Network Approaches to the Study of Corruption

Johannes Wachs

Supervisor: János Kertész

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Network Science

Central European University
Budapest, Hungary

2019

RESEARCHER DECLARATION

I Johannes Wachs certify that I am the author of the work Network Approaches to the Study of Corruption. I certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, the contributions of others. The thesis contains no materials accepted for any other degrees in any other institutions. The copyright of this work rests with its author. Quotation from it is permitted, provided that full acknowledgement is made. This work may not be reproduced without my prior written consent.

Statement of inclusion of joint work

I confirm that Chapter 3 is based on a paper which was written in collaboration with Taha Yasseri, Balázs Lengyel, and János Kertész. I conceived of the idea to relate social capital with local corruption risk outcomes by comparing data from procurement contracts and the online social network. Dr. Kertész and I conceived the details of the implementation. Dr. Lengyel and I collected data on municipalities. Dr. Yasseri and I developed the methods used. I carried out the analyses of the data. All authors contributed to the writing of the paper on which the chapter is based and gave final approval for publication. Dr. Kertész endorses this statement with his signature below.

I confirm that Chapter 4 reuses a plot from a working paper written in collaboration with Mihály Fazekas on the impact of corruption on public procurement market structure. The plot, which I drafted, serves as a prototypical example of the representation of procurement markets as networks. The remaining contents of the paper are independent of the contents of Chapter 4. Dr. Fazekas endorses this statement with his signature below.

I confirm that Chapter 5 is based on a working paper which was written in collaboration with János Kertész. I conceived of the idea to use a co-bidding network framework to study collusion. Dr. Kertész and I collaborated on developing and improving the methods used in the paper. I collected the datasets used and implemented the methods. Dr. Kertész and I and both contributed to the writing of the paper on which the chapter is based and gave final approval for submission. Dr. Kertész endorses this statement with his signature below.

Signature of PhD Candidate:

April, 2019

Signature of Dr. János Kertész, endorsing statement of joint work:

April, 2019

Signature of Dr. Mihály Fazekas, endorsing statement of joint work:

April, 2019

Abstract

Though corruption is a broad notion encompassing many kinds of behavior, it always has a relational aspect. Consider how a driver bribes a policeman, how a minister steers a contract to build a hospital to his son-in-law's construction company, how two managers from different firms agree to avoid competition in a region, or how a regulator goes easy on a potential future employer during an audit. The observation that interactions between people, firms, and institutions are where corruption happens is not a new one, but certainly merits further investigation. A better understanding of the relationship between the networks that these connections form and corruption can explain why corruption is so difficult to defeat.

This thesis applies the methods of network science to the study of corruption and its relationship with markets and society. I argue that corruption emerges from specific patterns of interactions that can productively be described using networks. The dyads of actors engaging in a corrupt behavior, the driver and policeman, minister and son-in-law, etc., are embedded in networks of social relations that facilitate corruption. Within this framework, the thesis addresses several questions about corruption. Why does corruption persist in certain communities? How does corruption relate to the organization of markets? How does corruption emerge when it depends on cooperation in highly adverse circumstances? I address these questions empirically using newly available micro-level data on corruption risks in public procurement.

Starting with a study of Hungarian towns, I relate corruption risk in local government contracts to the structure of their social networks. I find that fragmented towns have higher corruption risk, while towns with residents that have diverse connections have less. This suggests that corruption is embedded in the social networks of places. Next I zoom out to the national level, comparing the procurement markets, conceptualized as networks of issuers and winners, of different EU countries. I find a strong relationship between centralization and corruption risk. On the other hand, heterogeneity in market responses to changes in government across the EU suggests that corruption can be organized in many different ways. Finally, I investigate cartels, or groups of

firms that illegally agree to avoid competition. By drawing networks of firms that bid for the same contracts I highlight niches in markets where cartels are more likely to thrive.

ACKNOWLEDGEMENTS

I would like to acknowledge several individuals who have, whether they realize it or not, significantly influenced this work. First I must thank my advisor János Kertész for his continued support and confidence. His guidance has been invaluable and his patience is much appreciated.

I thank my most immediate colleagues David Deritei, Orsi Vásárhelyi, and Tamer Khraisha for their commiseration and support throughout the PhD. They and the rest of the Department of Network and Data Science at CEU, including Olga Peredi, the other PhD students, and the various postdocs and memorable visitors, provided a wonderful environment to learn and work through turbulent social and political times.

I am also grateful to my colleagues at the Government Transparency Institute, Misi Fazekas and Ági Czibik, my host at the Oxford Internet Institute, Taha Yasseri, and friends and collaborators including Bálint Daróczy, Ágnes Horvát, Dorottya Szalay, Anna May, Casey Tompkins, Theresa Gessler, Gergő Tóth, and Balázs Lengyel. I thank Ancsa Hannák for suggesting that a PhD can be a worthwhile thing to do, and then encouraging me during its most difficult moments.

I am lucky to have two families supporting me. I thank both the Czémán and Wachs clans for their unconditional love and support. My mother Kristin and father Karl know that they have given me endless encouragement and I hope they know how thankful I am for it. What they may not know is how important the example they set has been to me. I hope this work reflects that fact.

Most importantly I want to thank Zsófi. She alone knows what went into this thesis, and her tireless encouragement always kept me on the tracks. She is my inspiration and partner in all things. I dedicate this thesis to her.

Contents

List of Tables

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Corruption is a major cause of suffering around the world. It slows economic growth [1], stifles innovation [2], and increases inequality [3]. The social impacts of pervasive corruption in a society, reflected for example in distrust of government and strangers [4] and in the low quality in government services [5], suggest that corruption reinforces itself in a kind of feedback loop. Feeling that society is rigged, individuals adapt to their circumstances and play by the local rules. Like many other social scientific problems, understanding the emergence or persistence of corruption is difficult because it manifests both in individuals and societies.

This dichotomy is one reason why there are multiple definitions of corruption. The World Bank and Transparency International define corruption as "the abuse of public or corporate office for private gain" and "the abuse of entrusted power for private gain," respectively [6, 7, 8]. This definition provides a good benchmark to use when deciding whether an individual's behavior should be considered corrupt. Recently, scholars of government and institutions have focused on a more macro-oriented definition of good governance, namely: "the degree to which the exercise of public authority follows the principle of universalism or impartiality" [9]. The presence of partiality or particularism in the decision making of public actors is a useful definition of corruption because it defines a norm of behavior that can be applied to various contexts [10].

These two perspectives have had some success in quantifying causes and effects of corruption, building measures of the prevalence of corruption in countries and regions, and proposing policy interventions to improve the control of corruption in society. They share several important core concepts, for example both reference corruption as a kind of behavior existing at the intersection of the public and private domains. Both refer to the exploitation of power or advantage, presumably to the detriment of a wider group of people.

The definition of corruption based on the lack of impartiality has become increasingly relevant as researchers of corruption focus on grand corruption, distinct from petty corruption. Corruption is deemed petty when it is impersonal and transactional - for example when someone bribes a driving instructor to pass an exam or a policeman to avoid a speeding ticket. Grand corruption refers to coordinated, organized behavior to siphon resources to specific groups or networks of people [11, 12]. When examples of grand corruption are discovered they are typically front page headlines, as for example in the case of Petrobras, Brazil's state-owned petroleum company, whose executives were found to have taken nearly $10 billion of bribes and kickbacks from a group of 16 large construction firms in exchange for awarding them overpriced contracts [13, 14]. Over 160 people were arrested and 93 convicted, including the former President of Brazil.

Such a large conspiracy with such significant payoffs to its participants is only possible as a collective effort of many people. Despite having a larger target to aim at, it is difficult for authorities to combat grand corruption. This is because grand corruption is often organized in a sophisticated way, with its actors having specific responsibilities and roles, and connections organized in a way to limit their vulnerability as a whole [15]. Organized crime groups [16] and the September 11th hijackers [17] structured themselves in a similar manner. Grand corruption is a difficult topic for academic research for these same reasons and also because it is unlikely that large conspiracies can be understood in terms of a sequence of "abuses of public or corporate office for private gain". Somehow such conspiracies are much more complex social outcomes than the sum of the individual behaviors of their members.

Throughout this thesis when we speak about corruption, we will be referring to grand corruption and its organization, rather than instances of petty corruption. Though the prevalence of bribery and its acceptance in society are likely correlated with the prevalence of grand corruption, it is not essential to its functioning.

While framing corruption in terms of the norms and rules of society may be more applicable to the study of grand corruption than the transactional view of corruption focusing on bribery, it suggests an over-socialized description of corruption. In an extreme interpretation, the environment determines the actions of individuals, who merely "internalize norms and seek [to conform] to the expectations of others" [18] and so participate in corruption. Such a perspective not only fails to give a satisfying answer to the question of why certain places have developed a good control of corruption while others have not, but is also empirically unsupported. We will see that there is significant variation in the prevalence of corruption as we measure it within countries, even at the level of towns. It also has limited ability to explain how the prevalence of corruption in a place might change over time. Indeed, the Petrobras scandal offers some hope for Brazil: in the end powerful people were imprisoned for their corruption.

As they focus on individuals and societies, respectively, both definitions avoid mentioning that corruption, like nearly all socioeconomic activity, happens between actors. In this thesis we consider corruption as a networked phenomenon [19]. It is a property of the interactions between groups of actors such as people, firms, institutions, which evolves as these interactions change. In his 1985 paper on the embeddedness of economic action in social structure, Mark Granovetter suggested that "force and fraud are most efficiently pursued by teams, and the structure of these teams requires a level of internal trust–"honor among thieves"– that usually follows preexisting lines of relationship." In other words, Granovetter is suggesting that particular social structures or networks are required to scale corruption to high levels. We do not discard the productive micro-level definitions of corruption by the World Bank or Mungiu-Pippidi, but we do change the way they are applied.

We will demonstrate that this approach can describe in novel ways the organization and roots of corruption at various scales from towns to nations. It complements micro and macro-level frames of corruption by considering what happens in between.

Theory aside, why might networks be a useful in the study of corruption, for example in its measurement, detection, or diagnosis? Grand corruption, as its name suggests, requires organization and coordination. Such organization may manifest explicitly as a social network of specific actors, say members of parliament and heads of firms. It should also leave fingerprints in the interactions between firms and institutions doing business and exchanging money. Crucially, several recent developments make it possible to examine empirically the relationship between corruption and the networks of actors. The growth of the internet and its use by governments has created a wealth of fine-grained administrative data on interactions between the public and private sectors.

One distinguished example of such data is information on public contracting or procurement markets, which accounts for upwards of 20% of GDP in OECD countries [20]. Recently, researchers have developed ways to measure corruption risk in the award of such contracts [21, 22]. This data and measurement approach provides us with the micro-level data to undertake a network-based analysis of the phenomenon of corruption. We argue that this data, at present, offers the best possible approach to an evidence-based, data-driven anti-corruption research program as proposed by Mungiu-Pippidi in 2017 [23].

Nearly in parallel, researchers studying biological [24], ecological [25], social [26], economic [27, 28], and spatial [29] phenomena have made fundamental advances in their fields by using a network perspective. A by-product of these minor revolutions in different physical and social sciences is the emergence of set of tried and tested methods for studying networks within data. These efforts extend the pioneering work of sociologists and anthropologists [30, 31, 32], who have been working on social networks since the 1940s, to new contexts and larger scales[1]. This recent work on networks forms an emerging field of research called network science, which has created a rich tool set for the study of complex systems.

This thesis proposes to leverage these developments to extend the state of the art in corruption research. Specifically, we claim that corruption risk can effectively be measured at the micro-level using administrative public contracting data, and that by using network science methods we can describe the organization or structure of corruption. We do this in three contexts: relating social networks and corruption outcomes of towns, describing the relationship between corruption risk and market structures at the scale of countries, and analyzing the emergence of illegal cartel behavior among firms. In all three chapters we provide both a theoretical contribution to the understanding of corruption and policy implications.

First we describe how the structure of social networks are related with the prevalence of corruption in local government contracting using data from Hungarian settlements. By linking social structure with corruption we strengthen our claim that corruption is in general a networked phenomenon.

Next we zoom out to the level European countries, mapping their procurement markets as bipartite networks of public institutions and the firms they contract with. These networks have rich structure related to their level of corruption risk. By observing how corruption risk is distributed in these networks and how actors with high risk scores respond to political shocks, we highlight novel distinctions between the organization of corruption in different countries. Among EU countries we find significant heterogeneities, for example that in some countries corruption risk is concentrated in the core of the procurement markets, while in others it is more common in the periphery.

Finally, we shift focus to cartels in procurement markets, transferring the perspective we have developed to study corruption to a different but related economic problem. Cartels are groups of firms which illegally avoid competition to maximize their profits. We propose a framework to map markets of competing firms using contracting data, to identify groups of frequently inter-

[1]For an excellent history of research on social networks we refer the reader to the review of Freeman [33]. For an overview of the current relationships between the different traditions of network analysis, we refer the reader to Hidalgo's review [34].

acting firms, and to measure their potential for forming cartels.

The remainder of the thesis is structured as follows:

- Chapter 2: Here we review past work on the conceptualization and measurement of corruption, introducing measures based on both perceptions and administrative data and comparing them. We also review the literature on network aspects of corruption.

- Chapter 3: In this chapter we relate the social networks of Hungarian towns using data from an online social media portal with the amount of corruption risk in their local governments.

- Chapter 4: In this chapter we quantify corruption at the national level using data on public contracts awarded by member states of the European Union. By mapping procurement markets as networks, we can examine the distribution of corruption risk in different countries, and observe how they react to political shocks.

- Chapter 5: In this chapter we apply network methods to the problem of cartels, transferring the principles developed in earlier chapters to a domain adjacent to corruption. We consider the emergence of illegal cooperation among firms in procurement markets using records of their bidding behavior. We observe a "hot spots" in the network topology where collusion appears much more sustainable.

We conclude by tying our findings together, presenting some of their diagnostic interpretations, and suggesting future avenues of research.

CHAPTER 2

Related Work

In this chapter we discuss relevant previous work on corruption, including its measurement and its relational aspects, in order to orient the thesis. Following a brief survey of classical studies of corruption, we review how corruption is studied using experiments and models. These results give us insights into how corruption functions at smaller scales and suggests potential mechanisms to investigate. Next comes a review of how corruption is measured, considering surveys, and indicators derived from administrative data. We then survey findings on the causes and consequences of corruption, noting how corruption is measured in each result. With these results in mind, we argue that measuring corruption using indicators derived from administrative data is the most promising way to study the organization and structure of corruption because of its granularity. Correlations between such indicators and other measures of corruption demonstrate the validity of our chosen approach. Finally, we review applications of network methods to the study of corruption and crime more generally.

As suggested in the introduction of this thesis, corruption is a long-studied topic of interest to many branches of the social sciences. The result is that corruption has been studied from a variety of viewpoints, reflecting broad trends in how different fields have productively conceptualized human behavior at various times [35]. Before reviewing the most recent theoretical and empirical approaches to the study of corruption, we briefly highlight two such viewpoints, framing them as micro- and macro- oriented perspectives.

A major *micro*-oriented perspective on the study of corruption is the principal-agent framework [36]. In this framework, the principal represents an actor in charge of monitoring agents, seeking to block or limit corruption, referring sometimes to a high-level policymaker or to the public as a whole. The agents are individual bureaucrats, citizens, or politicians who weigh incentives to engage in corruption or to follow the rules. In a seminal work on this framework Klitgaard [**?**] describes corruption with a formula at the agent level: ”corruption equals monopoly plus discretion minus accountability”. Schleifer and Vishny [37] model the decision of bureaucrats to take bribes as a cost-benefit analysis of rational agents.

The results of such models offer immediate policy recommendations, for instance to increase accountability when an agent occupies a monopoly position. More generally speaking, this approach to corruption emphasizes necessary conditions for the control of corruption, for instance the existence of constraints in policy-making. It falls short in describing sufficient conditions for effective control of corruption. The principal-agent framework suffers from several flaws, such as the assumption that a clean principal exists or can be created (for example by establishing an anti-corruption agency), or that marginal changes to incentives can change a thoroughly corrupt equilibrium [38]. Mungiu-Pippidi and Dadasov suggest that a principal-agent framework is only useful when “corruption is an exception and the broader norm is ethical universalism.” [39] In other words, the principal-agent framework is an under-socialized approach to corruption, neglecting the norms and context that corrupt actions are embedded in.

Researchers adopting a *macro*-orientation to the study of corruption often seek to compare and explain the differences in outcomes of countries using structural or institutional factors. Structural factors include the level of development or education, or the legacy of a country’s history (for instance as a colony) [40]. Work by Treisman, for instance, shows a significant positive relationship between how long as place has been democratic and its control of corruption [41]. In general wealthier countries with an education citizenry have less corruption [42]. Institutional factors refer rather to the current legal organization of a place, for instance its constitution, and its politics. There is for example some evidence that political competition is an important ingredient to effective control of corruption [43]. Given, however, the heterogeneity of corruption outcomes at sub-national levels [44], there is a kind of natural resolution limit to entirely macro-based approaches to the study of corruption. We argue that such an approach suffers from an over-socialized perspective which struggles to explain how countries actually overcome corruption.

In the rest of this chapter we present previous experiments, models, and empirical studies that further demonstrate the potentially networked nature of corruption in a variety of contexts and at different scales. Again we emphasize that we do not discard the perspectives we describe above, but rather highlight how they can be enhanced by an alternative perspective.

2.1 Experiments

A major limitation of the social sciences has been the inherent difficulty of carrying out experiments to test hypotheses about the social world. In the natural sciences, experiments are a crucial ingredient of many research projects. Often, social scientists can only use observational data to test their ideas in the context of the social world. An alternative to using observational data is to test hypotheses about the social world in artificial or specific contexts. The former refers to laboratory experiments, in which participants are put into situations emulating the real world and their reactions to stimuli are observed [45]. Using computers and remote participation, the potential scale of such experiments has vastly increased, sometimes involving hundreds of simultaneous participants [46]. The latter notion refers to randomized control trials, in which a sample of a population (of individuals, towns, regions) is split into treatment and control groups, an intervention is applied to the treatment group, and outcomes are tracked [47]. Such methods offer convincing evidence of the effect of intervention, at great cost. Randomized control trials are expensive to implement and carry on over time. Both kinds of experiments have been applied to the study of corruption [48].

Lab experiments have been used to test the hypothesis that corruption is a cultural phenomenon by having participants from different cultures play games in which there is economic incentive to cheat or behave in a corrupt manner. Cameron et al. [49] find high cross-cultural variation (comparing subjects from Australia, India, Indonesia, and Singapore) in the propensity to punish corruption, and less variation in actually engaging in corruption. This finding suggests one reason why social networks may have an important role to play in corruption: actors across cultures may not be more or less willing to engage in corruption, but rather face different consequences depending on their alters. Other experiments show that when individuals are given opportunities to take corrupt actions without consequences in one round of a game, they are more likely to be dishonest when they are unsure about the consequences. This suggests that experience with corruption increases willingness to engage in corruption, again suggesting a role for networks in the spread of corruption [50, 51].

A third example of the study of corruption in a laboratory environment dramatically underscores the collaborative nature of corruption [52]. Weisel and Shalvi have pairs of subjects play a six-sided dice rolling game. The first player rolls a die, observes its outcome privately, and reports it to his partner. The partner then rolls his die, also observing the outcome in private. The players are paid proportionally to the value of their roll if and only if both players *report* having rolled the same number. In the results, both the frequency of matches and the frequency of high numbers (fives and sixes) reported by the players is many times greater than what would be expected if both players were honest. The amount of dishonest reporting is also significantly higher in this two player game than in a similar one player game. These results suggest that collaboration can facilitate dishonest behavior, perhaps via the diffusion of responsibility. On the other hand, the classic study of Kahneman, Knetsch and Thaler [53] suggests that individuals will reject unfair distributions at cost to themselves, so corrupt collaborations likely require a careful distribution of resources to be effective.

Despite their artificial nature, these experiments provide insights into potential underlying mechanisms of corruption. The studies we have highlighted suggest the importance of other people in an actor's choice to be corrupt. Others can be tolerant of corruption, they can provide examples, and they can even be accomplices.

There are several notable large-scale field experiments relevant to our work. Most prominent is the work of Olken in Indonesian villages [54]. Olken designed a series of interventions applied in a randomized way to over 600 villages that were about to start building roads using funds from a nationwide infrastructure project. A random sample of villages were informed that their implementation of the project would be audited by the central government, and that the results would be read publicly in an open forum in the village. This random intervention tests the effectiveness of top-down checks on corruption. Olken was also interested in the potential for bottom-up methods to combat corruption, so in another subset of the villages, he organized public accountability meetings with the project officials. Within this subset Olken ran another experiment: allowing residents of some villages to relay anonymous information about the project which would be read aloud at the meetings. These two interventions test the ability of grassroots organization to fight corruption.

In order to measure corruption in the delivery of the road construction projects, Olken hired a team of engineers and surveyors who generated independent estimates of the costs of projects. Comparing these estimates on a line-item basis with the observed outcomes, Olken found that the roads built cost 27% more on average than what the engineers estimated. Top-down auditing decreases the discrepancy between between cost and the independent estimates by nearly a third to 19%. The bottom-up grassroots organizing had no significant effect on corruption. Following up with a household survey, Olken found that family members of local government officials were significantly more likely to have been employed on the road project in audited towns. This suggests that the guarantee of a top-down audit of expenditures had a substitution effect - instead of reporting higher prices and pocketing the difference, corrupt leaders hired their family members. Contrasted with the insignificance of the grassroots intervention, this final observation highlights the importance of distinguishing between different kinds of social connections when studying corruption.

Bertrand et al. provide a second example of the use of randomized control trials [55] to study corruption. Participants in the study, run in India, are tasked with obtaining a driver's license. A third of the participants are offered a bonus if they obtain the license quickly, the second third are given free driving lessons, and a remaining third serve as a control. While members of both treated groups are more likely to obtain a license than members of the control group, members of the so-called bonus group accomplish this by paying bribes to third-party agents. These agents are a constant presence in Indian bureaucracy - nominally they are hired to stand in line on behalf of clients.

Many of members of the bonus group obtain a license without taking an exam, and are found to have significantly worse driving skills in a surprise exam - a good example of an externality of corruption. The agents arranging the corrupt transfer play an interesting role: they insure that license seekers do not interact directly with bureaucrats in the extra-legal process. They are by definition brokers - the crucial network connections that facilitate corruption in this environment.

In some rare cases, governments carry out randomized audits of public works and records. The most notorious example comes from Brazil, where a 2003 federal government program introduced lottery-based audits of municipalities [56, 57]. Researchers have used this data to quantify corruption at the municipal level in Brazil and to measure how corruption impacts incumbent electoral performance, and what impact local media have. A recent study shows that the electoral effects of revealed corruption spill over into neighboring towns [58]. Voters in towns neighboring a corruption scandal will punish politicians of the same party as the neighboring mayor in their own election.

Well-designed randomized control experiments can measure interesting causal relationships between variables relevant to corruption. However they do have some significant limitations. They are expensive and are difficult to scale. By definition they cannot compare effects in different contexts or environments unless the treatment itself is the difference. Randomized audits of public works seem to be cost-effective, as demonstrated by Olken's work in Indonesian villages and in the case of Brazilian municipalities, but we do not know an example of a central government body instituting random audits of its own actions at a large scale (consider for instance that corruption in Petrobras was going on at the same time as the introduction of the randomize audits of municipalities in Brazil).

2.2 Models

Purely abstract models of social systems can provide useful ideas to researchers of social phenomena [59]. In this section we describe some previous work on corruption which applies abstract models or agent-based simulations. Economic Nobel Laureate Jean Tirole created a model of the persistence of collective reputation to suggest that once bad behavior becomes a structural phenomenon, it is difficult to get rid of [60], echoing the experimental work cited above. Lambsdorff [61] points out that the informal mechanisms used to enforce corrupt agreements and notion of mutually-assured-destruction lock partners in corrupt deals together. Ferrali [15] models the spread of corruption as a game played on a network, finding that modular networks are eventually more corrupt than more mixed ones. These models and others suggest that it is valuable to consider the durability of corrupt partnerships and the importance of social connections in the evolution or spread of corruption. In other words, these models suggest that corrupt actors stick together.

Agent-based models are also useful tools to understand the emergence of macro-patterns from micro-behavior [62]. One under-explored aspect of corruption is how a corrupt society in which corruption is the rule might transition to one where it is rather the exception. Indeed, this has happened several times in history. It seems unlikely that the level of corruption in a society simply decreases in a linear manner. An agent-based model of citizens and bureaucrats by Hammond [63] demonstrates how phase-transitions between highly corrupt and very clean states might occur. In Hammond's model, randomly interacting citizens and bureaucrats play games in which mutually corrupt behavior is rewarded, while asymmetric corrupt behavior risks punishment and loss. Likely such models can be enriched by incorporating a more realistic network-based structure of interactions. In any case, Hammond's findings again suggest that it is worth thinking about corruption as a complex phenomenon.

2.3 Measuring Corruption

The most widely known measures of corruption are taken at the national level. They are used in the evaluation and comparisons of countries, and are important for several reasons. Within countries they can drive politics by shaming governments who fall in the global rankings. Internationally they spur competition between governments, not only for reasons of pride but also because investors use these rankings to decide where to put their capital. They also play an important role in driving awareness of corruption as a problem in society. Most national measures of corruption are based on survey data and rely on perceptions of corruption. In this section we first present several such measures which are commonly used both in research and policy to study corruption. We highlight some of the research findings on the causes and effects of corruption based on these measures. We then highlight some of their shortcomings and suggest an alternative based on administrative data which has become more popular in recent years. We then compare the indicators using European data, previewing our analysis in Chapter 4.

2.3.1 Survey and Perception-based Measures

The two measures of corruption perception that have been around the longest are Transparency International's (TI) Corruption Perceptions Index (CPI) [64, 65], available since 1995, and the World Bank's (WB) Worldwide Governance Indicators (WGI), specifically its Control of Corruption component (CoC) [66, 67], available since 1996. Both are composite indicators, drawing on a variety of sources. Both mix surveys of representative samples of the population in a country with targeted surveys of expats, firm managers, and NGOs. Both use expert testimony measure to assess different dimensions of corruption. While the CPI the average of indicators, the WGI applies a method known as an "unobserved component model" which decreases the weighting of outlier scores. The two measures are highly correlated (above .9).

As not all data sources are available for every country, and because data sources can change year to year, there are substantial problems with comparing the results of the CPI and WGI from year to year (though this is more of a concern for the CPI) [68]. Ironically, Heywood and Rose note that there is a distinct *lack* of variance in countries scores over time for both the CPI and WGI. For both measures, an ordinary-least-squares model predicting 2011 scores using only 2000 scores explains over 89% of the variance in the 2011 CPI and 86% of the variance in the 2011 WGI. Given the innate measurement errors of perception-based indicators, the fit is almost too good. Indeed Hawken and Munck, investigating the CPI between 1995 to 2009 find that a significant amount of variation within country scores over time comes from choice of experts and evaluators who decide about the inclusion or exclusion of sources [69]. Adjustment made since may have improved the situation at the margin, but aggregated perception-based indicators still have to contend with problems of sampling bias, the difficulty of measuring errors, and the issue that aggregation increases the distance between measurement and solutions [70].

One recent innovation addressing some of issues with both the WGI and CPI is the Bayesian Corruption Index (BCI) [71]. The BCI begins with the WGI and uses Bayesian methods to quantify how the error at the level of individual components aggregates into the composite measure. It also considers, unlike the

WGI, that individual components have correlation between the years. By estimating these correlations, the BCI can describe shifts of corruption perception from year to year more precisely than the original WGI. The BCI is highly correlated with the WGI (.95), but offers a way to quantify error, and is significantly less correlated over time within countries (.35). Though the BCI does not solve the issues of correlations in the errors of components of the WGI, it offers an interesting alternative measure, especially for describing changes in corruption over time.

As alternatives to these methods we also consider the Varieties of Democracy (V-DEM) indicator of political corruption, which is a composite index built entirely from indicators coded by country experts [72], and the Quality of Government Institute's (QoG) European Quality of Government Index (EQI), based entirely on surveys conducted in the subnational regions of Europe [73].

Despite these attempts to address specific issues with perception-based measures of corruption, there is evidence that such measures will always suffer from certain innate flaws. Olken followed up his study of corruption in Indonesian villages discussed above with surveys of the villagers [74]. He found that villagers could perceive significantly higher levels of corruption when the road project in their village had higher missing expenditures. However, the strength of this relationship was weak: a 10% increase in missing expenditures increases the probability a villager believes that the project was corrupt by only 0.8%. More importantly, Olken finds significant biases in perceptions. In villages with higher ethnic heterogeneity (often suggested as an important correlate of corruption [1]), perceptions of corruption were significantly higher while actual missing expenditures were lower. Social cohesion, measured using participation in social activities, is related to lower perceptions of corruption but higher missing expenditures. We will revisit this latter relationship in depth in Chapter 3. These biases highlight significant issues with using local perceptions to measure corruption [75]. We next present some alternatives.

2.3.2 Administrative Data-based Measures

We now highlight several approaches to measuring corruption that can complement the use of perception-based indicators. Generally speaking, the approaches we describe are based on the observations of outcomes, for instance the construction of public roads, and their comparison to a benchmark. Unexplained shortfalls in outcomes or deviations from the ideal benchmark are considered to be the residue of corruption. This general framework is increasingly relevant and applicable given the explosion of data available on the activities of public institutions and the widespread adoption of information and communications technologies (ICT) in the public sector [76]. Bureaucracies around the world are incorporating transparency and opening their procedures to the public [77].

One strand of research uses data on legal proceedings to measure corruption. For example, Glaeser and Saks [78] compare US states using data on federal corruption convictions, finding that increases in levels of education are related to decreases in corruption. This approach, like the Brazilian audit studies described above, relies on the presence of an independent source of data on corruption, in this case the federal government prosecuting corruptions in the different US states.

Golden and Picci measure corruption by tracking the difference between existing public infrastructure and money spent on that infrastructure at the level of Italian regions using accounting principles [79]. They find, to take just one example, that large cities in southern Italy have significantly higher construction costs for public works than their counterparts in northern Italy. In the private sector they observe the opposite effect. This is some sense a macro version of Olken's approach to measuring corruption by tracking missing expenditures [54].

