Community Detection as a Screening Device for Bid-Rigging: the case of Brazilian Procurement Tenders

Eduardo P.S. Fiuza *
IPEA

October 27, 2017

Please do not quote

Abstract

The present study proposes community detection methods — borrowed from the Social Network Analysis (SNA) literature — as a screening device for bid-rigging in Brazilian Federal procurement auctions. Underlying to this approach is the fact that not all the bidding ring members show up in all the auctions where that ring acts, so the Antitrust Authority has to compile information from all the auctions and then try to uncover each bidder's membership by observing their behavior across many auctions and lots, and also follow their presences in the auctions and how they interact with the other bidders along time. In the SNA language, *"communities, or clusters, are groups of vertices having higher probability of being connected to each other than to members of other groups, though other patterns are possible"* ((Fortunato and Hric, 2016)). We follow this approach as a first approximation to identify in a more formalized and structured way the groups of bidders that act in a concerted fashion across procurement auctions. We apply these techniques to a sample of tenders of pharmaceutical goods from a Brazilian Federal Procurement Data Warehouse.

Keywords: Bidding Rings; Bid-rigging; Collusion; Public Procurement; Social Network Analysis; Community Detection

JEL: L41; L14; H57; L13

*Av Pres. Antonio Carlos, 51 - 17o andar; Rio de Janeiro-RJ 20.020-010 - Brazil. E-mail: eduardo DOT fiuza AT ipea DOT gov DOT br.

Introduction

Bidding rings are collusive arrangements whose members deviate their bids from their competitive levels so as to maximize the joint profit of their members[1]. The figure of a knockout auction among the members illustrates that they pick the winner to be their member with the greatest value in the auction. This picked winner, in turn, has to repay their allies, either by means of side-payments or by taking turns in winning different auctions along time or across different buyers or objects. Side payments may be disguised through subcontracts or by having the ally join forces in a consortium, which translates into joint bidding in an auction. In a classic paper on bidding rings, (McAfee and McMillan, 1992) study the decision of bidders between taking part in a bidding ring or bidding competitively, depending on the transfer mechanisms available to the ring and on the one indeed utilized by them.

Now, what if participating in a bidding ring does not require the presence of all the members in a same auction?

In fact, participation costs may preclude the active members from bidding, as they do not have any prospect of winning the object. Thus, belonging to a bidding ring may be manifested by the member's absence, rather than their presence placing phony bids.

Participation may therefore be an outcome of the collusive strategy. For example, in the first Brazilian case of cartel denounced by a whistleblower (the first leniency agreement in Brazil), the final report states that the winning bidder chosen by lottery would choose the other firms who would charge higher prices[2].

In fact, the antitrust literature applied to auctions identify three strategies to implement collusive arrangements in auctions (see (Tóth, Fazekas, Czibik, and Tóth, 2014)):

1. Withheld bids
2. Losing (courtesy) bids
3. Joint bids

Many economic tests are applied to detect bid anomalies in order to screen for bid-rigging. Here we list some of the approaches that do not assume a prior knowledge of investigated bidding-rings. Some examples are:

[1] *"A bidding ring is a collection of bidders who collude in an auction in order to gain greater surplus by depressing competition"* — (Asker, 2010)

[2] Administrative procedure 08012.001826/2003-10, vol.32, p.142, or sheet 8271.

- (Kawai and Nakabayashi, 2015), in a sample of Japanese national government construction auctions, first shortlisted auctions where all (initial) bids fail to meet the reserve price and were rebid. They find out the identity of the lowest bidder is more persistent than expected, in contrast to a close to 50/50 odds that the third lowest bidder outbids the second. This suggests the presence of bidding rings.

