Chapter 4
Predicting Corruption Convictions Among Brazilian Representatives Through a Voting-History Based Network

Tiago Colliri and Liang Zhao

Abstract While analyzing voting data concerning almost 30 years of legislative work from Brazilian representatives, we have noticed the formation of some sort of "corruption neighborhoods" in the resulting congresspeople network, indicating the possible existence of an (until then) unsuspected relationship between voting history and convictions of corruption or other financial crimes among legislators. This finding has motivated us to develop a predictive model for assessing the chances of a representative for being convicted of corruption or other financial crimes in the future, solely based on how similar are his past votes and the voting record of already convicted congresspeople. In this study, we present the main results obtained from this investigation.

4.1 Introduction

Corruption affects the society negatively in several ways, from holding back businesses [1], to the waste of public spending and investment [2] and the weakening of democratic systems [3, 4]. Predicting the incidence of corruption, specially at the individual level, is a challenging task, because criminals constantly develop increasingly advanced mechanisms to cover their infractions. More recently, governments around the world have been taking measures to increase their transparency by making public administration data accessible to the population. This phenomenon had encouraged researchers and members of the society to develop new mechanisms to monitor and analyze such data, through multidisciplinary approaches, thus contributing to increase the levels of accountability in the public sector [5].

Network-based techniques have already proven successful in the analysis of politics-related data, such as in the legislators' relations through bill co-sponsorship

T. Colliri (✉)
Institute of Mathematics and Computer Science, University of Sao Paulo, Sao Carlos, Brazil
e-mail: tcolliri@usp.br

L. Zhao
Faculty of Philosophy, Science, and Letters, University of Sao Paulo, Ribeirao Preto, Brazil
e-mail: zhao@usp.br

O. M. Granados and J. R. Nicolás-Carlock (eds.), *Corruption Networks*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-030-81484-7_4

data [6, 7] and through roll-call voting data [8–12]. Additionally, this type of approach has also been applied on the analysis of crimes-related data, such as in the correlation between social capital and the risk of corruption in local governments contracts [13] and in the identification of missing links among the members of an Italian mafia group [14]. Another interesting work, still related to this type of application, used link prediction techniques to uncover hidden relations among Brazilian politicians cited on corruption scandals [15]. In a more recent work, a network-based approach has been applied for modeling the dynamics of a major corruption scandal occurred in Mexico involving embezzlement activities, contributing to provide systematic evidence on which corporate characteristics are likely to signal corruption in public procurement [16].

The term *complex network* refers to a large scale *graph* with non-trivial connection patterns [17]. Some examples of complex networks in the real world include the internet [18], biological neural networks [19], food chains [20], blood distribution networks [21] and power grid distribution networks [22]. Additionally, there are also various network-based models designed to perform *machine learning* related tasks, such as *clustering* [23–25], *classification* [26–29] and *regression* [30, 31]. More recently, there was the introduction of *temporal networks*, which allows the inclusion of the *time* dimension in the study of graphs. Examples of temporal networks that can be found in the real world include social networks, one-to-many information dissemination (e.g., emails or blogs), cell-biology networks, brain networks, traffic networks, and mobile communication networks [32].

In this work, we start by using a network-based approach to build a temporal network from voting data regarding almost 30 years of legislative work in the Brazilian House of Representatives. Each node in this network corresponds to a different legislator, who has voted at least once in the House during that period, and the edges between each pair of nodes are created according to the voting history similarity between them. Afterwards, we investigate whether this built network can be used to predict cases of conviction for corruption or other financial crimes among the congresspeople, based on the emerging topological structure formed by convicted congresspersons and their neighbors in the network. The dataset used in the application has been created especially for this work, comprising the votes of a total of 2,455 congresspeople and 3,407 legislative voting sessions, from 1991 until 2019. The results obtained from this study were originally published as part of the analyses made in [33] and, for this revised version, we have updated the figures, as well as inserted a new one, as Fig. 4.4 in Sect. 4.3, to display the legislators network used in the corruption prediction task. Additionally, we have also added Tables 4.1 and 4.2, in Sect. 4.2, to help demonstrate how the legislators temporal network is built, by using a simple voting dataset as example.

Table 4.1 Illustration showing voting sessions outcome and respective resulting legislators' temporal network slices, based on the voting history similarity

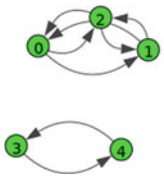
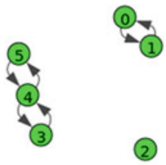
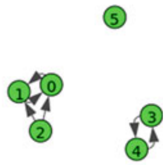
	Votes		
Voting Session	**1**	**2**	**3**
Legislator 0	No	Obstruction	Yes
Legislator 1	No	Obstruction	Yes
Legislator 2	No	Yes	Yes
Legislator 3	Yes	-	-
Legislator 4	Yes	No	Yes
Legislator 5	-	No	No
Resulting Network			

Table 4.2 Final values of weight matrix $W^{(t)}$ in the example voting dataset ($t = 3$). The values in the main diagonal are changed to 0 for avoiding self-loops in the built network

	0	1	2	3	4	5
0	0	3	1	–1	–3	0
1	3	0	1	–1	–3	0
2	1	1	0	–1	–3	0
3	–1	–1	–1	0	1	0
4	–3	–3	–3	1	0	0
5	0	0	0	0	0	0

4.2 Methodology

The methodology used in this study is described below. All network analyses performed in this study were implemented by using the igraph [34] and Teneto [35] Python packages.