One recent approach by Mungiu-Pippidi and Dadasov [39] quantifies control of corruption at the national level by measuring the quality of formal and informal institutions which govern the mechanisms by which corruption works in practice. For instance, high administrative burdens, measured by indicators of red tape in domestic bureaucratic regulations, create opportunities for corruption via selective enforcement. Another indicator of this *Index of Public Integrity* is the independence of the judiciary - which, in theory, constrains corruption by the threat of legal intervention. Though supported by theory, such measures cannot explain the significant variation in corruption within polities, where rules are the same and outcomes are significantly diffferent.

Another approach to detect corruption or fraud in large-scale administrative data is to compare the observed distribution of digits in public documents (for instance prices) against benchmarks of "natural" distributions of digits such as Benford's law [80]. Benford's law is based on the simple observation that the first digit of numbers found in administrative tables usually do not follow a uniform distribution. The digit 1 is significantly more likely to occur as the first digit of a number than the digit 9, for instance. Researchers use deviations from such statistical laws as evidence phenomena such as voting fraud [80] and the manipulation of national statistics [81].

These methods to measure corruption demonstrate significant improvements over the perception-based indicators in terms of bias and detail. They all however, depend on specific data which limits their potential for use in comparative studies, and tend to have limited granularity. One recently developed set of methods to measure corruption using data from public procurement contracts addresses many of these concerns. We now present this method, compare it with perception-based indicators, and use it in the rest of the thesis.

Public procurement is the process by which public institutions buy goods and services from the private sector. Such transactions represent a significant share of GDP in both developed and developing countries. The OECD estimates that between 10 and 20% of GDP is spent annually on procurement among its members [20]. Procurement is a significant locus of corruption according to many qualitative measures [82]. Indeed, by virtue of the fact that procurement accounts for a significant amount of the money moving from the public coffers to private bank accounts, it stands to reason that it is one of the major playing fields for actors engaging in grand corruption.

How does corruption in public procurement work? Best practices recommended by the EU [83] and international organizations such as the World Bank [84] suggest that free and fair competition for public contracts to provide goods or services offers the public the best value for money. In this context corruption manifests as the favoring of certain private firms to the detriment of the public good. In practice, corrupt officials adopt a variety of corruption strategies to restrict competition [85]. A favored firm, confident that rivals have been excluded can charge monopoly prices. It is often the case that administrative data on public contracts, which in many jurisdictions must be published, contain markers that such corruption strategies may have been employed. The automated detection of these markers or red-flags in the administrative data of contracts, pioneered by Fazekas [85, 22, 86], offers an objective, micro-level proxy of corruption risk in the behavior of government bodies.

We will provide a more thorough overview of corruption strategies and their corresponding risk indicators in Chapter 3. For now we highlight one such indicator: whether the contract awarded attracted only a single bidder. Single bidding is an outcome of the contract process without any competition. Of course this may happen for a variety of reasons: there may only have been one interested firm, for example. But we will see that aggregated over time and space, perhaps with adjustments made for the kind of good or service being procured, the tendency of contracts to be awarded to single bidders by an institution, town, region, or country has significantly related to other conceptualizations of corruption.

We emphasize that such risk indicators are not proof of corrupt behavior. However, they provide a suggestive indicator that can be used by the authorities and policymakers. The European Court of Auditors has indicated that procurement-based risk indicators of corruption are useful measures of "undetected fraud" [87], signaling to authorities that they are a valid approach to finding candidates for investigation. These indicators are also increasingly popular among academics researching the causes and consequences of corruption. They have been used to study the impact of meritocracy on the quality of government [88], the effect of political competition on the prevalence of corruption in municipalities [43], and the impact of discretion on the quality of bureaucratic outcomes [89].

Beyond the “objectivity” of such indicators, they have several advantages. They are micro-level, quantifying risk at the transactional level. This enables comparisons between regions and institutions at a far more granular level than is feasible with surveys. We exploit this advantage in Chapter 3, in which we measure the corruption risk of Hungarian settlements using such an approach. Data on procurement is also consistently improving time, with good data available in some jurisdictions going back as much as fifteen years.

2.3.3 Comparison of procurement-based indicators with perception-based indicators

We now compare a simple public-procurement indicator-based measure of corruption risk with various alternative measures of corruption risk based on surveys. We use data from Tenders Electronic Daily (TED), the European Union’s portal for public procurement notices and awards. This dataset will be the focus of the analysis in Chapter 4. All tenders estimated above a certain threshold (roughly 5 million Euros for public works contracts and 200 thousand Euros for services) issued by government bodies in the European Union must be posted in this database. As a simple indicator of corruption risk, we calculate the single bidding rate of contracts awarded from 2008 to 2016 by each EU country, visualized in Figure 2.1.

How does this measure correlate with the previously discussed survey-based measures of corruption risk? We find that national single bidding rates correlate significantly with a variety of corruption indicators discussed above, ranging in absolute value from .65 to .72.

We will return to the national-level data in Chapter 4, in which we apply network science methods to describe the distribution of corruption risk in public procurement markets.

2.4 Corruption as Networked Phenomenon

We now turn our attention to past work on corruption, or criminal behavior more general, which employs network methods or perspective. Though we have mentioned network-based interpretations of the findings of related work

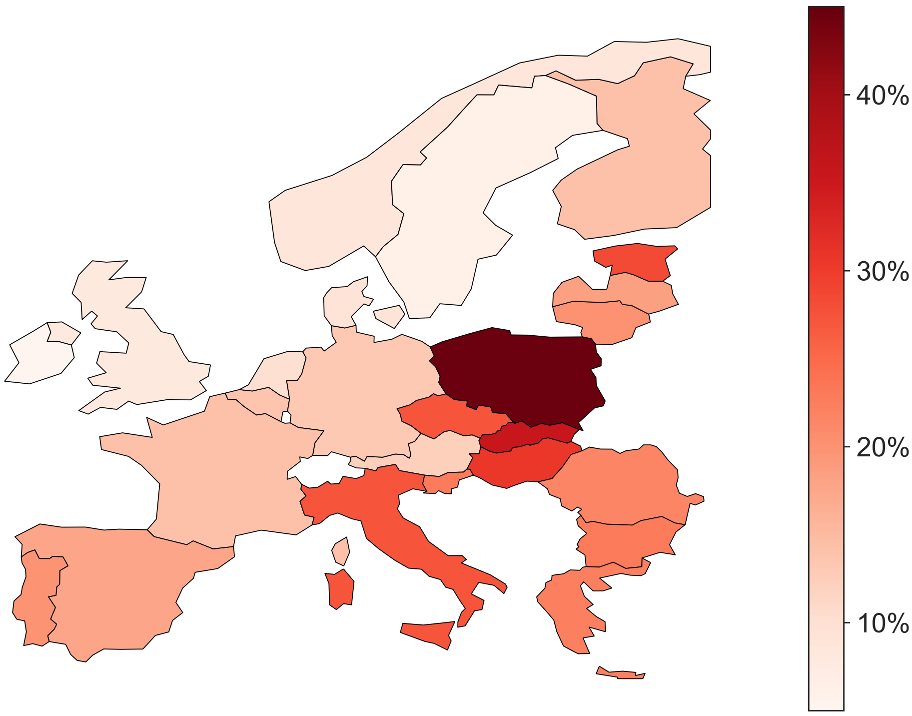

Figure 2.1. *Single bidding rates of EU countries, 2008-2016.*

in the previous sections, the works that we discuss now take an explicit network or relational approach to the study of corruption.

Many studies of organized crime, for instance the mafia represent the structure of the organization as a network [90, 16], especially when transnational organizations are studied [91]. Network structure of contacts and command hierarchy reveal how these organizations function, how they structure themselves to be resilient to turncoats or attacks from rival organizations [92, 93]. Similarly, networks provide a valuable perspective on the operation of terrorist cells [17].

White collar crime has also been studied using network methods. In an example from Canada, the diffuse network of actors responsible for various accounting procedures facilitated a significant corruption ring involving top members of a major political party [94]. Researchers have used email data from Enron, a major US Fortune 500 company which filed for bankruptcy in 2001 amid allegations of fraud and criminal conspiracy by its executives, to study patterns of communication during a crisis and criminal cover-up [95].

In a more recent article, Ribeiro et al. investigated the temporal evolution of the network of co-conspirators in Brazilian political scandals over a 27 year period [14]. By the time of the Petrobras scandal, a giant connected component

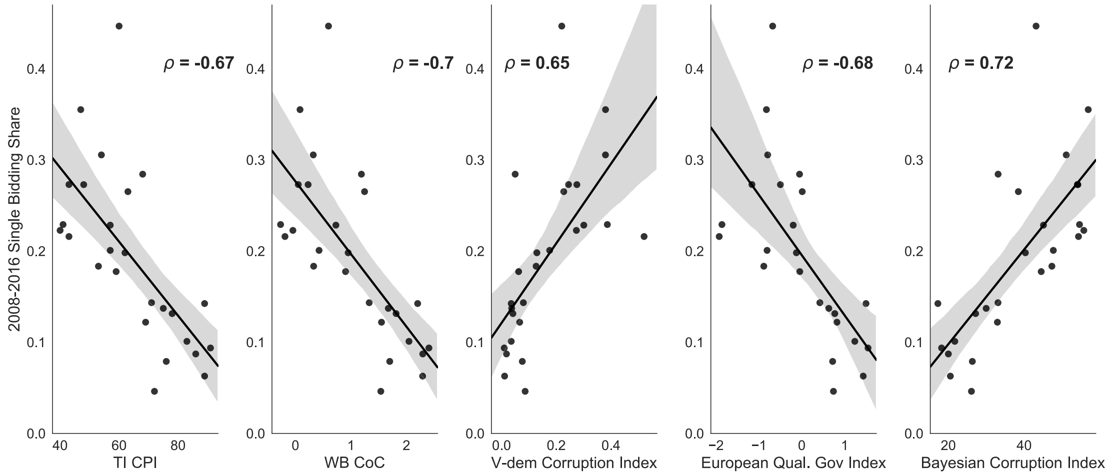

Figure 2.2. *Correlation of national-level single bidding rates 2008-2016 with other commonly used measures of corruption in 2013. Where the correlation between the single bidding rate and the corruption indicator is negative the corruption indicator measures control of corruption and higher values indicate better outcomes.*

had emerged in the network of co-conspiracy, connecting a large majority of individuals. More broadly, network methods have been used to understand the co-occurrence of different sorts of criminality [96, 97]. For example, criminals often specialize in certain kinds of crimes (for instance financial crimes such as fraud) and tend to carry out new kinds of crime with accomplices who already have some experience.

CHAPTER 3

SOCIAL NETWORKS AND CORRUPTION

Corruption is a social plague: gains accrue to small groups, while its costs are borne by everyone. Significant variation in its level between and within countries suggests a relationship between social structure and the prevalence of corruption, yet, large scale empirical studies thereof have been missing due to lack of data. In this chapter we relate the structural characteristics of social capital of settlements with corruption in their local governments. Using datasets from Hungary, we quantify corruption risk by suppressed competition and lack of transparency in the settlement's awarded public contracts. We characterize social capital using social network data from a popular online platform. Controlling for social, economic, and political factors, we find that settlements with fragmented social networks, indicating an excess of *bonding social capital* have higher corruption risk and settlements with more diverse external connectivity, suggesting a surplus of *bridging social capital* are less exposed to corruption. We interpret fragmentation as fostering in-group favoritism and conformity, which increase corruption, while diversity facilitates impartiality in public life and stifles corruption[1].

3.1 Prelude

Corruption is widely recognized to affect adversely social and economic outcomes of societies [99], yet it is difficult to fight [10]. Though education and in-

[1]A stand-alone version of this chapter is due to be published in Royal Society Open Science [98]. A preprint is available at https://arxiv.org/abs/1810.05485.

come seem to decrease corruption [78], it persists even under highly developed, democratic conditions, however, showing significant regional differences [44]. Researchers often relate corruption to social aspects of society such as segregation [100], interpersonal trust [101], civic-mindedness [102], and community engagement [103]. These approaches build on the insight that corruption is a collective outcome of a community shaped by the interactions among individuals [38], suggesting that differences in social capital, especially in the network structure, may help explain the persistence of corruption and the observed differences in its levels.

The concept of social capital or the "connections among individuals – social networks and the norms of reciprocity and trustworthiness that arise from them" [103] is usually applied to understand behavior of individuals [104]. Yet city or country-level aggregations have also proven useful [103], for example in studying economic development and prosperity [105]. As a communal quantity, social capital is a sort of public good embedded in a social network [106] of a settlement. Given the aforementioned relationship between corruption and social capital, it is therefore, as suggested above, natural to expect that the structure of social capital at the settlement level has considerable impact on the scale of corruption in that community. Despite significant interest in the network aspects of corruption [90] and recent experimental evidence that corruption has collaborative roots [52], less is known about how the patterns of connectivity of a whole society influences the general level of corruption in its government.

Previous studies of relating social capital and corruption [107, 108] have been constrained by two empirical challenges: the difficulty of measuring corruption and the lack of data on network structure at the settlement level. Corruption is one of the most hidden type of crimes therefore it is difficult to estimate its extent in general, even with significant local information. For example, an audit study of corruption in rural Indonesia road construction finds that villager perceptions of corruption are significantly distorted by factors such as ethnic diversity compared to objective measures of corruption [74].

As outlined in Chapter 2, many studies measure corruption using national or regional surveys [10] and suffer from the subjectivity of corruption perceptions [75]. Other studies use data on the frequency of investigations and convictions of politicians [78], in which a source of bias may be that in places where corruption is prevalent the judiciary is more likely to be corrupt itself [8]. Recent efforts to clean and standardize large datasets on public procurement [22] have been very helpful in this context as their study can lead to new, more objective indicators of corruption risk.

In the absence of direct network data, researchers often quantify social capital using proxies such as rates of voting, donating blood, and volunteering [102]. As these rates are themselves related to the underlying social networks, they indicate the relevance of social capital and trust instead of explaining the causes of corruption in terms of network structure. Mapping out the social capital at the level of settlements using traditional tools is a formidable task. Fortunately, recent developments in information-communication technologies and their increasing popularity present large data sets containing relevant information. For example, data from online social networks and cellphone records have been used to relate connectivity and socio-economic outcomes [109, 110, 111, 112, 113].

In this Chapter we propose to characterize the level of corruption risk in settlements in terms of their social capital using two sources of micro-level data from Hungary. We quantify the structural characteristics of settlements' social capital using complete data from "iWiW", a now defunct online social network once used by approximately 40% of the adult Hungarian population [114]. We measure corruption risk using administrative data on public procurement contracts over a period of eight years [85].

Public procurement contracts constitute a major channel of public funds to private hands and are highly vulnerable to corruption [22]. Recently, a set of corruption risk indicators have been derived from public contract data, for example, counting how often contracts attract only a single bidder. Averaged to the regional or national levels, these contract-based corruption risk measures have been shown to correlate with corruption perception surveys [22], quality of government indicators [88], and higher cost outcomes for internationally comparable goods such as CAT machines [115]. In the Hungarian case, we find that settlements involved in a recent corruption scandal [116] have significantly higher corruption risk in their contracts.

Putnam distinguishes between two structural categories of social capital: *bonding* and *bridging* social capital [103], and we expect that these have different impacts on corruption risk. Bonding social capital is based on the phenomenon of closure in a social network, describing the extent to which people form dense, homogeneous communities. Such communities have benefits: members share high levels of trust and can count on each other in times of crisis. They can also be confident that members who defy the norms of the community will be censured [117]. The homogeneity of such tight-knit communities is often based on ethnicity, religion, or class [118], indicating possible drawbacks to bonding social capital: homogeneity facilitates conformity and implies exclusion of outsiders [119]. Solidarity can reach the extent that insiders will protect each other even if norms from a wider context are broken, in some cases even if crimes are committed. Sophisticated criminal organizations like the Mafia, members of which may regularly be faced with great incentives to "flip", rely on bonding rituals, ethnic homogeneity, and family ties to enforce solidarity and in-group trust [120, 90]. The negative effects of excessive bonding social capital on society are not limited to crime and corruption. Entrepreneurs embedded in dense networks are disadvantaged because of pressure to employ under-qualified relatives [121], while ethnically homogeneous groups of traders are more likely to overprice financial assets held by their co-ethnics, causing financial bubbles to form [122].

Bridging social capital, on the other hand, refers to the connections between people from different social groups. Such ties are valuable for their ability to convey novel information [123] and exposure to diverse perspectives, though they do not serve as reliable sources of support in hard times. Previous work shows, for instance, that immigrants in the Netherlands with bridging connections outside their ethnic group have significantly higher incomes and employment rates [124]. But bridging social capital is not only thought to be useful for the resources it allocates. Using an agent-based model, Macy and Skvoretz showed how trust emerged among densely connected neighbors and diffused in a social network via weak ties [62], implying that low bridging social capital restricted trust to within-group interactions. Indeed, empirical evidence showed that ethnic groups in diverse communities with more bridging social capital evaluate each other more positively [125].

The two concepts of bonding and bridging social capital exist in tension with each other. They reflect, to quote Portes, "Durkheim's distinction between mechanical solidarity, based on social homogeneity and tight personal bonds, and organic solidarity, based on role differentiation, impersonal norms, and an extensive division of labor." [126]. A settlement in which mutual cooperation relies excessively on mechanical solidarity will tend towards norms of in-group favoritism or particularism [10]. Individuals in such a society will tend to make choices, for example in the allocation of public resources, distinguishing between insiders and outsiders based on a feeling of security rather than trust [127]. In contrast, when cooperation is built on impersonality general trust facilitates impartial outcomes.

We therefore pose two hypotheses relating bonding and bridging social capital to local corruption risk. The first (H1) is that excess bonding social capital, indicating the potential presence of norms of in-group favoritism in a settlement is correlated with higher corruption risk in its government. The second (H2) is that a high level of bridging capital, including connections to other settlements, is correlated with lower levels of corruption risk because it fosters impersonal and universalistic norms. Where mechanical solidarity or bonding social capital dominate organic solidarity or bridging social capital, universalistic norms under which public markets are thought to function best are unsustainable. These hypotheses suggest why corruption is so difficult to fight: it is embedded in the social network of a place.

Previous work using survey data is in accord with our hypotheses. Harris finds a significant positive relationship between excess bonding social capital, measured using surveys, and corruption across over 200 countries [107]. In a comparative study of the 50 US states, Knack finds that residents in states with higher census response and volunteering rate their governments' performances more highly [128]. He finds no such effect for rates of membership in social clubs, a more exclusive form of socialization than volunteering. Paccagnella and Sestito find that in regions with high electoral turnout and blood donation rates, Italian schoolchildren cheat less frequently on standardized tests. In schools with greater ethnic homogeneity and with hometown teachers, cheating is more frequent [129]. These case studies and indirect evidences give some support the above hypotheses, however, there is need for studies based on more direct data at multiple levels.

We find significant evidence for our hypotheses using multivariate regression models to relate corruption risk and structural aspects of social capital. Hungarian settlements with fragmented social networks, which we interpret as evidence of excess bonding social capital, have higher corruption risk in their public procurement contracts. On the other hand, if the typical resident of a settlement has more diverse connections, especially over the boundaries of their own settlements, then local corruption risk is lower. These results hold controlling for several potential confounders including economic prosperity, education, demographics, and political competitiveness.

3.2 Empirical Setting and Methods

3.2.1 Public contracting

Recall that in OECD economies procurement typically accounts for between 10 to 20% of GDP [20] covering everything from school lunches to hospital beds and highway construction. The complexity of the contracts and the relative inelasticity of the government's demand for goods make them a prime target for corruption [130].

Our framing of corruption in this context, described in Chapter 2, relies on a benchmark of non-corrupt behavior. Contracts are supposed to be awarded using impartial market mechanisms [131]: open and fair competition for a contract is considered the best way to ensure that the government makes purchases of good quality at the lowest cost. Usually, an issuer of a contract publishes a call for bids from the private sector, setting a deadline for submissions leaving enough time for broad participation. Companies submit sealed offers, including a price. The company offering to provide the good or service for the lowest price, meeting the standards set in the call for bids, wins the contract.

Measuring settlement corruption risk in contracting

Corruption in public contracting typically involves the restriction of competition. If corrupt bureaucrats wish to award a contract to a favored firm, they must somehow exclude other firms from participating in the competition for the contract. We quantify this phenomenon at the contract level by tracking the presence of elementary corruption indicators, signals we can extract from metadata suggesting that competition may have been curbed [22]. These quantitative indicators [85], deduced from qualitative work on corruption in public contracting, are the fingerprints of techniques used to steer contracts towards preferred firms. We consider eight such elementary indicators, defined in Table 3.1.

From these eight elementary indicators we define two measures of contract corruption risk. *Closed procedure or single bidding (C_{csb}):* Did the contract attract only a single bid **or** was the contract awarded by some procedure besides an open call for bids, for example by direct negotiation with a firm or by an invitation-only auction? In terms of the indicators defined above:

$$C_{csb} = max(C_{singlebid}, C_{closedproc})$$

Corruption Risk Index (CRI): Following [85], we average all eight elementary indicators defined in Table 3.1 for each contract.

$$CRI = \frac{1}{8}(C_{singlebid} + C_{closedproc} + C_{nocall} + C_{eligcrit} + C_{bidtime} \\ + C_{nonprice} + C_{callmod} + C_{decidetime})$$

These indicator-based measures of corruption risk have been related to traditional measures of corruption at the regional and national levels. Among EU countries, similar indicators are correlated ($\rho \approx .5$) with both the World Bank's Control of Corruption rankings and Transparency International's Corruption Perceptions Index [22]. We propose that our indicators supplement these perception-based measures with more objective data at a micro-scale.

Our indicators also predict cost overruns and price inflation in European infrastructure projects [115]. At the micro-level, public bodies issuing high corruption risk contracts are significantly more likely to award contracts to new

Indicator and Symbol	Values	Indicator Definition
Single bidder $C_{singlebid}$	$\{0,1\}$	1 if a single firm submits an offer.
Closed procedure $C_{closedproc}$	$\{0,1\}$	1 if the contract was awarded directly to a firm or by invite-only competition.
No call for bids C_{nocall}	$\{0,1\}$	1 if no call for bids was published in the official procurement journal.
Long eligibility criteria $C_{eligcrit}$	$\{0,1\}$	1 if the length in characters of the eligibility criteria for firms to participate in the tender is above the market average[2].
Extreme decision period $C_{decidetime}$	$\{0,1\}$	1 if the award was made within 5 days of the deadline or more than 100 days following.
Short time to submit bids $C_{bidtime}$	$\{0,.5,1\}$	1 if the number of days between the call and submission deadline is less than 5, 0.5 if between 5 and 15.
Non-price criteria $C_{nonprice}$	$\{0,1\}$	1 if non-price criteria are used to evaluate bids.
Call for bids modified $C_{callmod}$	$\{0,1\}$	1 if the call for bids was modified.

Table 3.1. *Elementary indicators of public contract corruption risk.*

companies after a change in government [133]. Finally, evidence from the US suggests that firms making campaign contributions are awarded contracts with higher corruption risk [134].

Each indicator quantifies different ways bureaucrats have excluded competitors in qualitative work on ground truth corruption cases from around the EU [85]. We stress that while no individual indicator or composite measure can credibly suggest that an individual contract was awarded by a corrupt process, aggregated over many contracts issued by the same institution these indicators map highly suggestive patterns. This point is an important motivation for filtering out towns awarding less than five contracts a year. We now describe the corruption techniques behind each indicator in greater detail.

- Single bidder ($C_{singlebid}$) is an outcome: was the contract awarded in a competition attracting only a single offer. At once this captures the success of the other corruption techniques used to restrict competition, and covers those cases in which some other unobserved strategy was used successfully.

- Closed procedure ($C_{closedproc}$ indicates when the contracting authority has decided to award a contract by direct negotiation with a firm or via an invitation-only bidding process. This decision can be used to completely subvert competition. A sophisticated bureaucrat may invite uninterested firms or other firms controlled by friendly actors besides the target firm in order to present a pretense of competition.

- No call for bids (C_{nocall}) indicates when, in the case that the contract was awarded via an open competition, no contract announcement or call for bids was published in the official procurement journal. A corrupt official can greatly decrease the chance of non-favored firms participating by limiting access to information.

- Long eligibility criteria ($C_{eligcrit}$) captures how bureaucrats can box out specific firms by adding requirements to participation criteria. By including many such restrictions (regarding previous experience, company size, qualifications), a corrupt bureaucrat can systematically exclude non-favored firms.

- Extreme decision period ($C_{decidetime}$ highlights suspicious activity between the end of a competition and the decision to award a contract. If the decision period is extremely short, this suggests that the decision to award a specific firm was premeditated, and that the bids were not carefully checked. If the decision period is very long, it may indicate that legal challenges about the contract may be delaying the award decision.

- Short time to submit bids ($C_{bidtime}$) indicates that favored firms may have been tipped off about a competition for tenders ahead of the public announcement. By leaving only a short time between the announcement and the award for non-favored firms, the corrupt official makes it very difficult to submit a bid. It is important to remember that bids are complex legal documents, including at times cost estimates, schematics, and references.

- Non-price criteria ($C_{nonprice}$) tracks the share of non-price related or subjective criteria in the evaluation of bids. For instance, a corrupt bureaucrat may reject a lower cost bid if, according to a subjective criteria of the quality of a bid, it is less favorably evaluated than that of a higher cost bid of a favored firm.

- Call for bids modified ($C_{callmod}$) checks to see if a call for bids was modified between the initial announcment and the deadline. This potential corruption strategy closely emulates $C_{bidtime}$ in that a corrupt official can suddenly change the specifications or rules of a tender shortly before the deadline.

Local Government Contracting Data

We examine 20,524 municipal government contracts from the period of 2006-2014 issued by Hungarian settlements awarding at least five contracts a year on average. We exclude towns issuing fewer contracts because we are interested in systematic patterns of corruption over a sustained period of time. Our indicators applied to individual contracts are only noisy measures of corruption - it is rather the consistent observation of red flags in contracting over time that suggests that a town has a significant problem with persistent corruption risk. Our results are robust to including towns issuing at least one contract per year on average, reported in the SI Table 7.4.

Our goal is to quantify the overall level of corruption risk in a settlement over the full period for which we have data. We create two such scores by averaging the risk indicators defined above over all contracts issued by the settlement. We arrive at two measures of settlement corruption risk: the rate at which a settlement issued closed-procedure or single bid contracts (C_{csb}), and the average Corruption Risk Index (CRI) score of its contracts.

There are 169 settlements in Hungary meeting the minimum contracting criterion, excluding Budapest. We exclude Budapest for two reasons: it is a severe outlier in size and economic importance and because of its unique governance structure. Budapest is split into 23 districts, each with its own local government and mayor. It also has a city-wide government and mayor. As iWiW does not

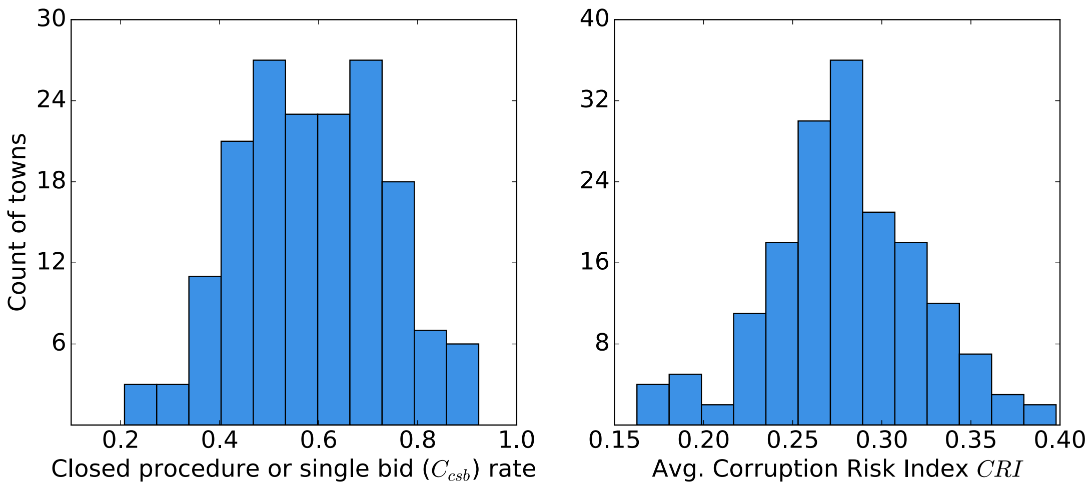

Figure 3.1. *Distributions of average contract corruption risk indicators across Hungarian settlements.*

distinguish between districts and that many contracting decisions are taken at the district level, we judged that we could not reasonably compare the full city with other settlements in Hungary.

We plot the distributions of the settlement corruption risk scores in Figure 3.1. We note that there is substantial variation across settlements: some award over 90% of their contracts either via a closed procedure or to a single bidding supplier, while others do so less than 25% of the time.

As a test of the validity of our settlement-level measures of corruption risk, we check them against a near-ground truth case of corruption. In 2018, OLAF, the European anti-fraud agency reported that 35 Hungarian local government public lighting contracts awarded between 2010 and 2014 contained "serious irregularities" [116, 135]. Elios, the company winning these contracts, was owned at that time by the son-in-law of the Hungarian Prime Minister. The contracts are considered to be overpriced and the Hungarian government was appealed for initiating an investigation, which has already started.

