a thick edge representing frequent message exchange; or bigger nodes representing taller people and smaller nodes representing shorter people.

As stated before, a network is formed by actors and their interactions. Public procurement therefore also forms a network, as the public institutions and private companies sign contracts with each other. Public procurement creates a network of monetary flows. In the following Section, we will briefly familiarize ourselves with the public procurement practice.

## 2.5 Public procurement

---

The “acquisition of goods and services by government or public sector organizations” is called public procurement (Lindskog et al., 2010). On average, public procurement accounts for around 15% of the GDP in the OECD countries<sup>13</sup> and therefore represents a substantial part of any country’s economy (OECD, 2007; Lindskog et al., 2010). For private companies, government contracts represent very lucrative and stable business opportunities, and also create new business opportunities for their suppliers (OECD, 2007). Acquiring a contract with the government is therefore very advantageous for the private sector. Two to three key actors participate in public procurement. The two most essential ones are the procurer, the procuring governmental organization, and the supplier, the private company offering services to the procurer. Optionally an intermediary can be used by the procurer or supplier, such as a consultant (OECD, 2009). The process of governmental procuring consists of several phases. The following four key phases were presented in Strand et al. (2011): pre-award, award, post award and litigation and complaint (if applicable). In the pre-award phase, the procurer identifies the requirements of the procurement and forms the “request for tender” (RfT). A RfT is a document where the procurer specifies the requirements for the procured services, as well as the anticipated price of the procurement. After created and published, the request is open for proposals by the suppliers. Proposals include the method and costs of the services the supplier can offer. An RfT cannot be altered once it has been published (Lindskog et al., 2010). Therefore it has to be designed carefully by the procurer. In the award phase, the suppliers engage in presenting their proposals in a competitive procedure<sup>14</sup>. It means that “contracts are awarded by comparing offers from a certain number of [suppliers] in order to establish which one can provide the most favorable terms for delivering the government’s requirements.” (OECD, 2007). While exclusively the merits of the supplier’s proposal (e.g., offered price, quality) should be considered, the procurer evaluates the proposals and chooses a winner. Finally, the procurer informs the participating suppliers about the decision in the post award phase. If applicable, suppliers may thereafter choose to complain or litigate about the procedure. Once the procedure

---

<sup>13</sup> Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, South Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States

<sup>14</sup> More types of procedures can be followed in a public procurement, including some that are non-competitive. In these cases the suppliers are approached by the procurers. The competitive procedure however, is the most common one.

---

is finished, the procurer and the winning supplier sign a contract in order to seal the deal. Each public procurement procedure is treated as a new task and therefore requires a new competitive process (Lindskog et al., 2010). There are numerous ways that the participants can disrupt the procedure in order to gain unlawful advantages. The next Section will introduce the key concepts that we use in analyzing unlawful behavior in public procurement. In Section 2.7. we elaborate on how this unlawful behavior manifests itself within public procurement.

## 2.6 Corrupt behavior

*Corrupt behavior* is an umbrella term that we use throughout the thesis to refer to three types of unlawful activities, namely corruption, collusion and conflict of interest. *Corruption* mostly refers to bribery. *Collusion* corresponds to cartel agreements between institutions, and *conflict of interest* is a state when conflicting professional or personal interests of a public official can influence his public duties (Begley et al., 2009). Corrupt behavior can occur in both the private and in public sector.

### Corruption

One of the definitions of corruption is the “abuse of entrusted power for private gain” (Lindskog et al., 2010). Corruption mostly refers to bribing. A bribe usually takes the form of money or a gift in exchange for favors. The act of paying a bribe is being referred to as *active corruption*, while the act of receiving or asking for a bribe is a *passive corruption*. Any form of corruption undermines the system of laws and regulations, diminishes authority and eventually can lead to disruptions within the competitive market (Wempe, 2002). The structure and actors in corruption schemes vary between and within domains and countries with respect to different actors or values (Wempe, 2002). Corruption charges are often pressed against public officials or politicians (political corruption), but also take place in the private sector.

### Collusion

The probably better-known term for collusion is *cartel forming*. The actors in this kind of secretive agreement are mostly parties that aim to control the market by dividing its parts between each other, setting prices or limiting production. Collusion is an illegal practice that works against the principles of a competitive market by disrupting the competition between the colluding companies, as well as hindering other companies’ ability to compete by preventing

---

them from entering the market (Mizruchi, 1996). Experimental research by Muren and Pyddoke (1999) suggests that in order to build and maintain mutual trust among the members of a collusive agreement, communication plays an important role in collusion. A form of creating information channels between different companies arises when companies share one or more directors on their executive or control boards. Hereby, they create a so-called *interlock*. Even if this practice is legal, there have been discussions on the potential collusive effects it can have. However, the research has delivered inconsistent results, confirming as well as disproving the possibility that an interlock can lead to collusive practices (Mizruchi, 1996). Therefore, even if interlocks themselves are not an indicator of collusion, they are a direct indicator of an intercompany relationship and can be relevant when combined with other collusion indicators.

### **Conflict of interest**

Exploiting one's own position in order to satisfy competing professional or personal interests can take many forms. For example a government official's privately owned business can create a conflict of interest while executing his or her duties in the public sector. Another common form is preferential treatment like contracting a long-term friend, family member or an associate mainly because of the mutual relationship and not because of their expertise or skills (OECD, 2007). A public official does not necessarily have to behave corruptly in a conflict of interest situation. However, the situation holds the potential for abusing entrusted power (Genckaya, 2009).

In all its forms, corrupt behavior creates an undue advantage for some of the participating parties, whether it is reaching business or private goals.

### **2.7 Indicators of corrupt behavior within public procurement**

Because of the many business advantages of acquiring a governmental contract, the competition between suppliers can be fierce (OECD, 2007). Therefore suppliers do not always play by the rules set by the legal boundaries. Indicators that may signal the possibility of corrupt behavior in the public procurement process are called *red flags*. There are numerous types of red flags that can be looked for at each of the stages of the procurement process.

The public procurement process may attract corrupt behavior at any of its stages, which may involve any of the participants. The whole spectrum of public officials, from procurement

---

officials to high-level politicians, can be affected by corrupt behavior. At the initial (pre-award) stage, a corrupt procurer can choose to tailor the Request for Tender towards the products of a certain supplier by specifying unnecessary features offered only by that supplier. In this way, other bidders are excluded from the competition. *Too narrow contract specifications* are a red flag (OECD, 2007).

The award stage can also be hijacked by collusive practices of suppliers. This oftentimes takes the form of *bid rigging*. Members of a collusive agreement choose the winner of a certain procurement process and they all bid accordingly, for example by offering significantly higher prices than the pre-chosen winner, or by lowering the number of bidders in the tendering. The members of the collusive agreement can then take turns in winning future procurements. *Rotation of winning bidders* is another red flag as well as *low competitiveness in the bidding* for tenders (OECD, 2010). As each tender is a new task open for competition, *repeated awards to the same company over a period of time* also raises a red flag (OECD, 2010). Another red flag is risen by so-called *revolving door*, that occurs when a “person with previous or current government experience who also has held, or currently holds, a professional position in the private sector where they can reasonably be expected to influence [...] public policy decisions” (CRP, 2013). Revolving door can therefore result in a conflict of interest.

It is however important to note that a red flag can also occur due to other causes than corruption and while they point to issues within the procurement process itself, they do not necessarily signal corrupt behavior in the procurement (Kenny & Musatova, 2010). As a counter example, low competitiveness in the procurement and high volumes of contracts towards one company can also be caused by the uniqueness of the product that the company delivers (Dančíková, personal communication, June 6, 2013). The proper evaluation therefore requires domain knowledge and context in order to be interpreted accurately.

Nevertheless, it is generally difficult to detect patterns of corrupt behavior (OECD, 2007). This poses a problem in stopping these practices. Corrupt behavior in public procurement results into inefficient spending of public funds. It also distorts the competitiveness of the procurement market by restraining access to the market to other companies. As a result, many companies may in the end give up on taking part in the public procurement resulting in limited companies procuring for the institutions (OECD, 2007).

---

It is obvious that a healthy competitive market is crucial for effective public spending. Empirical studies conducted on public procurement contracts show that, the number of bids in the competitive process influences the costs of the procurement. Fostering competition by minimizing corrupt behavior on the procurement market may therefore help to decrease the expenses of public procurement contracts and make spending of public resources more efficient.

## **2.9 Corrupt behavior in Slovak public procurement**

According to a the 2012 Transparency International research called “Corruption Perceptions Index”, that uses a set of surveys to score how corrupt the public sector is perceived to be on a scale from 0 (corrupt) to 100 (not corrupt), Slovakia scored 46. In comparison, The Netherlands have scored 84. (Transparency International, 2012). Public procurement is being carried out by public officials and is therefore a part of the public sector. According to a report on public procurement based on TED<sup>15</sup> data set, surveys and in-depth interviews prepared for European Commission, the competitiveness on the Slovak public procurement market is the smallest in the EU. The median number of bids placed in Slovak tenders is only 1, while the median in the EU is 4 bids per tender. Countries with the most bids per tender are Spain and Germany where the median is 6 bids per tender. As one of the main results of corrupt behavior is the distortion of the competition, the lack of competitiveness in the Slovak public procurement market may be indicator of corrupt behavior within this market. Low competitiveness in the bidding is also one of the red flags of corruption (Figure 5).

---

<sup>15</sup> TED (Tenders Electronic Daily) is the online version of the 'Supplement to the Official Journal of the European Union', dedicated to European public procurement. It is maintained by the EU Commission, contains detailed information on about 540 000 individual purchases of 1,4 trillion euro in value by authorities in 30 countries over the 2006-2010 period.