4.2.1 Database

The data for our analysis are collected from the official website of the Brazilian House of Representatives [36]. The obtained datasets comprise the outcome of all voting sessions of legislative bills deliberated in the House, from May 22, 1991 until Feb 14, 2019. As preprocessing, we have performed a thorough data cleansing in the database to detect and fix possible errors, such as duplicated names among legislators and also misprints. Each session comprises the following information: the bill to be considered, the voting date, and the representatives who have attended the session and voted. Additionally, the following information is provided for each voter:

- ID (a unique number for each congressperson),
- Full Name,
- Political Party, and
- Vote.

The voting system in the Brazilian House consists of four types of votes:

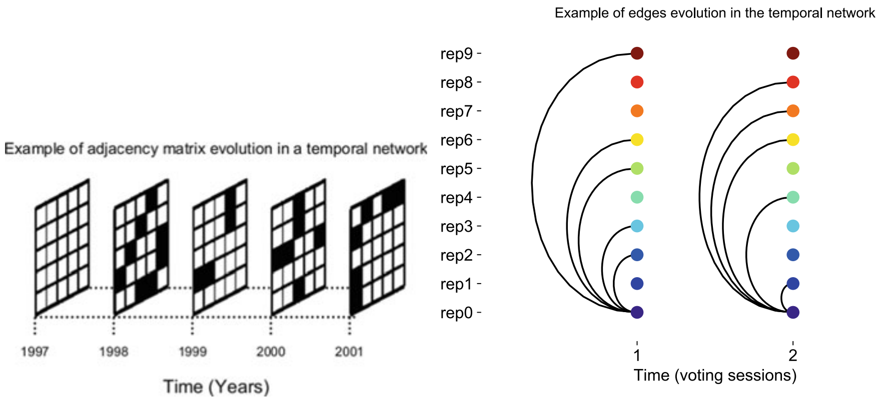
- *Yes*: if the representative approves the bill;
- *No*: if the representative disapproves the bill;
- *Abstention*: if the representative deliberately chooses to not take part in the voting;
- *Obstruction*: analogous to abstention, except that abstention counts for *quorum* effects, while obstruction does not count for it.

The final database contains the voting outcome from a total of 1,656,547 votes from 3,407 sessions, and 2,455 different congresspersons. To perform our analyses, we map each of the voting sessions to a static network, where each node represents a legislator who attended that session and the edges between them are created according to their respective voting history similarity, pairwise. The Brazilian House currently has 513 seats, hence this is the maximum number of nodes in the legislators network resulting from each voting session.

Additionally, we also made an extensive research in several media vehicles to collect data concerning which congresspersons in the database have already been officially convicted for corruption or other financial crimes, such as: embezzlement, improbity, misappropriation of public funds, money laundering, peculation, or crimes against the Public Administration. This task resulted in the identification of 21 legislators who have been convicted for corruption and 12 legislators who have been convicted for other financial crimes, out of the total 2,455 congresspeople in the database. This additional information was confirmed from legitimate Brazilian judiciary sources as well, such as the Federal Supreme Court (Supremo Tribunal Federal) [37].

4.2.2 Legislators Temporal Network Generation

A network can be characterized as graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of nodes and $\mathcal{E}$ is a set of tuples representing the edges between each pair of nodes $(i, j) : i, j \in \mathcal{V}$. The edges in $\mathcal{E}$ are usually provided in the form of a square matrix M, with size equal to the number of nodes in the network and binary values, in case of unweighted graphs. In order to convert a list of static networks into a temporal network $\mathcal{G} = \{G^{(t=0)}, G^{(t=1)}, G^{(t=2)}, \ldots, G^{(t=n)}\}$, where t is the time step, one then needs to insert a new dimension $\mathcal{D}$ in the formal network definition, such that it becomes $\mathcal{G} =$



(a) Adjacency matrix evolution (b) Edges evolution

Fig. 4.1 **a** Example of the adjacency matrix evolution in a temporal network whose dimension $\mathcal{D}$ is measured in terms of years. **b** Edges evolution. Illustration showing the graphlets evolution in time (in this work measured in terms of voting sessions). When $t = 1$, node 0 is connected to nodes 2, 3, 5, 6 and 9. Following, when $t = 2$, it disconnects from nodes 2, 3, 5 and 9 and becomes connected to nodes 1, 4, 7 and 8

$(\mathcal{V}, \mathcal{E}, \mathcal{D})$, where $\mathcal{D}$ represents the temporal network slices. This process is illustrated in Fig. 4.1. One way of achieving this is by generating a matrix for representing the edges in $\mathcal{E}$ as a *triplet* $(i, j, t) : i, j \in \mathcal{V}, t \in \mathcal{D}$, also known as dynamic *graphlets* [38] (Fig. 4.1b). The final result from this conversion process is a *multilayer network*, with each layer representing a static temporal slice of a single main graph (Fig. 4.1a) and, for cases when the dimension $\mathcal{D}$ is a set of indices ordered by time, one can also refer to this graph as a *temporal network* [35].