These cases provide a useful test of our corruption risk indicators. There is compelling evidence that settlements implicated in the scandal have, at least once, rigged a public procurement contract to favor a connected firm. We compare the average corruption risk indicators of the 35 settlements that awarded lighting contracts to Elios in the period in question with all other settlements in our sample in Figure 3.2. Using a Mann-Whitney U-test, we find that settlements involved in the scandal have significantly higher rates of corruption risk according to both measures (64% vs 58% C_{csb} rate, U = 1385, p=.033; .30 vs .28

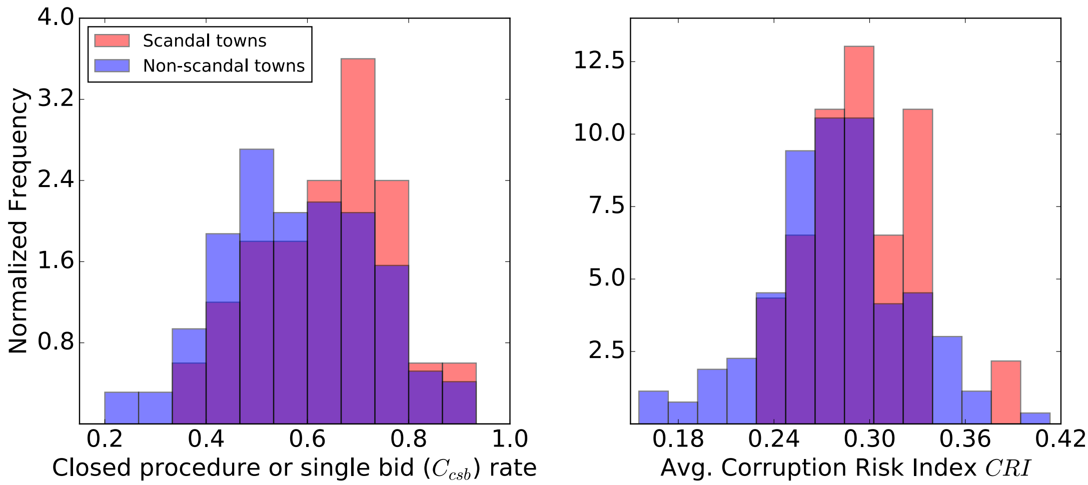

Figure 3.2. *Distributions of average contract corruption risk indicators for settlements involved in the Elios scandal compared with all other settlements. Settlements involved in the scandal have significantly higher average corruption risk in their contracting than their counterparts.*

average CRI, U = 1397, p=.037).

3.2.2 Measuring social capital

iWiW was popular online social network operating in Hungary from 2004 to 2013. At its peak it boasted over 3.5 million active users (out of a population of around 10 million) and was among the top 3 most visited sites in the country. After a period of sustained popularity, it finally collapsed in 2013 as competitors, including Facebook, conquered the market. The increasing tendency of users to leave led to cascades in the social network, highlighting the networked nature of the site [136, 137]. Geographic proximity is a major positive predictor of the likelihood of friendship ties on iWiW, and connections between settlements reflect historical administrative boundaries and geographical barriers [114].

The iWiW network consists of users as nodes and mutually acknowledged friendship ties between users as links. Data from iWiW includes information on each user's settlement, selected from a menu. We used an anonymized version of the data to insure privacy (see the Appendix for a more thorough discussion about data handling). We consider all nodes and links in the network present at the end of 2012, during the peak of its use and before the most significant period of turnover on the site leading to its collapse to define our measures of bonding and bridging social capital. We consider this aggregate network rather than an evolving network from year to year because we do not have in our data (in contrast with the repeated links that can be observed in cellular phone call data) and because we are interested in a long-run characterization of the social network structure of settlements. We describe steps we took to clean the data and the distribution of use rates at the settlement level in the SI.

Despite valid concerns about the representativity of data taken from online social networks [138], studies indicate data from online social networks offer a useful picture of the social capital of their users [139, 140]. As adoption of online social networks increases, they become increasingly useful for the study of the social structures [141]. In any case, we control for possible confounding factors including settlement average income, rate of iWiW use, and share of the population over 60 in our models.

Fragmentation

Our first settlement-level network measure, *fragmentation*, quantifies the extent to which people in the settlement form densely connected and well separated communities. We do not consider the links residents of a settlement have with other settlements. Fragmentation measures a settlement's bonding social capital. Before we proceed, we note that settlement will always be used to refer to a municipality, while community refers to a group of nodes detected in the iWiW social network of a settlement using a network science algorithm, in other words a subset of the nodes of the town which are densely connected.

We measure fragmentation of the settlement's internal social network using a community detection method to identify communities of highly connected nodes. We use the Louvain algorithm [142], a popular and efficient method leading to a partition of the network. We measure the quality of the partition, the tendency of edges to be within rather than between the detected communities, using modularity [143]. Given a social network of users in a settlement S and a partition of the network's nodes into K communities, the modularity $Q(S)$ of the partition of the network can be written as:

$$Q(S) = \sum_{k=1}^{K} \left[\frac{L_k^w}{L} - \left(\frac{L_k}{L} \right)^2 \right],$$

where L is the total number of edges in the considered network, L_k is the number of edges adjacent to members of community k, and L_k^w is the number of edges within community k.

As modularity is highly dependent on the size and density of the network [144], we scale each settlement's modularity score in order to make valid comparisons between the settlements. Following Sah et al. [145], we divide each settlement's modularity score by the theoretical maximum modularity $Q_{max}(S)$ that the given partition could achieve, namely if all edges were within communities.

$$Q_{\max}(S) = \sum_{k=1}^{K} \left[\frac{L_k}{L} - \left(\frac{L_k}{L} \right)^2 \right].$$

We then define the *fragmentation* F_S of a settlement S as the quotient

$$F_S = Q(S)/Q_{\max}(S). \tag{3.1}$$

Fragmentation measures the tendency of individuals to belong to distinct communities within a settlement. A fragmented settlement consists of tightly-knit communities that are weakly connected. Both the excess of connections within and the rarity of connections between communities in fragmented networks are relevant to our theoretical framing of the origins of corruption as they indicate excess bonding social capital. The high density of connections within a community facilitates the enforcement of reciprocity, while having few connections between communities fosters particularism.

To better understand the concept of fragmentation, we compare two settlements, one at the 90th percentile of fragmentation (settlement A) and the other at the 10th percentile (settlement B). The two settlements have populations of roughly 10,000 and have iWiW user rates between 30 and 35%. We randomly sample 300 users from both social networks for the sake of visualization and plot their connections in Figure 3.3. Settlement A is clearly more fragmented than settlement B. We also show the full adjacency matrices of the networks of these settlements, grouping nodes by their detected communities into blocks on the diagonal shaded in red. We label each community by the share of its edges staying within the community. The fragmented settlement has a clear over-representation of within-community edges.

Diversity

Past research on online social networks noted that users connect with people from a variety of focal experiences in their life-course [140]. For example a user may connect with her schoolmates, university classmates, coworkers, family, and friends from environments including social clubs, sports teams, or religious communities. We measure the diversity of an individual user's network by the (lack of) overlap between these foci. In this case, we do not restrict our attention to edges between users from the same settlement.

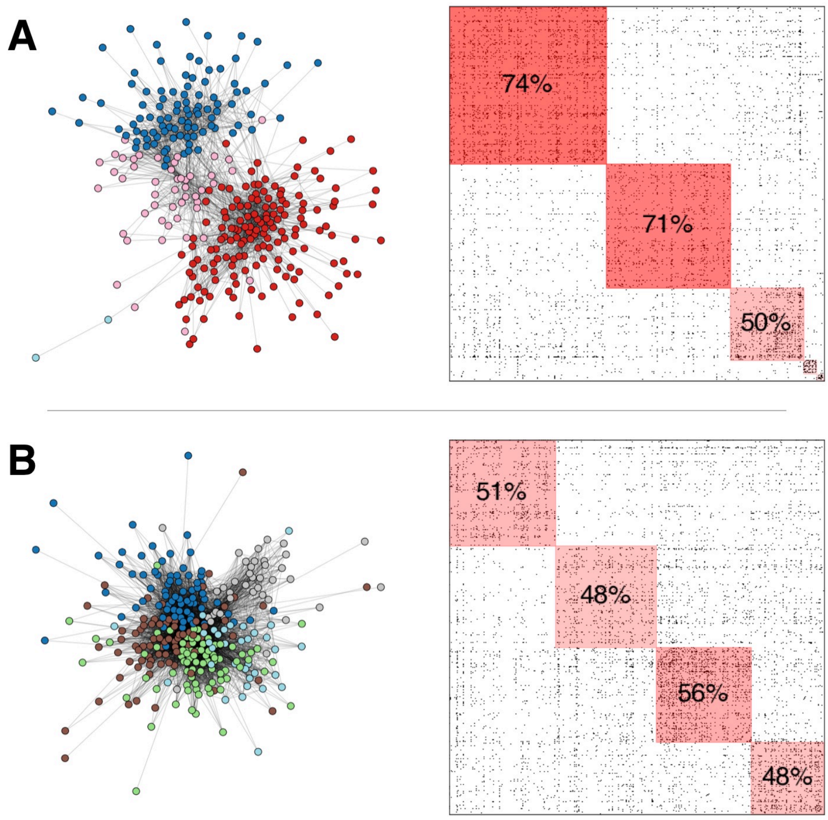

Figure 3.3. *Sampled social networks and adjacency matrices of high (A) and low (B) fragmentation settlements. Node colors indicate membership in communities. In the adjacency matrices, percentages indicate the share edges staying within each community. In the fragmented settlement, communities have a significantly fewer connections with other communities.*

Following Brooks et al. [140], we consider the connections amongst the friends of a focal user or ego, i.e. the ego network without the ego. We then detect communities in the resulting network using the Louvain algorithm [142]. We can assume that members of a community of alters share some common context. We measure the separation of these communities of alters using modularity. Low modularity indicates that a user's connections tend to know each other, and that the user's different spheres of life involve the same people. High modularity indicates that the ego has a bridging role between weakly connected communities, and so we refer to such users as having high diversity in their social networks. We show examples of low and high diversity users with networks of similar sizes in Figure 3.4.

We aggregate this user level measure to a measure of settlement diversity D_S by averaging each user's modularity score:

$$D_S = \frac{1}{|S|} \sum_{i \in S} Q(\{\text{alters}_i\}),$$

where $|S|$ is the number of nodes in the settlement S and $\{\text{alters}_i\}$ is the subgraph of the alters of node $i \in S$. This measure captures the typical diversity of social perspectives that the members of the settlement access. At the settlement level this measure captures bridging social capital.

Settlement diversity is positively correlated with share of the population graduating from high school ($\rho \approx 0.62$) and average income ($\rho \approx 0.55$). Fragmentation and diversity are positively correlated ($\rho \approx 0.46$), which is not surprising given that both are calculated using network modularity. However, the ego-focus and, more importantly, the inclusion of inter-settlement edges of the diversity measure distinguish it from fragmentation, see the SI. Despite this correlation, we observe that they predict different corruption outcomes.

3.2.3 Models

The primary aim of our chapter is to relate bonding and bridging social capital in settlements to corruption risk in their public contracts. Our hypothesis H1 is related to excess bonding social capital, measured by fragmentation while hypothesis H2 refers surplus in bridging social capital, measured by diversity. We predict average contract corruption risk at the settlement level using Ordinary Least Squares (OLS) multiple regressions of the following form:

$$C_S = \beta_1 * F_S + \beta_2 * D_S + X_S * \theta + \epsilon_S$$

Where C_S is one of two corruption risk indicators, averaged at the settlement level, F_S is the settlement's fragmentation, D_S is the settlement's diversity, X_S is

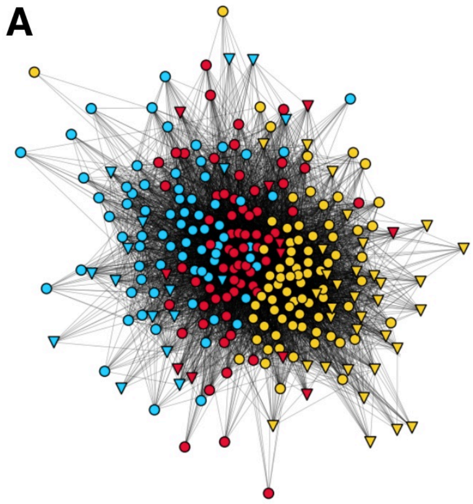

Figure 3.4. *Ego networks with low (A) and high (B) diversity, respectively. Colors indicate membership in detected communities in the ego network. Circles denote users from the same settlement as the ego, while triangles mark users from elsewhere. The high diversity user's network has clusters of alters mostly from different settlements.*

a matrix of control variables defined below, and ϵ_S is an error term. The β-s are scalar and θ a vector of unknown parameters.

We include a variety of control variables in our regressions. Past research has found significant relationships between wealth, education, employment and corruption [10] so we control for settlement average income, its share of high school graduates, the presence of a university in the settlement, and its unemployment and inactivity rates. As demographic features of settlements may influence the measured social network we include total population, rate of iWiW use, and share of the population over 60 in our models [146]. These socio-economic controls indicate 2011 levels when possible, as 2011 was the most recent Hungarian census. We also control for the settlement's mayor's average victory margin in the 2002, 2006, 2010 elections as a proxy for the level of political competition in the settlement, which has been found to be positively related with local quality of government [43]. Finally, we include a geographic feature of the settlements: the minimum travel distance in minutes from the capital, Budapest. Past work indicates that distance from central authorities predicts higher rates of corruption [147]. We present additional details on the control variables in the SI. For the sake of comparison we fit a baseline model including only the control terms.

Implicit in our modeling framework is our choice to aggregate the social network measures, corruption risk scores, and controls of settlements into a single snapshot. Contracts range from 2006 to 2014, iWiW friendships from 2002 to 2012, and controls are set at 2011 levels (corresponding to the last Hungarian census). As we are studying the relationship between social structure and corruption, both long-run phenomena, we claim that this represents sufficient temporal overlap.

3.3 Results

We summarize our findings in Table 3.2. We see that there is a significant relationship between social network structure and both dependent variables measuring corruption. More fragmentation consistently predicts more corruption, while more diversity consistently predicts less corruption. In both cases adding the network features significantly improves the adjusted R^2 of the model. Moreover, comparing the coefficients, we see that the social network features have effect sizes comparable to that of any social, political, or economic control. We present the full models in the SI, including the intermediate models containing only one network feature. All models pass a variance inflation factor (VIF) test for feature collinearity, see the SI.

We visualize the effects of our network variables in Figure 3.5. We plot model predicted rate of closed procedure or single bid contract awards (C_{csb}) including 90% confidence intervals for varying levels of fragmentation and average ego diversity. As the variables are standardized, the units can be interpreted as standard deviations from the mean (at 0). We observe that, all else equal, our model predicts that going from one standard deviation below average fragmentation to one standard deviation above average, increases C_{csb} by about one half of a standard deviation. Diversity has a stronger effect in the other direction: the same change (from one standard deviation below average to one above average) induces a full standard deviation decrease in the corruption indicator. The effect of the network features on C_{CRI} is similar. In the SI, we present an ANOVA feature importance test that indicates the significance of both network-based features.

3.4 Discussion

In this chapter we used data from an online social network and a collection of public procurement contracts to relate the social capital of Hungarian settlements to the corruption in its local government. To our knowledge, this study is the first to study social aspects of corruption using large-scale social network data.

Dependent variable:	% Closed or single bid. (1)	(2)	Average CRI (3)	(4)
Fragmentation		**0.263*****		**0.207****
(Bonding social capital)		(0.097)		(0.092)
Diversity		$-$**0.553*****		$-$**0.551*****
(Bridging social capital)		(0.176)		(0.168)
Income/capita	$-$0.262	$-$0.277*	$-$0.075	$-$0.096
	(0.169)	(0.162)	(0.161)	(0.155)
N contracts (log)	$-$0.313*	$-$0.314*	$-$0.685***	$-$0.697***
	(0.171)	(0.165)	(0.162)	(0.158)
Population (log)	$-$0.180	0.020	0.118	0.335**
	(0.143)	(0.166)	(0.136)	(0.159)
Rate iWiW use	0.045	0.037	0.122	0.107
	(0.137)	(0.132)	(0.130)	(0.126)
Mayor victory margin	0.278***	0.255***	0.303***	0.281***
	(0.089)	(0.086)	(0.085)	(0.082)
% high school grads	0.166	0.374*	$-$0.176	0.040
	(0.190)	(0.199)	(0.181)	(0.190)
Distance to Budapest	$-$0.021	$-$0.198*	0.061	$-$0.112
	(0.104)	(0.112)	(0.099)	(0.107)
Share of pop. inactive	$-$0.797***	$-$0.805***	$-$0.716***	$-$0.754***
	(0.229)	(0.229)	(0.218)	(0.219)
Unemployment Rate	0.239**	0.262**	0.299***	0.320***
	(0.118)	(0.113)	(0.112)	(0.108)
% population 60+	0.501***	0.491***	0.500***	0.503***
	(0.163)	(0.158)	(0.155)	(0.151)
Has university	0.351	0.294	0.431**	0.352*
	(0.220)	(0.221)	(0.210)	(0.211)
Constant	1.245*	1.206*	2.779***	2.790***
	(0.725)	(0.702)	(0.689)	(0.671)
Observations	169	169	169	169
Adjusted R^2	0.163	**0.230**	0.183	**0.243**
F Statistic	3.967***	4.859***	4.419***	5.142***

Table 3.2. *Settlement-level regression results predicting two corruption risk indicators. For both dependent variables, the first columns (1) and (3) correspond to the base model, predicting corruption risk using only control variables, and the second columns (2) and (4) show results, when the social network features are included. Note that all features are standardized with mean 0 and standard deviation 1. Significance thresholds:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

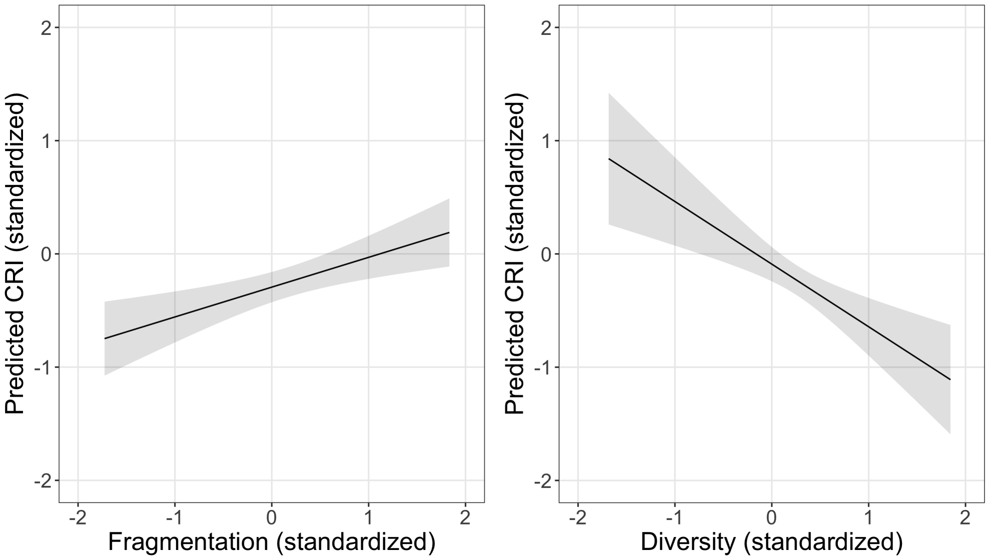

Figure 3.5. *Plots of marginal effects of the key social capital variables and their predicted impact on a settlement's rate of closed procedure or single bidder contract awards; shaded regions represent 90% confidence intervals are indicated. As the variables are standardized, unit changes on either axis can be interpreted as standard deviation changes. Fragmentation (on the left), quantifying excess bonding social capital in a community, predicts higher corruption risk, while diversity (on the right) predicts lower corruption risk.*

We introduced measures to quantify excess bonding and bridging social capital at the settlement level from online social network data. We found that settlements with high bonding social capital tend to award contracts with higher corruption risk. We also found that settlements with high bridging social capital, tend to award lower corruption risk contracts. Social capital measures add substantive predictive power to models of corruption outcomes, above baseline models controlling for other socio-economic factors such as average income, education, political competition, and demography.

We recognize several limitations to our approach. An inherent challenge in the research of corruption is that proven cases are rare, and so our measures can only track risk or suspicion of corruption. Moreover, we assume that steering contracts to certain firms by bureaucrats indicates corruption - but it may happen that bureaucrats make socially optimal decisions using their local knowledge of markets and discretion [148].

It is also clearly the case that iWiW is not a full map of social relations in Hungary and its users do not make up a representative sample of the population. Finally, we do not claim to have found a causal link between social capital and corruption risk. Besides the potential of omitted variable bias, it is highly likely that corruption also influences social capital in the long run [149].

Despite these limitations we believe that our findings are valuable. Above all, our novel, data-based, settlement-level approach provides new evidence for the old hypothesis that corruption is a structural phenomenon. Our finding that social structure relates to corruption risk suggests, for example, why appointing an ombudsman in a corrupt place rarely improves corruption outcomes [10] and why anti-corruption laws can backfire if they conflict with prevailing social norms [150].

That is not to say that fighting corruption is futile. Rather we believe our findings suggest that top-down efforts are unlikely to work unless they impact social capital or other significant covariates of our model like political competition. Our conclusions hint at potential mechanisms which sustain corruption. Factors such as racial segregation or economic inequality which may drive fragmentation are ideal targets for policy interventions [101].

CHAPTER 4

Corruption and Procurement Markets

4.1 Prelude

In this chapter we map public procurement markets as bipartite networks of issuers (alternatively buyers, public institutions) and winners (suppliers, firms) of contracts. Using corruption risk indicators, we study the relationship between corruption risk and micro-, meso-, and macro-level structure of markets. We find that corruption risk is related to various structural properties of the market captured by its network structure, and that important heterogeneities exist among EU countries. We present findings using a dataset of over four million contracts awarded across the European Union between 2008 and 2016[1].

Bipartite networks have been used to study a wide variety of phenomena involving two distinct sets of actors, for example buyers and sellers in markets [151], flowers and their pollinators [152], and cities and industries [153]. As public procurement markets consist of two kinds interacting of actors (issuers and winners), they fit into this paradigm. We create networks of buyers and suppliers at the national level and describe their structure using the tools of network science.

As in the case of unipartite networks, empirical bipartite networks often exhibit certain regularities that distinguish them from random networks. For example the nodes in such networks tend to have heterogeneous degree distri-

[1]This chapter partially draws on work that is under review at a journal at the time of submission of the thesis [86]. A preprint is available at http://www.govtransparency.eu/wp-content/uploads/2017/09/Fazekas_Wachs_Skuhrovec_CorruptionNetwork_Structure_in_CZ_HU_2017.pdf.

butions, meaning that certain nodes sometimes referred to as hubs have orders of magnitude more connections than other nodes. We show that procurement markets are no exception to this observed pattern, having heterogeneous degree distributions among both winners and issuers.

Some of the most interesting empirical regularities observed in real-world networks extend beyond the node level, describing structures at so-called mesoscopic scales. One example of a structure commonly found in empirical networks is the so called core-periphery structure. By filtering a network for its most important connections, referred to as the core of the network, using the network topology we can simultaneously highlight key actors and measure the centralization of a network [154, 155]. Core-finding methods have been used to analyze networks of trade [156], success in creative industries [157], and political organization [158]. We apply a method to our data that considers both the weight of connections between issuers and winners, and the bipartite nature of our networks. Besides distinguishing important actors in the market, the relative size of the core and the share of contracts between core members is a measure of the centralization of a market. Indeed we find that there is a significant variation in the share of contracts between core issuers and firms across the EU countries.

Another example of mesoscopic network structure that occurs frequently in empirical networks is *modularity*. As we saw in the previous chapter, social networks are often modular [143] because of factors such as homophily or social segregation [159]. Modularity in a network derived from market interactions merits its own interpretation. It may indicate the degree of specialization in a market - i.e. if firms tend focus on providing a narrower set of goods and services - or the effects of geography. Using a method adapted especially for bipartite networks, we partition procurement markets into communities and find that they have significant modularity. We then analyze the distribution of corruption risk across the different modules of a market. Corruption risk may be concentrated in a few specific sectors, or it may be spread rather evenly throughout the entire market.

Combined with the overall prevalence of corruption risk in a country, we are able to describe in much more detail the shape of corruption risk and its organization using these methods. While two countries may have the same overall level of corruption risk in their public contracting, its distribution can be different. In one case the corruption risk may be highly clustered in certain communities and more common in the periphery of the market. In another, corruption may be evenly spread around the market, with no specific hot-spots of corruption, and on average more common in its core. Such an perspective offers a powerful diagnostic tool to researchers of corruption. Certainly effective prescriptions to combat corruption would be different in these two cases.

To emphasize the dynamics of corruption risk, we also analyze how changes in government effect high corruption risk winners in the core of the network. We test the survival of winning firms across two years intervals, finding that in some countries winners are more stable across changes of government, while in others they are less likely to survive. More abstractly speaking, this analysis quantifies the political organization of corruption in the different EU countries.

We proceed by introducing the data, describing how it can be represented as a bipartite network, defining measures of the resulting networks, and comparing network structure with corruption risk indicators. We find significant differences between countries, beyond what can be explained by differences in the overall prevalence of corruption risk. We conclude with a discussion of these findings.

4.2 Data

Our data is collected from *Tenders Electronic Daily*[2] (TED) - the official journal of public procurement contracts of the European Union. TED includes both calls for tenders and award announcements of contracts from all member states in the European Union. TED estimates the total value of tenders published in a year is worth approximately 420 billion Euro - a significant share of EU GDP (2.8% in 2016). EU law requires, with some rare exceptions, that all public procurement contracts estimated in value above a certain threshold issued by member states be reported on the website. In 2016 these thresholds were 135,000 Euro for service and supply contracts, and 5.2 million Euro for works contracts. Though a significant amount of procurement takes place below this threshold (and in most cases is recorded in national or regional level data portals), we use only TED data above the threshold in order to maximize cross-national comparability. One drawback to our use of TED is that corrupt actors have significant incentives to issue contracts just under the reporting threshold [160], leading to an underestimation of corruption risk in certain countries or markets.

Our dataset consists of 4,098,711 contracts awarded between 2008 and 2016 from 26 member states of the EU[3]. We exclude contracts awarded by the European Institutions, for example the European Commission and European Parliament, as we are interested in a comparison between countries. As issuers and winners of public contracts are named by raw text string at the contract level, and not by unique identifiers (i.e. by EU VAT ID or other national tax ids), we

[2] https://ted.europa.eu/

[3] We exclude Luxembourg because it issues relatively very few contracts, and Croatia because it joined the EU only in 2013.

first created an algorithm to identify and merge aliases referring to the same issuers and winners, country by country. In the computer science literature, such a task is known as deduplication or record-linkage. Deduplication in general is a difficult problem because comparing all pairs of N potentially distinct entities will require in general on the order of N^2 comparison. Our method follows the overall approach outlined by Christen [161], consisting of five steps:

- Preprocessing: we process at the contract level each issuer and winner's name, street address, postal code, and country fields by lower-casing all characters and removing punctuation and superfluous white space.

- Choosing a measure of similarity: we quantify the similarity of two entities using an active learning approach [162] implemented in the *Dedupe* Python programming library [163]. For each field in a record, the method defines several measures of similarity. For instance, the text names of two winners can be compared using a measure of string edit distance (counting for instance the number of insertions and deletions required to transform one string into another), the overlap of shared words, or the similarity of the first characters in each word. The algorithm selects a small sample of pairs of records and calculates a large number of similarity measures. It presents the user with those examples of pairs of records which it is most unsure about. The algorithm learns both which similarity measures are productive in the classification of pairs of records and how and in what proportion they should be combined to achieve the best accuracy. By labeling 100 such examples, the user can greatly increase the accuracy of the deduplication procedure. We ran this algorithm and manually classified 100 examples for both winners and issuers for each country in our dataset.

- Choosing records to compare: As mentioned, there are many possible records to compare. Comparing all pairs of 1,000 records for example would require making roughly 500,000 comparisons. France alone has 364,125 unique winner names. Moreover, since the method of comparing records learned in the previous step requires the calculation of multiple measures of similarity, it is important to reduce the space of searches in a clever way. This is accomplished via a technique called blocking: the records are split into groups based on simple features such as sharing the same first three characters in their name (i.e. Microsoft and Microsystems would be placed in the same block) or sharing the same first character in the first two tokens of their names (i.e. Air France and Air Finland would be placed in the same block). Blocking is not a strict partition: records can be placed in multiple blocks. Using the same sample of manually labeled candidate matches, the computer program learns which blocks are most effective for maximizing the trade-off between accuracy and the number of comparisons that need to be calculated.

- Grouping similar records: Once the similarity of records for all pairs of issuers or winners in all blocks have been calculated, we use a hierarchical clustering method to determine which records should be identified as duplicates.

- Selecting a threshold in the clustering: Again, relying on the manually checked sample of records we select a threshold that weighs false positives (i.e. two records which do not refer to the same entity in reality, but are merged) and false negatives (i.e. two records which refer to the same entity in reality but are not merged) equally. The overall accuracy of our approach, while varying for issuers and winners and country by country, exceeds 90% on the sample of difficult to classify cases identified by the software. The number of unique winners in France, for example, decreases from 364,125 to 200,584 after the procedure.

Equipped with this deduplicated set of issuers and winners, we can, with significantly greater confidence, represent the interactions between actors as a network. We are also interested, of course, in quantify corruption risk at the contract level, as we did in the previous chapter. Unfortunately, data inconsistency across both time and countries in the TED data limit us in this regard. After an evaluation of the various potential corruption risk indicators, their stability over time within countries, and the rate at which data was missing, we concluded that a single indicator - single bidding - was the most appropriate measure of corruption risk. We have already presented the correlations between this measure, aggregated over time, and other perception-based measures of corruption risk in Chapter 2, finding a national level correlation between the rate of single bidding and measures such as Transparency International's Corruption Perceptions Index and the World Bank's Control of Corruption measure of between .6 and .7.

4.3 Markets as Bipartite Networks

Given a collection of contracts awarded by issuers to winners from a given country in a year, we construct a network as follows. The nodes consist of issuers and winners of public contracts. An issuer and winner are connected by an edge if they are in a contracting relationship that year. The edge has a weight, initialized at 1, that increases in the count of contracts entered between

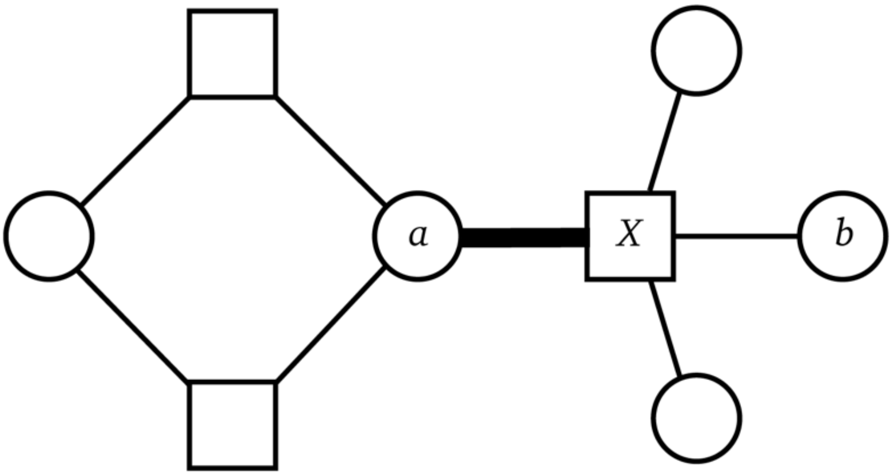

Figure 4.1. *A toy example of a public procurement market represented as a bipartite network. Squares represent issuers and circles represent winners of public procurement contracts. For instance, winners a and b are connected by edges with issuer X. The edge between a and X is thicker than the edge between X and b, indicating that a has won more contracts from X than b. Adapted from [86].*

the issuer and winner. We note that this network is *bipartite*, that is consisting of two nodes sets within which there are no edges, i.e. there are no connections between two winners or two issuers[4]. Each edge also encodes the single bidding rate of contracts between the given issuer and winner. We visualize a toy example of such a bipartite network in Figure 4.1.