Figure 2.25 Number of offers per country  
(median, mean in lines)

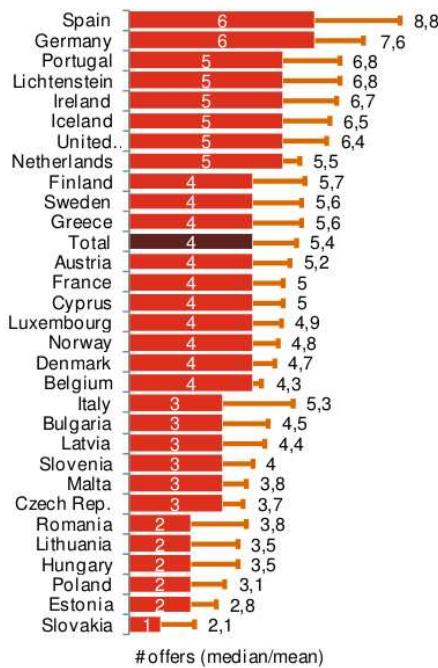


Figure 5. Median bids per tender in EU countries. Slovakia has the lowest median.  
Reprinted from the study “Public procurement in Europe” by Strand et. al., 2011

Qualitative journalistic research (interviews) also indicates, that the Requests for Tenders also tend to contain too narrow and unreasonable requests for some of purchases. For example while procuring CT scans, one of the hospitals explicitly required that the scanner uses water cooling instead of cooling by air, while there was only one company in Slovakia that could fulfill this requirement. However, this requirement was considered unimportant by a radiology specialist in the interview (Dančíková, 2012b). Another supplier of medical equipment compared some of the requirements in Requests for Tenders to the requirement that the “steering wheel has a diameter of 37 cm” (Beer & Slovak, 2003, translated by author) while buying a car. He also added, that companies that do not find a way to get the requirements of the Request for Tender tailored to their products do mostly not participate on public procurement (Beer & Slovak, 2003). These are strong allegations, however they remain difficult to prove. Nevertheless, narrow requirements in the Request for Tender are listed among the red flags of corruption. Because of the reasons listed

---

above, we assume that corrupted behavior is in some way present in the Slovak public procurement network.

In the following Section, we briefly touch upon literature that explores corrupt behavior using social network analysis.

## 2.9 Social Network Analysis of corruption

Up to now, there has been some research into various types of corrupt activities through social network analysis. As “systematic corruption requires a durable participation of different agents, any form of systematic corruption [...] implies establishing durable social relationships.” (Garay et al., 2010) Social relationships define social networks. SNA methodology provides a useful framework for examining these networks. Aven (2012) studied the effects of corruption on the communication network of Enron<sup>16</sup> employees that engaged in fraudulent accounting practices. Multiple projects (involving multiple employees) within the company were marked as corrupt as the result of an investigation by the U.S. Securities and Exchange Commission<sup>17</sup>. The data source used for this analysis was the email communication between multiple actors that were participating in these projects. The research revealed that, compared to the non-corrupt networks within the organization, the corrupt network showed a decreased frequency of information exchange between the actors, due to the risk of being discovered while sharing information. The paper also pointed out a larger number of central actors, as well as overall low interconnectedness of the network. This implied a higher control over the information flow in the network in the hands of fewer agents. In the book *Illicit Networks, Reconfiguring States*, Garay et al. (2010) used SNA to examine how unlawful groups such as drug cartels or paramilitary groups “reach agreements with authorities in order to infiltrate State institutions” through corruption and bribery (Garay et al., 2010). The connections between the individual actors were based on relationships documented in legal files. Here, network centrality measures were applied to emphasize important actors within the networks. Similarly, Lauchs et al. (2011) identified key actors in a long lasting corrupt network within the Queensland Police Force by using centrality measures.

---

<sup>16</sup> Enron was an American energy, commodities, and services company based Texas. It went bankrupt in 2001.

<sup>17</sup> U.S. Securities and Exchange Commission serves to protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation.

---

As far as we know, no previous research applies social network analysis on public procurement network. This can have numerous reasons, one of them being that the data to conduct such research is still sparse. In the next Section, we introduce our research questions.

## **2.10 Research questions**

In line with the theory introduced above, we have formulated four research questions to demonstrate how SNA can be used in journalistic research on public procurement:

- (1) *Do measures of the position of an actor within a network point out individuals that are important to the information flow in the public procurement network?*
- (2) *Does the social network graph expose conflict of interest situations?*
- (3) *Does the social network graph expose interlocking directorates?*
- (4) *Does the social network graph display differences in the magnitude of the flow of contracts between governmental institutions and private companies?*
- (5) *Does visualization of the social network graph in time reveal possibly corruptive patterns in awarding contracts between governmental institutions and private companies?*

Research question 1 is based on previous research on corrupted social network, where centrality measures were applied to point out important actors in the network. To answer the questions 2, 3, 4 and 5 we investigate how red flags manifest themselves in a public procurement social network. In the following Chapter we introduce the methodology we have used in order to answer these questions.

---

### **3. Methodology**

The thesis is primarily concerned with whether social network analysis can be utilized in the journalistic research on public procurement, as well whether we can detect red flags by utilizing this method. In this Chapter, we will introduce methodology used in order to answer our research questions.

#### **3.1 Research methods**

In order to answer the central research question and its sub questions, we will be using a descriptive method of research, namely a case study. A case study aims to understand dynamics of a phenomenon by investigating it within its real-life context and it is appropriate to use especially when there is little knowledge about a certain phenomenon (Halinen & Törnroos, 2005). As no prior research on the dynamics of a social network of public procurement exists and we will be analyzing a phenomenon in its real-life context, this method is appropriate for our research. Afterwards, social network analysis will be applied on the data in order to visualize and analyze the Slovak procurement network in the years 2005-2012. The decision to focus on utilizing social network analysis was made because of the research gap in using this method in the analysis of public procurement data.

#### **3.2 Variables**

Because the focus of SNA is on relationships between actors, both actors and their relationships are considered as variables. The networks consist of three types of actors. These will be government institutions, private companies and individuals in the top management positions in the companies. The attributes of the private companies and the government institutions will include for example the name of the institution, the type (private or governmental), or the sum of money they gave out or acquired in the procurement process, in order to visually distinguish the “big players” and “big spenders” in the social network graph. The only attribute of the individuals will be the name and one of the companies he or she works for. It is not necessary to include all of the companies, as he or she will be connected to other people only if he or she works in the top management of the same company. It will be therefore

---

easy to get this information from the social network graph. Other quantitative attributes as for example centrality measures of the actors will be created during the analysis.

### 3.3 Study design

In our case study we apply SNA on the public procurement network in order to search for red flags that can signal corrupt behavior in the public procurement. To answer our research questions, we will create two types of network. These networks contain different variables and therefore require a different approach to the analysis.

#### Institutions Network

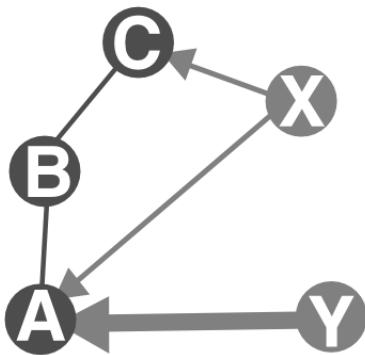
In the *institutions network*, we will use the data from the procurement contracts to visualize the flow of contracts from governmental institutions to private companies. The information acquired from the chamber of commerce will be used to connect the private company to each other through their directors. The network will consist of two government institutions and private companies. This will make it a bi-modal network. Unfortunately, Gephi is not able to analyze multimodal networks and it is therefore not possible to calculate the quantitative aspects of the network<sup>18</sup>. The focus will therefore lie on the visual representation of the network, or the the social network graph. This network will be used to answer the research question 3, 4 and 5.

#### *An example:*

Company A has 10 contracts with government institution Y (thick edge) and 2 contracts with government institution X (thin edge). It also shares a director with company B. Company B shares a director with company C and company C has 2 contracts with government institution X (see Figure 6).

---

<sup>18</sup> There are of course other tools that are capable of analysing multimodal networks. For the purposes of this thesis I will however not go beyond analysis that can be done with Gephi.



*Figure 6. Example social network graph in the companies network*

### Individuals Network

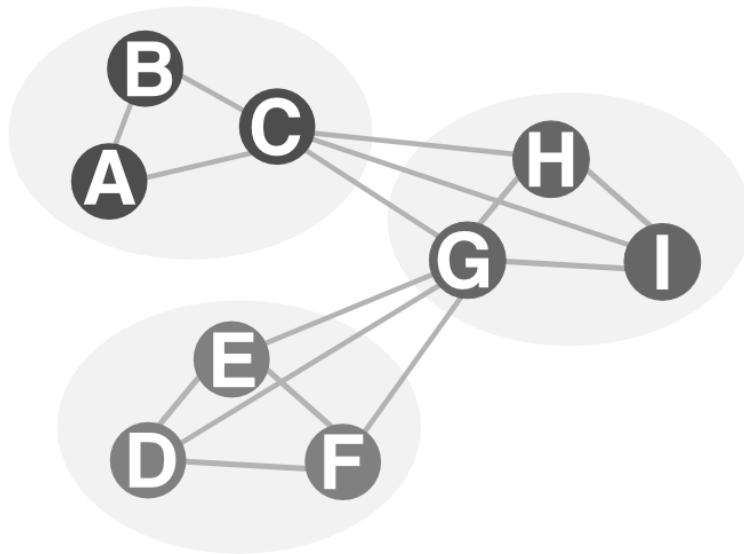
The second network will bring the relationships between *individuals* into perspective. In this network we will use the information from the chamber of commerce to visualize the relationships between the companies top management (or directors). In this network the focus will lay on the most connected individuals within the network through centrality analysis. The network will be created by connecting the personal network of each director in the government institutions. By “personal network” we mean people that occupied top management positions within the same companies as these directors since the year 2006. We can look for interlocking directorates by visualizing the relationships directors from governmental institutions may have outside the network of the government institution with other private companies. Three types of individuals will be present in the network. 1) directors of the governmental institutions; 2) directors of private companies that are doing business with the institutions; 3) directors of private companies that are not doing business with the institutions. These types will be color coded in the network. This network will be used to answer the research questions 1 and 2.

#### *An example:*

A,B and C are directors in a governmental institution AB & C. C is also a director in the CHG & I private company that does not procure for the governmental institutions. However, G is one of the directors in a private company GDE & F that does procure for the governmental institution. Despite that the governmental institution and the private

---

company that does procure for them do not have a direct link, G can serve as hub for their interaction (see Figure 7).



*Figure 6.* Example social network graph in the individuals network

In the following section, we will have a brief look at some reports focusing on the corrupt behavior within the public sector and public procurement in the EU, to explain the belief, that corrupt behavior is present in the Slovak public procurement.