In this work, we start by converting each voting session in the database into a static network, based on their "date" attribute, in ascending order. For each session, a square votes matrix $M^{(t)}$, of size $d \times d$, is created, where t is the session's sequential number, or time step, and d is the total number of legislators who attended that session. Hence, as mentioned earlier, the maximum possible value for d is 513, because this is the total number of seats in the Brazilian House. Each element $M_{ij}^{(t)}$ has a binary value, assuming:

$$M_{ij}^{(t)} = \begin{cases} 1, & \text{if congresspersons } i \text{ and } j \text{ voted alike in session } t, \\ -1, & \text{otherwise.} \end{cases} \tag{4.1}$$

The values of each voting matrix $M^{(t)}$ are accumulated in a distinct weight matrix $W^{(t)}$, whose size is equal to the total number of legislators who have registered at least one vote in the House, until session t. Therefore, each element $W_{ij}^{(t)}$ from this matrix returns the accumulated weight between congresspersons i and j, or the *voting history similarity* between them, until session t. Formally, the current value of each element $W_{ij}^{(t)}$ is yielded by:

$$W_{ij}^{(t)} = \sum_{t} M_{ij}^{(t)}. \tag{4.2}$$

Note that, from Eq. 4.2, the value of $W_{ij}^{(t)}$ will range from $-t$, in case congressperson i have always voted different from the representative j, until $+t$, when congressperson i have always voted similarly to the representative j, until session t. In this study, we are assuming that, in the former case, legislators i and j have complete opposite political views, while, in the latter case, legislators i and j are very politically aligned, up to the instant t.

The static slices $G^{(t)}$, for each voting session t, which will form the legislators temporal network $\mathcal{G}$, are obtained from the values in weight matrix $W^{(t)}$. For this end, we opt to connect each legislator to the one(s) with the highest weight(s) associated to him in $W^{(t)}$, i.e., each legislator will be connected to the representative(s) who are most politically aligned to him, in terms of their voting history similarity, up to the instant t. Thus, the edges between each pair of nodes representing congresspersons i and j, in each temporal network slice $G^{(t)}$, are yielded by:

$$G_{ij}^{(t)} = \begin{cases} 1, & \text{if } W_{ij}^{(t)} = \max\limits_{\forall x \in W_i^{(t)}} x \\ 0, & \text{otherwise} . \end{cases} \tag{4.3}$$

Note that, from Eq. 4.3, most of the nodes in $G^{(t)}$ will have only one outbound edge, connecting it to the node most politically aligned to it. The only exceptions for this rule are the situations when $\max_{\forall x \in W_i^{(t)}} x$ returns more than one element, which in this case will result in two or more outbound edges originated from node i.

One can also consider, in this step of the technique, the possibility of binning the legislators voting similarities by predetermined time slices, such as per year or per presidential term, for instance. In our case, we have performed some preliminary tests considering this possibility, and the results from these tests indicated that, for this specific database, it is necessary to process a large number of voting sessions before a clear topological pattern emerges in the temporal network. A possible reason for this may be in the fact that most representatives serve for only 4 years, which is the term length in Brazil, and only few of them get reelected. Additionally, we noted many cases of officially elected congresspeople leaving their seats for interim successors, to run for higher political positions, such as governors or senators. This specific feature in our database also contributes to prevent the formation of a clear voting similarity pattern in the legislators temporal network's topology, in the short-term and medium-term. Therefore, we have opted for taking the legislators' complete voting history into consideration for generating the edges in the network.

In order to illustrate how the temporal network $\mathcal{G}$ is generated, let us now consider a simple dataset, comprising only 3 voting sessions, each one attended by 5 legislators (Table 4.1). In voting session 1, legislators 0, 1 and 2 voted "No", while legislators 3 and 4 voted "Yes". This outcome results in a network with two components, with all nodes being connected with equal weights (+1, in this case). In session 2, legislator

2 votes different from legislators 0 and 1, and legislator 3 is replaced by legislator 5, which votes similar to legislator 4. Now, the temporal network is updated, with legislator 2 isolating from the others (because his current voting similarity score is 0), and a new node is inserted to represent legislator 5, connected to legislator 4. Note that the node from legislator 3 still remains in the temporal network, in the same position it was before, since its voting similarity score remains unchanged. In the third voting session, legislators 0, 1 and 2 again vote alike, so legislator 2 reconnects to legislators 0 and 1 in the network, with a similarity score of +1 with both of them. Legislators 0 and 1 reach a maximum similarity score of +3 with each other, since they voted alike in all three sessions, so they are reciprocally connected. Legislator 5 is the only one who votes "No" in this session, which results in his isolation from the others since his voting similarity score now is 0. The final values of weight matrix $W^{(t)}$, i.e., when $n = 3$ in the example, are shown in Table 4.2.