We proceed by calculating a few summary statistics about these networks, seeing if they conform to some of the regularities observed in other empirical networks which distinguish them from random networks. We calculate the density of each network by dividing the number of edges present, over the number of edges in the complete bipartite graph with the same number of issuers and winners. We also calculate the averages and standard deviations of degrees of both nodesets. Finally, we calculate the Robins-Alexander clustering [164] of each network, a measure of local correlations in connectivity for bipartite networks, analogous to measures of triadic closure in monopartite networks such

[4]In the rare case that a public institution also wins contracts as a supplier entity, we split the institution into two nodes: one capturing its role as an issuer, the other as a winner of contracts.

as social networks.

Robins-Alexander clustering counts the number of four-cycles (C_4) in a network, defined as a path of edges between 4 nodes, starting and ending at the same node and visiting each node once. In the toy network visualized in Figure 4.1, the path starting and ending at a and visiting the nodes in the left part of the network is a four-cycle. The count of four-cycles is divided by the count of paths of length three, representing the number of potential four-cycles. This quotient is multiplied by 4 to account for the fact that each four cycle contains four paths of length 3. This measure is a direct generalization of the concept of the global clustering coefficient for monopartite networks [32, 165], which in social networks measures the likelihood that if an individual has two friends, those two individuals are also friends with each other. As there are no triangles in bipartite networks, Robins-Alexander clustering is a natural extension of the concept from triangles to squares.

We present some summary statistics of the networks of each country averaged over the period 2008-2016 in Table 4.1. We first note that the number of contracts awarded and nodes participating in the market varies significantly from country to country. Despite this heterogeneity, in other regards all procurement markets seem to share properties typical of other empirical networks. The markets are sparse, with the number of observed connections well below the number of possible connections, as indicated by the low density. The networks have significant Robins-Alexander clustering, suggesting interdependence of nodes close to one another in the network. For both winners and issuers, the variance of their degrees exceeds the average of their degrees, a signal that the degree distribution for both node sets is heterogeneous.

Country	# Contracts	# Winners	# Issuers	Density	R-A Clust.	$\mu(\text{Deg}_W)$	$\sigma(Deg_W)$	$\mu(Deg_I)$	$\sigma(Deg_I)$
AT	3314	1882	395	0.0033	0.03	1.8	2.3	8.4	22.7
BE	6674	3046	1039	0.0014	0.02	2.2	4.4	6.4	15.2
BG	8653	2150	484	0.0048	0.27	4.0	14.6	17.9	56.3
CY	916	403	64	0.0179	0.14	2.3	3.7	14.3	48.7
CZ	8030	2933	986	0.0017	0.04	2.7	7.4	8.1	32.7
DE	32339	15395	4049	0.0004	0.03	2.1	5.2	7.9	23.5
DK	4858	2099	539	0.0028	0.04	2.3	4.1	8.9	26.5
EE	1913	967	170	0.0083	0.08	2.0	2.6	11.0	28.9
ES	20035	7496	1765	0.0011	0.13	2.7	6.8	11.3	31.8
FI	6248	2750	578	0.0029	0.05	2.3	12.2	10.8	27.4
FR	120946	42562	6294	0.0003	0.08	2.8	11.5	19.3	51.4
GR	4246	2348	437	0.0031	0.12	1.8	2.2	9.7	44.0
HU	5700	2016	610	0.0026	0.08	2.8	5.9	9.4	27.4
IE	2713	1587	208	0.0056	0.03	1.7	2.3	13.1	51.3
IT	18249	7749	2434	0.0006	0.07	2.4	7.2	7.5	29.0
LT	9007	1368	272	0.0084	0.32	6.5	40.2	32.5	178.2
LV	9451	2148	262	0.0057	0.15	4.3	14.8	36.1	119.4
NL	6691	3579	1136	0.0013	0.02	1.8	2.6	6.0	14.7
NO	3479	1899	497	0.0031	0.04	1.8	2.7	7.0	12.6
PL	108886	19079	3649	0.0006	0.23	5.7	64.7	29.7	96.9
PT	2255	1052	334	0.0041	0.08	2.1	3.1	6.7	30.4
RO	19807	3503	939	0.0025	0.32	6.0	44.5	21.1	72.5
SE	9441	4721	724	0.0022	0.06	2.0	3.7	13.1	33.3
SI	6623	1268	448	0.0067	0.27	5.2	18.7	14.8	33.3
SK	2654	1068	344	0.0043	0.09	2.4	5.1	8.1	32.5
UK	32275	15577	2230	0.0006	0.04	2.1	3.7	14.6	50.0

Table 4.1. *Summary statistics of procurement market networks, averaged over 2008-2016. R-A Clust. refers to Robins-Alexander clustering, a measure of the local correlation of connectivity in bipartite networks, analogous to triadic closure in monopartite networks. The final four columns present the averages and standard deviations of winner and issuer degree (weighted by contract count), respectively.*

In order to investigate the heterogeneity of the degree distributions further, we plot distribution of the number of contracts awarded and won by each issuer and winner, respectively, pooled across the entire EU in Figure 4.2. Using the method described by Clauset et al. [166, 167] we estimate the best fit alpha of a power-law distribution for both distributions, estimating an alpha of 2.27 for winners and 2.16 for issuers. In both cases a statistical test between the goodness of fit of a powerlaw distribution compared with a lognormal distribution is inconclusive: neither one is significantly better. What is clear, however, is that both node sets have extremely heterogeneous distributions, with most winners and issuers participating in very few contracts, while a few winners and issuers participate in over a thousand contracts. This sort of heterogeneity is a typical characteristic of empirical networks and has significant implications for their structure.

We repeat this visualization and calculation at the country level in Figure 4.3. We observe remarkable regularity in the distributions of issuers and winners across countries. In all countries, both distributions are highly heterogeneous - the number of contracts issued or won by nodes ranges across several orders of magnitude.

So far we have observed several regularities in the network structure of the EU national procurement market networks. These indicate in some sense why analyzing these markets as networks may be a fruitful endeavour: structure in these networks deviates significantly from what one would expect in networks with the same number of nodes and edges and uniform random probability. Such networks do not have significant clustering, nor do they have significant heterogeneity in their degree distributions. We now proceed with a study of these networks that is theoretically relevant to market structure and the distribution of corruption risk in them.

4.3.1 The Core of Procurement Markets

Given the size and complexity of these markets when modeled as networks, it is natural to ask if they can be filtered or simplified in a way that highlights the most central and significant actors and interactions. Networks often have such a center or *core* which is essential to its functioning [155]. It is unclear if public procurement markets have such a core-periphery structure. Likely this depends on the organization of the market, for example if the procurement contracts are awarded in a centralized or decentralized manner in a country. Detecting the core of procurement markets and measuring their relative sizes offers us a way to highlight important issuers and winners, and more generally how centralized the market is. This itself would be a valuable measure as we can revisit a major debate in the corruption literature, which asks if that centralization in

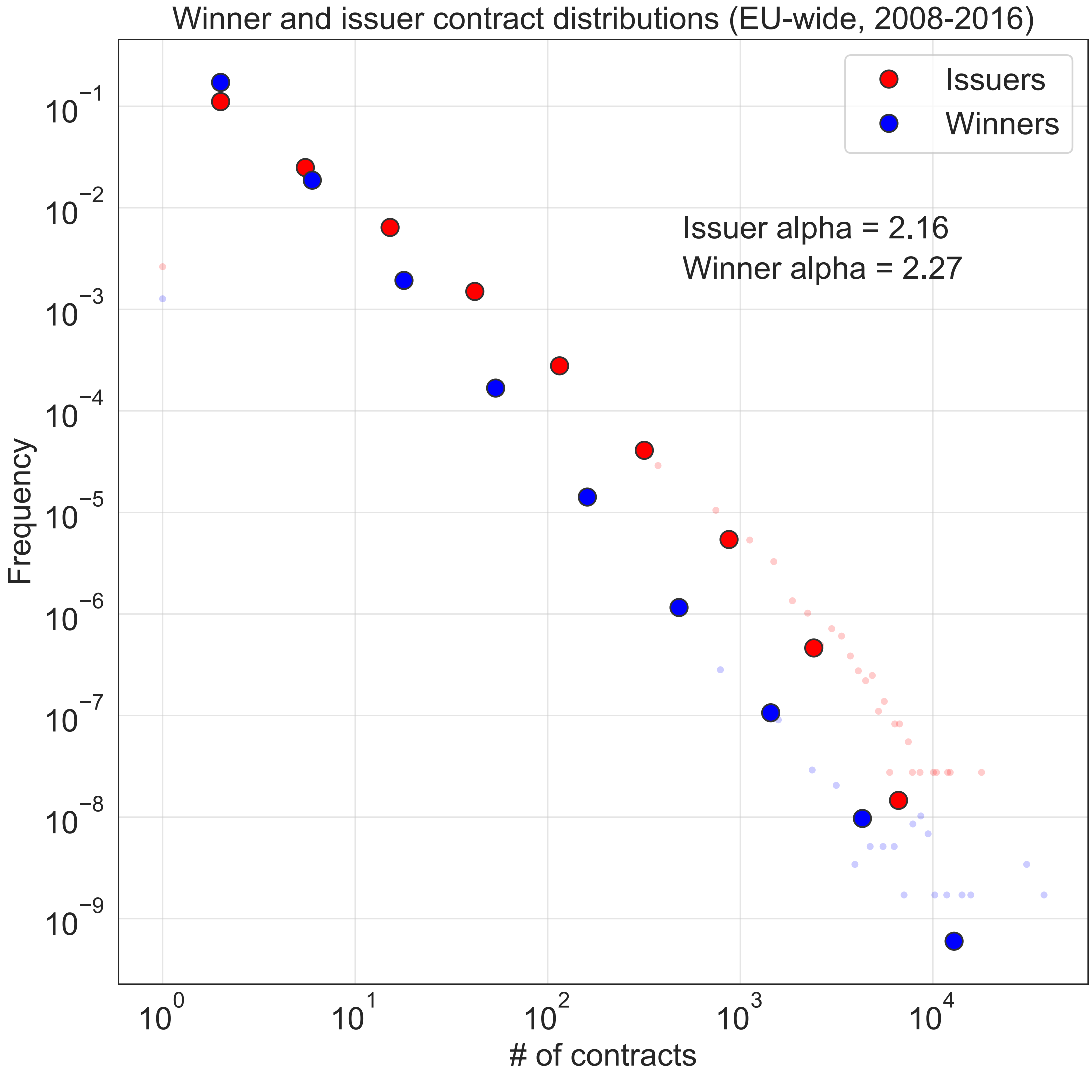

Figure 4.2. *The distribution of the number of contracts awarded and won, for issuers and winners, respectively, of procurement contracts across the entire EU. The distributions are plotted on a log-log scale. We report the alpha parameter of a power-law degree distribution fitted to both distributions in the plot.*

government is correlated with worse corruption outcomes, from a novel perspective.

Most comparative studies relating the centralization of government and corruption measured using perception-based indicators find a positive correlation [168, 169]: corruption is more prevalent in countries in which the responsibilities of government are concentrated in a few institutions. A theoretical

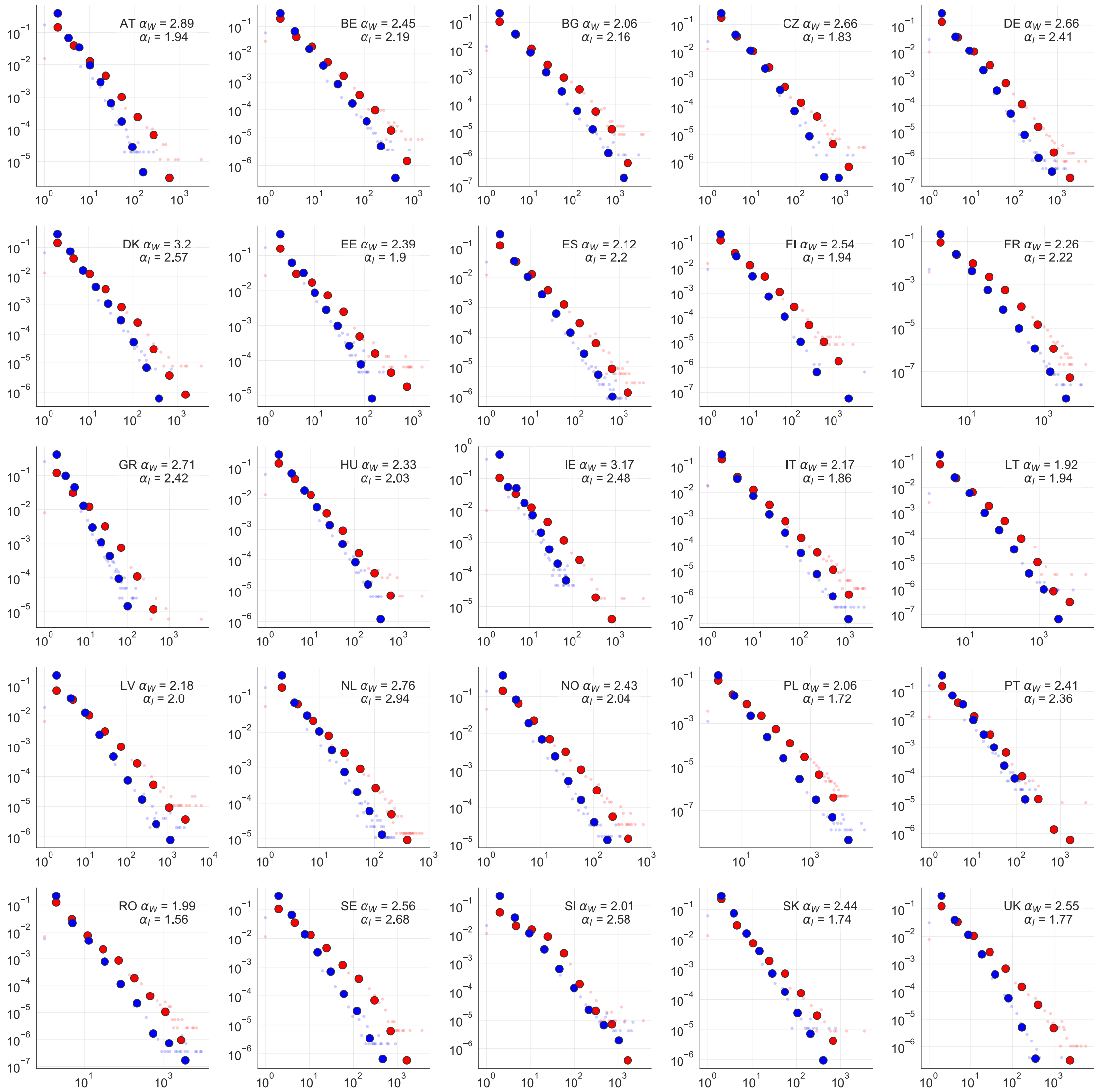

Figure 4.3. *The distribution of the number of contracts awarded and won, for issuers (in red) and winners (in blue), respectively, of procurement contracts in each country, aggregated from 2008 to 2016. The distributions are plotted on a log-log scale. We report the alpha parameter of a power-law degree distribution fitted to both distributions in the plot.*

model by Persson and Tabbellini proposes one possible mechanism, namely that centralization obscures political responsibility for corrupt outcomes [170]. In their model, politicians are reelected based on how well voters think they are doing their jobs. In a decentralized system responsibility is more closely tied to individuals, more directly linking good performance (abstaining from corruption) with rewards (reelection). In a centralized system it is harder for voters to observe which politicians are responsible for which outcomes. Yet Charron et al. find that within country variance of the EQI quality of governance indicator is not correlated with government centralization [44].

Going beyond the overall level of corruption risk in a country's procurement market, a decomposition of its participants into core and periphery allows us to test if corruption risk itself is centralized or decentralized in a given country. In the former case, the central government itself may be captured by corrupt interests [133], while in the latter case corruption may simply be an opportunistic action taken by less important actors. This is an important distinction which has been applied to the organization of corruption in organization [171] and among crooked police officers [172].

There are many methods to extract key nodes in a network. A popular approach which considers not only a node's individual importance, but that of its neighbors as well, is known as core decomposition [173, 155]. The general idea of core decomposition is to organize nodes of a given network into a hierarchical ranking according to their connectivity in an iterative way. Such methods have been used in biology [174], economics [175], and sociology [176]. We first introduce the concept of the and k-core of a monopartite, unweighted network. We then explain how to extend this concept to the particular case of weighted, bipartite networks, of which our mapping of public procurement markets are examples.

The core number of a node in a network is defined iteratively [173]. First we calculate the degree of each node. All nodes of degree 1 are assigned core number 1. We then remove all such nodes and recalculate the degree of all remaining nodes. In this new graph we assign core number 2 to any nodes with degree 1, and again remove them and repeat the procedure. The procedure ends when all nodes have been assigned a core number. The core number is a hierarchical ranking of the nodes which can be used to define the k-cores of a network. A k-core is the collection of nodes with core number at least k. Researchers are often interested in the maximal k-core of a network, defined as the highest value k for which any nodes have core number k [177].

One reason the k-core decomposition method is so popular is that there is an efficient algorithm to calculate the core number of all nodes in a network ($O(log(m))$ - where m is the number of edges in a network [178]. For a broad survey of network core decomposition methods and applications we refer to the survey paper by Malliaros et al. [179].

Applying a core decomposition method to public procurement markets require us to modify the definition of the original k-core algorithm in two ways: to consider the weights on the edges (encoding the frequency of the contracting relationship between an issuer and a winner) and the bipartite nature of the network. The latter factor is important because, as we saw in the previous section, the two node sets typically have distinct distributions.

In order to incorporate edge weights into our approach to finding the core of procurement market networks, we follow a similar approach Garas et al. in their extension of the k-core method to weighted networks [175]. Rather than iteratively pruning the network based on the degree of the nodes, we do the same using node strength, defined as the sum of weights on edges adjacent to a node. As all of the edge weights in our networks are integers (as they are counts of contract awards), we are able to repeat the same pruning process describe above for unweighted networks: remove all nodes with strength equal to 1, recalculate the strengths, and repeat - assigning the corresponding weighted core-numbers to each node as they are pruned.

This leaves a question: at what weighted k-core number do we consider a node to be a member of the periphery? By introducing node strengths based on heterogeneous weights, it is very likely that the maximal weighted k-core will be very small [175]. We also want to apply the same method to networks of different sizes and compare the results, suggesting that we need to consider a cutoff that is a function of the distribution of the strength of the nodes, for example its average. A final concern is that the distribution of node strengths is different for issuers and winners in a given network.

We address these concerns by using separate weighted core number cutoffs for the two node sets, namely their averages. In other words, a winner (respectively issuer) is considered to be in the core if its weighted-core number is greater than the average weighted degree of all winners (issuers) in the network. Instead of applying the same cutoff across all countries (and across node sets), this cutoff adapts to the size of the network and the distributions of connectivity observed in them. We visualize the core of the Hungarian procurement market over time in Figure 4.4, highlight connections with above average single bidding rates (relative to the whole market) in red.

It is worth noting several things from this visualization. The first is that our adaptation of the general idea of k-cores to bipartite, weighted procurement market networks returns relatively dense networks. In other applications of k-core methods the density of connections among the core nodes is often highlighted to indicate that these nodes are distinguished in the network not only because they have many connections, but because they interact with other such nodes [179]. Second, the size of the core seems stable, indicating that the method itself is not overly sensitive to perturbations in the networks from year to year. On the other hand the mesoscopic structure of the network does seem to undergo interesting changes over time - compare the modularity of the core

Figure 4.4. *The core of the Hungarian market from 2008 to 2016. Gray nodes denote issuers of contracts, while white nodes denote winners. Nodes are included in the core if they engage in many contracts with other highly active nodes, defined iteratively. Edges are colored red if the rate of single bidding on contracts between the issuer and winner exceeds the average single bidding rate observed in the whole market (including the periphery).*

in 2008 with its concentration in 2015, echoing earlier research on the centralization of Hungarian procurement since 2010 [21].

Finally, we note that the overall prevalence of high corruption risk edges in the core also seems to fluctuate. Recall that we color an edge red if the single bidding rate between the issuer and winner exceeds the total market average in that year. An increase in red edges in the core suggests that an increase in the frequency of single bidding in the core relative to the periphery. We first report summary statistics about the size of the cores in each country, averaged over the years in our data in Table 4.2.

In the rest of this section we probe these two aspects of the core of procurement markets: its size and the rate of single bidding of core contracts, relative to the whole market. The former allows us to quantify the centralization of procurement markets in a novel way, allowing us to revisit questions about the relationship between government centralization and corruption. The latter allows us to explore a potential heterogeneity in the prevalence of corruption across countries, namely if corruption risk is more common in the central actors of the procurement apparatus or in the periphery.

We plot the relationships between market centralization, measured as the share of all contracts which are between core issuers and winners, and both the single bidding rate, and the EQI QoG measurement in Figure 4.5. We find that the share of contracts in the core of the market predicts significantly higher corruption risk and lower quality of government. The correlation with the EQI quality of government measure is particularly high (.77). We interpret this as novel evidence in support of the theory that government centralization increases corruption.

In order to calculate the relative prevalence of single bidding in the core versus the periphery, we compare the observed relative prevalence of single bidding in the core to that of a null model with randomly shuffled single bidder labels. More specifically, we randomize single bidding at the contract level, then recalculate the rate of single bidding on contracts between core issuers and winners. In order to take into account market-specific effects (for instance markets for highly specialized services may have fewer firms, and so may have naturally higher rates of single bidding), we do not shuffle the single bidding labels freely across the entire market. We only permute labels within 2-digit CPV codes, which, as discussed in the previous chapters, are an EU-wide taxonomy of goods and services. Our measure of prevalence of single bidding in the core is the ratio of the observed single bidding in the core to the average of single bidding in the core in 1000 such randomized shuffles. In Figure 4.6 we highlight those countries with significant (at the 95% confidence interval of the 1000 randomizations) over or under-representation of single bidding in their cores.

Country	Core #Contracts	Share	#Winners	#Issuers	#Edges	%Single bid.
AT	696	0.21	111	25	168	0.14
BE	1680	0.25	188	92	382	0.14
BG	3923	0.44	162	61	1287	0.18
CY	390	0.42	32	3	41	0.39
CZ	2764	0.34	218	88	663	0.26
DE	6973	0.21	802	229	1827	0.16
DK	1342	0.27	147	45	323	0.11
EE	441	0.21	63	12	120	0.19
ES	7182	0.36	529	158	3206	0.18
FI	1712	0.27	167	50	661	0.15
FR	39462	0.32	2829	582	16345	0.14
GR	715	0.18	122	23	267	0.26
HU	1949	0.34	163	56	412	0.27
IE	682	0.24	91	9	111	0.02
IT	6958	0.38	462	156	1980	0.29
LT	5414	0.57	94	27	711	0.17
LV	4664	0.46	161	27	481	0.18
NL	1137	0.16	176	63	344	0.08
NO	552	0.15	87	28	233	0.06
PL	66963	0.61	896	455	14483	0.44
PT	653	0.26	67	20	116	0.22
RO	12087	0.59	165	106	2367	0.14
SE	1761	0.17	268	33	686	0.04
SI	2632	0.39	90	84	976	0.19
SK	934	0.34	68	28	144	0.37
UK	8511	0.25	1072	119	2058	0.08

Table 4.2. *The core statistics of each national market, averaged over 2008-2016. Core share refers to the share of overall contracts awarded that are between core issuers and winners.*

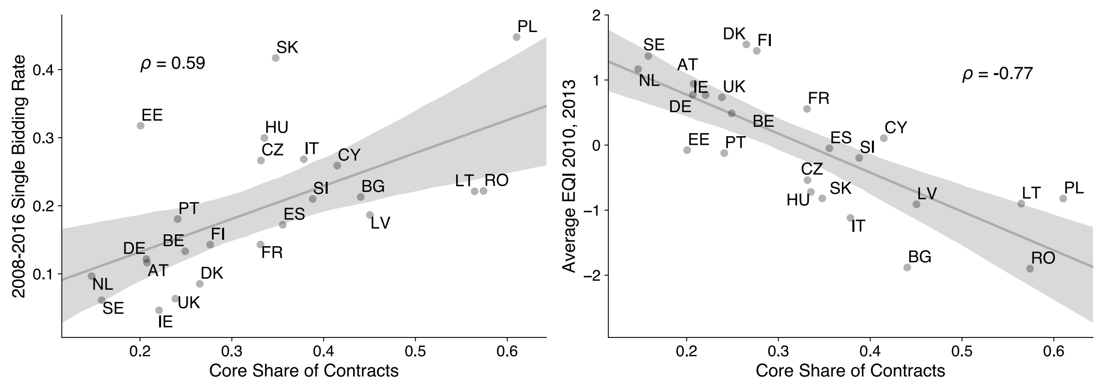

Figure 4.5. *Comparing the centralization of the procurement markets, measured as the share of all contracts which are between core issuers and winners, and measures of corruption risk at the national level. All measures are averaged over the years 2008-2016. We report Pearson correlations.*

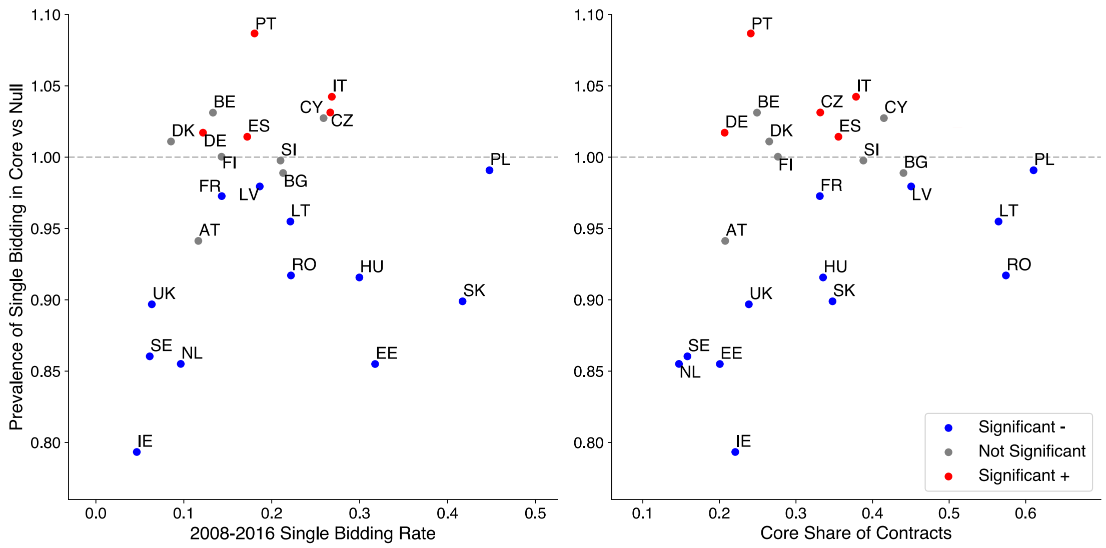

Figure 4.6. *Comparing the relative prevalence of single bidding in the core of EU procurement markets with their overall single bidding rates and the relative core sizes, respectively. Blue points are countries in which single bidding is less common among core contracts than expected under a sector-preserving null model. Red points are countries in which the opposite is true: single bidding is significantly more common in core contracts.*

We observe no clear linear relationship between a country's overall single bidding rate or its centralization and the tendency for single bidding to be more prevalent in the core or periphery. In Hungary and Slovakia, for instance, single bidding is more common in the periphery, while in Italy and Czechia, it is more common in the core of the market. In several countries including Bulgaria and Denmark, core rates of single bidding are not meaningfully different from the overall rate.

These findings contain important lessons for both policymakers and researchers of corruption. The concentration of corruption risk in the core or the periphery of a market suggests a very different organization of corruption in a society. Potential remedies likely depend on this distinction. For instance, if corruption is concentrated among core actors, as our findings suggest is the case in Italy and Czechia, checks on central government actors are likely lacking. When corruption is more common in the periphery, allocating more resources to policing corruption in local and regional governments or in geographically periphal areas may be more productive.

4.3.2 The Clustering of Corruption Risk in Markets

We now turn to a second question about the distribution of corruption risk in procurement markets: is corruption risk clustered? In other words: is corruption risk randomly distributed, or is it bunched up in certain parts of the network? We claim that answering this question is an important part of diagnosing the organization of corruption in a market and suggesting a remedy. If corruption risk is indeed clustered in specific parts of the market, targeted interventions in those areas will likely uncover useful information. In the case that corruption risk is randomly distributed, it would likely be more effective to reward whistleblowers.

How can we quantify the clustering of corruption risk in markets? We map this problem to a question of community detection, the process of grouping densely connected nodes in a network into modules [180], as we did for social networks in Chapter 3. Again we must adapt the methods to our specific circumstances: corruption risk is a property of edges, and not of nodes. If we were to group nodes and calculate the heterogeneity of corruption risk across groups based on node-level data, it is not clear how one would handle corruption risk on edges between groups. Moreover, these edges are distinguished as bridges between groups of nodes, and would likely differ in some qualitative way from their intra-group counterparts. Ignoring them would bias any measure of the heterogeneity of corruption risk across groups.