- Another approach — and a highly cited one – is due to (Bajari and Ye, 2003). Starting from a general auction model with asymmetric bidders, the authors state two conditions for bids to be considered non collusive: (1) they should be conditionally independent; (2) their equilibrium distribution should be exchangeable. For (1) a conditional independence test such as the one proposed by (Su and White, 2008) or the one from (Jun, Pinkse, and Wan, 2010) could be performed. Regarding (2), tests of equality between pairs of coefficients in reduced form bid functions are to be used. Finally, as a third option, the authors collect information on structural cost parameters from experts and from this prior they use Bayes' theorem to compare between collusive and competitive models of industry equilibrium.

- Screens based on bid variance ((Abrantes-Metz, Froeb, Geweke, and Taylor, 2006)), collusive markers ((Harrington, 2006) or Benford's law (see (Abrantes-Metz, Villas-Boas, and Judge, 2011))

Except for collusive markers such as market share stability (which does not provide a statistical test), these approaches rely on information provided by firms that participated in the auctions, i.e., they build on the existence of courtesy bids. But, as I emphasized above, the investigator should bear in mind that not always does he observe all the members of the cartel in a single bid. They may have to track bidders across several auctions, and a couple of issues come out:

1. The patterns of distribution of bidders across the auctions may not be easily detectable by just pooling their names; in other words, the withheld bids are not easily accounted for in the available tests

2. Cartel members may alternate wins across different products of a common portfolio

Screens have been actively used by competition agencies around the world, such as the Netherlands, Mexico, Brazil [3], Australia, Bulgaria, Chile, Estonia, Hungary,

[3]Brazilian screening program *Projeto Cérebro* was useful, for example, to unveil the functioning

India, Israel, Japan, South Korea, Latvia, Lithuania, Peru, Russia, Sweden, Taiwan, Turkey and Ukraine ((OECD, 2014)).

These remarks lead us to improve the identification of cartel members by applying complex network tools borrowed by graph theory. In particular, the present paper shall apply community detection methods to identify the community structure in a complex networks. Detecting one or more communities amounts to find a graph partitioning, i.e., segmenting a set of elements into 'natural subsets. According to (Kolaczyk and Csárdi, 2014), *"a partition $\mathcal{C} = \{C_1, ...C_K\}$ of a finite set S is a decomposition of S into K disjoint, nonempty subsets Ck such that bigcup$^K_{k=1} C_k = S$. In the analysis of network graphs, partitioning is a useful tool for finding, in an unsupervised fashion, subsets of vertices that demonstrate a 'cohesiveness' with respect to the underlying relational patterns"*.(p.59)

1 Structural estimaton

As (Fortunato and Hric, 2016) warns, identifying communities is an ill-defined problem. Among the various approaches from this literature — which includes spectral methods and methods based on modularity optimisation, partition similarity measures, consensus clustering, artificial benchmarks, and edge clustering – the most attractive approach is the use of statistical inference. It provides a powerful set of tools to tackle the problem of community detection . The standard approach is to fit a generative network model on the data.

The use of generative models to infer modular structure in networks has thus been gaining increased attention in recent years, due to its more general character, and because it allows the use of more principled methodology when compared to the most common methods ((Peixoto, 2014a)). The most popular generative model used for this purpose is the so-called stochastic block model (SBM), which generalizes the notion of "community structure" in that it accomodates not only assortative connections, but also arbitrary mixing patterns, such as bipartite, multipartite, hierarhical and core-periphery structures, as well as disconnected components. This versatility, combined with analytic tractability, has made the blockmodel a popular tool in a number of contexts. The task of detecting communities turns into a process of statistical inference of the parameters of the generative model given the observed data, allowing for the use of statistical analysis' robust structure. SBMs fall in the general class of random graph models and have a long tradition of study in the social

of a bidding ring that rigged bids for protheses, ortheses and other medical supplies, announced in June 2017 (Administrative Procedure 08700.003699/2017-31). See press release in English.