It is important to stress that, as it is shown in the example from Tables 4.1 and 4.2, when a new congressperson is inserted in the network – as in the case of new legislative elections or because of a resignation, for instance – he or she does not inherit any voting data from the congressperson who previously occupied its node in the network (or "seat" in the House, so to speak). For such cases, a new row and a new column is inserted in weight matrix $W^{(t)}$ to record the voting similarity between the new node, representing the new congressperson, and all the other elements in $W^{(t)}$. Another important aspect to be emphasized in the legislators' temporal network building process is that we opt to not take into account the attribute "political party" from the legislators when generating the edges. We have decided to proceed this way because our goal, in this study, is to capture the political affinities among congresspeople beyond their party affiliations, solely by considering their voting records. This decision can be justified inasmuch as there are currently 33 political parties officially in Brazil, and this exaggerated number of parties eventually results in a consistent attenuation of the ideological differences between them.

4.2.3 Corruption Prediction Model

Let us now proceed to describe how we assess whether a representative is more or less likely to be convicted of corruption or other financial crimes in the future by analyzing the built voting-history-based legislators network. To accomplish this task, we make use of link prediction techniques applied on a subgraph of the congresspeople network, formed from a subset of the nodes comprising only already convicted representatives and their respective neighbors, i.e., the congresspersons identified as the most politically aligned to them, according to our model. The prediction tests are performed in a supervised learning manner, in which we take the top n links predicted by each considered technique whose source node is a convicted congressperson, and label their target nodes as being convicted ones as well. Following, we provide more details regarding the prediction model.

The final weight matrix $W^{(t)}$ from the built legislators network has a total of 2,455 nodes, representing all congresspeople in our database who have voted at least once in the House, from 1991 until 2019. While browsing this final resulting network, we noticed that, oftentimes, the neighbor of a convicted congressperson was a convicted one as well, seemingly forming some sort of "corruption neighborhoods" in the built network, so to speak. This unexpected emerging pattern motivated us to investigate this aspect further, by applying algorithms which consider the network topological structure for predicting missing links. Below, we list the required steps for the corruption prediction model evaluated in this study.

1. Create a subgraph from the built legislators network resulting from matrix $W^{(t)}$, comprising only convicted congresspeople and their respective neighbors, from both incoming and outgoing edges.
2. Convert the network resulting from the subgraph to an *undirected* one, and delete all existing links between convicted labeled nodes.
3. Apply the link prediction technique on the network.
4. Take the top n link predictions whose source node is convicted and label their target nodes as convicted ones as well.

The accuracy achieved by each link prediction technique is evaluated in a supervised learning manner, by inspecting how many of its target nodes are indeed labeled as convicted ones in the database. Please note that, by making use of link prediction techniques in this task, we are thus considering the legislators network topological structure for conviction prediction purposes.

The link prediction problem is often studied in social networks, and can be formulated as "to what extent can one predict which links will occur in a network based on older or partial network data" [39]. This task can be categorized into two different types: *missing link prediction*, which involves predicting links based on an incomplete or damaged version of a network, and *future link prediction*, which involves predicting links in a future snapshot of the network based on its current state [40]. The predictors, by their turn, can be categorized into two different groups, as listed below.

- *Local predictors*: based on the neighborhoods of the source and target nodes (e.g., Cosine [41, 42], Maximum and Minimum Overlap [43], NMeasure [44] and Pearson [45]).
- *Global predictors*: based on measures that evaluate how the source and target nodes may be related in the network, even if they do not share any common neighbors (e.g., Katz [46] and Rooted PageRank [47]).

In this study, we test the following 6 link prediction techniques in the corruption prediction task: Cosine, MinOverlap, NMeasure, Pearson, Rooted PageRank and Random (for comparison purposes). All prediction techniques are implemented by using the tool introduced in [40], with default parameter values.

4.3 Results

We start this section by presenting an example of one of the networks resulting from our analysis, in Fig. 4.2. This network is built from data of the voting session for bill PEC 77/2003, occurred on September 19, 2017, which was attended by all 513 Brazilian congresspeople. The outbound edges connect each congressperson to the one most politically aligned to him, based on their voting history similarity. Therefore we can say that the hubs in the network, within this context, are congresspersons who represent a "local majority" in terms of voting, or the majority within a local neighborhood. We have also applied a community detection technique in this figure, by using the fast greedy algorithm [48], through the tool introduced in [34], for the sake of highlighting the emerging "voting aggregation patterns" among congresspeople in the Brazilian House of Representatives. These voting patterns likely occur due to aggregation factors among the legislators, such as political parties, civic associations or interest groups.