---

## 4. Data

### 4.1 Collection and content

We have acquired data sets from Slovak public procurement and Slovak chamber of commerce. The data sets were used for the creation of the networks, but they contain qualitative information as well. Decision to use the data from Slovakia was made for several reasons. The initial one is the Slovak background of the author. Another one is the questionable “health” of the Slovak public procurement market. However, the most important factor is the availability of the data. In Slovakia it is mandatory to publish governmental contracts online. Transparency International Slovakia is active in collecting and cleaning the published governmental contracts and putting them into data sets that are freely available online.<sup>19</sup> For the analysis, four different data sets are used.

#### Transparency data set

Three data sets come from the official database of public contracts in Slovakia and consists of data on public procurement contracts collected from 2005 to 2012. These will be further referred to as “Transparency data sets”. Publishing each contract online has been made mandatory by a change in procurement law in only the year 2011. Data sets created from these contracts are however often incomplete and inconsistent. Each institution is responsible for posting its own contracts. As a result we encounter many different forms of publishing these contracts, whether we talk about file formats or document structures. This makes it difficult to automatically create structured data sets. Therefore, data collected from the Slovak tender website are used.<sup>20</sup> These are less prone to errors because of a normalized format. They also cover a longer period of time. Even if this data is not a complete set of public procurement contracts, these contracts account for most of funds spent in procurements in Slovakia. Table 1 shows the amounts of published contracts per year.

---

<sup>19</sup> [www.tender.sme.sk](http://www.tender.sme.sk)

<sup>20</sup> [www.uvo.gov.sk/evestnik](http://www.uvo.gov.sk/evestnik)

---

**Table 1.** Contracts published per year in the Transparency international data set

year	number of contracts	year	number of contracts
2005	9373	2009	4323
2006	7842	2010	4919
2007	3690	2011	4655
2008	2825	2012	5420
<b>total</b>	<b>43086</b>		

Table 2 shows an example of the data. Each row represents one contract. *Competition type* is the type of procurement process (e.g. public competition), *award criteria* indicate the reasons for the contract to be awarded to a certain supplier (e.g. lowest price). An order can contain multiple contracts with multiple suppliers. Table 2a contains descriptive statistics on the contract values.

**Table 2.** Example data of the first Transparency international data set

id	contract key	bulletin date	contract name	competition	competition	contract type id	contract type
				type id	type		
2049	2009-2854-1	1/17/2009	Ultrazvukový prístroj pre kardiológiu	7	verejná súťaž <sup>21</sup>	5	tovary <sup>22</sup>
42403	2012-43411-1	10/2/2012	Biologické srdcové chlopne	7	verejná súťaž	5	tovary
42413	2012-43411-6	10/2/2012	Biologické srdcové chlopne	7	verejná súťaž	5	tovary
42399	2012-43411-7	10/2/2012	Biologické srdcové chlopne	7	verejná súťaž	5	tovary
42408	2012-43411-8	10/2/2012	Biologické srdcové chlopne	7	verejná súťaž	5	tovary

---

<sup>21</sup> public competition

<sup>22</sup> goods

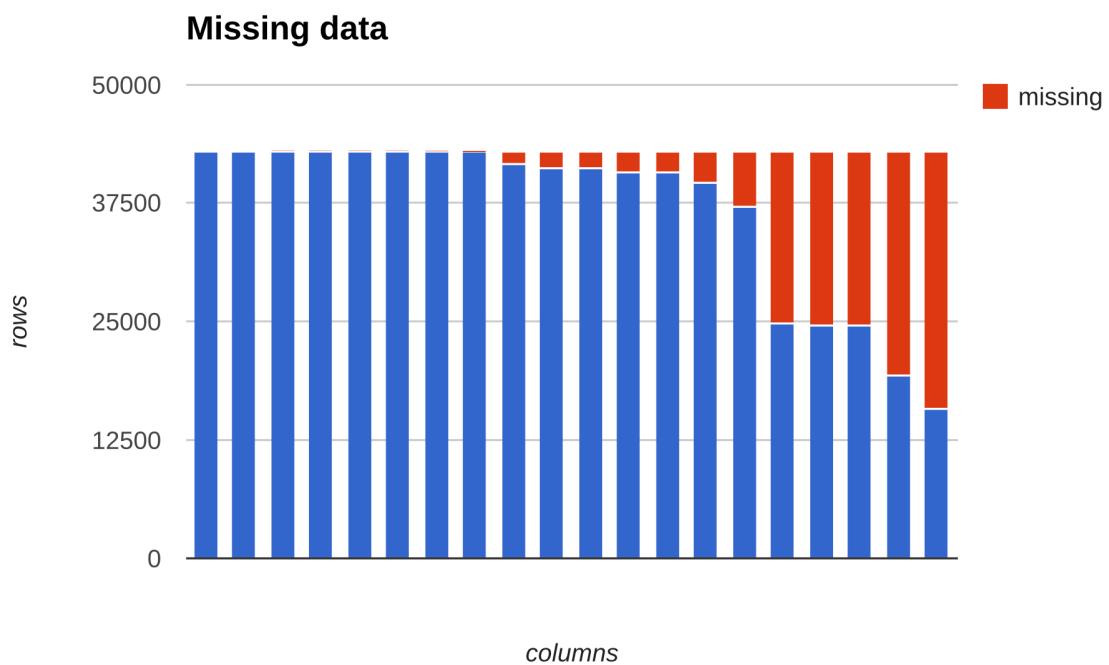
award criteria id	award criteria	id procurer	id supplier	number of		number of contracts	order value	contract value
				bids	contracts			
4	najnižšia cena <sup>23</sup>	35971126	31349935	1	1	234192.95	234192.95	
4	najnižšia cena	35971126	35849258		9	2213576	245952.88	
4	najnižšia cena	35971126	602175		9	2213576	245952.88	
4	najnižšia cena	35971126	35757329		9	2213576	245952.88	
4	najnižšia cena	35971126	35757329		9	2213576	245952.88	

**Table 2a.** Descriptive statistics on the contract values (rounded)

N	42,971 (+ 115 missing)
<b>EUR</b>	
mean	740,998
std dev	22,537,000
median	63,333
min	1
max	420,417,000

Figure 8 contains an overview of data missing in all 20 columns. We encounter the most missing data in the columns (in descending order): number of bids, contract type, order value, award criteria and contract name.

<sup>23</sup> lowest price



*Figure 8.* Missing data in the Transparency international data set

The second and the third Transparency data sets contain detailed information about procurer and supplier institutions, as for example names of the institutions, region they reside in or in the case of the procuring institutions also procuring officials name (Table 3, 4).

**Table 3.** Example data of the second Transparency international data set (procurers)

<b>id</b>	<b>procurer_region</b>	<b>procurer_name</b>	<b>procurer_contact_email</b>	<b>procurer_contact_person</b>
35971126	Bratislava III	Národný ústav srdcových a cievnych chorôb; a.s.	obstaravanie@nusch.sk	Elena Lukáčová
35971126	Bratislava III	Národný ústav srdcových a cievnych chorôb; a.s.	maria.duricova@nusch.sk	Mária Ďuricová
36601284	Košice II	Východoslovenský ústav srdcových a cievnych chorôb; a.s.	zkudlacova@vusch.sk	Ing. Eduard Adamčík
36601284	Košice II	Východoslovenský ústav srdcových a cievnych chorôb; a.s.	dkapakova@vusch.sk	Ing. Marián Alb
36644331	Banská Bystrica	Stredoslovenský ústav srdcových a cievnych chorôb a.s.	valentinyova.jozefa@suscch.eu	Ing. Jozefa Valentínyová

---

**Table 4.** Example data of the third Transparency international data set (suppliers)

<b>id</b>	<b>supplier_region</b>	<b>supplier_name</b>
31349935	Bratislava II	S&T Slovakia s. r. o.
35849258	Bratislava IV	Biomedica Slovakia; s.r.o.
612791	Trenčín	UNOMED; spol. s r.o.
602175	Bratislava II	TIMED; s.r.o.
35757329	Bratislava IV	Operatíva; medicínska spoločnosť s.r.o.

### Chamber of Commerce data set

The fourth data set was created by parsing the information from the official Slovak Chamber of Commerce website. This will be further on referred to as “Chamber of Commerce data set”. In this data set, each row represents an entity (a person or a company, as sometimes companies are listed in the Chamber of Commerce as owners). It contains information on the senior management (name, name of the company and his or her relation to the company, address, etc.) (Table 5).

**Table 4.** Example data of the Chamber of Commerce data set

company								
<b>company id</b>	<b>name</b>	<b>relation</b>	<b>personal id</b>	<b>entity</b>	<b>name</b>	<b>address</b>	<b>from date</b>	<b>to date</b>
LUMEN							Petzwalova 46	
36529079	družstvo Nitra	štatutár	SKP1461794	person	Ján Lajda	Nitra	14.09.1998	
DAPEX							Široká 19 Nitra	
36535630	DAPEX s.r.o.	spoločník	SKC979052	company	Invest s.r.o.	949 05	28.05.2008	
Zvolenská 30								
36535630	DAPEX s.r.o.	spoločník	SKP1127427	person	Peter Hatiar	Nitra 949 11	16.03.2000	26.06.2001
Zvolenská 30								
36535630	DAPEX s.r.o.	štatutár	SKP1127427	person	Peter Hatiar	Nitra 949 11	16.03.2000	26.06.2001
Široká 19 Nitra								
36535630	DAPEX s.r.o.	spoločník	SKP1184681	person	Ján Križan	949 05	23.08.2008	

---

Information from the Chamber of Commerce are published online by the Ministry of Justice as a searchable database, but are not downloadable. This database was scraped and published within a project called “Firemny register”<sup>24</sup> (company register). One of the creators of the scraper willingly provided us with the Chamber of Commerce data in a *csv* format. It is important to note, that the data can also contain mistakes due to for example persons with the same name. This data set is used to create the *individuals network*.

## 4.2 Pre-processing

Because of the considerable size and complexity of the network, it would be complicated to analyze the whole network at once. Therefore, a sample is drawn from the network. The sample is drawn from the healthcare sector, because of small competition within healthcare procurement market and procurement practices that raise red flags of corrupted behavior, as for instance only one company bidding for a tender in 68% of the tenders (Dančíková, 2012). Three biggest governmental procurers within this market are chosen and examined closely. The average number of bids these institutions got in their tendering procedures in the years 2009 to 2012 are 1.0 and 1.1 for services and goods for hundreds of millions euro (Dančíková, 2012). Snowball sampling of the companies these institutions are doing business with is used in order to create two procurement networks of these institutions. Companies were removed from the Chamber of Commerce data set, as we are interested in the relationships between persons only. Table 6 contains descriptives of the sample.