Luckily there are several tried and tested methods to cluster the edges in a network into communities. Such "link communities" were originally used to cluster nodes into overlapping communities, i.e. nodes are members of all communities that their adjacent edges are assigned to. Specifically, we use the methods proposed by Evans and Lambiotte [181, 182] and Ahn, Bagrow, and Lehmann [183], based on line graphs. The overall idea proceeds as follows: transform the graph in question into a new graph in which the edges of the original graph are now nodes, connected if they were adjacent in the original graph. Then by running a standard node-oriented community detection algorithm, such as the Louvain method described in the previous chapter [142], to group the edges of the original graph.

The line graph L_G of a graph G with vertices V and edges E is defined as a graph of nodes from E, connected if they have share a vertex from V. As edges adjacent to high degree nodes of the the graph G will be significantly overrepresented in L_G, Evans and Lambiotte suggest to weigh the connections in $L(G)$ inversely proportional to the degree of their shared nodes in G. In other words, if x and y are edges in G sharing a node v, hence nodes in $L(G)$, the edge connecting x and y is given a weight equal to $1/(deg(v)-1)$. This adjustment [5] is motivated by the desire to maintain the consistency of the behavior of random walkers between G and L_G.

Given the network associated to a national procurement market, we transform the network into a line graph, and then apply the Louvain algorithm to cluster the edges. This assigns each contracting relationship, and indeed each contract within that relationship to a community. We list the average of the yearly modularities, a measure of the quality of the partition - that is the tendency of edges in the network to be within rather than between communities - defined in the previous chapter, calculated for each country in our dataset in Table 4.3. In all cases we see very high levels of modularity, indicating that procurement markets typically have distinct submarkets. This is a reasonable finding considering that these markets include contracts for goods and services ranging from road construction to school milk to IT consultancy services, and likely have significant geographic influences.

As an example, we plot the 2014 Hungarian market in Figure 4.7. For simplicity of visualization we consider only those nodes in the giant component and involved in at least three contracts, and we color nodes by the most frequent edge community they are adjacent to. We color the edges red if the single bidding rate of contracts between those actors exceeds the market average that year. From this visualization we observe that single bidding seems to be clustered, particularly between actors in the top cluster. A manual inspection of the prominent issuers and winners in this group indicates that this is the health

[5]Evans and Lambiotte suggest a further adjustment in case the original graph G is weighted which we leave for future work.

Country	Edge Clustering Modularity
SI	0.58
LT	0.60
BG	0.62
RO	0.62
PL	0.64
ES	0.66
FI	0.67
FR	0.68
NO	0.69
SE	0.69
HU	0.70
DK	0.70
CZ	0.70
EE	0.70
LV	0.71
DE	0.72
NL	0.72
SK	0.73
IT	0.73
BE	0.73
PT	0.73
AT	0.74
GR	0.75
UK	0.75
CY	0.76
IE	0.79

Table 4.3. *The average modularity of the edge-clustering of each national procurement market from 2008 to 2016. Large numbers indicate a more fragmented market topology.*

and pharmaceutical industry in Hungary. This is in significant contrast with the relatively few significant single-bidding edges seen in the top right cluster, containing several prominent food and drink providers. In contrast to the strong relationship between market centralization, measured by the share of contracts awarded in the core of the network, and overall market corruption risk, we find no significant relationship between market modularity and corruption risk.

Motivated by this observation of cluster-level heterogeneity of single bidding, we calculate, for each country-year procurement market, a measure of the inter-cluster variance of single bidding. We first calculate a size-weighted coefficient of variation of single bidding rates across edge clusters in each network. Seeking to find an appropriate benchmark, we randomize single bidding across the network and recalculate the same coefficient of variation measure. As before, we do not randomize single bidding across all contracts blindly, rather we shuffle the single bidding label only within level-2 CPV codes, as we did with the benchmark for the concentration of single bidding in the network cores. Again we do this in order to allow for differences in the levels of competition between sectors.

Given a partition of the edges of the network, we naturally have a partition of the underlying contracts. We then calculate the coefficient of variation of single bidding rates across the clusters, defined as the ratio of the standard deviation (σ_{SB}) of single bidding to the mean of single bidding (μ_{SB}) across clusters. Because the clusters vary significantly in size, we weight the contribution of each cluster to the standard deviation and mean of single bidding as follows:

$$\sigma^W_{SB} = \sqrt{\frac{\sum_{c \in C} |c|(sb_c - \mu^W_{SB})^2}{\frac{(|C|-1)}{|C|}\sum_{c \in C} |c|}},$$

and

$$\mu^W_{SB} = \frac{\sum_{c \in C} |c| sb_c}{\sum_{c\ inC} |c|},$$

where C is the set of contract clusters, c is a specific cluster, and sb_c is the rate of single bidding in the cluster c. The weighted clustering coefficient, which in our context we refer to as the clustering of single bidding, is simply the ratio σ^W_{SB}/μ^W_{SB}.

As suggested, we wish to compare the observed clustering of single bidding against a plausible null model of within-sector randomized single bidding. For each market we randomize the single bidding label within the CPV-2 codes 1000 times and recalculate the same weighted coefficient of variation. We scale the observed clustering of single bidding against the average of these 1000 randomizations to arrive at a comparable measure of the extent which corruption

Figure 4.7. *The 2014 Hungarian procurement market. We plot the largest connected component of the network, filtering out nodes involved in less than three contracts for the sake of visualization. Nodes are buyers and suppliers of contracts, connected by an edge if they contract with one another. Edges are colored red if the single bidding rate on the edge exceeds the average rate of single-bidding that year. The node colors denote membership in the same edge community. For the sake of visualization we assign each node to the edge community most common among its adjacent edges. Single bidding is significantly over-represented among the edges in the top left cluster, consisting of pharmaceutical and medical contracts.*

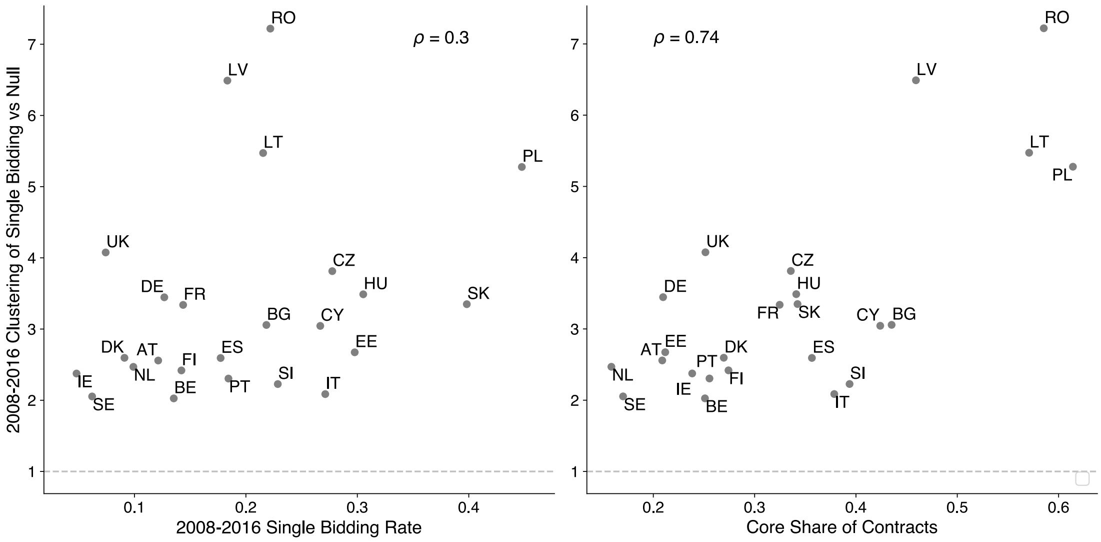

Figure 4.8. *The scaled clustering of single bidding in a market, compared with its overall single-bidding rate and centralization. Averaged over 2008-2016, with Pearson correlations.*

is clustered. We again average over all years for each country, reporting the resulting values in Table 4.4.

We plot the relationship between clustering of single bidding with both overall single bidding rates, and the relative size of the core of each country in Figure 4.8. We find only a weak relationship between the overall rate of single bidding and its tendency to cluster. We also observe a few interesting outliers: Romania, Latvia, Lithuania, and Poland have over five times the heterogeneity of single bidding across network clusters than expected under a conservative null model. Italy, with its relatively high rate of single bidding has one of lowest rates of clustering of single bidding. Among countries with lower single bidding rates, the UK, Germany, and France have relatively high rates of clustering, while Belgium and Sweden have relatively low rates. We find a stronger correlation between clustering of risk and centralization of procurement in the core of the network, driven mostly by the four outliers mentioned above.

We interpret these findings as indicating significant heterogeneity in the organization of corruption across the EU countries. In some countries corruption risk as measured by single bidding rates is highly concentrated in a few corners of the market, while in others it is much more evenly distributed. To underline the value of finding these heterogeneities, we claim that the right policy recommendations to counteract corruption depends crucially on understanding these patterns. As we saw in the visualization of the Hungarian market,

Country	Normalized Clustering of Single Bidding
BE	2.03
SE	2.05
IT	2.09
SI	2.23
PT	2.30
IE	2.38
FI	2.42
NL	2.47
AT	2.56
ES	2.59
DK	2.60
EE	2.67
CY	3.04
BG	3.06
FR	3.34
SK	3.35
DE	3.45
HU	3.49
CZ	3.81
UK	4.08
PL	5.28
LT	5.47
LV	6.49
RO	7.22

Table 4.4. *Average clustering of single bidding within edge-communities of EU countries, normalized by a randomized sector-preserving null model of single bidding. Higher numbers indicate that single bidding rates are more heterogeneous across clusters.*

one cluster clearly has higher rates of single bidding than do the others. Investigators, public or private, can immediately hone in on specific clusters in such a situation. In Italy, on the other hand, where the heterogeneity of single bidding across clusters is relatively low, a similar approach would have less value. There policymakers may be more interested in taking a decentralized approach to anti-corruption activities, for example by supporting whistleblowers and incentivizing the exposure of corruption by insiders.

4.3.3 Market Turnover and Change in Government

In the final section of this chapter, we analyze the effect of changes in government on procurement markets. This is another potential source of heterogeneity in the organization of corruption in different contexts. Past work has found significant evidence of political cycles in procurement in countries such as Russia [184]. Using the survival rate of procurement winners in the core, we can compare the turnover of frequent high single bid contract winners with other core firms across politically volatile and tranquil years. We find differences that again offer interesting theoretical insights and suggest different kinds of policy interventions.

We define change in government using data from the ParlGov database on cabinets [185], reported in Table 4.5. We say that a country experienced a change in government in a specific year, if the head of the government (for example the prime minister or chancellor) and all parties participating in the government changed. This is a rather conservative definition of change in government. For example, if a junior coalition party leaves the government or if a prime minister is replaced by a technocratic government with support by the same parties, we do not consider this a change in government. One drawback to this approach is that some countries, for example Germany and Austria, did not have, by our definition, a change of government in the span of our dataset. Despite our restrictive definition, some countries with more volatile political cycles (Italy or Czechia for example) had multiple changes of government. In any case, political turbulence is itself an important indicator of political health of a country, with past work indicating that political stability is correlated with lower levels of corruption [186]. On the other hand, extreme stability may indicate a lack of political competition and we observed in Chapter 3 that Hungarian settlements with weaker political competition had higher rates of corruption risk in their contracts. Though researchers have developed increasingly granular, social media-based measures of political volatility or "turbulence" [187], we stick with our top down approach.

Another limitation of our method of analysis is that we only consider central government changes. Many countries in Europe have significant federal

Country	Year
BG	2009, 2013, 2014
CY	2013
CZ	2009, 2010, 2013, 2014
DK	2011, 2015
ES	2011
FR	2012
GR	2009, 2012, 2015
HU	2010
IE	2011
IT	2006, 2008, 2011, 2013
LT	2012
NO	2013
PL	2007, 2015
PT	2011, 2015
RO	2008, 2012, 2015
SE	2006, 2014
SI	2012
SK	2006, 2010, 2012
UK	2010

Table 4.5. *Substantial changes of government in EU countries, 2006-2017. We code a change of government if the head of the government (i.e. chancellor, prime minister) changes, and if all parties involved in government change. Source: ParlGov database (Döring and Manow 2018)*

structures with significant devolution of political responsibilities. We mitigate limitation this by focusing our efforts on contracts in the core of the network, which are more likely to involve central government attention, directly or indirectly.

For each year from 2008 to 2014, we collect the core winners of each market, as defined in the previous section. We calculate their survival (as either a core or periphery winner of public contracts in that country) two years later. We use a two year gap to give room for potential changes of government to happen in the intervening year. We split the core winners into two groups: those with above and below average single bidding rates (where the average is calculated only among core contracts). We present the Hungarian case in Figure 4.9, observing a steep drop in the survival rate of core winners with high single bidding rates across the 2010 elections. This suggests a significant turnover of favored

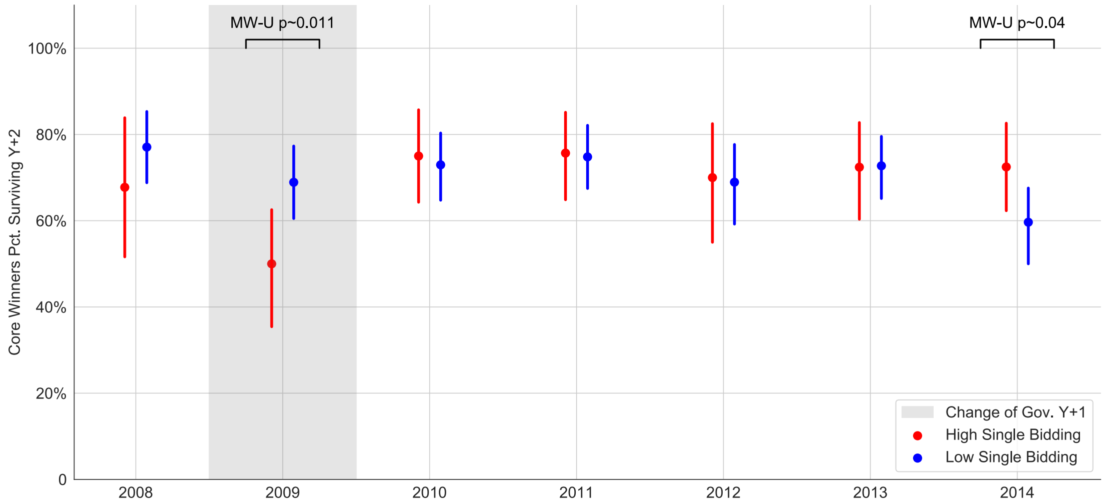

Figure 4.9. *The two year survival rates of Hungarian core winners, split up by above and below average single bidding rates. When significant, a Mann-Whitney U test of the difference in means is reported. 2009, the year preceding a change in government in Hungary is highlight in grey.*

firms in the Hungarian case, with particularist contract steering as a potential mechanism.

How do the other EU countries compare? We now outline a method to compare the differences in survival rates of winners with high and low single bidding rates across years of change of government and other years. Explicitly, we compare the ratio of the survival rates of high to low risk winners across change in governments, S^H_{CoG}/S^L_{CoG}, and the same ratio in other years: $S^H_{\neg CoG}/S^L_{\neg CoG}$. The ratio of these ratios can be interpreted as a kind of *survival premium* that high single bidder winners have across changes of government, relative to their low single bidder counterparts. We estimate the statistical significance of this premium using the bootstrap [188]. Specifically, we sample the survival outcomes of each of the four pools of winners (high and low single bidders and across or away from changes in government) with replacement and recalculate the ratio a thousand times for each country, generating a confidence interval for our estimate of the survival premium of high single bidder winners across changes in government. We interpret countries with the 95% confidence interval of the ratio entirely below 1 (as we shall soon see is the case for Hungary) as those countries in which corrupt arrangement are vulnerable to political turnover. Countries in which the interval is entirely above one have corruption organized in a way that is robust to political turnover. Finally, when the 95% confidence overlaps with 1, we say that corruption in the country is

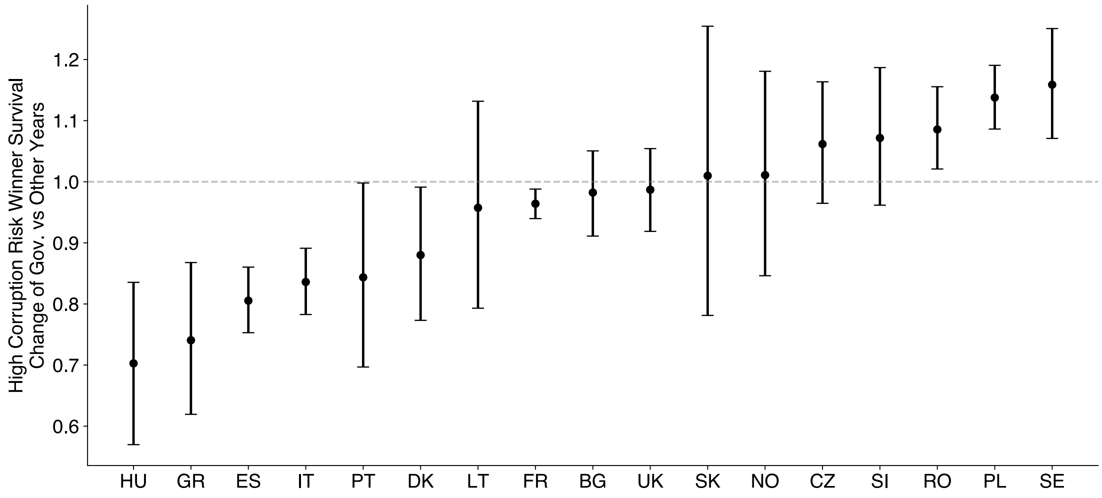

Figure 4.10. *The two-year survival premium of core winners with high rates of single bidding across change of government vs other years. Low values indicate that winners of relatively many single bid contracts are less likely to survive across a change in government than across other two year periods. High values indicate that such winners are rather robust to change in government.*

neither vulnerable nor robust to political turnover.

We plot the estimates and intervals for each country in Figure 4.10, sorted in increasing order. Echoing the previous sections we find some surprising differences between the countries in our sample. High corruption risk winners in Hungary, Greece, Spain, Italy, Portugal, Denmark and France (though the effect size of France is not large) have significantly more turnover compared to their low risk counterparts when there is a change in government than otherwise. The opposite is true for Romania, Poland, and Sweden. For the remainder of the countries, the effect is not significant.

That corrupt winners are robust to political volatility in Romania and Poland presents a puzzle. In the former socialist countries of Central Eastern Europe, corruption and state capture are often described as the results of an ideological competition between elites who seek to extract rents from the state, which Innes refers to as the “great electoral lottery” [189]. With such a framing one would expect more turnover of high corruption risk winners across changes in government, as the elite rewire the rent extracting connections between institutions and firms. An alternative framing could be that in Romania and Poland, corruption works along non-ideological lines, with corrupt actors simply engaging with whoever is in charge, and embedding themselves into the system in a way that ensures survival regardless of political outcome. This idea has some theoretical support in Romania, where MPs frequently and opportunistically switch parties [190]. Poland, however, is usually presented as a polarized society in which political competition is highly ideological [191]. Polish voters are especially likely to mobilize because of corruption perceptions [192]. This finding certainly merits further study.

On the other hand it is quite interesting that corruption risk implies increased volatility consistently among Hungary and the Mediterranean countries, in contrast with both the rest of Eastern Europe and Northern Europe. We briefly consider implications for policymakers given these heterogeneities. In countries in which corrupt firms are vulnerable to political turnover, it is likely that staggering local and regional elections may have a virtuous effect on the overall level of corruption, as suggested by previous work comparing the Czech Republic (where elections are entirely asynchronous) and Hungary (where elections are highly synchronized) [133]. By increasing volatility across levels of government, it makes it harder to organize corruption, for example because it is more likely that an important election is imminent at any given time and that voters can punish misbehaving political parties more quickly and consistently. In Romania and Poland, where we found the opposite tendency, it is rather likely that personal ties and brokers between firms and political parties are enabling the survival of high corruption risk winners across elections.

4.4 Discussion

In this chapter we studied the relationship between corruption risk and procurement market structure in EU countries. By observing centralization, clustering, and turnover of the market using network methods, we describe the organization of corruption in novel ways. Our approach highlights the multifaceted nature of corruption. With the exception of a strong correlation between centralization in a market and its rate of single bidding, we do not generally find clear relationships between the overall prevalence of corruption risk in a country and the structure of its procurement market. For example, in some countries corruption risk is more prominent in the core of the procurement market, while in others it is rather over-represented in the periphery. In all countries we found that corruption risk is distributed across sub-markets in a non-random way, but the extent to which this is true varies greatly. Finally, we found that winners of corruption contracts react very differently to changes in government in different countries around the EU.

Not only are these distinctions theoretically important, but they suggest practical implications for anti-corruption actors from activist to prosecutor. The network approach builds up from micro-level interactions to describe emergent structure - in this case highlighting significant national-level differences in procurement markets. They also suggest several avenues for further research. One could for instance focus on specific countries, just as this chapter has tended to use Hungary as an example. The hypotheses of country-level experts on the nature of corruption in specific societies could be tested using our framework. It could also be extended to compare non-EU countries, depending on the comparability of the data. Finally, we note that in future work a greater emphasis on the temporal evolution of these markets and their actors is merited. For instance, an analysis of market structure can give a much more real-time measure of change in corruption risk than can survey-based perception measures, which are often a lagging indicator of such effects. Such a perspective is crucial to discovering the causal mechanisms and directions behind the relationships we have discovered in this chapter, for if procurement market decentralization causes a decrease in corruption.

CHAPTER 5

CARTELS

Competing firms can increase profits by setting prices collectively, imposing significant costs on consumers. Such groups of firms are known as cartels and because this behavior is illegal, their operations are secretive and difficult to detect. Cartels do face a significant internal obstacle: members face short-run incentives to cheat. Here we present a network-based framework to detect potential cartels in bidding markets based on the idea that the likelihood that a group of firms can overcome this obstacle and sustain cooperation depends on the patterns of its interactions.

We create a network of firms based on their co-bidding behavior, detect interacting groups, and measure their cohesion and exclusivity, two group-level features of their collective behavior. Applied to a market for school milk, our method detects a known cartel and calculates that it has high cohesion and exclusivity. In a comprehensive set of nearly 150,000 public contracts awarded by the Republic of Georgia from 2011 to 2016, detected groups in the high cohesion and exclusivity region are significantly more likely to display traditional markers of cartel behavior. We replicate this relationship between group topology and the emergence of cooperation in a simulation model. Our method presents a scalable, unsupervised method to find groups of firms in bidding markets ideally positioned to form lasting cartels.[1].

This investigation contrasts with the previous chapters on corruption in that collusion is not a crime bridging the public and private sectors. In this sense the chapter and its findings represent a transfer of the network-based approach to the study of corruption to a related, parallel domain.

We do note several important similarities between "classic" corruption in

[1]This chapter partially draws on work that is under review at a journal at the time of submission of the thesis. No preprint is yet available.

procurement and collusion. The first is that both require coordination in secret. The second similarity is that both activities are illegal, so participants face similar dilemmas. Third, the two often coincide as cartel may develop insider connections with bureaucrats [193] - recall that the Petrobras scandal involved a cartel of firms on the private sector side. Finally, both activities impose costs on the public. Therefore, though collusion is not usually considered to be the same as corruption, a better understanding of the emergence and sustenance of illegal collusion can inform us about corruption as well[2].

5.1 Prelude

Cartels have been studied by economists since Adam Smith, who wrote that "people of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices." [194] Competition between firms decreases profits, and so they have incentive to collude by setting prices or production collectively [195]. Three significant forces make running such a cartel difficult. First, they are generally illegal and are aggressively policed by competition authorities, presenting a significant obstacle to coordination. The second challenge is that cartels face a classic collective action problem: individual cartel members have short-term incentive to deviate from the collusive agreement. In the language of game theory, defection is a dominant strategy in the one-shot game. Finally, effective collusion requires unanimity, meaning that a single defector can significantly diminish the profits of the other cooperating firms.

Even though cartels face these internal and external threats to their stability there are many examples of cartels in a variety of industries from financial products [196] to vitamins [197] operating successfully for long periods of time. Past research estimates that on average cartels increase prices by 20-30% [198] and that in a given year a functioning cartel has a roughly 10% chance of being discovered [199], indicating that cartels are costly and that deterrence can be improved. In this chapter we propose a novel approach to the issue of detecting cartels using network methods, built on the idea that the network of firm-firm interactions can reveal hot-spots in which cooperation is easier to sustain. Groups of firms in these distinguished positions interact repeatedly and exclusively amongst themselves, creating ideal conditions to overcome their internal collective action problem.

Economists have long studied the market conditions under which cartels thrive, namely how they enable a cartel to overcome the problems of "coor-

[2] A stand-alone version of this chapter is under review at a journal at the time of submission of the thesis.

dination, cheating, and entry." [200] For example coordination is easier in a smaller group with homogeneous firm sizes [201], while frequency of interaction facilitates punishment of defectors, and high costs to entry insulate the cartel from outsiders [202]. Another perspective on collusion is to consider that firms are playing a repeated prisoner's dilemma (PD) game [203]. When choosing to compete (equivalently defecting from the cartel in the language of the PD) firms charge low prices, while when they collude (equivalently cooperate), they charge high prices. Collective profit is maximized if everyone colludes, but when colluding firms are undersold by even a single competing firm they fare badly [204]. When players play multiple rounds of the PD, Axelrod demonstrated that cooperation can emerge as a winning strategy through learning and imitation [205], a finding which replicates in the context of firms setting prices in oligopoly [206]. In the PD and many other games in which collectively optimal actions are personally costly to players, altruistic cooperation emerges under a variety of conditions through mechanisms such as reciprocity and the altruistic punishment of defectors or cheaters [207].

Just as certain market conditions are known to favor cartels, researchers have observed that when players of the PD are arranged in some space or network which restricts their potential interactions, the potential for the emergence stable cooperation crucially depends on the structure of the space [208, 209, 210, 211]. For example, correlations in the spatial distribution of agents have been shown to facilitate cooperation [212]. To the best of our knowledge, this observation that local correlation of interactions have significant influence on the emergence of cooperation has not been applied to the problem of detecting cartels.

We propose to apply network science methods to identify groups of intensely interacting firms in a market and to screen them for collusive potential, based on their network topology and how it may facilitate collusion. We focus on the case of collusion in public contracting markets, in which public bodies buy goods and services from private firms. These markets are vulnerable to collusion because of the inelasticity and regularity of government's demand for certain goods [213, 214]. They are also large, accounting for between 10-20% of GDP in the OECD [20]. Contracts are commonly awarded via auction to the lowest bidder. In these markets cartels often engage in bid-rigging, coordinating their bids to mask their agreement to avoid competition [215].

Specifically, we use data on firms bidding for contracts to map in public contracting markets as networks of competing firms. We argue that such a network represents an embedding of the firms into a space which describes the competitive landscape of their industry or location, including its geography, technology, and scale. Within such co-bidding networks, we detect groups of firms whose local network topology are naturally conducive to sustaining collusion. Previous work on cartels has considered co-bidding networks of firms [216, 217, 218], though none have used the network topology to detect groups of firms within markets. In recent work on cartel screening, Imhof et al. have used patterns of bidding interactions between firms to study cartels [219]. More broadly, network methods have been fruitfully applied to problems in criminology including corruption [86, 14], the mafia [92], and the evolution of crime [97].

We first map the co-bidding market of the suppliers of public school milk in 1980s Ohio containing a known cartel case [220]. We note that the cartel firms occupy a distinguished position in the network: they frequently interact with one another, forming coherent links, and are relatively isolated from outside firms, forming an exclusive group. We then turn to a dataset of bids on nearly 150,000 contracts awarded in the Republic of Georgia from 2011 to 2016 worth roughly 5 billion US dollars. Using a greedy, bottom-up algorithm to detect overlapping groups of interacting nodes, we find that groups with cohesive and exclusive interactions have higher prices, are less likely to sue each other, and are more likely to have low variance in their bids and prices - classic screens for cartel behavior used by competition authorities around the world [221]. Finally, we simulate a market in which firms compete for randomly placed contracts when they are close in proximity, introducing spatial correlations to interactions. Firms see their competitors for the contract, and decide to compete or collude based on the previous actions of their partners and the frequency with which they meet them. In the resulting co-bidding network, detected groups with coherent and exclusive links successfully collude with much higher frequency.

5.2 Results

Our framework to find groups of firms that may be engaging in collusion consists of several steps. First we extract the co-bidding network of firms in a market, connecting two firms by an edge with a weight that increases as they more frequently bid for the same contract. We then identify groups of firms which frequently bid for the same contracts using a modified version of a popular overlapping community detection algorithm [222]. The method is greedy, and the function to merge nodes into groups has a penalty term for the number of nodes included, insuring that the groups detected remain small relative to the size of the market. Finally, we calculate topological features of the groups: their *coherence* and *exclusivity*. We suggest sustained collusion is more likely to emerge among high coherence and exclusivity groups because they offer the ideal conditions for firms to learn to cooperate and trust one another. We find evidence of this phenomenon in three settings: a dataset of school milk contracts with a known cartel, a dataset of virtually all contracts awarded in the Republic of Georgia over several years, and in a simulation model of contracting markets with spatial correlations.

5.2.1 The 1980s Ohio School Milk Market

We first analyze bidding data from the market for public school milk in 1980s Ohio [220]. Every summer school districts called for bids from dairies to provide school milk for the following academic year. Firms submitted sealed bids quoting a price in cents per pint. In 1993 representatives from two firms confessed to colluding with a third firm to rig bids for contracts in the Cincinnati area as part of a settlement. The third firm eventually settled out of court, paying significant civil penalties.

Previous work by Porter and Zona highlights irregularities in the bidding behavior of the suspected cartel firms compared to the rest of the market [220]. Exploiting specific features about the market for school milk, the authors created an econometric model to predict the bids of firms on contracts, including information on the capacity of firms, the specifications of the bids (i.e. whether drinking straws were required), and the physical distance between the firm and school. They found that the bids submitted by cartel members were often decreasing in distance - a highly suspicious fact given that a major cost in the supply of school milk is its transportation.

For each year from 1981-1990, inclusive, we created the co-bidding network of firms, connecting two firms based on the similarity of their bidding behavior. We apply our method to detect overlapping groups of interacting firms. We use a force layout algorithm to visualize the network in subplot A of Figure 5.1, highlighting the cartel firms in red and outlining the detected groups. For each group we calculate its coherence, the ratio of the geometric to arithmetic means of its edge weights [223] and its exclusivity, the ratio of strength within the group to the total strength of nodes in the group (including edges leaving the group). As features of groups of firms, coherence captures the consistency and intensity of interactions among firms in the group, while exclusivity quantifies the extent to which group interactions happen in isolation from the rest of the firms in the broader market.