In Fig. 4.3, we show the final resulting network from our analysis, originated from the final matrix $W^{(t)}$, with $t = 3,407$, i.e., the total number of voting sessions in the database. This graph comprises all 2,455 representatives in the database, who have voted at least once in the Brazilian House, from May 22, 1991 until February 14, 2019. The red nodes indicate the 33 congresspeople currently identified as convicted ones, in our research. While browsing the nodes of this network, we have noted that the neighbors of a convicted congressperson were oftentimes convicted ones as well. This unexpected emerging feature has motivated us to further investigate whether the nodes of convicted legislators indeed tend to stay close to each other in the network, thus forming some sort of "corruption neighborhoods", so to speak. For this end, we build a separated network, from a subgraph of this final network, containing only the nodes from convicted representatives, along with their respective neighbors, from both incoming and outgoing edges. It is worth noting that these neighbors may be convicted ones as well or not. As preprocessing, before applying the link prediction techniques, we convert this network to an *undirected* one and remove all existing links between two nodes labeled as convicted from the network (5 edges in total). The subgraph resulting from this preprocessing is shown in Fig. 4.4. It comprises 211 legislators, with 33 of them being convicted, and 1,374 edges.

In order to predict new conviction cases among representatives, we apply link prediction techniques in the network shown in Fig. 4.4. A total of 5 different link prediction techniques plus a Random method (for comparison purposes) were considered for this task. We take the top 10 links predicted by each technique, having convicted nodes as a source, and labeled their target nodes as being convicted ones as well. The accuracy of the model is assessed by inspecting how many of these predictions are correct, according to the convictions information in the database.

In Fig. 4.5, we display the accuracy achieved by each considered technique in the conviction prediction task. These results indicate that it is possible to achieve a high performance, in terms of corruption prediction, when considering the topological structure of the built legislators network. The Cosine, NMeasure and Pearson tech-

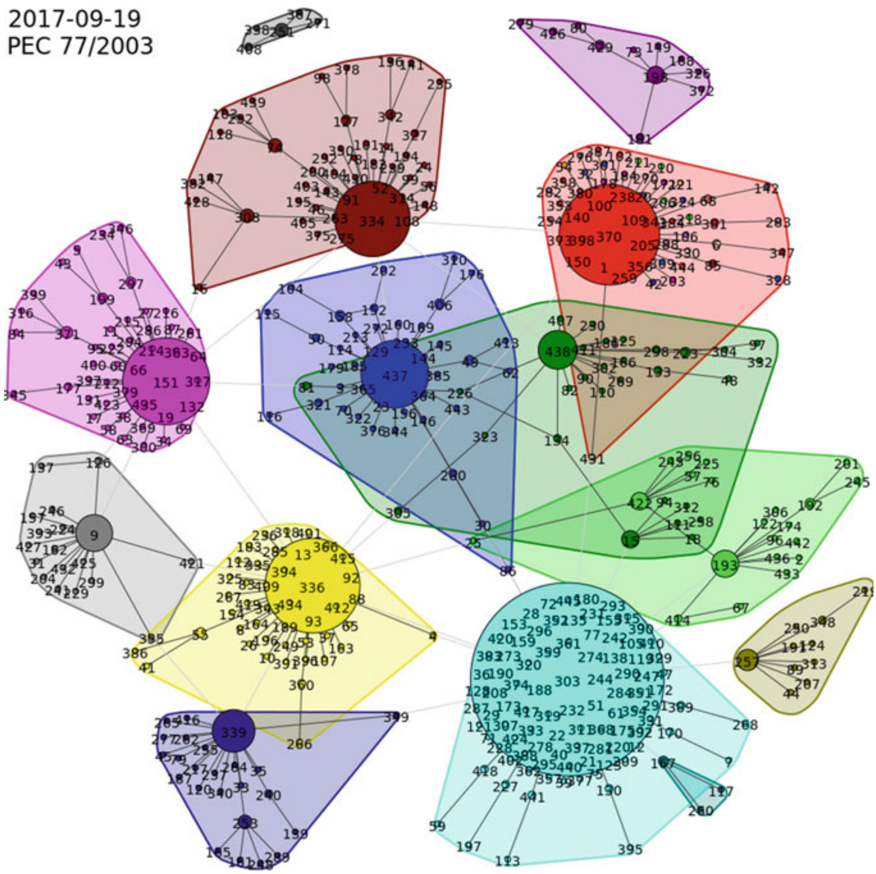


Fig. 4.2 Representation of one of the temporal network slices generated in our analysis, from data regarding voting session of bill PEC 77/2003, occurred on 2017-09-19. Each node represents one of the 513 congresspersons who have voted in this session, and the edges are created based on their voting history similarity. The communities were detected through the fast greedy algorithm. This graph was converted to an undirected one only for visibility purposes

niques were able to achieve an accuracy of 0.9 in this task, i.e., 9 out of the 10 links predicted by these techniques turned out to be correct, with their target nodes being convicted congresspersons. Rooted PageRank and MinOverlap techniques have presented slightly lower performances in this task, with an accuracy of 0.7 and 0.5, respectively. Note that the Random technique was not able to predict not even one link correctly in this task, which indicates both a higher level of difficulty involved in it and also that prediction techniques based on the graph topological structure are considerably more appropriate for this end.