sciences and computer science ((Karrer and Newman, 2011)). In fitting blockmodels to empirical network data as a way of discovering block structure in real networks, the social networks literature refer to them as *a posteriori blockmodeling*. As (Karrer and Newman, 2011) point out, however, the simple SBM does not work well in many applications to real-world networks. It does not account for the degree heterogeneity of most real networks, so it does a poor job at describing the group structure of many of them. Therefore, the authors proposed the Degree-Corrected SBM (DCSBM), in which the degrees of the vertices are kept constant, on average, via the introduction of additional suitable parameters. The most important drawback of this type of approach is the need to specify the number q of groups beforehand, which is usually unknown for real networks. A statistically principled solution for this problem is model selection is the minimum description length principle (MDL), which predicates that the best choice of model which fits a given data is the one which most compresses it, i.e. minimizes the total amount of information required to describe it ((Peixoto, 2013)).

SBMs are very versatile: they can be extended to a variety of contexts, e.g., directed networks ((Peixoto, 2014c)), networks with weighted edges ((Aicher, Jacobs, and Clauset, 2015)), with overlapping communities ((Airoldi, Blei, Fienberg, and Xing, 2008)), with multiple layers ((Peixoto, 2015)), with annotations ((Clauset, Moore, and Newman, 2007), (Hric, Peixoto, and Fortunato, 2016)). Besides, the procedure can be applied to any network model with group structure, not necessarily SBMs. The choice between alternative models can be done via model selection. A posteriori block modelling is not among the fastest techniques available. Networks with millions of vertices and edges could be investigated this way, but very large networks remain out of reach. Fortunately, many networks of interest can be attacked. The biggest problem of this class of methods, i.e., the determination of the number of clusters, seems to be solvable ((Hric, Peixoto, and Fortunato, 2016)).

For the present application, we estimate a DCSBM Maximum Likelihood on data from DW Comprasnet (Siasg) — the main repository of public procurement information in Brazil – using Peixoto's Python package **graph-tool** (`https://graph-tool.skewed.de/static/doc/dev/community.html`).

2 The model[4]

2.1 Background for SBM

Given a partition $\mathbf{b} = \{b_i\}$ of the network into B groups, where $b_i \in [0, B-1]$ is the group membership of node i, we define a model that generates a network $\mathbf{G}$ with a probability $P(\mathbf{G}|\theta, \mathbf{b})$, where θ are additional model parameters. Thus, if we observe a network $\mathbf{G}$, the likelihood that it was generated by a given partition $\mathbf{b}$ is obtained by the Bayesian posterior:

$$P(\mathbf{b}|\mathbf{G}) = \frac{\sum_\theta P(\mathbf{G}|\theta, \mathbf{b}) P(\theta, \mathbf{b})}{P(\mathbf{G})} \tag{1}$$

where

- $P(\theta, \mathbf{b})$ is the prior likelihood of the model parameters, and
- $P(\mathbf{G}) = \sum_\theta P(\mathbf{G}|\theta, \mathbf{b}) P(\theta, \mathbf{b})$ is called the model evidence

But there is only one choice of θ compatible with the generated network, such that the equation above simplifies to

$$P(\mathbf{b}|\mathbf{G}) = \frac{P(\mathbf{G}|\theta, \mathbf{b}) P(\theta, \mathbf{b})}{P(\mathbf{G})} \tag{2}$$

with θ above the only choice compatible with $\mathbf{b}$ and $\mathbf{G}$. The inference consists in either finding a network partition that maximizes 2, or sampling different partitions according to its posterior probability. The latter is called a minimization of the description length. In fact, we can write 2 as

$$P(\mathbf{b}|\mathbf{G}) = \frac{exp(-\Sigma)}{P(\mathbf{G})} \tag{3}$$

where

$$\Sigma = -lnP(\mathbf{G}|\theta, \mathbf{b}) - lnP(\theta, \mathbf{b}) \tag{4}$$

is the so-called **description length** of the network $\mathbf{G}$.Therefore choosing to maximize the posterior likelihood of 2 is fully equivalent to minimizing the description length. This approach corresponds to an implementation of Occam's razor, whereby the simplest model is selected, among all options with the same explanatory power. The selection is based on the statistical evidence available, and therefore will not overfit, i.e., mistake stochastic fluctuations for the actual structure.