We plot the distribution of groups across all ten years in the coherence-exclusivity space in subplot B of Figure 5.1. In the first plot we show the distribution groups detected in 100 null models for each yeah in which bidding behavior was randomized. Specifically, the null model shuffles bidders between contracts, such that each firm bids on the same number of contracts and each contract receives the same number of bids. In the second plot we show the ob-

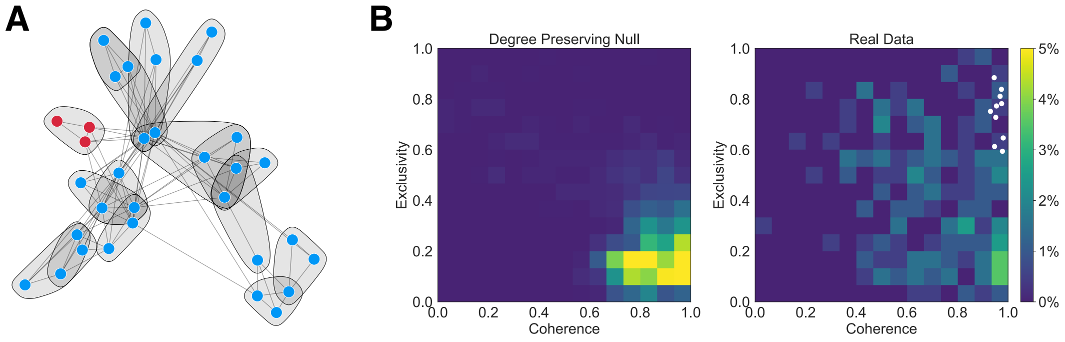

Figure 5.1. *A. Ohio school milk market co-bidding network, 1986. Overlapping groups are detected using our algorithm. Red nodes are member of the alleged cartel operating near Cincinnati. We exclude firms participating in less than 3 auctions for the purposes of visualization. B. Coarsened two-dimensional histograms of groups detected in all Ohio school milk networks 1980-1990 in the coherence-exclusivity space. The first plot shows the distribution of the groups detected in 100 bid-degree preserving null models of each of the 10 years. The second plot shows the real distribution of groups. The cartel group's position is marked by white circles. The cartel group has both high exclusivity and coherence.*

served distributions, indicating the position of the cartel firms (which our group detection algorithm identified as a group in each year) with white circles. We note two phenomena: the first is that groups in the empirical network have significantly higher exclusivity and coherence than what would be expected if the bids were random, while the second is that the high coherence and exclusivity regime is sparsely populated in both the empirical data and the null model.

5.2.2 Georgian Public Procurement Markets

We now turn to data from a much larger procurement market covering a wide range of goods and services. Specifically we collected virtually all public contracts from the Republic of Georgia from 2011 to 2016. The data consists of nearly 150,000 contracts bid on by nearly 15,000 unique firms with total value roughly five billion US dollars. As with the Ohio dataset, we observe the bids and bidder identities for each contract. We again apply our method to detect overlapping groups and calculate their coherence and exclusivity for each year. In the analysis that follows we consider only groups of firms identified from the co-bidding network that exclusively bid on at least 30 contracts in a given year in order to focus on significantly interacting firms. Our findings are robust to

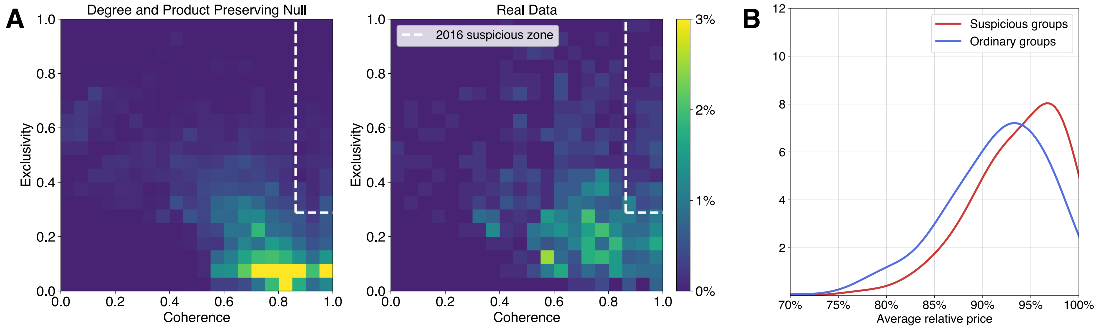

Figure 5.2. *A. The distribution of groups in the cohesion-exclusivity feature space detected in a product-type and bid-degree preserving null model compared with groups detected in observed data from Georgian procurement markets from 2011-2016. We label groups of firms as suspicious if its coherence and exclusivity are in the 80th percentile of the null model outcomes of the year in which they are detected. We highlight 2016's suspicious zone in white. B. Distribution of average relative prices of contracts bid on by suspicious groups and ordinary groups, 2011-2016. Suspicious groups consistently win more expensive contracts.*

consider a range of cutoffs, which report in the SI.

To compare our data against a plausible null model, we created randomized networks from data by shuffling the contracts firms bid on within specific product classes. This insures that firms bidding exclusively on school milk contracts do not bid on software consulting contracts in our null model. We use the resulting distributions of group cohesion and exclusivity from the null model to create thresholds for labeling groups from the empirical network as suspicious. We consider a group from the empirical co-bidding network suspicious if its coherence and exclusivity exceed the 80th percentile of coherence and exclusivity of groups in the null model in the same year.

We visualize the distributions of groups in the coherence and exclusivity space for the randomized and empirical data in subplot A of Figure 5.2. We plot the data from all years in the same visualization, and highlight the suspicious zone of high coherence and exclusivity calculated for 2016.

As there are no confirmed ground truth cartels in our dataset, we validate our claim that groups in the suspicious zone are operating under conditions that facilitate collusion using four measures. First we consider the average cost of contracts won by the group when they were the only participants. As contracts are announced with a reserve price, we can scale each contract's cost outcome to enable comparisons between contracts. We plot the distribution of relative prices for contracts won by suspicious groups versus all other groups in subplot B of Figure 5.2. We confirm that groups in the suspicious zone are winning more expensive contracts, confirmed by a Mann-Whitney U test shown in Table 5.1.

Next, we calculated two price and bid based screens for collusion from the literature. The first is the price coefficient of variation of contracts won by the group [221], measuring the extent a group's prices are both high and stable. This screen is based on the theoretical observation that when prices are set collectively, it is costly to coordinate price changes [224]. It aligns with empirical observations of real cartels [225] and has been used extensively by competition authorities [196]. Specifically, the price coefficient of variation CV^G_{price} of a group G is defined in terms of the average cost of contracts C cornered by the group, μ^G_C, and the standard deviation σ^G_C:

$$CV^G_{price} = \frac{\sigma^G_C}{\mu^G_C}$$

The second cartel screen we apply is the average of the coefficient of variation of bids on each contract for which only group firms submitted bids [226]. Previous research has shown that the fake bids submitted by losing members of the cartel tend to closely hug the winning bid. For each contract c bid on exclusively by members of a group, we calculate the coefficient of variation in the bids:

$$CV^c_{bidding} = \frac{\sigma_c}{\mu_c}$$

We average over all contracts C cornered by a group G to obtain its bidding coefficient of variation:

$$CV^G_{bidding} = \mu_{c \in C}\left(CV^c_{bidding}\right)$$

We say that a group of firms has a low bidding coefficient of variation if it is less than one standard deviation below the market average. A Mann-Whitney U test, shown in Table 5.1, indicates that groups in the suspicious zone are significantly more likely to have lower $CV^c_{bidding}$ and CV^G_{price}.

We carry out one more test of our method using data on bid protests. Bid protests are legal actions by firms against contracts awarded by procurement authorities. Firms can protest, for example, if the contract was not advertised in the proper venue, or if they believe criteria to participate in an auction unfairly excluded them. We collected data on which firms protested which contracts, including the firm to which the contract was awarded. We argue that colluding firms would never protest the contracts won by their cartel partners, while one may expecting intensely competing firms to frequently protest each others' winnings. For each group we check if any contract awarded to a group member was protested by another group member that year. We find that suspicious

	Suspicious Groups Mean (St. Dev.)	Ordinary Groups Mean (St. Dev.)	Differences MW U (p-value)
Avg. Rel. Price	0.938	0.914	30211***
	(0.046)	(0.053)	(p<0.001)
Avg. CV^{G}_{price}	0.098	0.117	33470***
	(0.055)	(0.059)	(p<0.001)
Avg. $CV^{G}_{bidding}$	0.047	0.055	32306***
	(0.056)	(0.038)	(p<0.001)
Bid Protest Rate	0.134	0.237	37516*
	(0.341)	(0.425)	(p≈.011)

Table 5.1. *Cartel screens applied to suspicious and ordinary groups of firms detected in the Georgia procurement market, 2011-2016. Cartel groups have higher average relative prices, are more likely to have a low average coefficient of variation on bids for a contract, and are less likely to legally protest the winnings of other group members.* * $p < .05$, ** $p < .01$,*** $p < .001$

groups are half as likely to have such internal protests - a statistically significant difference shared in Table 5.1.

Suspicious groups detected by our methods are more likely to manifest the four collusive markers we have measured than their non-suspicious counterparts. Though this is no proof of collusion, it does indicate that many of the groups of firms that competition authorities might be interested in investigating based on their behavioral patterns exist in the same high coherence and exclusivity zone as the Ohio school milk cartel. In the next section we present a simple simulation model of a procurement market with spatial correlations which replicates our observation that collusion is more common among cohesive and exclusive groups.

5.2.3 Simulation Model

We simulated a market of interacting firms placed uniformly at random in the unit square. The location of firms can be interpreted as their physical location or as a more abstract position in a space of product similarities (for instance firms supplying computer hardware might be closer to one another). Contracts, also located randomly, attract bids from nearby firms, introducing spatial correlation to the interactions between firms. We assume that firms participating in an auction know the other participants. Each firm must decide whether to cooperate or compete for the contract using two factors: the firm's memory of the previous action of the other firms, and the frequency by which they have met the same firms in the recent past.

For the first factor, the focal firm recalls the previous decision made by the other firms it is meeting using a proportional tit-for-tat strategy [205]. The second factor increases the likelihood of cooperation when the other firms have been. The decision to collude depends on the product of these two: familiarity and experiences of reciprocity are essential to start and sustain collusion [227]. In order to keep the model as simple as possible, we do not introduce a price mechanism or consider who wins a given auction. We seek to demonstrate that random spatial correlations can create environments with locations heterogeneously favorable to collusion. We present the precise parameters and initial conditions in the section on data and methods.

We simulated 5000 instances of our model, each time initializing a new market with randomly placed firms and contracts. In each instance we award 2,000 contracts, discarding the data from the first 1,000 contracts as burn-in. As before, we constructed the co-bidding networks of firms, detected groups in them, and plotted their distributions in Figure 5.3, subplot A.

For each group, we calculated the rate at which members unanimously cooperated on a contract, in other words the relatively frequency of successful collusion among the group. We plot the distribution of this frequency across the coherence-exclusivity space in subplot B. In agreement with our empirical evidence, we find that collusion is significantly more likely to emerge among groups in the region of high coherence and exclusivity.

As discussed earlier in the article, economic theory and empirical observation suggests that there are certain environments in which cartels are more likely to emerge [200]. Inspired by the literature on evolutionary game theory, we considered a simple model of cooperation based games played between agents embedded in space [212]. Our findings support the notion that the co-bidding network captures localized market conditions, which in turn govern the likelihood and effectiveness of emergent cooperation. Interestingly, in the case when a market is significantly governed by the locations of firms in physical space, for example in the Ohio milk market, our model has the potential to be calibrated with geographical data.

5.3 Discussion

In this chapter we developed a framework to find groups of firms in public contracting markets and to screen them for collusive markers. Testing our method on a ground truth case, a large scale market without known collusion, and a simple model of such markets, we find that collusion seems more likely to emerge among groups of firms with cohesive and exclusive interac-

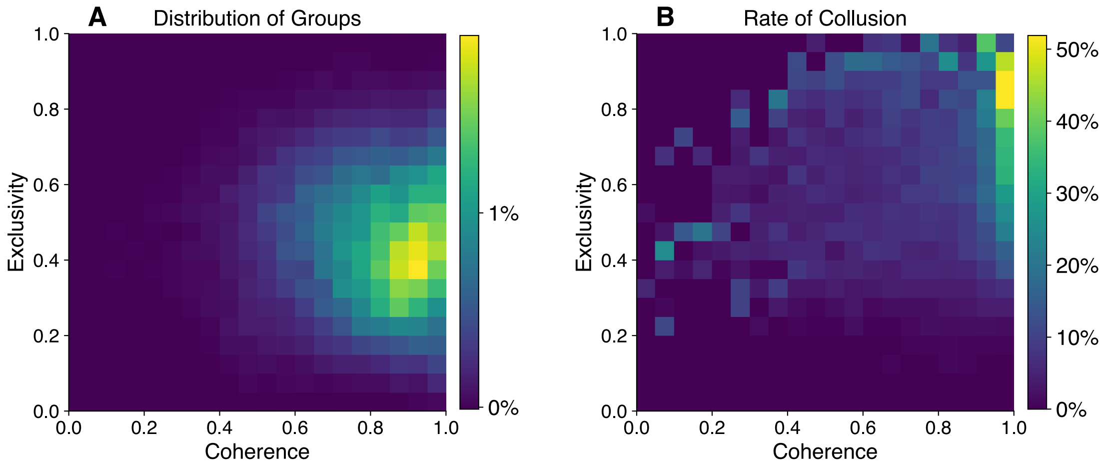

Figure 5.3. *Simulation model results over 5000 simulations. A) the distribution of groups observed from the resulting co-bidding networks in the coherence-exclusivity feature space. B) the rate of collusion by groups with given coherence-exclusivity. The model suggests that high coherence and exclusivity groups are not common, but that they have significantly higher rates of cooperation.*

tions. Groups occupying such distinguished places in the broader market have found a niche with conditions ripe for the emergence of cooperation.

We must acknowledge that our approach to cartel detection is only suggestive - it cannot prove that a group of firms are engaged in collusion. Rather we propose that our method be used to narrow down a large space of possibilities, into a shorter list of candidates for investigation. Authorities can then apply classical screens for evidence of illegal cooperation [228, 12], for example by observing abnormal stability in prices or market shares [221, 226], or by comparing observed behavior against a model of competitive behavior [229, 213]. More granular data required for these tests for collusion can be collected once a key subset of firms is identified, at significantly lower cost. Another advantage of our approach is that it does not rely on information from whistleblowers to highlight a candidate group of firms, avoiding a potential source of bias in the cartel literature [230]. We also acknowledge that there are other cartel strategies in public contracting markets beside bid-rigging, for instance when firms agree to stay out of each others' markets entirely.

In our model we do not consider the idea that some firms might simply be honest and refuse to form a cartel even in optimal conditions, nor do we consider how fear of prosecution might influence the choice to collude, for instance in our model. Though the illegality of collusion adds an additional obstacle to the emergence of cooperation among firms, the empirical observation that cartel life spans are heterogeneous suggests that many firms are willing to collude, but that only certain environments are conducive to cartels [200].

In spite of the limitations, we note that our method can be applied to other questions about cartels. For example, what does the co-bidding network look like near a cartel when it is born compared to when it dies? The inner-workings of potential cartels would surely be reflected in network structure of the market. Observed cartels have operated by methods such as rotating the winner [231], by side payments to losing firms [213], and some even run internal auctions to optimize their profits [232]. Observing the relationship between the procedure by which contracts are awarded, for example to the trimmed-average bidder in Italian road contracts [233] or by randomly chosen open or sealed bid procedures in timber auctions [234], and network structure may also reveal whether firms are competing or colluding. The specifics of a market and manner by which contracts are awarded matters a great deal to how collusion might evolve [235]. Certain rules make it easier for firms to collude or easier to detect collusion [233, 234].

We are confident that our approach can be applied to these cases in which we have extra information about the rules of a market. It is likely still the case that certain patterns of interaction are effective markers of collusion and that networks provide a useful map of such interactions.

5.4 Methods and Data

5.4.1 Co-bidding networks, group detection, and group features

We define a public contracting market's *co-bidding network* as a projection of a bipartite network onto the set of firms active in the market. Specifically, we form a bipartite network of contracts and firms bidding on them, then create a network of firms which bid for the same contract. We weight the connections based on the similarity of co-bidding behavior between firms using Jaccard similarity. Specifically, firm A and firm B are connected by a link with weight equal to the overlap of the contracts they bid on:

$$w_{A,B} = \frac{|c_A \cap c_B|}{|c_A \cup c_B|},$$

where c_A (c_B) is the set of contracts of A (B) with at least one other bidder and $|\cdot|$ is the cardinality of a set.

Given a co-bidding network our aim is to extract groups of nodes which may be analyzed for cartel activity. Groups should be communities in the network sense: there should be more interactions within the group than leaving the group. The case in question suggests several other criteria for our algorithm. Groups should be small, as cooperation becomes more difficult to sustain with more participants. Firms might be present in more than one part of the market, so we should consider overlapping groups.

We adapt a bottom-up method for community detection which merges nodes into groups by local optimization of a fitness function from previous work by Lancichinetti, Fortunato, and Kertész (hence: LFK) [222]. We define the fitness f_G of a group of nodes G in a network as:

$$f_G = \frac{s_{in}^G}{\left(s_{in}^G + s_{out}^G\right)^\alpha \times |G|^\beta},$$

where s_{in}^G and s_{out}^G denote the strength (the sum of weights) of edges within the group and adjacent to the group, respectively. $|G|$ is the size of the group, and α and β are free parameters which control the size of the groups found. When α is increased, additional strength is penalized, while β penalizes the number of group members independently of their strength. We set both parameters to 1.5. Increasing α insures that new nodes added to a group interact primarily within the group, while increasing β restricts the size of the groups we detect, in line with the stylized facts about cartels from the economics literature that lasting cartels are small and frequently interacting [200]. We report the sizes of the groups found in the empirical cases in the SI.

Given such a fitness function of a group of nodes in a co-bidding network, we can define the fitness of a node n relative to a group by calculating the difference in fitness of the group with n and without it:

$$f_G^n = f_{G+\{n\}} - f_G$$

With this node-level measure of fitness we can define our group detection algorithm. For each node in the network:

- select a node n and initialize a group containing only n,
- select the neighbor of n with the largest fitness and, if it has positive fitness, add it to the group.
- repeat until no nodes adjacent to the group have positive fitness.

In this way we find groups in the network which are overlapping, small (tuned by the parameters), with more weight among themselves than with non-group members. It is possible to save significant computational time by initializing new groups only for nodes that have not been included in a group before. In contrast with the LFK method we do not recalculate the individual fitness of all nodes in the group following the inclusion of a new node. In this sense our adaptation is greedy and not iterative, saving computational time.

Once groups have been extracted from a market's co-bidding network, we then define topological features of each group that may suggest that the firms could form a cartel. The first measure is the *coherence* [223] of a group C_G, the ratio of the geometric and arithmetic means of the edges weights among group members, measuring the balance and overall frequency of interactions among the group members:

$$C_G = \frac{\left(\prod_{l_G} w_l\right)^{1/|l_G|}}{\frac{\sum_{l \in G} w_l}{|l_G|}}$$

The second measure is *exclusivity* , the ratio of strength within the group over the total strength of the group, excluding on edges to non-group members, measuring the group's relative isolation in the broader market:

$$E_G = \frac{s_{in}^G}{\left(s_{in}^G + s_{out}^G\right)}$$

5.4.2 Null models

In both empirical cases, we created null models of the market to capture the extent to which groups of certain cohesion and exclusivity emerge by chance. For the Ohio school milk data we shuffled the bidders across all contracts, preserving the number of bidders each contract received, and the number of contracts each firm bid on. In Georgia we repeated the same procedure with an additional restriction: firm bids were only shuffled among contracts with the same 2-digit Common Procurement Vocabulary (CPV) code [132]. CPV codes describe the type of good or service being contracted, from road repair to medicine. By restricting the random shuffling of bids by CPV code, we create a randomized version of the broader market which preserves the tendency of firms providing similar products to interact.

5.4.3 Agent Based Model

In this section we describe the specific parameters of our simulated model of a spatially embedded contracting market. Each simulated market was initialized with 50 firms and 75 issuers of contracts (analogous to school districts) placed uniformly at random in the unit square. We then play 2,000 rounds corresponding to contract awards.

In each round a randomly selected issuer releases a contract C placed nearby (at a position drawn from a 2-d normal distribution centered on the issuer with standard deviation .3). Firms participate in the competition for the contract if they are within .1 distance of the contract (if no firms are close enough, the distance for inclusion is extended by .1 repeatedly until at least one firm participates). The set of firms participating, F, is known to all firms.

Each firm must then decide to collude or compete. Collusion is successful if all firms collude. Each firm f considers two pieces of information about the other firms in its decision making process, its *memory* of previous interactions with each other firm, and the relative *frequency* with which it meets with the others. It recalls the decision made by the other firms the previous time they met (initialized randomly) and calculates the share of previous round cooperators:

$$f_{memory} = \frac{\sum_{\hat{f} \in F \setminus f} \delta_{\hat{f}}^{C_{prev}}}{\|F \setminus f\|},$$

where $\delta_{\hat{f}}^{C_{prev}}$ equal 1 if $\hat{f}$ cooperated the last time it encountered f. This is the proportional (compared to the absolute) generalization of the tit-for-tat strategy to multi-agent games [205].

Next, f considers how often, in the last k contracts it was participating in, the current other firms were a subset of the participating firms. If this is true at least two-thirds of the time, the firm considers the other firms it meets as familiar.

$$f_{frequency} = \begin{cases} 1, & \text{if } \dfrac{\sum_{i=1}^{k} \delta_{F \subset F(C_f^i)}}{k} \geq 2/3 \\ 0, & \text{otherwise} \end{cases}$$

where $F(C_f^i)$ denotes the firms participating in the i'th previous contract of firm f. $f_{frequency}$ increases as f tends to meet the same firms. We set k to 10.

The focal firm's decision to collude or compete depends on the product of these two factors:

$$f_{compete} = \begin{cases} 1, & \text{if } f_{memory} * f_{frequency} \geq .9 \\ 0, & \text{otherwise} \end{cases}$$

Finally, we add noise to the system by allowing a .1% chance that a firm spontaneously colludes. In our model agents do not learn or track the outcome of their actions - they only react to their most recent memory of other firms and the frequency by which they meet. After 2,000 contracts are awarded, we end the simulation and discard the outcomes of the first 1,000 contracts as burn-in.

5.4.4 Datasets

The Ohio school milk data was generously provided by Porter and Zona [220]. The data consists of a significant share of all school-milk procurement contracts from 1980s Ohio provided to Porter and Zona by the State of Ohio. Porter and Zona served as expert witnesses in a trial against the suspect cartel. There are several other significant examples of cartels in public school milk markets in the US during the 1980s, for example in Florida and Texas [213, 236].

We collected the Georgian contracts dataset from the centralized procurement portal of the State Procurement Agency (SPA) of Georgia[3], including all contracts awarded through the portal between 2011 and 2016. Contracts are awarded to the lowest bidder in a sealed-bid auction. Each contract includes a product category (CPV code [132], which we use for the null model, and a reserve price, the maximum price that the public buyer would pay for the good or service, which we use to normalize prices. The procurement portal also reports bid protests: these are legal disputes of participants in the procurement process against the agency issuing a contract. For example, a firm may protest that it was unfairly excluded from the competition for a contract.

[3]https://tenders.procurement.gov.ge/

CHAPTER 6

CONCLUSION

This thesis set out to demonstrate the utility of a network perspective on corruption. Noting that previous approaches to the study of corruption are either often overly atomic or overly structural, we framed corruption as an emergent phenomenon occurring between actors within the networks of their interactions. We leveraged the recent proliferation of data on public procurement to create transaction-level measures of corruption risk, which we embed in social and economic contexts. Our approach is validated by the observation that bad behavior at the dyadic level (i.e. between a corruption issuer and winner or among colluding firms) is related to and to some extent predicted by the network the dyad is embedded in [19].

We have seen that social networks in a place relate to the prevalence of corruption in its local government. The findings of Chapter 3 suggest that certain modes of social organization facilitate corrupt behavior. These modes are increasingly observable and measurable at the level of societies because of the vast amount of data created by the use of modern telecommunication services. This suggests why corruption is a stubborn phenomenon, and why most interventions don't work.

At the national level, we saw in Chapter 4 that corruption risk is distributed in very different ways across EU countries, nearly independent of the overall level of corruption risk observed in a country. In some countries with high levels of corruption risk, corruption is significantly concentrated in the relationships between core issuers and winners, while in other countries corruption is more concentrated among peripheral actors. Corruption risk tends to cluster in all countries in our analysis, though the extent to which it does varies greatly. Finally, we also observe significant heterogeneity in the response of corrupt relationships to political shocks. These differences have significant implications for anti-corruption strategies, casting doubt on global solutions. Indeed one could say that while countries with little amounts of corruption are similar, each highly corrupt country is corrupt in its own way.

Finally, we presented a method to detect potential hot-spots for the emergence of illegal collusion in competitive markets using data on bidding. Using two empirical cases and a simulated model we showed how groups occupying specific positions in the co-bidding network topology are uniquely able to sustain the cooperation needed to maintain a cartel. Our framework demonstrates a way to reduce the vast complexity of a market to highlight groups of firms worth investigating.

Given previous work on the networked nature of criminal conspiracies, it is perhaps no surprise that a study of corruption and collusion in procurement markets found significant relationships between network structure and bad behavior. We argue that the specific methods developed and analyses carried out offer actionable insights into how corruption works in different contexts. They also offer a kind of blueprint for future analyses, showing how procurement data can be analyzed using network methods.

Future Work

The future for data-driven anti-corruption research looks bright. Data quality and access are generally improving. New sources of data will go a long way to addressing some of the shortcomings and limitations of the work in this thesis. For example, data on company owners and board members and the economic and social relations of politicians and regulators have potential to extend the scope of the work presented in this thesis. Network methods are clearly applicable in these contexts as well. For instance, by tracking social network connections of firm leaders to people in power, one could measure the extent of political corruption in a country by the effect of such connections on profitability.

A major challenge in corruption research using big data will be to carry out causal inference. Though data collected at large scales has many advantages, for instance allowing us to observe whole markets across significant periods of time, it seems unrealistic to carry out experimental studies at the same scale, especially without the participation of government bodies. Further research is needed to extend methods of causal inference, for instance as often applied to panel data by economists, to the setting of networks.

In the absence of causal identification, the scientific value of the methods developed in this thesis can be tested in other ways. If network analyses of procurement based risk indicators can predict corruption cases that authorities, researchers, or journalists can confirm, that would lend additional credibility to our approach. One can also strengthen the validity of these methods by finding evidence of other kinds of white-collar crime occurring among distinguished actors, following the classical adage that “where there is smoke, there is fire”.

The involvement of governments in anti-corruption research presents both opportunities and dangers. Public actors willing to experiment with rules or enforcement can help overcome problems of causal interpretation inherent in the study of observational data. Such collaborations also have the greatest potential for real-world impact, clearly. However, if researchers of corruption focus their attention too much on such collaborations, they would introduce a significant bias to our understanding of corruption. Indeed, where corruption is endemic, it is unlikely that relevant government bodies would be willing to participate in effective anti-corruption studies. Engagement with super-national actors with some independent authority in certain locations, for example with the World Bank in its procurement-based development projects or the EU and it cohesion funds, would overcome this issue to some extent. Certainly, the first step in this direction would be to convince individuals in these organizations of the value of network methods in the study of corruption. We hope that this thesis will be useful in this regard.

