# How Congressmen Connect:
# Analyzing Voting and Donation Networks
# in the Brazilian Congress

**Victor S. Bursztyn**[1,2], **Marcelo Granja Nunes**[1], **Daniel R. Figueiredo**[1]

[1]Programa de Engenharia de Sistemas e Computação/COPPE
Universidade Federal do Rio de Janeiro (UFRJ) – Brazil

[2]EMC Brazil Research & Development Center
Parque Tecnológico, Ilha do Fundão – Rio de Janeiro, RJ – Brazil

`{vbursztyn,mgranja}@cos.ufrj.br, daniel@land.ufrj.br`

***Abstract.*** *The goal of this work is to analyze two of the most central activities in the life of a congressman: raising funds and voting bills. We investigate the Brazilian Congress to shed light on the relationships between the donations received by congressmen elected in 2014 and their voting behaviors during the year of 2015. We merged publicly available data obtained from the Brazilian House of Representatives and the Superior Electoral Court (TSE) in order to create a tripartite network containing campaign donors, elected congressmen, and legal bills. Using this data, we create two projected networks having congressmen as nodes and links given as follows: 1) congressmen who received donations from the same donors (donation network); and 2) congressmen who voted in accordance to each other on legal bills (voting network). After characterizing these networks, we propose three fundamental questions on the behavior of congressmen that could benefit from the methods and concepts provided by Network Science. Finally, we analyze the results and compare them to general domain knowledge.*

## 1. Introduction

In November 2014, the influential Brazilian newspaper Folha de São Paulo [Folha de São Paulo 2014] reported that the top 10 corporate donors accounted for 70% of the funds raised by deputies elected earlier that year. Political scientists have long debated about possible negative aspects of electoral campaigns funded by large private companies. Notably, a study conducted by Datafolha suggests that the new House of Representatives' strong position in favor of those donations opposes to the position of the Senate and, more worth-noticing, to public opinion – 79% of the population was contrary in July 2015, according to Datafolha's polls [O Globo 2015].

Almost one year after the election, in September 2015, Brazil's Federal Supreme Court (STF) ruled that the Constitution does not allow corporations to donate to electoral campaigns, meaning that corporate actors are no longer allowed in upcoming elections [BBC 2015]. Nevertheless, the availability of donations data and the fact that they were legal for the most part of 2015 provide an opportunity – maybe a historical one – to shed light on the dynamics between congressmen funding and their behavior in Brazil's legislature.

According to David Lazer [Lazer 2011], the use of Network Science by political scientists is fairly recent, despite the long-lasting notion that power is an intrinsically relational phenomenon: it rises from the capacity that one actor has to affect other actors. Over the past two decades, a series of works focused on studying power and networks of legislative institutions in countries with pioneering transparency initiatives. Previous works have studied parliaments using public data such as bill co-sponsorship [Zhang et al. 2008][Fowler 2006], committee participation [Porter 2006], roll call voting [Dal Maso et al. 2014], social media [Peng et al. 2014], or voting advice applications [Etter et al. 2014]. Some of the goals of these works were to characterize congressmen communities, predict issue voting results, and verify ideological consistency over time. We observed an opportunity to apply similar methods to the national context, specifically on the Brazilian House of Representatives.

This work expects to deliver the following contributions: first, we created a unified dataset by integrating data regarding the Brazilian House of Representatives previously scattered across different sources. This resulted in a tripartite network with the following types of nodes: campaign donors, deputies elected in 2014, and legal bills voted during the year of 2015. Moreover, for each bill we mine the voting position of each congressman and for each donor we mine the set of congressmen to whom donations were given. Based on this data, we can measure the similarity between congressmen through two points of view: 1) congressmen who received donations from the same donors (donation network); and 2) congressmen who voted the same positions on legal bills (voting network). To the best of our knowledge, we could not find any work that has analyzed these two aspects together, nor a dataset connecting all three domains (donors, congressmen and votings). Therefore, we stress that the dataset itself might be a valuable contribution as it could enable other researchers to explore such relationships or to unveil other phenomena.