[4]This subsection is strongly based on (Peixoto, 2014b).

2.2 Nested Degree-Corrected SBM for weighted directed graphs

Despite its generality, the traditional model assumes that the edges are placed randomly inside each group, and as such the nodes that belong to the same group have very similar degrees. This is often a poor model for networks with a highly heterogeneous degree distribution. The *degree-corrected* SBM ((Karrer and Newman, 2011) extends the traditional model to add the degree sequence $k = k_i$ of the graph as an additional set of parameters.

Proceeding further, we note that the SBM cannot be used to find relatively small groups in very large networks: the maximum number of groups that can be found scales as $B_{max} \sim \sqrt{N}$, where N is the number of nodes in the network, if Bayesian inference is performed. In order to circumvent this, we need to replace the noninformative priors used by a hierarchy of priors and hyperpriors, which amounts to a *nested SBM*, where the groups themselves are clustered into groups, and the matrix e of edge counts is generated from another SBM, and so on recursively ((Peixoto, 2014c)).

Last, but not least,we notice that our dataset contains edges with a very skewed distribution of weights. The SBM can be extended to cover these cases by treating edge weights as covariates that are sampled from some distribution conditioned on the node partition ((Aicher, Jacobs, and Clauset, 2015); (Peixoto, 2017)), i.e.,

$$P\left(\mathbf{x}, \mathbf{G}|\mathbf{b}\right) = P\left(\mathbf{x}|\mathbf{G}, \mathbf{b}\right) P\left(\mathbf{G}|\mathbf{b}\right) \tag{5}$$

where $P\left(\mathbf{G}|\mathbf{b}\right)$ is the likelihood of the unweighted SBM described previously, and $P\left(\mathbf{x}|\mathbf{G}, \mathbf{b}\right)$ is the integrated likelihood of the edge weights:

$$P\left(\mathbf{x}, \mathbf{G}|\mathbf{b}\right) = \prod_{r \leq s} \int P\left(\mathbf{x}_{rs}|\gamma\right) P\left(\gamma\right) d\gamma \tag{6}$$

where $P\left(\mathbf{x}_{rs}|\gamma\right)$ is some model for the weights $\mathbf{x}_{rs}$ between groups (r, s), conditioned on some parameter γ, sampled from its prior $P\left(\gamma\right)$. The posterior partition distribution is then simply

$$P\left(\mathbf{b}|\mathbf{G}, \mathbf{x}\right) = \frac{P\left(\mathbf{x}|\mathbf{G}, \mathbf{b}\right) P\left(\mathbf{G}|\mathbf{b}\right) P\left(\mathbf{b}\right)}{P\left(\mathbf{G}, \mathbf{x}\right)}, \tag{7}$$

which can be sampled from, or maximized, just like with the unweighted case, but will use the information on the weights to guide the partitions. For the data in hand, we initially use a discrete binomial distribution for the weights, although other suitable distributions may be also tested in the next versions.

3 The data

Data Warehouse Comprasnet is the greatest repository of public procurement data in Brazil. It collects information on procurement tenders, procuring agencies, winning bidders, individual bids, contracts and their amendments, inspections, suppliers' identities, auctioneers, approving personnel, dates of publication of bidding documents, auction timestamps, and payments authorized and cleared, among others. We collected the bidders participating in tenders for all medical goods (drugs, vaccines, serum, and medical equipment and material)[5]. Second, we take as directed edges the pairs pointing from each participating bidder in a lot to the winner[6] — a lot comprises a number of units of a same good, defined by a detailed procurement catalog code description. Using a directed graph allows us to focus on the relationships that really matter: if one bidder lets another win, this is the most obvious manner that they have to repay one another for winning a previous or contemporaneous lot. In fact, that is how (Ishii, 2009) model collusion in a Japanese road paving cartel.