**Table 6.** Descriptive statistics on the contract values of the sample

N	1235 (+ 0 missing)
EUR	
mean	441,396
std dev	1,239,778
median	142,285
min	381
max	28,140,776

---

<sup>24</sup> [www.firemny-register.sk](http://www.firemny-register.sk)

---

### 4.3 Network

By combining the data a network of procurers and suppliers is created. To create the *Company network*, first the three Transparency data sets are used. Here we are matching the *procurer id*'s with *supplier id*'s while both, procurer and companies, create nodes. Relationships (edges) between a company and a procurer represent the flow of procurement contracts the two actors signed with each other. The thickness (weight) of the edge indicates the number of contracts they share. The size of the company node represents the amount of money that has flowed into the company from the procuring institutions (*contract value*). The relationship is directed from the institution to the company. A relationship between companies is formed if they share a director in their top management. These edges are undirected. By creating directed and undirected edges in the network, in the analysis we can measure the number of government institutions the company does business with (indegree) as well as how many relationships a certain company builds with other companies (by subtracting the indegree from the degree). The government institutions only build relationships with companies and therefore create only one type of edges. To connect the companies to each other, we use Chamber of Commerce data set and match *company ids* through *personal ids*. We can match the data sets through *company ids* in the Chamber of commerce data set and *procurer ids* and *supplier ids* in the Transparency data sets, as these are unique Chamber of Commerce numbers.

To create the *Individuals network*, only the Chamber of commerce dataset is used. A relationship between the individuals is formed if they are both in the top management of the same company. In other words, we match the *personal ids* through *company ids*, while the *personal id*'s forms nodes and the *company id*'s form edges. This kind of relationship is undirected and no frequency is taken into account.

For the visualization of social network graph as well as for the calculations of the quantitative properties of the network (e.g. centrality measures), an open source tool called *Gephi* was used.

For the purpose of this thesis, we have chosen those red flags that fall into the scope of the data we are working with. With the data at our disposal we can for example neither examine any sudden rise in the wealth of a government official, nor can we uncover private relationships

---

between public officials and private sector employees. We can however visualize contract flow, contract value, and connect suppliers to procurers and institutions sharing directors.

In the following Chapter, we will have a look at the results of our analysis as well as give answers to our research questions.

---

## 5. Experiment and results

In our case study, we have examined a network of the three biggest government spenders in the Slovak health care sector. These are three hospitals that specialize in cardiovascular and circulatory illnesses. We have used three names in order to anonymize the hospitals: “National”, “Central” and “Eastern”. The analysis has provided us answers to our research questions. It was unfortunately not possible to include the Eastern hospital into the individuals network due to missing data.

### 5.1 Companies network

#### Introduction

The purpose of this experiment is to examine flow of contracts and company interconnectedness. Through directed edges, we measure the amount of contracts awarded to the private companies as well as localize private companies that got awarded the most contracts. The amount of contracts will then be expressed as weighted degree. Through undirected edges, we can localize companies that have management ties to other companies within this network. These measures will provide us with answers to two of our research questions, namely:

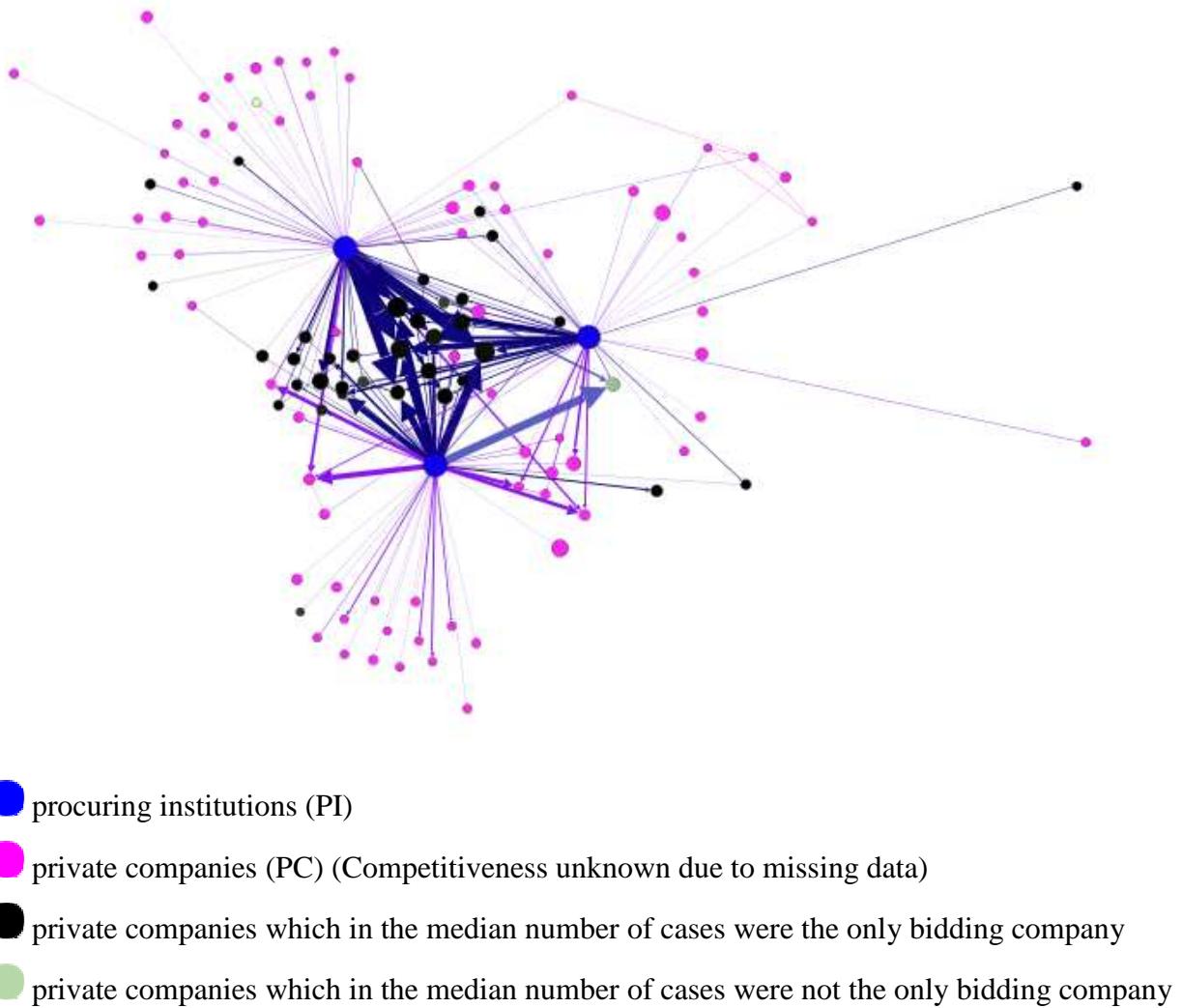
*Does the social network graph display differences in the magnitude of the flow of contracts between governmental institutions and private companies?*

And:

*Does the social network graph expose interlocking directorates?*

#### Results

The social network graph of the institutions reveals a structure of the flow of contracts between the governmental institutions and the companies (Figure 9).



*Figure 9.* Companies social network graph. Contains 122 nodes and 195 edges.

Not only does the social network graph reveal the magnitude of the flow of contracts between the procuring institutions (represented by the thickness of an edge) and private companies, but also shows the concentration of contracts to a group of companies that provide their services to all the three governmental institutions (Figure 10). The color of the nodes reveals, that most of these companies were the only bidding company in most of the procurement processes (black nodes).

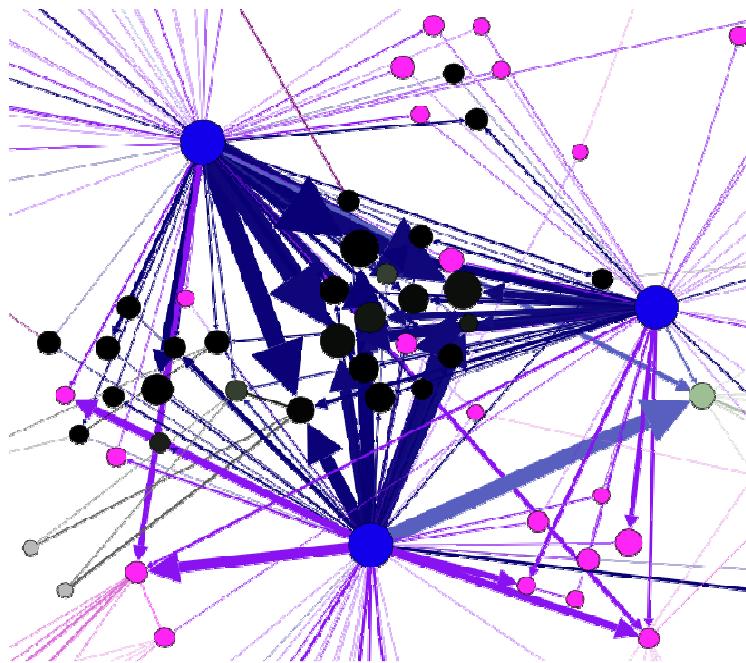
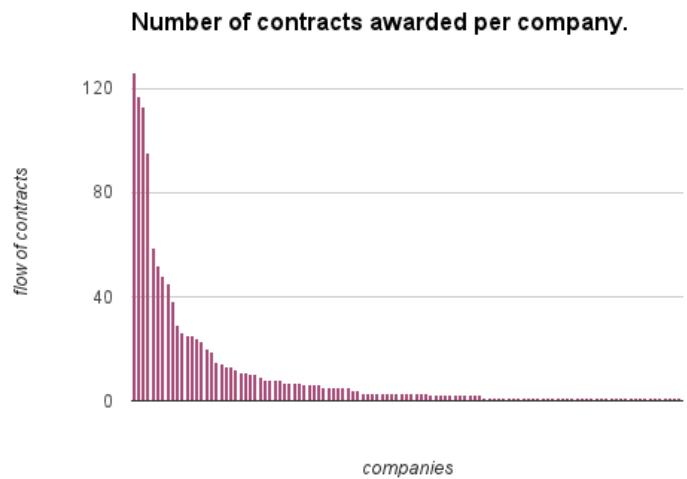
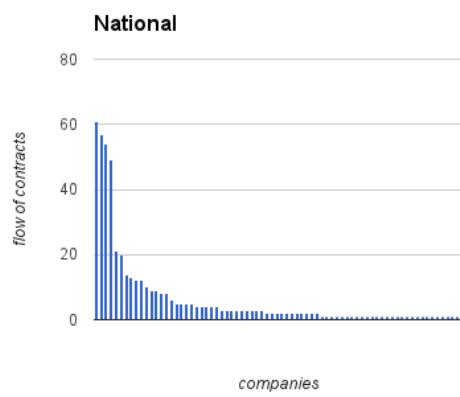


Figure 10. Companies awarded the most contracts. The thicker the edge from the blue node, the more contracts awarded.