To better understand the reason why some link prediction techniques have performed better than others in the corruption prediction task, we show, in Fig. 4.6, a

Fig. 4.3 Representation of the final network resulting from our analysis, with all 2,455 representatives in the database, who voted at least once in the Brazilian House of Representatives, from May 22, 1991 until Feb 14, 2019. The edges are created according to their voting record similarity. Convicted representatives are denoted by the red color (33 in total)

comparison between the top 10 links predicted by Cosine and Rooted PageRank techniques. Note that the target nodes from the links predicted by Cosine technique are all nearer their respective source nodes while, on the other hand, the target nodes from the links predicted by Rooted PageRank technique are often more far from their respective source nodes. The explanation for this difference is in the fact that techniques such as Pearson, Cosine and NMeasure are all *local predictors*, i.e., they take into account the neighborhood from each pair of nodes for assessing the possibility of existing a hidden link between them in the network. Thus, the fact that such type of predictors have performed better in the corruption prediction task is a strong sign that our initial hypothesis regarding the seemingly existence of "corruption neighborhoods" in the legislators network is correct, pointing in the direction that there is indeed a correlation between voting history and convictions of corruption or other financial crimes among Brazilian congresspeople.

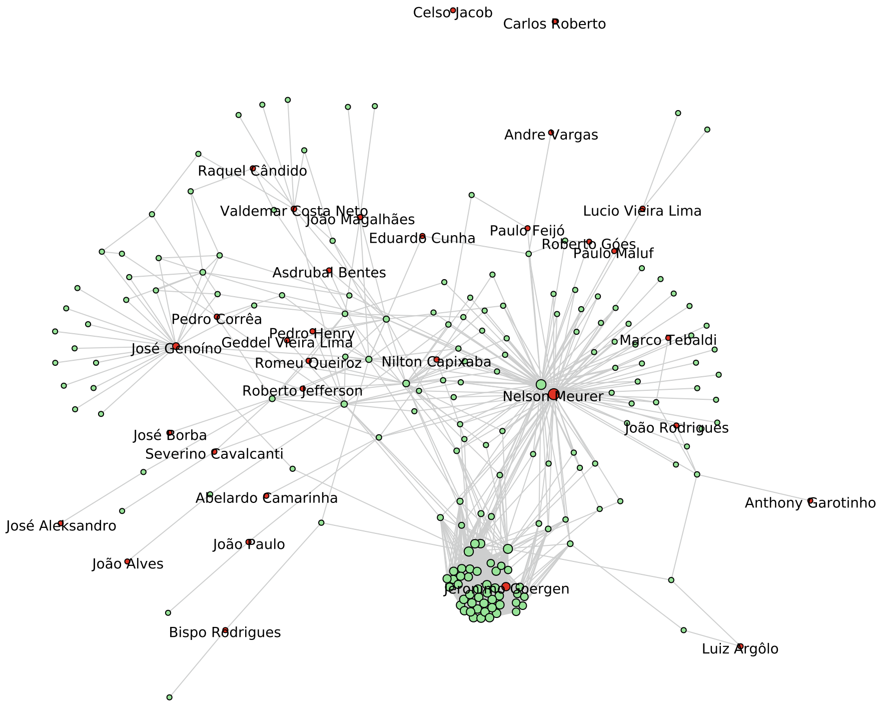


Fig. 4.4 Subgraph of the consolidated network, shown in Fig. 4.3, with only the 33 convicted representatives (in red) and their respective neighbors. The network is converted to an undirected one and links between two convicted congresspeople are suppressed (5 links in total), before applying the link prediction techniques. We opted for displaying only the names of legislators who were identified as officially convicted in this figure

4.4 Final Remarks

In this work, we introduce a technique to generate a temporal network from legislative voting data, and apply it on a dataset specially built for this study, comprising almost 30 years of votes from Brazilian congresspeople. We also investigate the possibility of using the resulting network for predicting corruption-conviction cases among legislators, by using link prediction techniques that consider the topological structure emerging from convicted congresspersons and their neighbors in the network. The obtained results in this task are encouraging, especially the ones achieved by techniques based on local predictors, i.e., based on the neighborhoods of the source and target nodes, which is an indication that there is indeed a correlation between voting history similarity and corruption convictions among Brazilian representatives.