We understand thus that examining loser-winner relationships is better than examining simple co-bidding (i.e., collect all pairs of bidders that participate in a same auction lot) In fact, phantom bidders may show up in the tenders and never win a single lot. On the other hand, this simple metrics is unable to uncover other more subtle arrangements. For example, as we mentioned before, subcontracting is also a means of disguising side-payments, and yet no subcontractor is ever recorded in DW Comprasnet. Another limitation of our present data is that bidding rings may pervade other levels of government: bidders may alternate awards across Federal, State and Municipal tenders; adding State tender portals is therefore the next step in our broader research project.Last, but not least, both across other levels of government and within a same level,the classification of goods may differ (see (Carvalho, de Paiva, da Rocha, and Mendes, 2014)).

The caveats above apply only in part to the results presented herein. In fact, we study here tenders for pharmaceuticals; since 2004 the classification of the objects

[5] We may aggregate the catalog codes into the so-called Descriptive Standards for Materials (PDM in Portuguese). A PDM item ordinarily corresponds to an active ingredient, such as beclometasone dipropionate, but it may as well comprise a class of goods, such as coagulation factor concentrates, human imunoglobulin or albumin, radiological contrasts, parenteral solutions or lung surfactants.

[6] It is worth noting that in the bidding data the bidder's identification is the taxpayer ID, which describes the plant or warehouse that will ship the merchandise and therefore issue the invoice. We aggregated all plants and warehouses to their respective parent firms by truncating their taxpayer IDs from 16 to 8 digits. This, however, is still insufficient to take into account cross-ownership. We might have a solution for that in the next version of this paper, if we are able to successfully include partners' information in another layer of the network.

follow a strict Standard for Description of Materials (PDM). We do not expect misclassification for this particular subset of tendered goods. Still, some of the goods tendered are also purchased by States and Municipalities, and neither we nor *Projeto Cérebro* have merged those data to the Federal ones; a semantic modelling is now in course; a shortcut is the Health Price Databank, from the Ministry of Health; States and Municipalities inform manually their awards and prices obtained, and classify the goods according to the Federal Procurement Material Classification (Catmat); we may use it for training the model. Another constraint to be borne in mind is for collecting the edges between pairs of bidders, we can only rely on electronic auctions (*pregões*). Notice that our sample span from 2008, when electronic auctions became the main procurement format in awarded value, until 2017. See Table Number of Lots - Government Pharmaceutical Purchases Through Federal Platform.

Table 1: Number of Lots - Government Pharmaceutical Purchases Through Federal Platform

Ref Year	Invited Tender	Restricted Tender	Open Tender	International Open Tender	Hybrid Auction	Negotiated Procedure	Ineligibility of Tender
2000	10.330	5.056	5.370	24	2	18.101	217
2001	19.026	7.812	8.992	45	2.197	38.632	426
2002	15.152	7.508	8.573	26	4.594	33.517	306
2003	20.286	7.204	7.825	7	9.180	29.083	156
2004	20.940	7.147	8.353	3	13.953	25.025	109
2005	11.131	1.724	4.445		20.983	21.253	175
2006	6.071	397	279		33.480	18.469	133
2007	3.718	134		3	32.746	16.523	113
2008	984				41.399	17.982	89
2009	141			6	48.958	19.523	113
2010	66			6	43.739	16.419	116
2011	140			7	50.841	13.565	150
2012	160			1	50.438	14.729	226
2013					55.146	16.639	258
2014					59.146	12.334	214
2015				1	59.902	11.474	303
2016					60.665	10.804	207

Source: Data Warehouse Comprasnet.

4 Preliminary results

4.1 SBM versus DCSBM

Our first results refer to both the traditional SBM and the Degree-Corrected SBM (DCSBM). The former found 89 blocks or communities, while the latter found 78. The entropy difference was -3.2×10^4, which demonstrates an enormous improvement of fit due to the degree-correction.