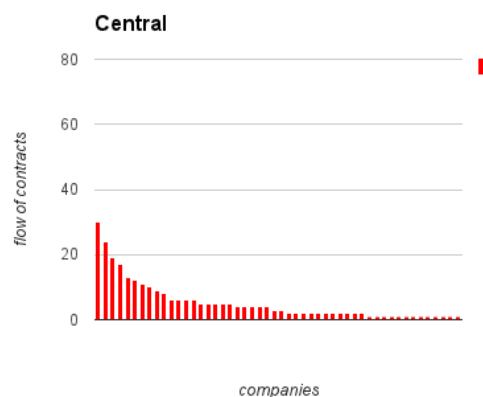
On top of the visual qualitative evaluation, we have applied quantitative operations on the degree measures within the network. The weighted degree of a procuring institution has pointed out the most active companies in the network (Figure 11). The frequency of contracts (weight of an edge) reveals the intensity of the flow of contracts between companies and institutions (Figures 12, 13, and 14). There is a relatively small number of companies that have a stronger relationships with the governmental institutions than the others.



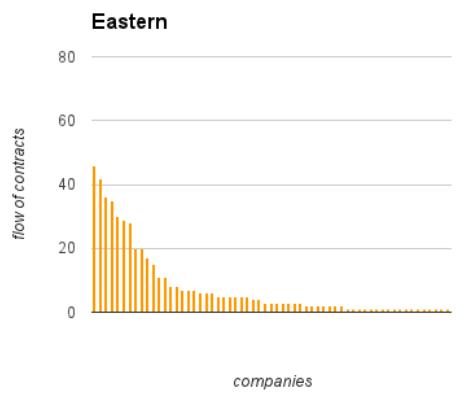
*Figure 11.* Number of contracts awarded per company by all three institutions



*Figure 12.* Number of contracts awarded per company by the National hospital



*Figure 13.* Number of contracts awarded per company by the Central hospital



*Figure 14.* Number of contracts awarded per company by the Eastern hospital

---

Social network visualization as well as degree measurements can point out a strong flow of contracts between governmental institutions and certain private companies.

After hiding the edges representing the flow of contracts, the graph reveals multiple networks of interlocked companies and hereby provides an overview of information flow between companies. As we can see in Figure 15, there are numerous interlocks present in this network. An interesting one represents the network of health insurance companies. Four of them acquired procurement contracts within this network. They are also well connected to each other.

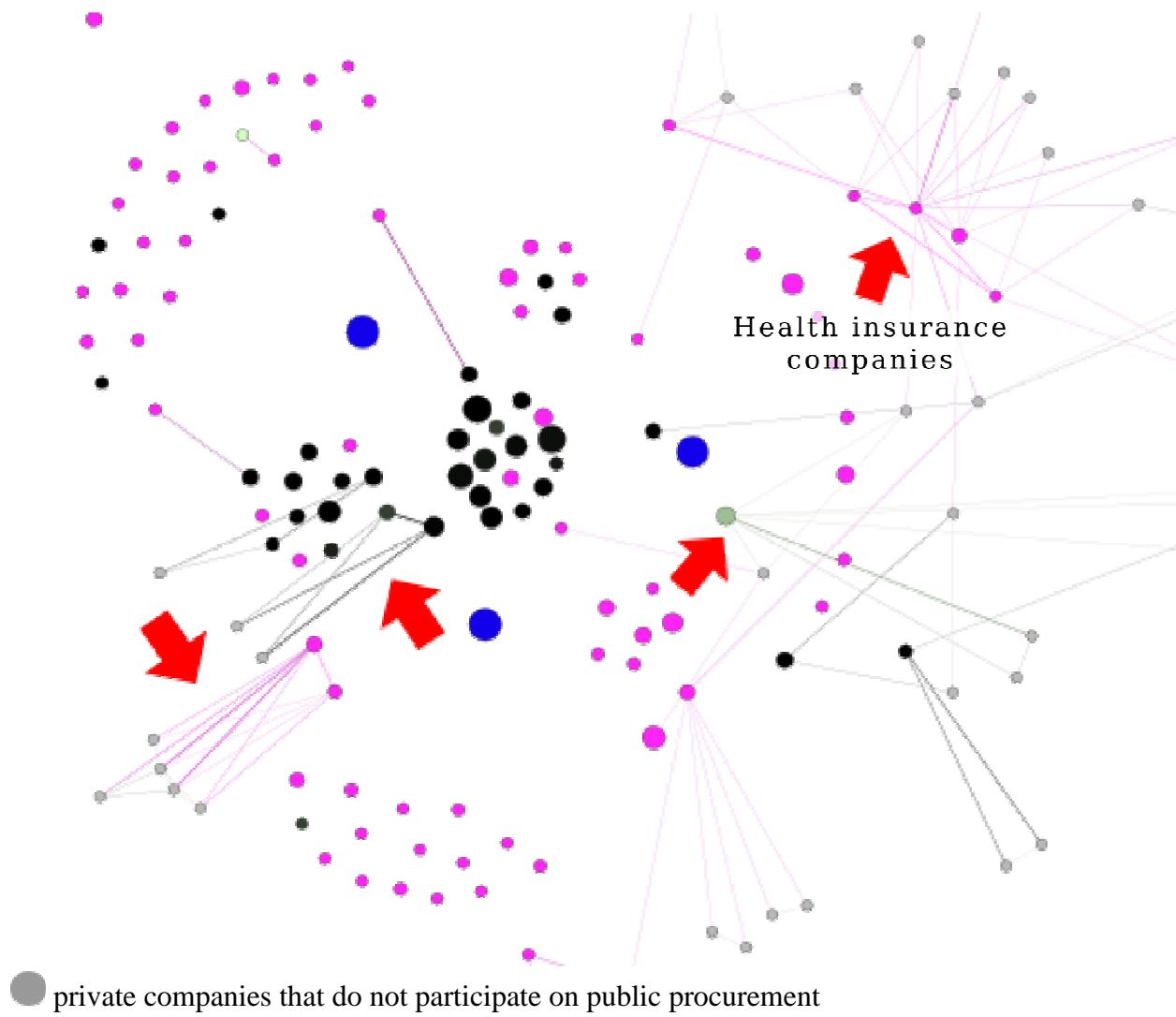


Figure 15. Relationships between the companies after hiding the contracts edges.

---

## Conclusion

In this experiment, we have examined whether we can use the social network graph to display the differences in the magnitude of the flow of contracts between governmental institutions and private companies. This has proven to be possible. The social network graph has revealed the most active companies in the network through the thickness of an edge pointed from the public institution to the company. We have then used weighted degree in order to point out individual companies that acquired the most contracts. Table 7 shows data of the top 10 most active companies in the network. The first column represents the amount of contracts awarded to the company (weighted in-degree). The third column represents the average number of companies that took part on the competitions for the tenders. It is notable, that 7 of the top 10 most active companies were often the only company that placed a bid on the tender, except for company 30 and two unknown competition. These 7 companies therefore demonstrate two red flags, namely low competitiveness in the bidding and repeated awards to the same company over a period of time.

**Table 7.** *Most active companies in the network*

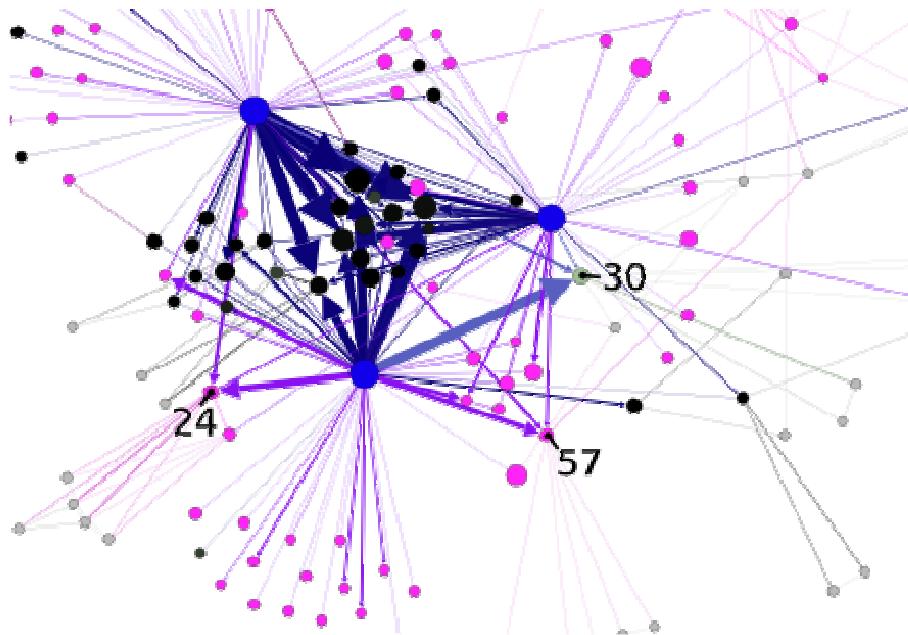
Nr.	ID	Nr. of contracts	Average Competition
1	75	126	1.25
2	01	117	1
3	29	113	1.2
4	91	95	1.07
5	30	59	4
6	58	52	1.35
7	24	48	unknown
8	02	45	1.4
9	57	38	unknown
10	80	29	1

In this experiment, we have also explored whether the social network graph can exposed interlocking directorates. This also proved to be possible. Table 8 represents the top 10 most interlocked companies within the network.