BIBLIOGRAPHY

[1] Mauro, P.: Corruption and growth. The Quarterly Journal of Economics **110**(3), 681–712 (1995)

[2] Rodríguez-Pose, A., Di Cataldo, M.: Quality of government and innovative performance in the regions of europe. Journal of Economic Geography **15**(4), 673–706 (2014)

[3] Gupta, S., Davoodi, H., Alonso-Terme, R.: Does corruption affect income inequality and poverty? Economics of governance **3**(1), 23–45 (2002)

[4] Rothstein, B.: Corruption and social trust: Why the fish rots from the head down. social research **80**(4), 1009–1032 (2013)

[5] Lambsdorff, J.: How corruption in government affects public welfare: A review of theory. Technical report, CeGE Discussion Paper (2001)

[6] Nye, J.S.: Corruption and political development: A cost-benefit analysis. American political science review **61**(2), 417–427 (1967)

[7] Rose-Ackerman, S.: Corruption: A Study in Political Economy. Academic Press, ??? (1978)

[8] Transparency International: Global corruption report 2007. Technical report (2007)

[9] Rothstein, B.O., Teorell, J.A.: What is quality of government? a theory of impartial government institutions. Governance **21**(2), 165–190 (2008)

[10] Mungiu-Pippidi, A.: Controlling corruption through collective action. Journal of Democracy **24**(1), 101–115 (2013)

[11] Mungiu-Pippidi, A.: Corruption: Diagnosis and Treatment. Democracy **17**(3) (2006)

[12] Fazekas, M., Bence, T.: Assessing the potential for detecting collusion in Swedish public procurement. Technical report, Swedish Competition Authority (2016)

[13] Watts, J.: Operation car wash: Is this the biggest corruption scandal in history. The Guardian **1**(06), 2017 (2017)

[14] Ribeiro, H.V., Alves, L.G., Martins, A.F., Lenzi, E.K., Perc, M.: The dynamical structure of political corruption networks. Journal of Complex Networks (2018)

[15] Ferrali, R.: Corruption as a criminal network. PhD thesis, Princeton University (2018)

[16] Calderoni, F.: Strategic positioning in mafia networks. In: Third Annual Illicit Networks Workshop, pp. 1–21 (2011). Équipe de recherche sur la délinquance en réseau

[17] Krebs, V.E.: Mapping networks of terrorist cells. Connections **24**(3), 43–52 (2002)

[18] Wrong, D.H.: The oversocialized conception of man in modern sociology. American sociological review, 183–193 (1961)

[19] Granovetter, M.: Economic action and social structure: The problem of embeddedness. American Journal of Sociology **91**(3), 481–510 (1985)

[20] OECD.Stat: Government at a Glance - 2017 edition: Public procurement. `https://stats.oecd.org/Index.aspx?QueryId=78413`. Accessed: 2018-08-09 (2017)

[21] Fazekas, M., Tóth, I.J.: From corruption to state capture: A new analytical framework with empirical applications from hungary. Political Research Quarterly **69**(2), 320–334 (2016)

[22] Fazekas, M., Kocsis, G.: Uncovering high-level corruption: Cross-national objective corruption risk indicators using public procurement data. British Journal of Political Science, 1–10 (2017)

[23] Mungiu-Pippidi, A.: The time has come for evidence-based anticorruption. Nature: Human Behavior (2017)

[24] Alm, E., Arkin, A.P.: Biological networks. Current opinion in structural biology **13**(2), 193–202 (2003)

[25] Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., Edwards, F., Figueroa, D., Jacob, U., Jones, J.I., *et al.*: Ecological networks–beyond food webs. Journal of Animal Ecology **78**(1), 253–269 (2009)

[26] Borgatti, S.P., Everett, M.G., Johnson, J.C.: Analyzing Social Networks. Sage, (2018)

[27] Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., White, D.R.: Economic networks: The new challenges. Science **325**(5939), 422–425 (2009)

[28] Jackson, M.O.: Social and Economic Networks. Princeton university press, (2010)

[29] Barthélemy, M.: Spatial networks. Physics Reports **499**(1-3), 1–101 (2011)

[30] Moreno, J.L.: Foundations of sociometry: An introduction. Sociometry, 15–35 (1941)

[31] Freeman, L.C.: Centrality in social networks conceptual clarification. Social networks **1**(3), 215–239 (1978)

[32] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications vol. 8. Cambridge university press, (1994)

[33] Freeman, L.C.: Some antecedents of social network analysis. Connections **19**(1), 39–42 (1996)

[34] Hidalgo, C.A.: Disconnected, fragmented, or united? a trans-disciplinary review of network science. Applied Network Science **1**(1), 6 (2016)

[35] Elster, J.: Explaining Social Behavior: More Nuts and Bolts for the Social Sciences. Cambridge University Press, ??? (2015)

[36] Klitgaard, R.: Controlling Corruption. Univ of California Press, ??? (1988)

[37] Shleifer, A., Vishny, R.W.: Corruption. The quarterly journal of economics **108**(3), 599–617 (1993)

[38] Persson, A., Rothstein, B., Teorell, J.: Why anticorruption reforms fail—systemic corruption as a collective action problem. Governance **26**(3), 449–471 (2013)

[39] Mungiu-Pippidi, A., Dadašov, R.: Measuring control of corruption by a new index of public integrity. European Journal on Criminal Policy and Research **22**(3), 415–438 (2016)

[40] Acemoglu, D., Johnson, S., Robinson, J.A.: Reversal of fortune: Geography and institutions in the making of the modern world income distribution. The Quarterly journal of economics **117**(4), 1231–1294 (2002)

[41] Treisman, D.: The causes of corruption: a cross-national study. Journal of public economics **76**(3), 399–457 (2000)

[42] Treisman, D.: What have we learned about the causes of corruption from ten years of cross-national empirical research? Annu. Rev. Polit. Sci. **10**, 211–244 (2007)

[43] Broms, R., Dahlström, C., Fazekas, M.: Procurement and competition in swedish municipalities. (2017)

[44] Charron, N., Dijkstra, L., Lapuente, V.: Regional governance matters: Quality of government within european union member states. Regional Studies **48**(1), 68–90 (2014)

[45] Camerer, C.F., Loewenstein, G., Rabin, M.: Advances in Behavioral Economics. Princeton University Press, (2011)

[46] Arechar, A.A., Gächter, S., Molleman, L.: Conducting interactive experiments online. Experimental economics **21**(1), 99–131 (2018)

[47] Duflo, E., Glennerster, R., Kremer, M.: Using randomization in development economics research: A toolkit. Handbook of development economics **4**, 3895–3962 (2007)

[48] Abbink, K.: Laboratory experiments on corruption. International handbook on the economics of corruption (2006)

[49] Propensities to engage in and punish corrupt behavior: Experimental evidence from Australia, India, Indonesia and Singapore. Journal of Public Economics **93**(7-8), 843–851 (2009). doi:10.1016/j.jpubeco.2009.03.004

[50] Shalvi, S.: Corruption corrupts. Nature **531**, 456–457 (2016)

[51] Gächter, S., Schulz, J.F.: Intrinsic honesy and the prevalencce of rule violations across societies. Nature **531**(7595), 1–11 (2016). doi:10.1038/nature17160

[52] Weisel, O., Shalvi, S.: The collaborative roots of corruption. Proceedings of the National Academy of Sciences **112**(34), 10651–10656 (2015)

[53] Kahneman, D., Knetsch, J.L., Thaler, R.H.: Fairness and the assumptions of economics. Journal of business, 285–300 (1986)

[54] Olken, B.A.: Monitoring corruption: evidence from a field experiment in indonesia. Journal of Political Economy **115**(2), 200–249 (2007)

[55] Bertrand, M., Djankov, S., Hanna, R., Mullainathan, S.: Obtaining a driver's license in india: an experimental approach to studying corruption. The Quarterly Journal of Economics **122**(4), 1639–1676 (2007)

[56] Ferraz, C., Finan, F.: Exposing corrupt politicians: the effects of brazil's publicly released audits on electoral outcomes. The Quarterly Journal of Economics **123**(2), 703–745 (2008)

[57] Muço, A.: On Anti-Corruption Tools in Developing Countries. Stockholm School of Economics, (2017)

[58] Muço, A.: Learn from thy neighbour: Do voters associate corruption with political parties. Technical report, mimeo (2016)

[59] Miller, J.H., Page, S.E.: Complex Adaptive Systems: An Introduction to Computational Models of Social Life vol. 17. Princeton university press, (2009)

[60] Tirole, J.: A Theory of Collective Reputations. Review of Economic Studies **63**(1), 1–22 (1996). doi:10.2307/2298112

[61] Lambsdorff, J.G.: How Confidence Facilitates Illegal Transactions: An Empirical Approach. The American Journal of Economics and Sociology **61**(4), 829–853 (2002)

[62] Macy, M.W., Skvoretz, J.: The evolution of trust and cooperation between strangers: A computational model. American Sociological Review, 638–660 (1998)

[63] Hammond, R.: Endogenous Transition Dynamics in Corruption: An Agent-Based Computer Model. Working Paper, Center on Social and Economic Dynamics, Brookings Institute (19) (2000)

[64] Transparency International: Transparency international corruption perceptions index. Technical report. data retrieved from `https://www.transparency.org/research/cpi/overview`

[65] Lambsdorff, J.G.: 2002 corruption perceptions index. Global Corruption Report, 262–322 (2003)

[66] The World Bank: World Bank Worldwide Governance Indicators. data retrieved from https://info.worldbank.org/governance/wgi/index.aspx#home

[67] Kaufmann, D., Kraay, A., Mastruzzi, M.: Governance Matters IV: Governance Indicators for 1996-2004. The World Bank, (2005)

[68] Heywood, P.M., Rose, J.: "close but no cigar": the measurement of corruption. Journal of Public Policy **34**(3), 507–529 (2014)

[69] Hawken, A., Munck, G.L.: Do you know your data? measurement validity in corruption research. Technical report, Working Paper - School of Public Policy, Pepperdine University (2009)

[70] Urra, F.-J.: Assessing corruption an analytical review of corruption measurement and its problems: Perception, error and utility. Technical report, Working Paper - Edmund A. Walsh School of Foreign Service (2007)

[71] Standaert, S.: Divining the level of corruption: a bayesian state-space approach. Journal of Comparative Economics **43**(3), 782–803 (2015)

[72] Coppedge, M., Gerring, J., Lindberg, S.I., Skaaning, S.-E., Teorell, J., Altman, D., Andersson, F., Bernhard, M., Fish, M.S., Glynn, A., et al.: V-Dem Codebook v8. data retrieved from https://www.v-dem.net/en/reference/version-8-apr-2018/ (2017)

[73] Rothstein, B., Charron, N., Lapuente, V.: Quality of Government and Corruption from a European Perspective: a Comparative Study on the Quality of Government in EU Regions. Edward Elgar Publishing, (2013)

[74] Olken, B.A.: Corruption perceptions vs. corruption reality. Journal of Public economics **93**(7-8), 950–964 (2009)

[75] Torsello, D., Venard, B.: The anthropology of corruption. Journal of Management Inquiry **25**(1), 34–54 (2016)

[76] Bertot, J.C., Jaeger, P.T., Grimes, J.M.: Using icts to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. Government information quarterly **27**(3), 264–271 (2010)

[77] Kornberger, M., Meyer, R.E., Brandtner, C., Höllerer, M.A.: When bureaucracy meets the crowd: Studying "open government" in the vienna city administration. Organization Studies **38**(2), 179–200 (2017)

[78] Glaeser, E.L., Saks, R.E.: Corruption in America. Journal of Public Economics **90**(6-7), 1053–1072 (2006)

[79] Golden, M.A., Picci, L.: Proposal for a new measure of corruption, illustrated with italian data. Economics & Politics **17**(1), 37–75 (2005)

[80] Beber, B., Scacco, A.: What the numbers say: A digit-based test for election fraud. Political analysis **20**(2), 211–234 (2012)

[81] Rauch, B., Göttsche, M., Brähler, G., Kronfeld, T.: Deficit versus social statistics: empirical evidence for the effectiveness of benford's law. Applied Economics Letters **21**(3), 147–151 (2014)

[82] Integrity in Public Procurement: Good Practice from A to Z. Technical report, OECD (2007)

[83] European Commission: Public procurement - guidance for practitioners. Technical report (2018). Accessed: 2019-02-15

[84] Group, W.B.: The World Bank Group and Public Procurement : An Independent Evaluation. https://openknowledge.worldbank.org/handle/10986/21434. Accessed: 2019-01-15 (2014)

[85] Fazekas, M., Tóth, I.J., King, L.P.: An objective corruption risk index using public procurement data. European Journal on Criminal Policy and Research **22**(3), 369–397 (2016)

[86] Fazekas, M., Skuhrovec, J., Wachs, J.: Corruption, government turnover, and public contracting market structure–insights using network analysis and objective corruption proxies. Technical report, Government Transparency Institute Working Papers. (2017)

[87] European Court of Auditors: Fighting fraud in EU spending: action needed. Technical report (2019)

[88] Charron, N., Dahlström, C., Fazekas, M., Lapuente, V.: Careers, connections, and corruption risks: Investigating the impact of bureaucratic meritocracy on public procurement processes. The Journal of Politics **79**(1), 89–104 (2017)

[89] Guerakar, E., Meyersson, E.: State discretion, political connections and public procurement: Evidence from turkey. Technical report, Working paper, Stockholm Institute for Transition Economics (2016)

[90] Gambetta, D.: The Sicilian Mafia: the Business of Private Protection. Harvard University Press, (1996)

[91] Williams, P.: Transnational criminal networks. Networks and netwars: the future of terror, crime, and militancy **1382**, 61 (2001)

[92] Agreste, S., Catanese, S., De Meo, P., Ferrara, E., Fiumara, G.: Network structure and resilience of mafia syndicates. Information Sciences **351**, 30–47 (2016)

[93] DellaPosta, D.: Network closure and integration in the mid-20th century american mafia. Social Networks **51**, 148–157 (2017)

[94] Neu, D., Everett, J., Rahaman, A.S., Martinez, D.: Accounting and networks of corruption. Accounting, Organizations and Society **38**(6-7), 505–524 (2013)

[95] Diesner, J., Frantz, T.L., Carley, K.M.: Communication networks from the enron email corpus "it's always about the people. enron is no different". Computational & Mathematical Organization Theory **11**(3), 201–228 (2005)

[96] Tumminello, M., Edling, C., Liljeros, F., Mantegna, R.N., Sarnecki, J.: The phenomenology of specialization of criminal suspects. PloS One **8**(5), 64703 (2013)

[97] Rostami, A., Mondani, H.: The complexity of crime network data: A case study of its consequences for crime control and the study of networks. PloS one **10**(3), 0119309 (2015)

[98] Wachs, J., Yasseri, T., Lengyel, B., Kertész, J.: Social capital predicts corruption risk in towns. Royal Society Open Science (2019)

[99] Mauro, P.: The persistence of corruption and slow economic growth. IMF staff papers **51**(1), 1–18 (2004)

[100] Alesina, A., Zhuravskaya, E.: Segregation and the quality of government in a cross section of countries. American Economic Review **101**(5), 1872–1911 (2011)

[101] Rothstein, B., Uslaner, E.M.: All for all: Equality, corruption, and social trust. World Politics **58**(1), 41–72 (2005)

[102] Guiso, L., Sapienza, P., Zingales, L.: Civic capital as the missing link. In: Handbook of Social Economics vol. 1, pp. 417–480. Elsevier, (2011)

[103] Putnam, R.D.: Bowling Alone: The Collapse and Revival of American Community. Simon and Schuster, (2001)

[104] Portes, A.: Social capital: Its origins and applications in modern sociology. Annual Review of Sociology **24**(1), 1–24 (1998)

[105] Westlund, H., Adam, F.: Social capital and economic performance: A meta-analysis of 65 studies. European Planning Studie s **18**(6), 893–919 (2010)

[106] Lin, N.: Social Capital: A Theory of Social Structure and Action vol. 19. Cambridge University Press, (2002)

[107] Harris, D.: Bonding social capital and corruption: A cross-national empirical analysis. Technical report, University of Cambridge, Department of Land Economics (2007)

[108] Bjørnskov, C.: Corruption and social capital. Technical report, University of Aarhus, Aarhus School of Business, Department of Economics (2003)

[109] Eagle, N., Macy, M., Claxton, R.: Network diversity and economic development. Science **328**(5981), 1029–1031 (2010)

[110] Bokányi, E., Kondor, D., Dobos, L., Sebők, T., Stéger, J., Csabai, I., Vattay, G.: Race, religion and the city: twitter word frequency patterns reveal dominant demographic dimensions in the united states. Palgrave Communications **2**, 16010 (2016)

[111] Mamei, M., Pancotto, F., De Nadai, M., Lepri, B., Vescovi, M., Zambonelli, F., Pentland, A.: Is social capital associated with synchronization in human communication? an analysis of italian call records and measures of civic engagement. EPJ Data Science **7**(1), 25 (2018)

[112] Norbutas, L., Corten, R.: Network structure and economic prosperity in municipalities: A large-scale test of social capital theory using social media data. Social Networks **52**, 120–134 (2018)

[113] Bailey, M., Cao, R., Kuchler, T., Stroebel, J., Wong, A., *et al.*: Social connectedness: Measurement, determinants, and effects. Journal of Economic Perspectives **32**(3), 259–80 (2018)

[114] Lengyel, B., Varga, A., Ságvári, B., Jakobi, Á., Kertész, J.: Geographies of an online social network. PloS one **10**(9), 0137248 (2015)

[115] Fazekas, M., Tóth, B.: The extent and cost of corruption in transport infrastructure. new evidence from europe. Transportation Research Part A: Policy and Practice **113**, 35–54 (2018)

[116] European Commission: The OLAF Report 2017, p. 15 (2018). Available at URL: https://ec.europa.eu/anti-fraud/sites/antifraud/files/olaf_report_2017_en.pdf

[117] Fearon, J.D., Laitin, D.D.: Explaining interethnic cooperation. American political science review **90**(4), 715–735 (1996)

[118] Coleman, J.S.: Social capital in the creation of human capital. American Journal of Sociology, 95–120 (1988)

[119] Portes, A.: The two meanings of social capital. Sociological Forum **15**, 1–12 (2000)

[120] Erickson, B.H.: Secret societies and social structure. Social Forces **60**(1), 188–210 (1981)

[121] Geertz, C.: Peddlers and Princes: Social Development and Economic Change in Two Indonesian Towns vol. 318. University of Chicago Press, (1963)

[122] Levine, S.S., Apfelbaum, E.P., Bernard, M., Bartelt, V.L., Zajac, E.J., Stark, D.: Ethnic diversity deflates price bubbles. Proceedings of the National Academy of Sciences **111**(52), 18524–18529 (2014)

[123] Granovetter, M.S.: The strength of weak ties. American Journal of Sociology **78**(6), 1360–1380 (1973)

[124] Lancee, B.: The Economic Returns of Immigrants' Bonding and Bridging Social Capital: The Case of the Netherlands. International Migration Review **44**(1), 202–226 (2010)

[125] Laurence, J.: The effect of ethnic diversity and community disadvantage on social cohesion: A multi-level analysis of social capital and interethnic relations in UK communities. European Sociological Review **27**(1), 70–89 (2009)

[126] Portes, A.: Downsides of social capital. Proceedings of the National Academy of Sciences **111**(52), 18407–18408 (2014)

[127] Yamagishi, T.: Trust: The Evolutionary Game of Mind and Society. Springer, (2011)

[128] Knack, S.: Social capital and the quality of government: Evidence from the states. American Journal of Political Science, 772–785 (2002)

[129] Paccagnella, M., Sestito, P.: School cheating and social capital. Education Economics **22**(4), 367–388 (2014)

[130] PwC EU Services: Identifying and Reducing Corruption in Public Procurement in the EU. Technical report, European Commission (2013)

[131] Weber, M.: Economy and Society: An Outline of Interpretive Sociology vol. 1. Univ of California Press, (1978)

[132] Attström, K., Kröber, R., Junclaus, M.: Review of the Function of the CPV Codes/System. Technical report, European Commission (2012)

[133] Fazekas, M., Skuhrovec, J., Wachs, J.: Corruption, government turnover, and public contracting market structure–insights using network analysis and objective corruption proxies. GTI Working Paper Series (2) (2017)

[134] Fazekas, M., Ferrali, R., Wachs, J.: Institutional quality, campaign contributions, and favouritism in us federal government contracting. GTI Working Paper Series (1) (2018)

[135] Norman, L., Komuves, A.: EU Fraud Office Finds Irregularities in Projects Linked to Hungarian Leader's Son-in-Law. The Wall Street Journal (2018)

[136] Török, J., Kertész, J.: Cascading collapse of online social networks. Scientific Reports **7**(1), 16743 (2017)

[137] Lorincz, L., Koltai, J., Gyor, A.F., Takacs, K., et al.: Collapse of an online social network: The blame on social capital. Technical report, Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences (2017)

[138] Török, J., Murase, Y., Jo, H.-H., Kertész, J., Kaski, K.: What big data tells: sampling the social network by communication channels. Physical Review E **94**(5), 052319 (2016)

[139] Williams, D.: On and off the'net: Scales for social capital in an online era. Journal of Computer-Mediated Communication **11**(2), 593–628 (2006)

[140] Brooks, B., Hogan, B., Ellison, N., Lampe, C., Vitak, J.: Assessing structural correlates to social capital in facebook ego networks. Social Networks **38**, 1–15 (2014)

[141] Dunbar, R.I., Arnaboldi, V., Conti, M., Passarella, A.: The structure of online social networks mirrors those in the offline world. Social Networks **43**, 39–47 (2015)

[142] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10), 10008 (2008)

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

[144] Santo, F., Marc, B.: Resolution limit in community detection. Proceedings of the National Academy of Sciences **104**, 36–41 (2007)

[145] Sah, P., Leu, S.T., Cross, P.C., Hudson, P.J., Bansal, S.: Unraveling the disease consequences and mechanisms of modular structure in animal social networks. Proceedings of the National Academy of Sciences, 201613616 (2017)

[146] Pfeil, U., Arjan, R., Zaphiris, P.: Age differences in online social networking–a study of user profiles and the social capital divide among teenagers and older users in myspace. Computers in Human Behavior **25**(3), 643–654 (2009)

[147] Campante, Filipe R and Do, Quoc-Anh: Isolated capital cities, accountability, and corruption: Evidence from US states. American Economic Review **104**(8), 2456–81 (2014)

[148] Coviello, D., Guglielmo, A., Spagnolo, G.: The effect of discretion on procurement performance. Management Science **64**(2), 715–738 (2017)

[149] Richey, S.: The impact of corruption on social trust. American Politics Research **38**(4), 676–690 (2010)

[150] Acemoglu, D., Jackson, M.O.: Social norms and the enforcement of laws. Journal of the European Economic Association **15**(2), 245–295 (2017)

[151] Hernández, L., Vignes, A., Saba, S.: Trust or robustness? an ecological approach to the study of auction and bilateral markets. PloS one **13**(5), 0196206 (2018)

[152] Jordano, P., Bascompte, J., Olesen, J.M.: Invariant properties in coevolutionary networks of plant–animal interactions. Ecology letters **6**(1), 69–81 (2003)

[153] Bustos, S., Gomez, C., Hausmann, R., Hidalgo, C.A.: The dynamics of nestedness predicts the evolution of industrial ecosystems. PloS one **7**(11), 49393 (2012)

[154] Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: K-core organization of complex networks. Physical review letters **96**(4), 040601 (2006)

[155] Csermely, P., London, A., Wu, L.-Y., Uzzi, B.: Structure and dynamics of core/periphery networks. Journal of Complex Networks **1**(2), 93–123 (2013)

[156] Mahutga, M.C.: The persistence of structural inequality? a network analysis of international trade, 1965–2000. Social Forces **84**(4), 1863–1889 (2006)

[157] Cattani, G., Ferriani, S.: A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the hollywood film industry. Organization science **19**(6), 824–844 (2008)

[158] Barberá, P., Wang, N., Bonneau, R., Jost, J.T., Nagler, J., Tucker, J., González-Bailón, S.: The critical periphery in the growth of social protests. PloS one **10**(11), 0143611 (2015)

[159] Kossinets, G., Watts, D.J.: Origins of homophily in an evolving social network. American journal of sociology **115**(2), 405–450 (2009)

[160] Szucs, F.: Discretion and Corruption in Public Procurement (2017)

[161] Christen, P.: Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Springer, (2012)

[162] Sarawagi, S., Bhamidipaty, A.: Interactive deduplication using active learning. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 269–278 (2002). ACM

[163] Gregg, F., Eder, D.: Dedupe. https://github.com/datamade/dedupe (2015)

[164] Robins, G., Alexander, M.: Small worlds among interlocking directors: Network structure and distance in bipartite graphs. Computational & Mathematical Organization Theory **10**(1), 69–94 (2004)

[165] Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks. Nature **393**(6684), 440 (1998)

[166] Clauset, A., Shalizi, C.R., Newman, M.E.: Power-law distributions in empirical data. SIAM review **51**(4), 661–703 (2009)

[167] Alstott, J., Bullmore, E., Plenz, D.: powerlaw: a python package for analysis of heavy-tailed distributions. PloS one **9**(1), 85777 (2014)

[168] Fisman, R., Gatti, R.: Decentralization and corruption: evidence across countries. Journal of Public Economics **83**(3), 325–345 (2002)

[169] Gurgur, T., Shah, A.: Localization and Corruption: Panacea or Pandora's Box? The World Bank, (2005)

[170] Persson, T., Tabellini, G.E.: Political Economics: Explaining Economic Policy. MIT press, (2002)

[171] Pinto, J., Leana, C.R., Pil, F.K.: Corrupt organizations or organizations of corrupt individuals? two types of organization-level corruption. Academy of Management Review **33**(3), 685–709 (2008)

[172] Lauchs, M., Keast, R., Chamberlain, D.: Resilience of a corrupt police network: the first and second jokes in queensland. Crime, law and social change **57**(2), 195–207 (2012)

[173] Batagelj, V., Zaveršnik, M.: Generalized cores. arXiv preprint cs/0202039 (2002)

[174] Wuchty, S., Almaas, E.: Peeling the yeast protein network. Proteomics **5**(2), 444–449 (2005)

[175] Garas, A., Schweitzer, F., Havlin, S.: A k-shell decomposition method for weighted networks. New Journal of Physics **14**(8), 083030 (2012)

[176] Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J.: Structural diversity in social contagion. Proceedings of the National Academy of Sciences, 201116502 (2012)

[177] Alvarez-Hamelin, J.I., Dall'Asta, L., Barrat, A., Vespignani, A.: Large scale networks fingerprinting and visualization using the k-core decomposition. In: Advances in Neural Information Processing Systems, pp. 41–50 (2006)

[178] Batagelj, V., Zaversnik, M.: An o (m) algorithm for cores decomposition of networks. arXiv preprint cs/0310049 (2003)

[179] Malliaros, F., Giatsidis, C., Papadopoulos, A., Vazirgiannis, M.: The core decomposition of networks: Theory, algorithms and applications (2019)

[180] Fortunato, S.: Community detection in graphs. Physics reports **486**(3-5), 75–174 (2010)

[181] Evans, T., Lambiotte, R.: Line graphs, link partitions, and overlapping communities. Physical Review E **80**(1), 016105 (2009)

[182] Evans, T.S., Lambiotte, R.: Line graphs of weighted networks for overlapping communities. The European Physical Journal B **77**(2), 265–272 (2010)

[183] Ahn, Y.-Y., Bagrow, J.P., Lehmann, S.: Link communities reveal multiscale complexity in networks. nature **466**(7307), 761 (2010)

[184] Mironov, M., Zhuravskaya, E.: Corruption in procurement and the political cycle in tunneling: Evidence from financial transactions data. American Economic Journal: Economic Policy **8**(2), 287–321 (2016)

[185] Döring, H., Manow, P.: Parliament and government composition database (parlgov). An infrastructure for empirical information on parties, elections and governments in modern democracies. Version **10**(11), 6 (2010)

[186] Lederman, D., Loayza, N.V., Soares, R.R.: Accountability and corruption: Political institutions matter. Economics & Politics **17**(1), 1–35 (2005)

[187] Margetts, H., John, P., Hale, S., Yasseri, T.: Political Turbulence: How Social Media Shape Collective Action. Princeton University Press, (2015)

[188] Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. CRC press, (1994)

[189] Innes, A.: Party competition in postcommunist europe: The great electoral lottery. Comparative politics, 85–104 (2002)

[190] Fierăscu, S.I., Pârvu, M., Topîrceanu, A., Udrescu, M.: Exploring party switching in the post-1989 romanian politicians network from a complex network perspective. Romanian Journal of Political Science **18**(1), 108–136 (2018)

[191] Grzymalała-Busse, A., Innes, A.: Great expectations: The eu and domestic political competition in east central europe. East European Politics and Societies **17**(1), 64–73 (2003)

[192] Slomczynski, K., Shabad, G.: Perceptions of political party corruption and voting behaviour in poland. Party Politics **18**(6), 897–917 (2012)

[193] Luz, R.D., Spagnolo, G.: Leniency, collusion, corruption, and whistleblowing. Journal of Competition Law & Economics **13**(4), 729–766 (2017)

[194] Smith, A.: The Wealth of Nations. Bantam Classics, (2003)

[195] Marshall, R.C., Marx, L.M.: The Economics of Collusion: Cartels and Bidding Rings. Mit Press, (2012)

[196] Abrantes-Metz, R., Sokol, D.: Lessons from LIBOR for Detection and Deterrence of Cartel Wrongdoing. Antitrust Chronicle **11** (2012)

[197] Harrington, J.E., Skrzypacz, A.: Private monitoring and communication in cartels: Explaining recent collusive practices. American Economic Review **101**(6), 2425–49 (2011)

[198] Connor, J.M., Bolotova, Y.: Cartel overcharges: Survey and meta-analysis. International Journal of Industrial Organization **24**(6), 1109–1137 (2006)

[199] Combe, E., Monnier, C., Legal, R.: Cartels: The probability of getting caught in the european union. Technical Report 102, Bruges European Economic Research papers (2008)

[200] Levenstein, M.C., Suslow, V.Y.: What determines cartel success? Journal of economic literature **44**(1), 43–95 (2006)

[201] Compte, O., Jenny, F., Rey, P.: Capacity constraints, mergers and collusion. European Economic Review **46**(1), 1–29 (2002)

[202] Motta, M.: Competition Policy: Theory and Practice. Cambridge University Press, (2004)

[203] Kofman, F., Lawarrée, J.: A prisoner's dilemma model of collusion deterrence. Journal of Public Economics **59**, 117–136 (1996)

[204] Rapoport, A.: Prisoner's dilemma—recollections and observations. In: Game Theory as a Theory of a Conflict Resolution, pp. 17–34. Springer, (1974)

[205] Axelrod, R., Hamilton, W.D.: The evolution of cooperation. Science **211**(4489), 1390–1396 (1981)

[206] Dixon, H.D., Wallis, S., Moss, S.: Axelrod meets cournot: Oligopoly and the evolutionary metaphor. Computational Economics **20**(3), 139–156 (2002)

[207] Bowles, S., Gintis, H.: A Cooperative Species: Human Reciprocity and Its Evolution. Princeton University Press, (2011)

[208] Nowak, M.A., May, R.M.: The spatial dilemmas of evolution. International Journal of bifurcation and chaos **3**(01), 35–78 (1993)

[209] Santos, F.C., Santos, M.D., Pacheco, J.M.: Social diversity promotes the emergence of cooperation in public goods games. Nature **454**(7201), 213 (2008)

[210] Perc, M., Gómez-Gardeñes, J., Szolnoki, A., Floría, L.M., Moreno, Y.: Evolutionary dynamics of group interactions on structured populations: a review. Journal of the royal society interface **10**(80), 20120997 (2013)

[211] Battiston, F., Perc, M., Latora, V.: Determinants of public cooperation in multiplex networks. New Journal of Physics **19**(7), 073017 (2017)

[212] Számadó, S., Szalai, F., Scheuring, I.: The effect of dispersal and neighbourhood in games of cooperation. Journal of Theoretical Biology **253**(2), 221–227 (2008)

[213] Pesendorfer, M.: A study of collusion in first-price auctions. The Review of Economic Studies **67**(3), 381–411 (2000)

[214] Hüschelrath, K., Veith, T.: Cartel detection in procurement markets. Managerial and Decision Economics **35**(6), 404–422 (2014)

[215] Porter, R.H., Zona, J.D.: Detection of bid rigging in procurement auctions. Journal of political economy **101**(3), 518–538 (1993)

[216] Tóth, B., Fazekas, M., Czibik, Á., Tóth, I.J.: Toolkit for detecting collusive bidding in public procurement. Technical report, GTI Working Paper Series,(2) (2014)

[217] Reeves-Latour, M., Morselli, C.: Bid-rigging networks and state-corporate crime in the construction industry. Social Networks **51**, 158–170 (2017)

[218] Morselli, C., Ouellet, M.: Network similarity and collusion. Social Networks **55**, 21–30 (2018)

[219] Imhof, D., Karagök, Y., Rutz, S.: Screening for bid rigging—does it work? Journal of Competition Law & Economics **14**(2), 235–261 (2018)

[220] Porter, R.H., Zona, J.D.: Ohio school milk markets: An analysis of bidding. The Rand Journal of Economics, 263–288 (1999)

[221] Abrantes-Metz, R.M., Froeb, L.M., Geweke, J., Taylor, C.T.: A variance screen for collusion. International Journal of Industrial Organization **24**(3), 467–486 (2006)

[222] Lancichinetti, A., Fortunato, S., Kertész, J.: Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics **11**(3), 033015 (2009)

[223] Onnela, J.-P., Saramäki, J., Kertész, J., Kaski, K.: Intensity and coherence of motifs in weighted complex networks. Physical Review E **71**(6), 065103 (2005)

[224] LaCasse, C.: Bid rigging and the threat of government prosecution. The RAND Journal of Economics, 398–417 (1995)

[225] Bolotova, Y., Connor, J.M., Miller, D.J.: The impact of collusion on price behavior: Empirical results from two recent cases. International Journal of Industrial Organization **26**(6), 1290–1307 (2008)

[226] Imhof, D.: Simple statistical screens to detect bid rigging. Technical report, Université de Fribourg (2017)

[227] Hilbe, C., Chatterjee, K., Nowak, M.A.: Partners and rivals in direct reciprocity. Nature Human Behaviour, 1 (2018)

[228] Harrington, J.E.: Behavioral screening and the detection of cartels. In: European Competition Law Annual 2006: Enforcement of Prohibition of Cartels, (2006)

[229] Bajari, P., Ye, L.: Deciding between competition and collusion. Review of Economics and statistics **85**(4), 971–989 (2003)

[230] Allain, M.-L., Boyer, M., Ponssard, J.P.: The determination of optimal fines in cartel cases: Theory and practice. Technical Report 34, Centre Interuniversitatire de Recherche en Analyse des Organisations Discussion Paper (2011)

[231] Ishii, R.: Favor exchange in collusion: Empirical study of repeated procurement auctions in japan. International Journal of Industrial Organization **27**(2), 137–144 (2009)

[232] Asker, J.: A study of the internal organization of a bidding cartel. American Economic Review **100**(3), 724–62 (2010)

[233] Conley, T.G., Decarolis, F.: Detecting bidders groups in collusive auctions. American Economic Journal: Microeconomics **8**(2), 1–38 (2016)

[234] Athey, S., Levin, J., Seira, E.: Comparing open and sealed bid auctions: Evidence from timber auctions. The Quarterly Journal of Economics **126**(1), 207–257 (2011)

[235] Hendricks, K., Porter, R.H.: Collusion in auctions. Annales d'Economie et de Statistique, 217–230 (1989)

[236] Hewitt, C., McClave, J., Sibley, D.: Incumbency and bidding behavior in the dallas-ft. worth school milk market. Technical report, Mimeo. Austin, TX: University of Texas at Austin (1993)

[237] Cao, Q., Sirivianos, M., Yang, X., Pregueiro, T.: Aiding the detection of fake accounts in large scale social online services. In: Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation, pp. 15–15 (2012). USENIX Association

[238] Sung, H.-E.: Democracy and political corruption: A cross-national comparison. Crime, Law and Social Change **41**(2), 179–193 (2004)

[239] Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., et al.: Multivariate Analysis, (1999)

CHAPTER 7

APPENDICES

7.1 Social Networks and Corruption

7.1.1 Description of iWiW data

In line with previous work on iWiW we filtered the data used in our analysis. We use the data from the network at its peak activity in 2012. Out of roughly 4.5 million user accounts, we dropped the roughly 500,000 accounts with location outside of Hungary. We follow Lengyel et al. [114], we dropped the 193 users with more than 10,000 connections, arguing that such a large number of connections cannot represent social ties. We argue that this cutoff balances two concerns: it excludes those accounts with so many connections that it brings into question the nature of its connections, and we avoid truncating the tail of the distribution of social connectivity too much, allowing for sociality to range over several orders of magnitude. Many approaches to detect "fake" accounts in social network use the degree of a node as an important input [237].