Secondly, we explore these relationships posing and addressing three fundamental questions, briefly put as follows:

1) Homophily: How co-partisanship (congressmen from a same political party) and co-regionality (congressmen from the same region) reflect on campaign donations and on voting behaviors?

2) Partisan cohesion: How cohesive are political parties in terms of voting?

3) Influence prediction: If we use a list of top influencing congressmen as ground truth (i.e., from a broadly acknowledged report), could we retrieve that same list by mining voting network's topological features?

Beyond data collection and classic network characterization, another key contribution of this work is the proposal of a specific network-based methodology (and metrics) to address the above questions. In particular, our approach to study homophily and cohesion can be applied to any network where nodes are subject to a similarity metric and belong to a category (in our case, a political party or a geographical location).

The remainder of this paper is organized as follows. In Section 2 we describe our dataset; in Section 3 we explain how donation and voting networks are created and their associated characteristics; in Section 4 we analyze higher level network features; and finally, in Section 5, we discuss our findings as well as future directions.

## 2. Data Sources

We considered two main sources of publicly available data: 1) the Superior Electoral Court (TSE), where all official donations are available for bulk download; and 2) the House of Representatives, where all voting records for bills are open for access through an API. A process was developed in order to fetch those sources and to merge them into our tripartite network. Our resulting network was persisted in a Neo4J database, as illustrated in Figure 1. The final database records 52245 campaign donors, 456 elected congressmen and 234 bills (legal bills voted between January and November 2015, right after the new Congress was established). Our consolidated dataset is made available at http://github.com/vbursztyn/HowCongressmenConnect, in various formats.
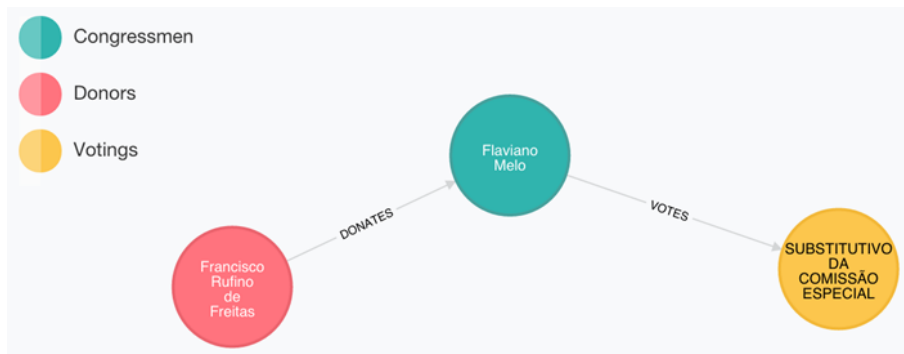


**Figure 1. A single example illustrating all types of nodes and links**

## 3. Donation and Voting Networks

From the tripartite network, we created two projected networks having congressmen as nodes and links as follows: 1) congressmen who received donations from the same donors (donation network); and 2) congressmen who voted the same positions on legal bills (voting network). In both networks, nodes are deputies and links are weighted similarities with values between zero (dissimilar) and one (identical). The metric used for similarity is the Jaccard Index, which is among the most frequent metrics in network applications. It is defined as the ratio between the cardinality of the intersection set and the cardinality of the union set. In our context, each congressman has a set of donors or a set of bills on which he has voted. Thus, for every pair of congressmen with a nonempty intersection of voting positions we have a similarity score in the voting network, and for every pair of congressmen with a nonempty intersection of donors we have a similarity score in the donation network.

Still, there are additional details on how this definition applies to each network. For the donation network, we only considered the presence of donors, disregarding the value of the donation. Although this application does flatten some information, it also prevents unwanted dissimilarity between candidates with a similar portfolio of donors but vastly different donations amounts. In essence, donors manage limited resources – regardless of their scales –, and the fact that a campaign attracted a person or a lobbyist to the point of making a donation is fairly significant.