**Table 8.** Most interlocked companies in the network

Nr.	ID	Nr. of shared directors	Nr. of companies connected to
1	56	39	32
2	08	45	27
3	00	27	16
4	30	13	7
5	52	8	7
6	57	7	6
7	24	21	6
8	51	7	5
9	41	7	5
10	74	7	5

Three companies (ID's 24, 30 and 57) can be found in the top 10 of both experiments. Their location in the network is shown in Figure 16.



*Figure 16.* Position of companies ID 30, 57 and 24 in the social network graph

We could consider these three companies the most flagged ones, but unfortunately, we have no competition data on 24 neither 57. Moreover, company 30 has an average competition score of 4, the best one in the network. We can therefore not rise the red flag of low competitiveness on these three companies.

---

## 5.2 Individuals network

### Introduction

The purpose of this experiment is to examine whether we can expose conflict of interest situations and point out actors important for the information flow within the network. Through visual evaluation of the social network graph we will locate conflict of interest situations. Betweenness centrality measures will then point out how important a certain individual is for the information flow in the network. These measures and evaluation will provide us with answers to two of our research questions, namely:

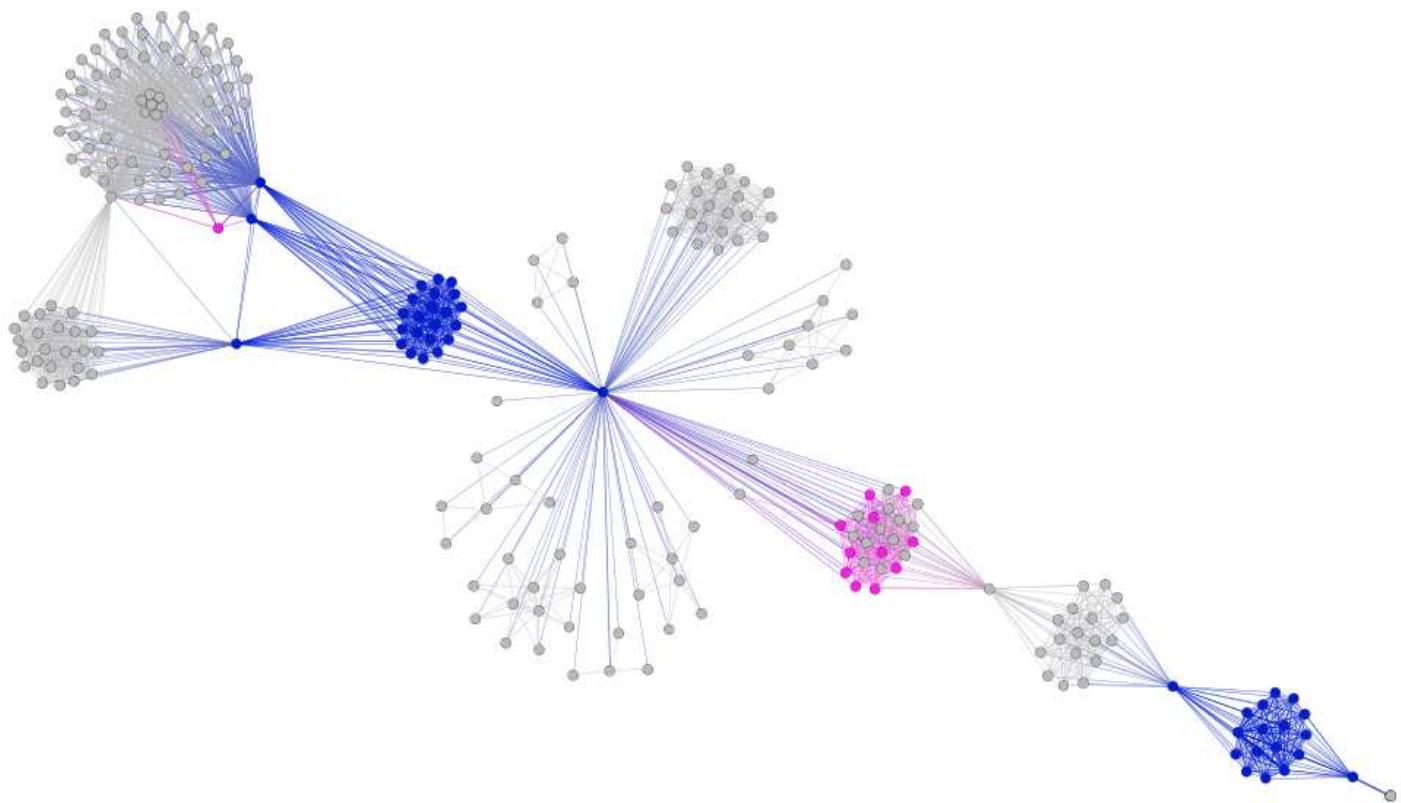
*Do measures of the position of an actor within a network point out individuals that are important to the information flow in the public procurement network?*

And

*Does the social network graph expose conflict of interest situations?*

### Results

Groups of nodes that are closely connected represent companies. This network consists of the network of only two of the hospitals (two groups of blue nodes) due to missing data. Due to the large amount of connections and individuals in these networks, Figure 17 is a combination of personal networks of each DPI. A personal network takes into account associated directors from each company that the DPI is or was active in. It is therefore possible, that the visualization of the whole network would uncover more connections.



- directors from procuring institutions (DPI)
- pink directors from private companies that *procure* for the institutions (DPC)
- grey directors from private companies that *do not procure* for the institutions (DPCN)

Figure 17. Individuals social network graph. Contains 241 nodes and 2148 edges.

The social network graph contains connections between the top management of the procuring institutions. Three DPI's were or are active in a company with a DPC. The social network graph has revealed, that two of them (DPI1, DPI2) have connections to an energy supply company that has contracts with the institution (Figure 18).

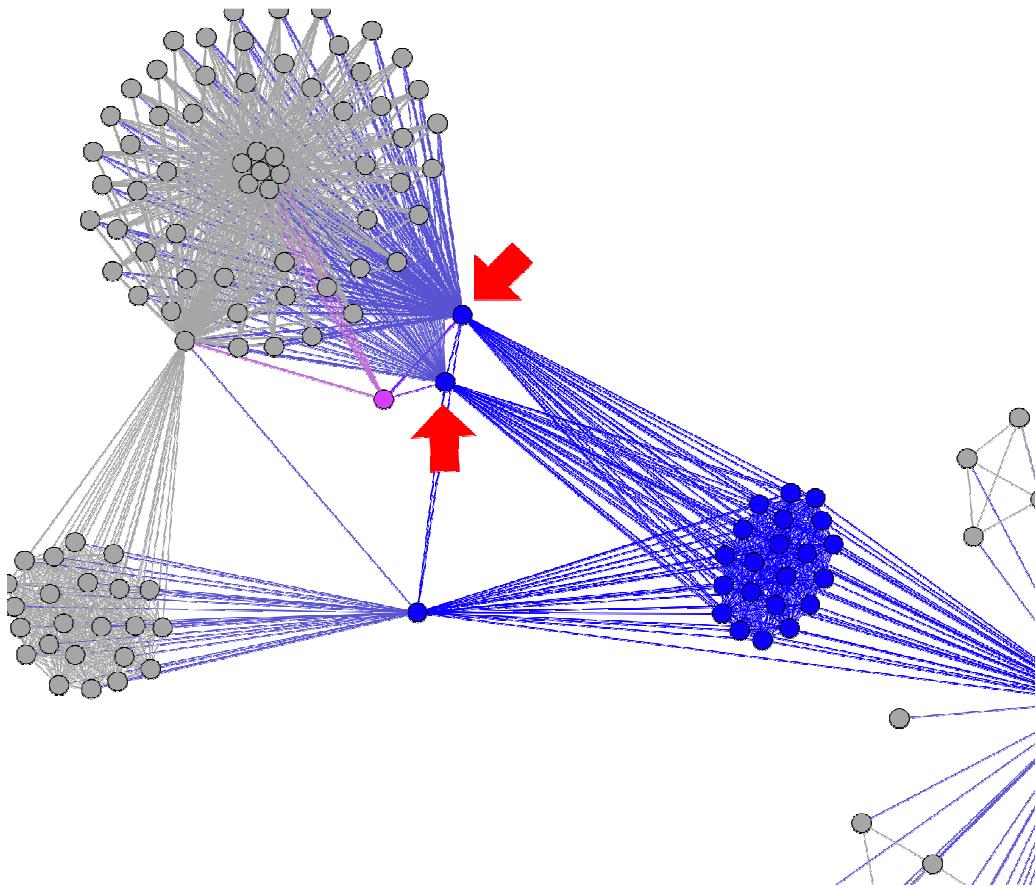


Figure 18. Position of DPI1 and DPI2 in the social network graph.

While these DPI1 and DPI2 were not active within these institutions in the same periods, the social network graph reveals a link between these individuals and the energy supply company. In the fourth case, the individual (DPI3) is active in a procuring institution at the same time as in a medical supply company that has contracts with this institution (Figure 19).