In Plot A of Figure 7.1 we plot the sensitivity of fragmentation and diversity to the maximum degree threshold. If we discard all users with more than 100 connections (compared to the 10,000 connection cutoff we use in our paper), fragmentation would be significantly higher and diversity significantly lower than the versions we use in the paper. However this is not a reasonable cutoff as nearly 10% of users have more than 500 connections (see Plot B, Figure 7.1). The settlement fragmentation and diversity measures are within 5% of the versions we use in the paper if the threshold is set at 500, 1000, or 2000 connections.

In Figure 7.2 we show the relationship between settlement population and the number of iWiW users listing their location in the settlement, and the share of the population registered to iWiW. As mentioned in the text, user privacy is a key concern. The anonymized iWiW data was made available to a consortium of researchers in Hungary, each of whom signed a non-disclosure agreement (NDA) to use the data for research purposes only. As a result, only settlement level aggregated data can be shared.

7.1.2 Relationship between fragmentation and diversity

Fragmentation and diversity, our measures of bonding and bridging social capital respectively, are positively and significantly correlated ($\rho \approx 0.46$). Though fragmentation considers only edges within the settlement and ego diversity includes external edges, both variables measure modularity in the network. However, according to our hypotheses, they are expected to capture different kinds of socialization. We found that despite their positive correlation these features have opposite relationships with our corruption risk measures: high fragmentation is positively and high diversity is negatively correlated with corruption risk. To test whether inter-settlement edges or the ego focus of diversity does more to distinguish the measure from fragmentation we recalculated the diversity considering only edges within the settlement. This alternative "internal" diversity measure is weakly correlated ($\rho \approx 0.28$) with fragmentation, and strongly correlated with diversity ($\rho \approx 0.72$). This suggests that both the connections to other settlements and the ego-focus of the diversity measure distinguish fragmented settlements from diverse ones.

7.1.3 Model covariates and controls

In this appendix section we present the settlement-level variables used as controls in our models. We also report their summary statistics. Note that in our models, we scale all features to have mean 0 and standard deviation 1. Our controls mostly refer to data from 2011, when the last large scale Hungarian census took place and the data are of highest quality.

- *Average income per capita (2011)*: Wealthier places tend to be less corrupt [10] as competition for limited resources is expected to create greater incentive to cheat. Data on median income or the income distribution at the settlement level were, to the best our knowledge, not available in Hungary.

- *Population (log)(2011)*: Larger cities may have different contracting needs, different political and social norms, and different network characteristics.

- *Number of contracts awarded (log)*: Settlements contracting more frequently may be more experienced and may follow better practices. As more people are involved in contracting, corruption may become more difficult.

- *Rate of iWiW use (2012)*: The rate of iWiW use both proxies for the economic development of the settlement and controls for differences in observed social network structure resulting from differences in access to the web. Previous work suggests that iWiW users, especially the early adopters, skew young and wealthy [114].

- *Average mayoral victory margin*: Measured across three elections (2002, 2006, 2010), this variable proxies for the lack of political competition in the settlement. The absence of political competition has been shown to correlate with corruption [43].

- *Share of population with at least a high school diploma (2011)*: Education is typically correlated with better control of corruption [101].

- *Share of working-age population inactive and unemployment rate (2011)*: Counting the long-term and short-term unemployed respectively, these variables quantify economic stagnation. The economic hardship connected with high unemployment is conjectured to worsen political corruption [238].

- *The minimum travel distance to Budapest, the capital city*: This variable captures the physical isolation of the settlement from the main economic, political, and social hub of the country. Past research has shown that geographic isolation reduces accountability and increases corruption [147].

- *Share of population over 60 years old (2011)*: This variable controls for the over-representation of the elderly. The elderly are underrepresented on online social networks and tend to use these platforms differently than younger users [146].

- *Whether the settlement has a university (2011)*: This variable controls for the presence of a place of higher education in the settlement, including local branches of universities headquartered elsewhere. this which inflates the number of young people, hence likely iWiW users in the settlement.

7.1.4 Model results, diagnostics, and feature importances

We also present models including only one of the two network measures in Table 7.2. The effect and significance of both features is preserved when the

Statistic	N	Mean	St. Dev.	Min	Max
Closed procedure or single bid.	169	0.59	0.15	0.21	0.92
Average CRI	169	0.28	0.04	0.16	0.40
Fragmentation	169	0.32	0.04	0.16	0.46
Avg. ego diversity	169	0.35	0.07	0.20	0.51
Income/capita (thous. HUF)	169	823.57	189.93	488.44	1,516.55
N contracts (log)	169	4.52	0.69	3.69	6.42
Population (log)	169	9.72	0.89	7.66	12.24
Rate iWiW use	169	0.33	0.06	0.18	0.46
Average mayoral victory margin	169	0.15	0.14	0.00	0.64
% high school graduates	169	47.23	10.22	25.70	76.80
Distance to Budapest (minutes)	169	114.00	54.34	22.55	228.57
Share of population inactive	169	0.30	0.04	0.20	0.40
Unemployment Rate	169	0.06	0.01	0.03	0.09
Share of population 60+	169	0.24	0.03	0.15	0.39
Has university	169	0.25	0.44	0	1

Table 7.1. *Descriptive statistics of key settlement-level variables and controls.*

other is excluded. Recall that all variables are standardized with mean 0 and standard deviation 1. This aids interpretation, for example: a one standard deviation increase in the settlement's mayor's average margin of victory increases corruption risk by roughly one quarter of a standard deviation.

The estimated coefficients of the control variables and their levels of statistical significance offer additional insight into the phenomenon of corruption risk. Wealthier settlements are in general less corrupt, though the effect is not significant for CRI. Rate of iWiW use is not related with corruption risk and this does not change when we include the social capital features. The average mayoral victory margin is a highly significant positive predictor of corruption risk. One potential explanation is that mayors, who do not face significant competition do not fear being voted out of office if they are corrupt. Similarly settlements that are far from Budapest, which our models predict to be significantly more corrupt, may be insulated from investigation by the central authorities simply by being out of the spotlight.

One potential source of bias in the coefficient estimates of multiple regression models is collinearity among the predictors. We test for multi-collinearity for each predictor using a variance inflation factor (VIF) test, defined as the ratio of variance in the full model over the variance of the single-predictor model. We run this diagnostic for each predictor used in models (2) and (4) in the main

Dependent variable:	% Closed or single bid.			
	(1)	(2)	(3)	(4)
Fragmentation (Bonding social capital)			0.233^{**} (0.099)	0.263^{***} (0.097)
Diversity (Bridging social capital)		-0.505^{***} (0.179)		-0.553^{***} (0.176)
Income/capita	-0.262 (0.169)	-0.295^{*} (0.166)	-0.243 (0.167)	-0.277^{*} (0.162)
N contracts (log)	-0.313^{*} (0.171)	-0.359^{**} (0.168)	-0.269 (0.169)	-0.314^{*} (0.165)
Population (log)	-0.180 (0.143)	0.083 (0.168)	-0.257^{*} (0.144)	0.020 (0.166)
Rate iWiW use	0.045 (0.137)	0.009 (0.134)	0.073 (0.135)	0.037 (0.132)
Mayor victory margin	0.278^{***} (0.089)	0.259^{***} (0.087)	0.276^{***} (0.088)	0.255^{***} (0.086)
% high school grads	0.166 (0.190)	0.397^{*} (0.203)	0.126 (0.188)	0.374^{*} (0.199)
Distance to Budapest	-0.021 (0.104)	-0.169 (0.114)	-0.035 (0.102)	-0.198^{*} (0.112)
Share of pop. inactive	-0.797^{***} (0.229)	-0.931^{***} (0.229)	-0.675^{***} (0.232)	-0.805^{***} (0.229)
Unemployment Rate	0.239^{**} (0.118)	0.253^{**} (0.115)	0.247^{**} (0.116)	0.262^{**} (0.113)
% population 60+	0.501^{***} (0.163)	0.546^{***} (0.160)	0.449^{***} (0.162)	0.491^{***} (0.158)
Has University	0.351 (0.220)	0.198 (0.222)	0.449^{**} (0.221)	0.294 (0.221)
Constant	1.245^{*} (0.725)	1.426^{**} (0.712)	1.036 (0.720)	1.206^{*} (0.702)
Observations	169	169	169	169
Adjusted R^2	0.163	0.198	0.186	0.230
F Statistic	3.967^{***}	4.460^{***}	4.207^{***}	4.859^{***}

Table 7.2. *Stepwise regressions. The effect and significance of the network features are preserved when including them only one at a time.* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Predictor	VIF
Fragmentation	1.407
Diversity	6.337
Income/capita	5.430
N contracts (log)	3.045
Population (log)	5.892
Rate iWiW use	2.885
Mayor victory margin	1.040
% high school grads	7.106
Share of pop. inactive	9.899
Unemployment Rate	2.360
Distance to Budapest	3.068
% population 60+	5.442
Has university	2.192

Table 7.3. *VIF scores for model predictors.*

text and report the results in Table 7.3. A popular rule of thumb is that VIF values under 10 denote acceptable levels of correlation between variables [239]. As it is near our limit, we reran our analyses without the "Share of population inactive" control variable, finding no substantive change in our results. The relevant model tables are available on request.

We show the relative variable importances of Model (6) (column 6 in Table 3.2), the fully specific model predicting average CRI, using an Analysis of Variance F-test in Figure 7.3. We include only terms with a significant ANOVA F-test. Though other features have stronger predictive power, the social network features are more useful in predicting corruption risk than economic variables like unemployment, inactivity, and average income.

Dependent variable:	% Closed or single bid.		Average CRI	
	(1)	(2)	(3)	(4)
Fragmentation (Bonding social capital)		0.143** (0.069)		0.140** (0.067)
Diversity (Bridging social capital)		−0.358*** (0.138)		−0.440*** (0.134)
Income/capita	−0.324** (0.131)	−0.351*** (0.129)	−0.323** (0.128)	−0.356*** (0.126)
N contracts (log)	−0.389*** (0.118)	−0.384*** (0.118)	−0.669*** (0.116)	−0.672*** (0.115)
Population (log)	−0.064 (0.112)	0.036 (0.131)	0.176 (0.110)	0.318** (0.128)
Rate iWiW use	0.042 (0.094)	−0.001 (0.094)	0.105 (0.092)	0.052 (0.092)
Mayor victory margin	0.176** (0.070)	0.173** (0.069)	0.174** (0.069)	0.169** (0.067)
% high school grads	0.170 (0.122)	0.348** (0.144)	−0.036 (0.120)	0.190 (0.140)
Distance to Budapest	−0.089 (0.078)	−0.204** (0.088)	0.048 (0.077)	−0.093 (0.086)
Share of pop. inactive	−0.456*** (0.138)	−0.440*** (0.138)	−0.430*** (0.135)	−0.422*** (0.134)
Unemployment Rate	0.058 (0.079)	0.064 (0.078)	−0.017 (0.078)	−0.011 (0.076)
% population 60+	0.358*** (0.108)	0.329*** (0.107)	0.283*** (0.106)	0.251** (0.104)
Has University	0.289 (0.204)	0.289 (0.208)	0.406** (0.200)	0.384* (0.202)
Constant	1.561*** (0.463)	1.540*** (0.464)	2.642*** (0.453)	2.652*** (0.451)
Observations	305	305	305	305
Adjusted R^2	0.106	0.129	0.143	0.175
F Statistic	4.271***	4.452***	5.628***	5.974***

Table 7.4. *Settlement-level regression results predicting two corruption risk indicators, including all towns issuing at least one contract a year on average from 2006 to 2014. Note that all features are standardized with mean 0 and standard deviation 1. Significance thresholds: *p<0.1; **p<0.05; ***p<0.01.*

Year	# Contracts	# Firms	Avg. # Bidders
1981	273	49	1.96
1982	287	43	1.96
1983	318	46	2.01
1984	339	55	1.94
1985	357	49	1.93
1986	378	49	2.02
1987	411	42	1.97
1988	419	41	1.83
1989	392	40	1.76
1990	331	43	1.73

Table 7.5. *Summary statistics of the Ohio school milk market by year.*

7.2 Cartels

7.2.1 Ohio School Milk Data

In this section of the appendix we report summary statistics about the Ohio school milk market in Table 7.5 and plot the co-bidding networks of the market annually in Figure 7.4. We plot the distributions of the sizes of groups detected in the co-bidding network by year in Figure 7.5.

Year	# Contracts	# Firms	Avg. # Bidders	Share Protested	Total Value(GEL)	Avg. Rel. Cost
2011	17,396	3,804	1.73	.003	1,327,227,101	.88
2012	18,575	4,048	1.75	.004	1,331,025,922	.88
2013	20,230	4,399	2.03	.012	1,665,836,758	.85
2014	22,122	4,884	2.02	.014	1,808,309,559	.86
2015	26,033	5,600	2.02	.023	2,302,110,968	.87
2016	28,092	6,191	2.15	.031	2,497,797,345	.87

Table 7.6. *Summary statistics of the public contracting market of the Republic of Georgia by year. Share protested refers to the share of contracts legally protested by firms, Avg. Rel. Cost and StDev. Rel. Cost refer to the average cost of a contract, scaled by the maximum reserve price, and the standard deviation of the same, respectively. 1 Georgian Lari (GEL) equals roughly .6 US Dollars from 2011-2014, then .45 in 2015-2016.*

7.2.2 Georgian Contracting Data

In Table 5.1 we report annual summary statistics on the Georgian public procurement market. In Figure 7.6 we plot the distribution of the sizes of the groups we detected in the Georgian co-bidding network each year.

	Suspicious Groups		Ordinary Groups		Differences	
Threshold = 20	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*M-W U*	*p-value*
Avg. Rel. Price	0.938	0.046	0.914	0.053	30211***	<0.001
Avg. CV^G_{price}	0.098	0.055	0.117	0.059	33470***	<0.001
Avg. $CV^G_{bidding}$	0.047	0.056	0.055	0.038	32306***	<0.001
Protest Rate	0.134	0.341	0.237	0.425	37516*	0.011
Threshold = 10	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*M-W U*	*p-value*
Avg. Rel. Price	0.955	0.045	0.924	0.056	140930***	<0.001
Avg. CV^G_{price}	0.075	0.067	0.105	0.068	158212***	<0.001
Avg. $CV^G_{bidding}$	0.034	0.062	0.050	0.056	167427***	<0.001
Protest Rate	0.077	0.266	0.113	0.317	226584	0.0791
Threshold = 5	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*M-W U*	*p-value*
Avg. Rel. Price	0.975	0.039	0.925	0.056	49270***	<0.001
Avg. CV^G_{price}	0.066	0.054	0.104	0.068	49663***	<0.001
Avg. $CV^G_{bidding}$	0.023	0.038	0.050	0.057	50594***	<0.001
Protest Rate	0.132	0.339	0.111	0.314	80909	0.0791

Table 7.7. *Robustness check of cartel screens applied to suspicious and ordinary groups of firms detected in the Georgia procurement market, 2011-2016. We vary the threshold of the minimum number of contracts bid on exclusively by members of the group (20, 10, 5, compared with 30 in the main text). We replicate the main findings in the text that cartel groups have higher average relative prices, and are more likely to have a low average coefficient of variation on bids for a contract. The finding that suspicious groups are more likely to legally protest the winnings of other group members is no longer statistically significant when we filter at 5 or 10 contracts.* * $p < .05$, ** $p < .01$,*** $p < .001$

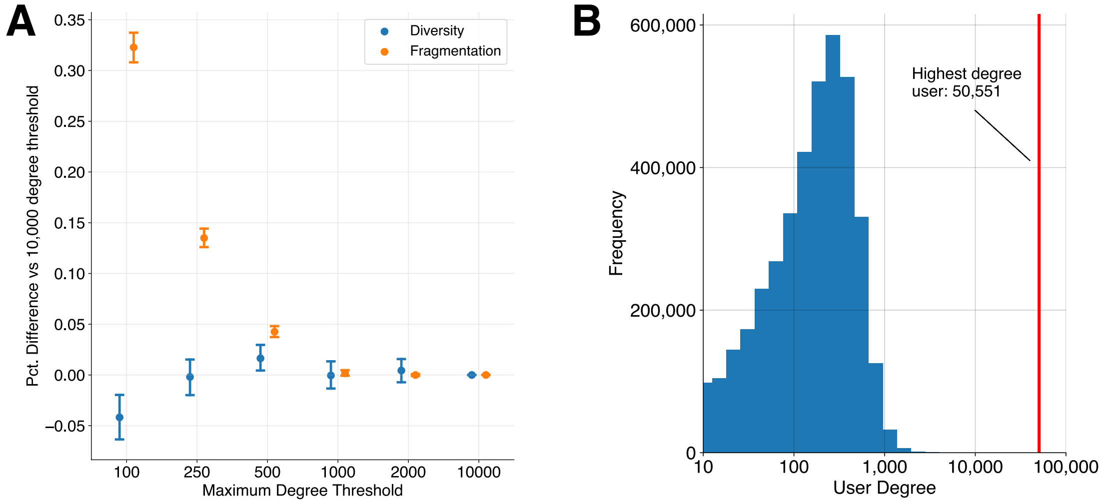

Figure 7.1. *A) The sensitivity of diversity and fragmentation to changing the maximum degree threshold, relative to the 10,000 degree threshold used in the paper. Error bars represent 95% confidence intervals. The measures are within 5% of the version we use in the paper for cutoffs at or above 500. B) The distribution of user connections on a log scale. Very few users (193) have more than 10,000 connections, while many (405,337) have more than 500.*

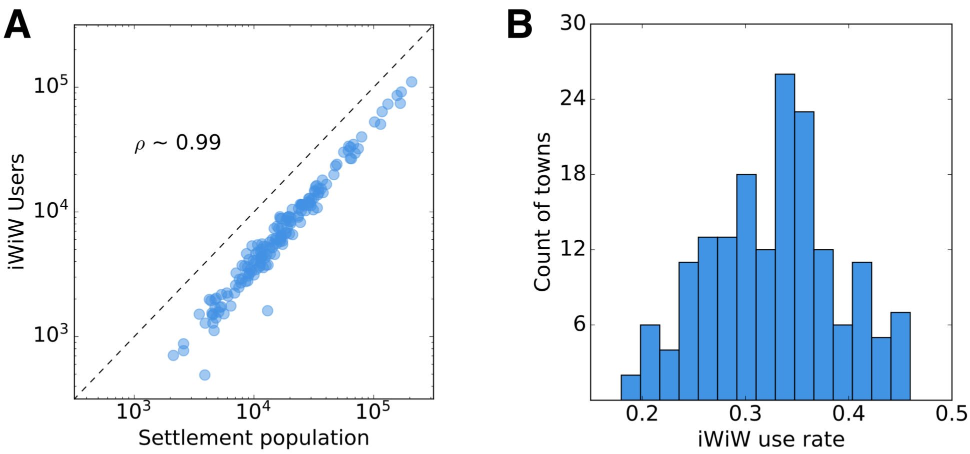

Figure 7.2. *A) Settlement population and number of iWiW users plotted on a log-log scale. B) iWiW use rate by settlements.*

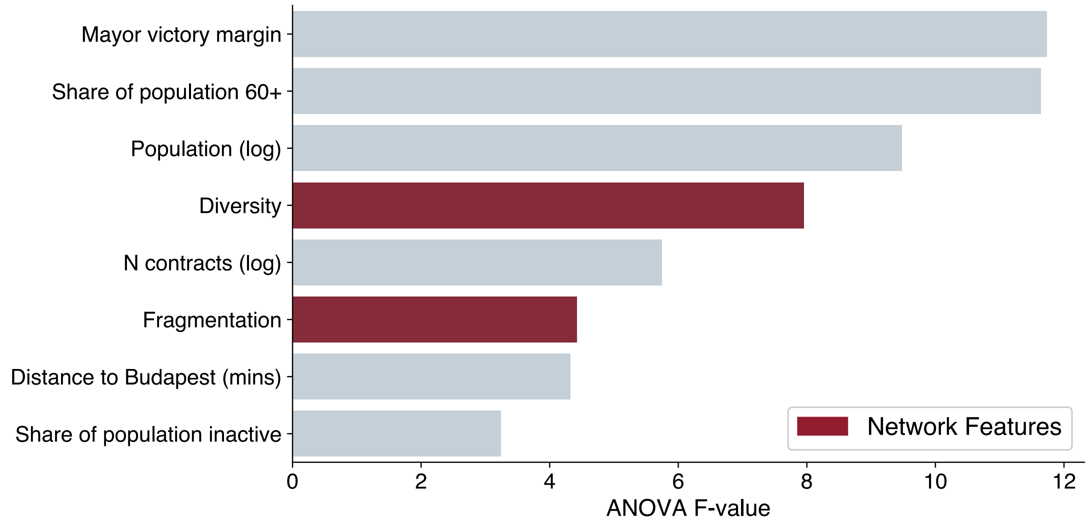

Figure 7.3. *Analysis of Variance F-test feature importances of OLS regression predicting average settlement CRI. We only include significant features, and highlight the network-based social capital measures.*

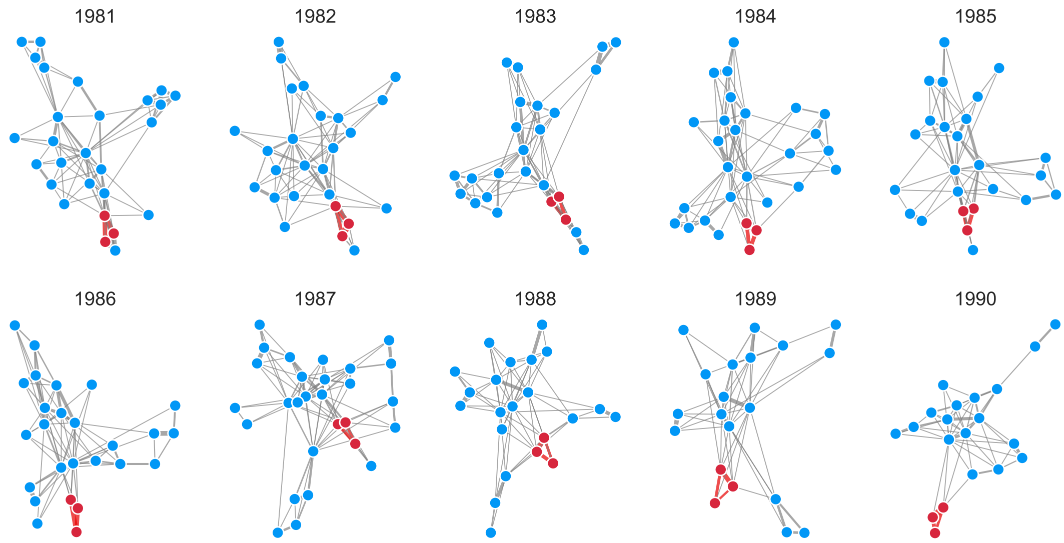

Figure 7.4. *Ohio school milk procurement market co-bidding networks, 1981-1990. Red nodes are members of the alleged cartel. For the purposes of visualization we filter out nodes participating in fewer than 5 auctions with other firms. Nodes are placed using a force-layout algorithm, with initial position equal to the final position of the nodes in the previous year.*

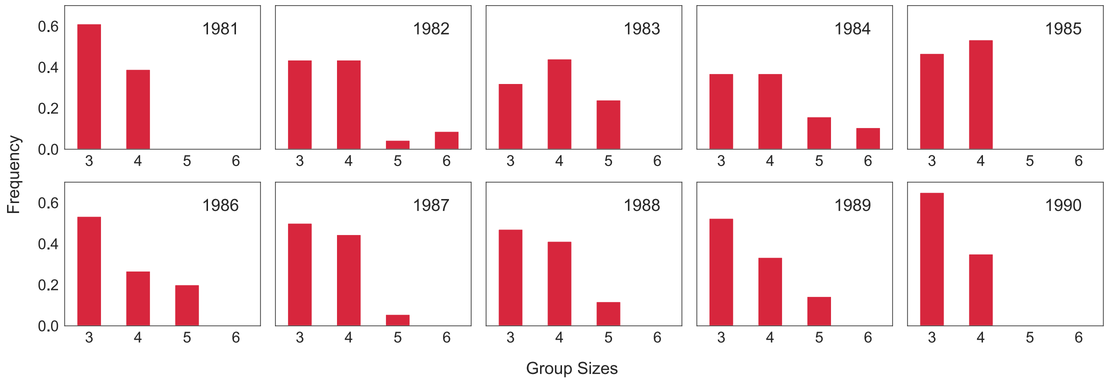

Figure 7.5. *Distributions of detected group sizes from the Ohio school milk contracting data, by year.*

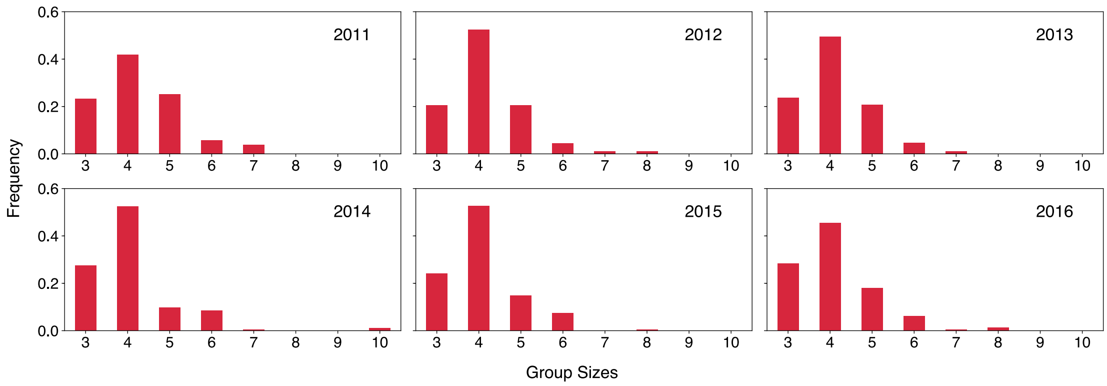

Figure 7.6. *Distributions of detected group sizes from the Georgian contracting data, by year.*