Although in this study we have focused our analyses on aspects regarding corruption and other financial crimes among legislators, it is important to highlight that the congresspeople network resulting from the voting history similarity between them

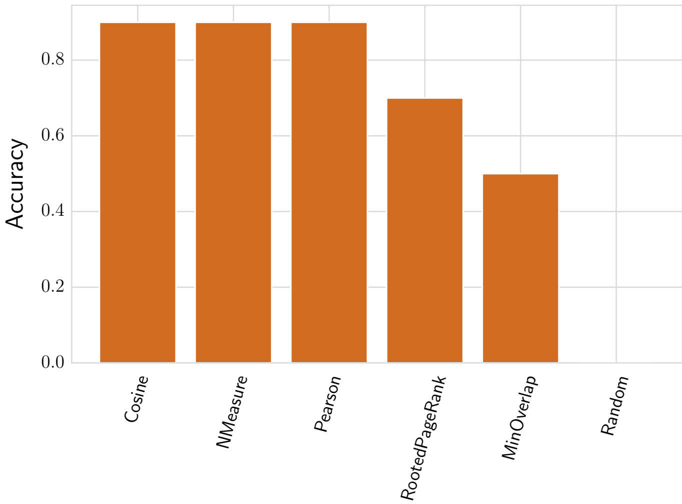


Fig. 4.5 Accuracy achieved by each link prediction technique in the conviction prediction task

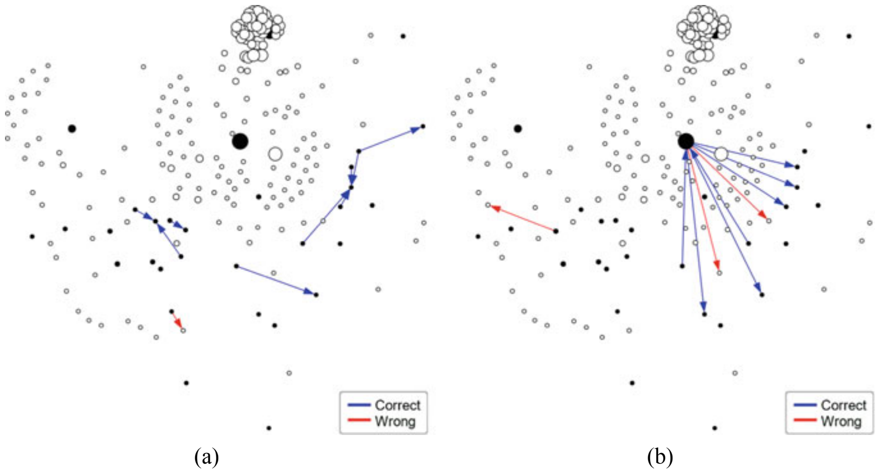


Fig. 4.6 Comparison between the link prediction outputs from **a** Cosine and **b** Rooted PageRank techniques. In **a** Links predicted by Cosine (9 correct, 1 wrong). In **b** Links predicted by Rooted PageRank (7 correct, 3 wrong). The black color denotes convicted legislators. A link prediction is considered correct when its target node is also labeled as convicted in the database. All other links are removed from the network for visibility purposes

may also be employed in other types of analyses, focused on the political aspect of the connections. Within this context, one can further explore, for instance, the emerging aggregations verified in the temporal network slices to perform analyses comparing the level of cohesive voting behavior between political parties, and how the changes in these aggregations may impact the legislative decision-making in the House. Or, as another possibility, one can also analyze the aggregation patterns in these network slices for evaluating and mapping the similarities and differences between political parties, in terms of legislative voting behavior.

Acknowledgements This work is supported in part by the Sao Paulo State Research Foundation (FAPESP) under grant numbers 2015/50122-0, the C4AI under grant number FAPESP/IBM/USP: 2019/07665-4, the Brazilian National Council for Scientific and Technological Development (CNPq) under grant number 303199/2019-9 and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

1. Wilhelm PG (2002) International validation of the corruption perceptions index: implications for business ethics and entrepreneurship education. J Bus Ethics 35(3):177–189
2. Tanzi V, Davoodi H (1998) Corruption, public investment, and growth. In: The Welfare State, Public Investment, and Growth. Springer, Berlin, pp 41–60
3. Linde J, Erlingsson GÓ (2013) The eroding effect of corruption on system support in s weden. Governance 26(4):585–603
4. Rose-Ackerman S (2013) Corruption: a study in political economy. Academic, New York
5. Hale TN (2008) Transparency, accountability, and global governance. In: Global governance, pp 73–94
6. Kirkland JH, Gross JH (2014) Measurement and theory in legislative networks: the evolving topology of congressional collaboration. Soc Netw 36(1):97–109
7. Neal ZP (2018) A sign of the times? Weak and strong polarization in the US Congress, 1973–2016. Social Networks
8. Andris C, Lee D, Hamilton MJ, Martino M, Gunning CE, Selden JA (2015) The rise of partisanship and super-cooperators in the U.S. House of representatives. PLoS ONE 10(4):1–14
9. Dal Maso C, Pompa G, Puliga M, Riotta G, Chessa A (2014) Voting behavior, coalitions and government strength through a complex network analysis. PLoS One 9:12
10. Moody J, Mucha PJ (2013) Portrait of political party polarization. Netw Sci 1(1):119–121
11. Waugh AS, Pei L, Fowler JH, Mucha PJ, Porter MA (2009) Party polarization in congress: a network science approach. arXiv preprint arXiv:0907.3509
12. Victor JN, Montgomery AH, Lubell M (2017) The Oxford handbook of political networks. Oxford University Press, Oxford
13. Wachs J, Yasseri T, Lengyel B, Kertész J (2019) Social capital predicts corruption risk in towns. R Soc Open Sci 6(4):182103
14. Berlusconi G, Calderoni F, Parolini N, Verani M, Piccardi C (2016) Link prediction in criminal networks: a tool for criminal intelligence analysis. PLoS One 11:4
15. Ribeiro HV, Alves LG, Martins AF, Lenzi EK, Perc M (2018) The dynamical structure of political corruption networks. J Complex Netw 6(6):989–1003
16. Luna-Pla I, Nicolás-Carlock JR (2020) Corruption and complexity: a scientific framework for the analysis of corruption networks. Appl Netw Sci 5(1):1–18
17. Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97

18. Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the internet topology. ACM SIGCOMM Comput Commun Rev 29(4)
19. Sporns O (2002) Network analysis, complexity, and brain function. Complexity 8(1):56–60
20. Montoya JM, Solé RV (2002) Small world patterns in food webs. J Theor Biol 214(3):405–412
21. West GB, Brown JH, Enquist BJ (2009) A general model for the structure, and allometry of plant vascular systems. Nature 400:125–126
22. Albert R, Albert I, Nakarado GL (2004) Structural vulnerability of the north American power grid. Phys Rev 69(2):025103
23. Liu W, Suzumura T, Ji H, Hu G (2018) Finding overlapping communities in multilayer networks. PLOS One 13(4):e0188747
24. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818
25. Silva TC, Zhao L (2012) Stochastic competitive learning in complex networks. IEEE Trans Neural Netw Learn Syst 23(3):385–398
26. Silva TC, Zhao L (2012) Network-based high level data classification. IEEE Trans Neural Netw Learn Syst 23(6):954–970
27. Colliri T, Ji D, Pan H, Zhao L (2018) A network-based high level data classification technique. In: 2018 international joint conference on neural networks (IJCNN). IEEE, pp 1–8
28. Carneiro MG, Zhao L (2017) Organizational data classification based on the importance concept of complex networks. IEEE Trans Neural Netw Learn Syst 29(8):3361–3373
29. Backes AR, Casanova D, Bruno OM (2013) Texture analysis and classification: a complex network-based approach. Inf Sci 219:168–180
30. Loglisci C, Malerba D (2017) Leveraging temporal autocorrelation of historical data for improving accuracy in network regression. Stat Anal Data Min: ASA Data Sci J 10(1):40–53
31. Gao X, An H, Fang W, Huang X, Li H, Zhong W, Ding Y (2014) Transmission of linear regression patterns between time series: from relationship in time series to complex networks. Phys Rev E 90(1):012818
32. Holme P, Saramäki J (2012) Temporal networks. Phys Rep 519(3):97–125
33. Colliri T, Zhao L (2019) Analyzing the bills-voting dynamics and predicting corruption-convictions among Brazilian congressmen through temporal networks. Sci Rep 9(1):1–11
34. Csardi G, Nepusz T et al (2006) The igraph software package for complex network research. Int J, Complex Syst 1695(5):1–9
35. Thompson WH, Brantefors P, Fransson P (2017) From static to temporal network theory: applications to functional brain connectivity. Netw Neurosci 1(2):69–99
36. Câmara (2019) Dados Abertos. [Accessed on December 1, 2019]
37. Federal ST (2019) Processos. https://portal.stf.jus.br/. [Accessed on December 1, 2019]
38. Hulovatyy Y, Chen H, Milenković T (2015) Exploring the structure and function of temporal networks with dynamic graphlets. Bioinformatics 31(12):i171–i180
39. Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inform Sci Technol 58(7):1019–1031
40. Guns R (2014) Link prediction. In: Measuring scholarly impact. Springer, Berlin, pp 35–55
41. Salton G, McGill MJ (1986) Introduction to modern information retrieval. McGraw-Hill Inc, New York
42. Spertus E, Sahami M, Buyukkokten O (2005) Evaluating similarity measures: a large-scale study in the Orkut social network. In: Proceedings of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining, pp 678–684
43. Esquivel AV, Rosvall M (2011) Compression of flow can reveal overlapping-module organization in networks. Phys Rev X 1(2):021025
44. Egghe L, Leydesdorff L (2009) The relation between Pearson's correlation coefficient r and Salton's cosine measure. J Am Soc Inform Sci Technol 60(5):1027–1036
45. Ahlgren P, Jarneving B, Rousseau R (2003) Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. J Am Soc Inform Sci Technol 54(6):550–560

46. Katz L (1953) A new status index derived from sociometric analysis. Psychometrika 18(1):39–43
47. Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: bringing order to the web. Technical report, Stanford InfoLab
48. Newman ME (2004) Fast algorithm for detecting community structure in networks. Phys Rev E 69(6):066133