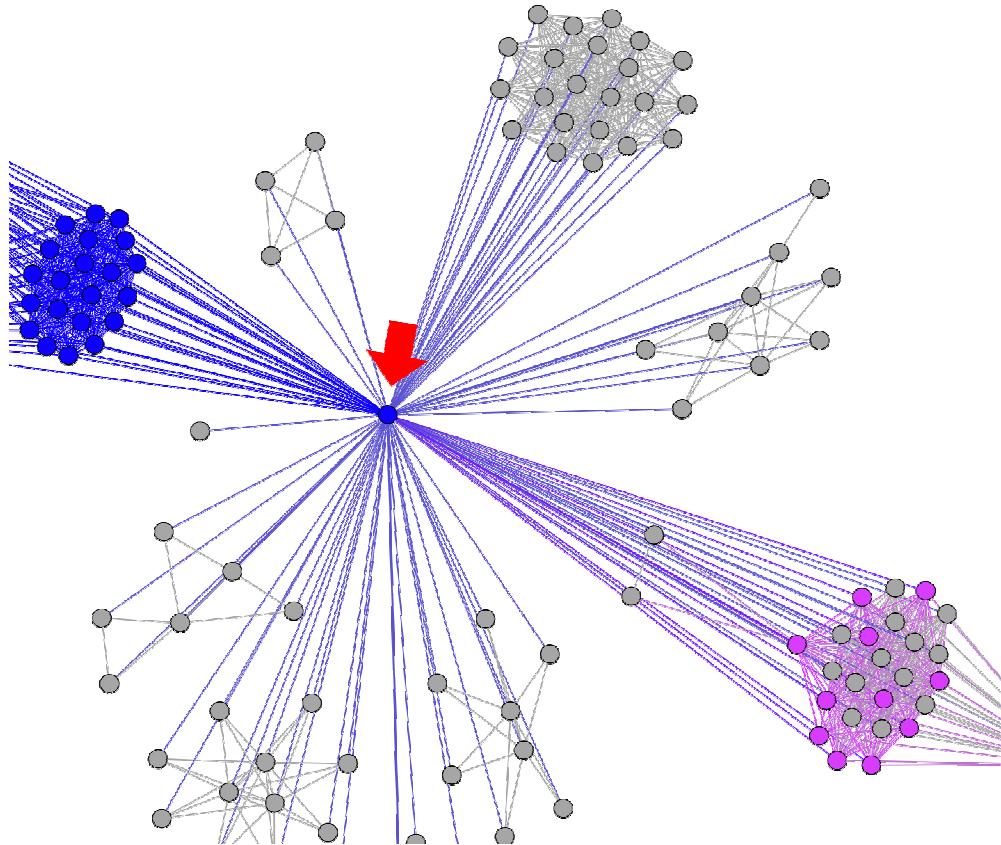
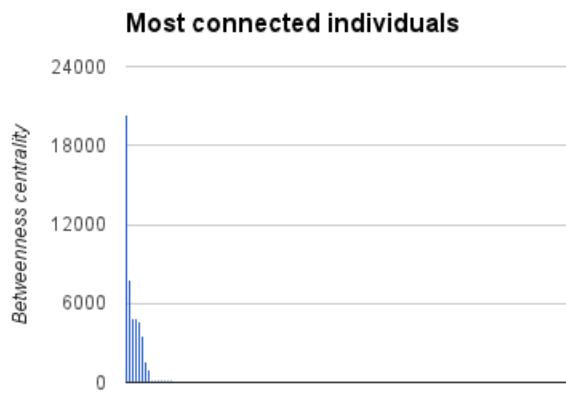


Figure 19. Position of DPI3 in the social network graph.

DPI3 is active in the institution since 2012. In 2012, the medical supply company acquired contracts in the value of about 641,500 EUR.

The centrality measures have pointed out a few individuals that are important for the information flow within the personal professional network of the DPI's. Betweenness centrality measures pointed out a small number of highly connected top managers within the procurement network sample (Figure 20). DPI3 is the most connected individual, largely due to the fact that he is or was active in numerous other private companies. DPI1 and DPI2 are also among the top 5 most connected individuals within this network. Table 9 gives an overview of the betweenness centrality measures of the top 5 most connected individuals and Figure 21 of their positions in the network.



*Figure 20.* Most connected individuals according to their betweenness centrality measures

**Table 9.** *Most connected individuals according to their betweenness centrality measures*

ID	Betweenness Centrality
DPI3	20,360.25
DPI4	7,842.65
DPI1	4,880.93
DPI2	4,880.93
DPI5	4,601.55

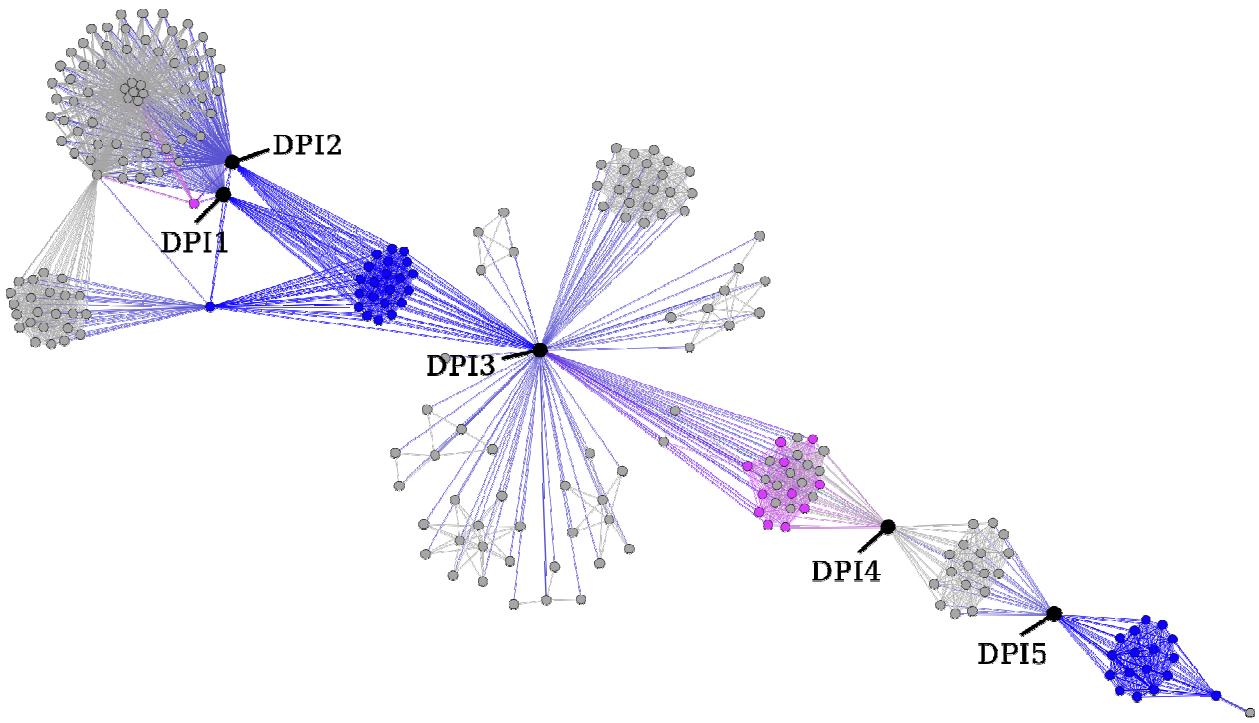


Figure 21. Positions of the top5 most connected individuals in the social network graph

## Conclusion

In this experiment, we have examined whether we can use the social network graph to expose conflict of interest situations as well as locate individuals that are important for the information flow within the network. Both has proven to be possible. The social network graph has exposed one conflict of interest situations. The revolving door of DPI3 raises a red flag. Betweenness centrality pointed out the most connected individuals within this network. We can assume that these individuals could serve as hub for communication between the interlocked companies and institutions.

---

## 6. Discussion

In our thesis, we have investigated, whether social network analysis is a method that can be applied to journalistic research on public procurement. We have argued, that journalists play an important role in supporting the principles of a democratic society. We then have explained the importance and relevance of utilizing new methods in journalistic research in the information society. We have introduced social network analysis, and we have explained some of the processes in public procurement. To answer our problem statement, we have formulated five research questions. We aimed to look for manifestations of the so-called red flags of corruption in the social network of public procurement, as well as for information flows within these networks. For our study, we have chosen the three biggest procuring medical institutions from Slovak public procurement market. We have visualized and analyzed the flow of contracts from 2005 to 2012 between these institutions and their suppliers. In addition, we have also visualized and analyzed personal professional networks of the top management of two of the medical institutions. Although we could locate red flags of corruption in these networks, we need to accent the fact, that these corruption indicators can also be raised by non corrupt behavior within the procurement market, such as the requirement of very specific goods and services. Medical procurement market is indeed one of these very specific markets, where for example repeatedly awarded contracts to the same company can indicate, that the company specializes in a marginal market. We are also unable to tell to what extent is the amount of contracts awarded to a certain company is considered suspicious. To answer this question, the investigation either needs to go deeper than social network analysis (e.g. interviews, field research, court cases). The research as well omitted the role clusters and clustering algorithms can play in these networks. These can also be an interesting addition to the ways we can look for corruption within procurement networks.

It would be very interesting to examine and compare procurement networks of different countries in order to disclose different patterns in the awarding of contracts. Due to limited competition on the tendering market, Slovak public procurement is also probably corrupt to a higher degree than other EU countries. A comparative case study between countries with low and high corruption perception index would therefore probably provide valuable insights in the

---

network structure of the global procurement market. We also recommend to do more statistical research that focuses on the trends in the different procurement markets (e.g. specialized versus not specialized). This however requires that the public sector gets more open than it is at the moment. Therefore, more open data and transparency in public spending needs to be demanded not only by journalists worldwide, but by researchers as well. Another interesting approach would be to visualize and analyze an already existing case of corrupt behavior within the public procurement network, for example by examining judicial files. In a well-documented case of corrupt behavior, the researcher could focus on particular relationships in the network and putting them in line with social network analysis theory. This would allow the researcher to use a more inductive approach to the research.

---

## 7. Conclusion and recommendations

Social network analysis proved to be useful in some aspect of the public procurement network analysis. Our first research question was: “*Do measures of the position of an actor within a network point out individuals that are important to the information flow in the public procurement network?*” Using the centrality measures, we could point out the most important individuals for the information flow within the network. These were individuals that created interlocked directorates between companies, as well as one revolving door situation. The revolving door situation gave answer to our second research question: “*Does the social network graph expose conflict of interest situations?*” We could localize a conflict of interest, when a director from a medical supply company was also a director in one of the medical institutions that was procuring goods for hundreds of thousands EUR from the same company. While conflict of interest is not a sign of corrupt behavior by itself, it holds a potential for abusing entrusted power. Our third research question: “*Does the social network graph expose interlocking directorates?*” also yielded a positive answer. The social network graph has revealed small networks of interlocked companies within the public procurement network. Even if interlocks are not a sign of collusion by themselves, they are an indicator of an intercompany relationship and can be relevant when in combination with other collusion indicators. We have also succeeded by visualizing the flow of contracts between governmental institutions and private companies in order to answer our fourth research question: “*Does the social network graph display differences in the magnitude of the flow of contracts between governmental institutions and private companies?*” The social network graph showed that some of the companies were acquiring many contracts from each of the institutions. We could identify and rank these companies through their degree measurements. The visualization has also revealed, that many of these companies were in most of the cases the only ones that took part in the bidding. We have found evidence for repeated awards to the same company over a period of time combined with low competitiveness in the bidding, which are both red flags. These red flags can be a sign of corruption as well as collusive practices. Unfortunately, we could not find any support for our fifth research question: “*Does visualization of the social network graph in time reveal possibly corruptive patterns in awarding contracts between governmental institutions and*

---

*private companies?"* The visualization in time did not reveal any patterns in awarding contracts in time and therefore also no bid rigging patterns (rotation of winning bidders). Hence, we could provide a positive answer to four from our five research question. We therefore believe, that social network analysis can be used in further research in other public procurement markets.