The plots for both models indicate a number of small isolated components, and an unclear and intricate mix of groups. See Figures 1 and 2

Figure 1: Traditional SBM
SBM.png

Source: DW Comprasnet. Own estimates.

Figure 2: Degree Corrected SBM

Source: DW Comprasnet. Own estimates.

4.2 Nested SBM with weighted edges

The Nested SBM allowing for different edge weights was also run with and without degree-corrected fit. The entropy difference decreased somewhat, to -1.9×10^4. The DCSBM without degree-correction found five hierarchical levels, while the degree-corrected version found six levels. See figures 3 and 4.

5 Interim Concluding Remarks

Detection of communities using tools from the literatures of social and complex networks is a promising approach for Antitrust Practitioners as a screening device for bidding rings. In fact, when the statistical tests most used in the Economics literature are applied, the enforcer has already some prior knowledge, either by some dennounce or whistleblowing, that a particular market has been affected. Auctioneers of several levels of government involved in public procurement were trained by the Brazilian Antitrust System in how to detect abnormal behaviors from bidders. It so happens that bid rotations and other schemes to distribute auction lots among members of a bidding ring may follow sophisticated patterns that the auctioneer and other offers may have difficulties in noticing, because he or she must take a distance from the ordinary day-to-day activities to be able to visualize the formation and functioning of such cartels. What we propose is that along a reasonable time frame the Antitrust Screening Officer collects the ordered pairs loser-winner in e-auction lots, whenever the information on all bidders' identities are available, and start testing for the existence of distinct communities. Once a small, active and cohesive community is detected, econometric tests may then be called for and applied to the distribution of bids in that subset of auctions. This is work in a very preliminary stage. Extensions available under scrutiny are:

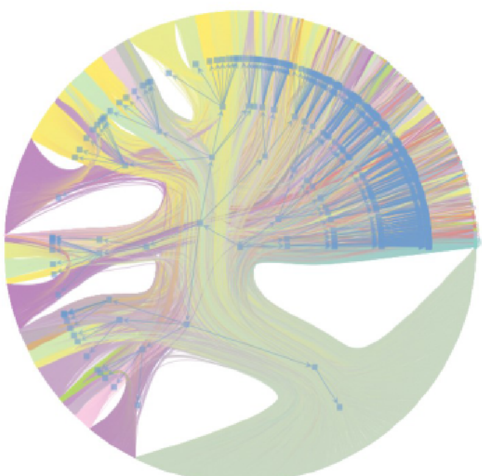
1. Run community detection for chunks of particular product codes which may uncover more cohesive bidding rings

2. Append to the data set other sources of information, especially from State and Municipality auctions, whenever viable

3. Introduce multilayered networks, such as cross-ownership or spatial proximity links (e.g. suppliers located in a same building or Zip Code).

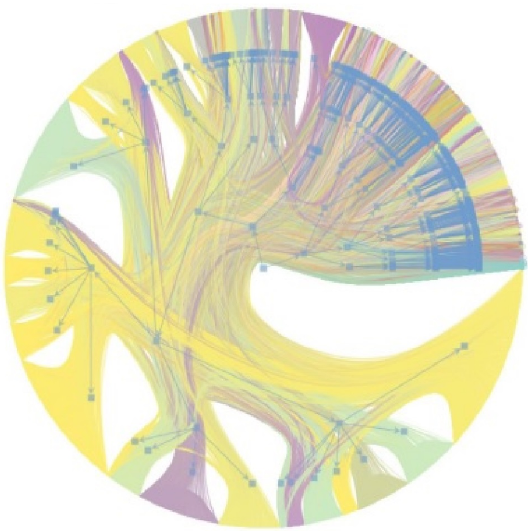
4. Split the sample into a sequence of years, model link formation and test for preferential attachment

Figure 3: Nested SBM, weighted edges
SBM-Weighted.jpg



Source: DW Comprasnet. Own estimates.