The added value of social network analysis methods is that it provides an unique overview of the relationships between the actors, where many other methods are focusing on the actors themselves. It also offers a range of metrics (e.g. degree measurements, centrality measurement and a lot more that are outside of the scope of this thesis) to investigate these relationships. The social network graph also provides the journalist with a quick overview of the various patterns and trends in the network ties without having to investigate them manually.

The method by itself is of course not sufficient in creating a story, but it does serve as a useful tool to find stories. Utilizing other, more journalistic methods as for example interviews and an in-depth research is necessary to acquire the whole picture. By revealing patterns in interaction, social network analysis of, in this case, public procurement can however provide a backbone on which a journalist can base his or her research. In our case study, we have focused on the flow of contracts. Nevertheless, networks are present in all social and institutional ties. Therefore, social network analysis can be applied in a much broader sense than analyzing public procurement. Our results have supported the notion that this method has its rightful place in the toolkit of a data journalist.

---

## Bibliography

Act on Public Procurement 2010 (Slovakia), Retrieved June 12, 2013 from:

[<http://www.scribd.com/doc/55518001/2010-Act-on-Public-Procurement-Slovakia>]

Aven, B. (2012), "The Effects of Corruption on Organizational Networks and Individual Behavior", Carnegie Mellon University, Pittsburgh, Retrieved May 30, 2012 from:

[<https://student-3k.tepper.cmu.edu/gsiadoc/WP/2012-E39.pdf>]

Begley, T.M et al. (2009), 'Network and cronyism: A social exchange analysis', Springer science

Beer, G., Slovak, K. (2003), 'Ako sa bačuje na onkológiu' [article], Retrieved May 23, 2013  
from: [<http://ekonomika.etrend.sk/ekonomika-slovensko/ako-sa-bacuje-na-onkologii.html>]

Boulding, K. (1966), 'The Economics of Knowledge and the Knowledge of Economics' in *The American Economic Review*, Vol. 56, No. 1/2., pp. 1-13

Borgatti, P. et al. (2009), 'Network analysis in the social sciences' in *Science 13*, Vol. 323, No. 5916 pp. 892-895

Breit, R. & Harrison, J. (2004), *If the Watchdog Bites, Is It Put Down?*

CRP (The Center for Responsive Politics), 2013, "Revolving Door: Methodology", Retreived  
June 14, 2013 from: [<http://www.opensecrets.org/revolving/methodology.php>]

Dančíková, Z. (2012a), "Ďakujeme, konkurenciu neberieme" [online article], Retreived June 6,  
2013 from: [<http://www.transparency.sk/sk/dakujeme-konkurenciu-neberieme/>]

Dančíková, Z. (2012b), "Ďakujeme, konkurenciu neberieme 2" [online article], Retreived June 6,  
2013 from: [<http://www.transparency.sk/sk/dakujeme-konkurenciu-neberieme-2/>]

Dahlgren, P. (1995), 'Mediating democracy' in *Television and the Public Sphere: Citizenship, Democracy and the Media*, SAGE, pp. 1-21

Freeman, L.C. (2011), 'The Development of Social Network Analysis - with an Emphasis on Recent Events' in *The SAGE Handbook of Social Network Analysis*, pp. 26-39

Garay et al. (2010), *Illicit Networks, Reconfiguring States* [online book], Metodo Foundation,  
available at: [<http://www.scribd.com/doc/118033325/Illicit-Networks-Reconfiguring-States>]

- 
- Genckaya, Ö. F. (2009), ‘Conflict of interest’, Retrieved August 5, 2013 from:  
[<http://www.coe.int/t/dghl/cooperation/economiccrime/corruption/projects/tyec/1062-TYEC%20Research%20-%20Conflict%20of%20Interest.pdf>]
- Gray, J. et al. (2011), *Data journalism handbook*, EJC & OKF, Retrieved May 30, 2013 from:  
[<http://datajournalismhandbook.org/>]
- Habermas, J. (1964), ‘The Public Sphere: An Encyclopedia Article’ in *New German Critique*, No. 3, pp. 49-55.
- Halinen, A. & Tornroos, J. A. (2003), “Using case methods in the study of contemporary business networks”, Journal of Business Research 58 pp. 1285 – 1297, Retreived May 30, 2012 from: [<http://www.sciencedirect.com/science/article/pii/S0148296304000499>]
- Johnson, J. A. & Reitzel, J. D. (2011), ‘Social network analysis in an operational environment: Defining the utility of a network approach for crime analysis using the Richmond City Police Department as a case study’, International Police Executive Symposium
- Kenny, C. & Musatova, M. (2010), ““Red flags of Corruption” in World Bank Projects: An Analysis of Infrastructure Contracts”, The World Bank, Retreived 13 June, 2013 from:  
[<http://elibrary.worldbank.org/docserver/download/5243.pdf?expires=1371205869&id=id&accname=guest&checksum=2F7B29C658A592A5E175371B3D0C08B9>]
- Khatri, N. et al. (2003), ‘Antecedents and Consequences of Cronyism in Organizations’ in: *Journal of Business Ethics* April 2003, Volume 43, Issue 4, pp 289-303
- Lauchs, M. A., et al. (2011) ‘Corrupt police networks : uncovering hidden relationship patterns, functions and roles’ in *Policing and Society*, 21(1), pp. 110-127, Retreived 3 June, 2013 from: [<http://eprints.qut.edu.au/41539/1/41539.pdf>]
- Lindskog, H. et al. (2010), ‘Corruption in public procurement and private sector purchasing’ in *Journal of Organisational and Social Change*, Vol. 7, No. 2, Intellect Ltd.
- Lippman, W. (1920), *Public Opinion*, Greenbook Publications, LLC
- Marineau, R. (1989), *Jacob Levy Moreno: 1889 - 1974 ; Father of Psychodrama, Sociometry, and Group*, Routledge
- McCombs, M. E. (2003), ‘The Agenda-Setting Role of the Mass Media in the Shaping of Public Opinion’

- 
- Meyer, P. (1991), *The New Precision Journalism*, Indiana University Press, available at:  
[<http://www.unc.edu/~pmeyer/book/>]
- Middleton, M. (2009), *Social responsibility in the media*, Oxford university
- Mizruchi, M. S. (1996), ‘What Do Interlocks Do? An Analysis, Critique, and Assessment of Research on Interlocking Directorates’ in *Annual Review of Sociology*, Vol. 22., pp. 271-298.
- OECD, (2007), ”Bribery in Public Procurement. Methods, Actors and counter-measures”, Retrieved June 3, 2013 from: [<http://www.oecd.org/investment/anti-bribery/anti-briberyconvention/44956834.pdf>]
- OECD, (2010), *Collusion and Corruption in Public Procurement*, Retrieved May 30, 2013 from: [<http://www.oecd.org/competition/cartels/46235399.pdf>]
- OKF, (2013), “What is Open Data?”, Retrieved May 30, 2013 from: [<http://okfn.org/opendata/>]
- Proulx, S. R. et al. (2005), ‘Network thinking in ecology and evolution’ in *Trends in Ecology and Evolution*, Vol. 20, No.6, pp. 346-353
- Rogers, S. (2012), ‘Anyone can do it. Data journalism is the new punk’, Retrieved April 10, 2013, from: [<http://www.guardian.co.uk/news/datablog/2012/may/24/data-journalism-punk>]
- Rocchini, C. (2012), *Centrality* [image], Retrieved May 23, 2013, from:  
[<https://en.wikipedia.org/wiki/File:Centrality.svg>]
- Strand, I. et al. (2011), ‘Public procurement in Europe’ [study], Retrieved May 26, 2013, from:  
[[http://ec.europa.eu/internal\\_market/publicprocurement/docs/modernising\\_rules/cost-effectiveness\\_en.pdf](http://ec.europa.eu/internal_market/publicprocurement/docs/modernising_rules/cost-effectiveness_en.pdf)]
- Stray, J. (2013), ‘Computational Journalism, Spring 2013, Lecture 5: Social Network Analysis’ [lecture], available at: [<http://www.youtube.com/watch?v=dq9-XIMQYfU>]
- Toffler, A. (1980), *The Third Wave*, Bantam Books, New York
- Transparency international (2013), ‘Engaging the private sector in the fight against corruption’, Retrieved May 23, 2013, from:  
[[http://www.transparency.org/whatwedo/activity/engaging\\_the\\_private\\_sector\\_in\\_the\\_fight\\_against\\_corruption](http://www.transparency.org/whatwedo/activity/engaging_the_private_sector_in_the_fight_against_corruption)]
- Transparency international (2012), “Corruption perceptions index 2012”, Retrieved June 5, 2013 from: [<http://cpi.transparency.org/cpi2012/results/>]

- 
- USAID (1999), *The Role of Media in Democracy: A Strategic Approach*, U.S. Agency for International Development, Washington
- Ware, T. G. (2006), “Addressing systematic and grand corruption at the local level: a toolkit for World Bank staff and Project Officials”, World Bank, Retrieved June 14, 2013 from: [\[http://www.cca.org.mx/funcionarios/toolkit/html/theme1/t1p1.htm\]](http://www.cca.org.mx/funcionarios/toolkit/html/theme1/t1p1.htm)
- Wasserman, S. & Faust, K. (1994), *Social network analysis: Methods and Applications*, Cambridge university press
- Webster, F. (2006), *Theories of the Information Society* (third edition), Routledge, Oxford, Retrieved April 25, from: [\[http://cryptome.org/2013/01/aaron-swartz/Information-Society-Theories.pdf\]](http://cryptome.org/2013/01/aaron-swartz/Information-Society-Theories.pdf)
- Wempe, J. (2002), ‘The worldwide fight against corruption: factors and actors.’ [lecture], Retrieved May 23, 2013, from: [\[http://bit.ly/10NReSw\]](http://bit.ly/10NReSw)
- Yang, W. L. et al. (2002), ‘AIMQ: a methodology for information quality assessment’ in *Information & Management*, Vol 40, Issue 2, pp. 133-146, Retrieved May 30, 2013, from: [\[http://www.sciencedirect.com/science/article/pii/S0378720602000435\]](http://www.sciencedirect.com/science/article/pii/S0378720602000435)