Figure 4: Nested Degree Corrected SBM, weighted edges
DCSBM-Weighted.jpg



Source: DW Comprasnet. Own estimates.

5. Use as community detection tools the ordinary algorithms consolidated in the network literature, such as Random Walk and Cluster Betweeness (present in packages such as **igraph**, available both for **R** and Python)[7].

References

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

Abrantes-Metz, R. M., S. B. Villas-Boas, and G. Judge (2011): "Tracking the Libor rate," *Applied Economics Letters*, 18(10), 893–899.

Aicher, C., A. Z. Jacobs, and A. Clauset (2015): "Learning latent block structure in weighted networks," *Journal of Complex Networks*, 3(2), 221–248.

Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008): "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, 9(Sep), 1981–2014.

Asker, J. (2010): "Bidding rings," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.

Bajari, P., and L. Ye (2003): "Deciding between competition and collusion," *The Review of Economics and Statistics*, 85(4), 971–989.

Carvalho, R., E. de Paiva, H. da Rocha, and G. Mendes (2014): "Using Clustering and Text Mining to Create a Reference Price Database," *Learning and NonLinear Models*, 12(2014), 38–52.

Clauset, A., C. Moore, and M. E. Newman (2007): *Structural inference of hierarchies in networks*pp. 1–13. Springer.

Fortunato, S., and D. Hric (2016): "Community detection in networks: A user guide," *Physics Reports*, 659, 1–44.

Harrington, J. E. (2006): "Behavioral screening and the detection of cartels," *European competition law annual*, pp. 234–251.

[7]Results may also be quite varied. Preliminary results using **igraph** for **R** in a subsample of auctions for the top 50 PDMs yielded 721 blocks with Cluster Betweenness and 291 using Walktrap. A lot of work is faced ahead for improving robustness.

Hric, D., T. P. Peixoto, and S. Fortunato (2016): "Network structure, metadata, and the prediction of missing nodes and annotations," *Physical Review X*, 6(3), 031038.

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

Jun, S. J., J. Pinkse, and Y. Wan (2010): "A consistent nonparametric test of affiliation in auction models," *Journal of Econometrics*, 159(1), 46–54.

Karrer, B., and M. E. Newman (2011): "Stochastic blockmodels and community structure in networks," *Physical Review E*, 83(1), 016107.

Kawai, K., and J. Nakabayashi (2015): "Detecting Large-Scale Collusion in Procurement Auctions," Mimeo.

Kolaczyk, E. D., and G. Csárdi (2014): *Statistical analysis of network data with R*, vol. 65. Springer.

McAfee, R. P., and J. McMillan (1992): "Bidding rings," *The American Economic Review*, pp. 579–599.

OECD (2014): "Ex officio cartel investigations and the use of screens to detect cartels," Discussion Paper DAF/COMP(2013)27, Organisation for Economic Cooperation and Development – Competition Committee, Paris.

Peixoto, T. P. (2013): "Parsimonious module inference in large networks," *Physical review letters*, 110(14), 148701.

——— (2014a): "Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models," *Physical Review E*, 89(1), 012804.

Peixoto, T. P. (2014b): "The graph-tool python library," *figshare*.

Peixoto, T. P. (2014c): "Hierarchical block structures and high-resolution model selection in large networks," *Physical Review X*, 4(1), 011047.

——— (2015): "Inferring the mesoscale structure of layered, edge-valued, and time-varying networks," *Physical Review E*, 92(4), 042807.

——— (2017): "Nonparametric weighted stochastic block models," *arXiv preprint arXiv:1708.01432*.

Su, L., and H. White (2008): "A nonparametric Hellinger metric test for conditional independence," *Econometric Theory*, 24(4), 829–864.

Tóth, B., M. Fazekas, Á. Czibik, and I. J. Tóth (2014): "Toolkit for detecting collusive bidding in public procurement. With examples from Hungary," Discussion Paper Working Paper series: CRC-WP/2014:02, Corruption Research Center Budapest.