In the voting network, we defined similarity considering the roll call votes made by elected congressmen on legal projects in 2015's agenda. Considering a pair of congressmen, the union set would be every bill they voted, and the intersection set would be

the bills in which their votes converged (options are: yes, no, abstention or obstruction). A special case to be mentioned is the speaker of the House of Representatives, who does not vote and therefore was excluded from this network.

The main features of these two networks are summarized in Table 1 below, where link weights have not been considered (and thus all links with non-zero weight are present in the network). Also, the voting attendance rate for deputies is shown on Figure 2.

**Table 1. Characterization of the two projected networks**

|  | Donation Network | Voting Network |
|---|---|---|
| Nodes count | 456 | 456 |
| Link count | 60492 | 182256 |
| Maximum degree | 323 | 454 |
| Minimum degree | 0 | 0 |
| Average degree | 132.66 | 399.68 |
| Density | 0.29 | 0.88 |
| Number of components | 20 | 1 |
| Relative size of the largest component | 96% | 100% |
| Global clustering index | 0.69 | 0.95 |
| Average local clustering index | 0.68 | 0.96 |
| Average distance | 1.77 | 1.12 |
| Pseudodiameter | 5.0 | 3.0 |
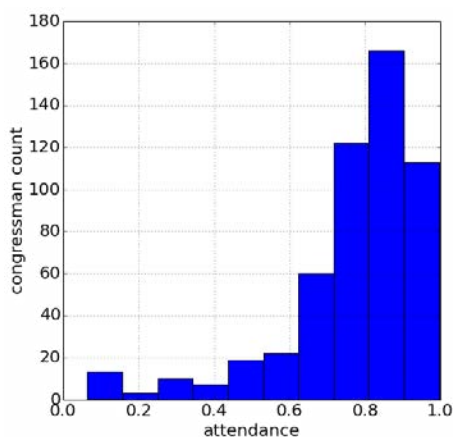


**Figure 2. Voting attendance rate for deputies**

Within those features, it is worth highlighting the average degree observed in both networks. In particular, the high connectivity in the voting network suggests that it is very unlikely for a congressman not to find a minimum degree of consensus with his peers, in a year's agenda. This happens because congressmen are intrinsically constrained: they must declare a position if participating on a voting section, which in turn is limited to four options (yes, no, abstention or obstruction). All alone, the fact that there are more donors than bills could make more likely that a random pair of congressmen shares a donor. However, voting in a limited range of options imposes a very low probability for not sharing at least one single position throughout the year (for a random pair of congressmen:

$0.75^{234} \approx 6 * 10^{-30}$). Values observed for features such as the clustering coefficient and the average distance between nodes can also be explained from these observations.

Regarding voting attendance, Figure 2 shows that most deputies have a relatively high attendance, while a few have high levels of absence. The majority of these absent deputies are replacements, i.e. deputies who occupied the seat of another deputy. Since these deputies joined the Congress at a later date, they tend to have attended fewer votings.

## 4. Domain Specific Network Features

After a summarized view on the two projected networks, we propose and address the three fundamental questions in the following subsections.

### 4.1. Homophily – How co-partisanship and co-regionality reflect on campaign donation and on voting behaviors?

According to Anmol Madan et al. [Madan et al. 2011], homophily is a fundamental phenomenon in social networks, which refers to the tendency of individuals to form relationships with others that have similar attributes. Let's define homophily, in this particular case, by means of two attributes: congressmen who belong to the same party (i.e., co-partisanship), and congressmen elected from the same region (i.e., co-regionality). Once homophily is contextualized in our political setting, then we may propose our central questions: can we detect homophily in the donation network and in the voting network? More than uncovering such tendencies, it would be interesting to actually measure how strongly each attribute (co-partisanship and co-regionality) acts in both networks. In order to address this issue, we developed the following approach:

Consider that for each network each congressman holds an average similarity score defined by the average weight of its links (i.e., $w_{avg}$). Alternatively, for that same congressman, we may also calculate an average similarity selecting only the links that exhibit homophily: $w_{avg}^p$ is the average score restricted to co-partisans, and $w_{avg}^r$ restricted to co-regionality. Thus, for each node $u$, there are three averages based on the Jaccard Indexes: $w_{avg}(u)$, $w_{avg}^p(u)$ and $w_{avg}^r(u)$. Suppose our focus is on node $u$, then it is possible to calculate two similarity gains associated with our two homophily hypothesis:

w.r.t. co-partisanship: $\text{gain}^{party}(u) = g^p(u) = (w_{avg}^p(u) - w_{avg}(u))/(w_{avg}(u))$

w.r.t. co-regionality: $\text{gain}^{region}(u) = g^r(u) = (w_{avg}^r(u) - w_{avg}(u))/(w_{avg}(u))$

We can apply this similarity gain metric for every node in the network, for each network separately. Figure 3 shows the rank order of similarity gains in the donation network: green for co-partisanship, red for co-regionality. For instance, regarding co-regionality (red) congressman Mendonça Filho (DEM-PE) has a similarity gain of 1.43, Marco Antônio Cabral (PMDB-RJ) a gain of 1.13 and Tiririca (PR-SP) a negative gain of -0.04. It means that, on average, Mendonça Filho and Marco Antônio Cabral share more donors with congressmen from their respective states (AL and RJ) if compared to their respective average similarities to all peers. Conversely, Tiririca shares more donors with all of his peers, on average, than he shares with his peers from SP. This observation conforms to domain knowledge: Tiririca is not originally from SP and may have donors from other regions, that do not donate to other SP deputies.
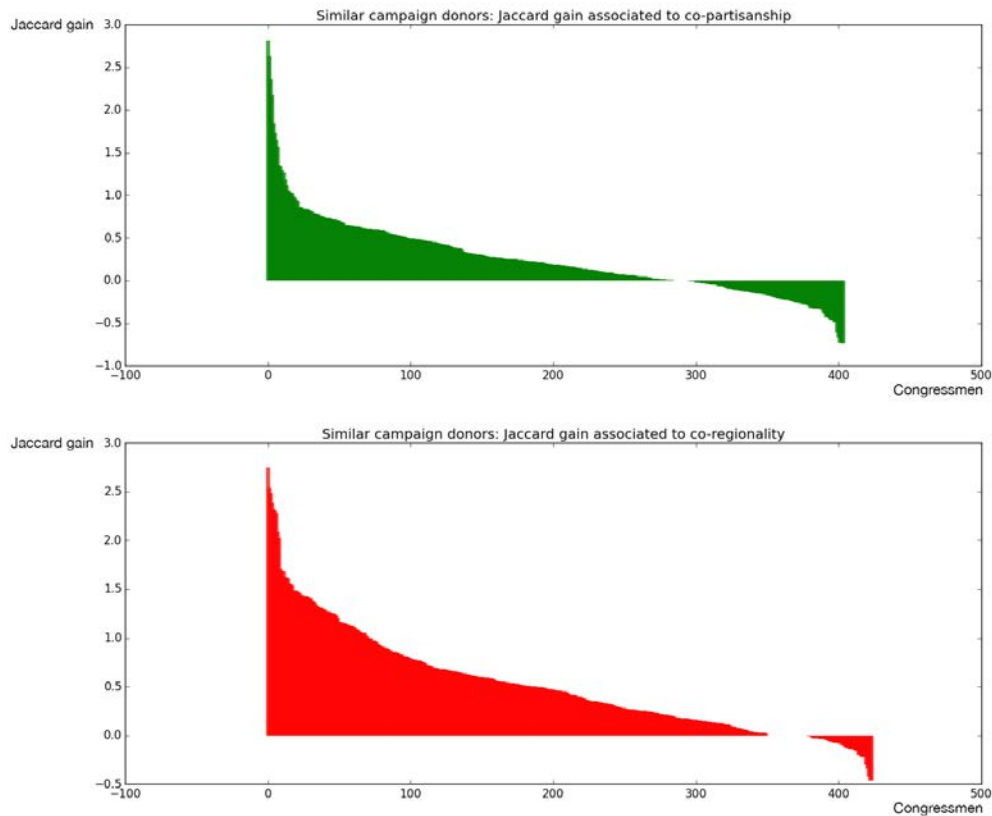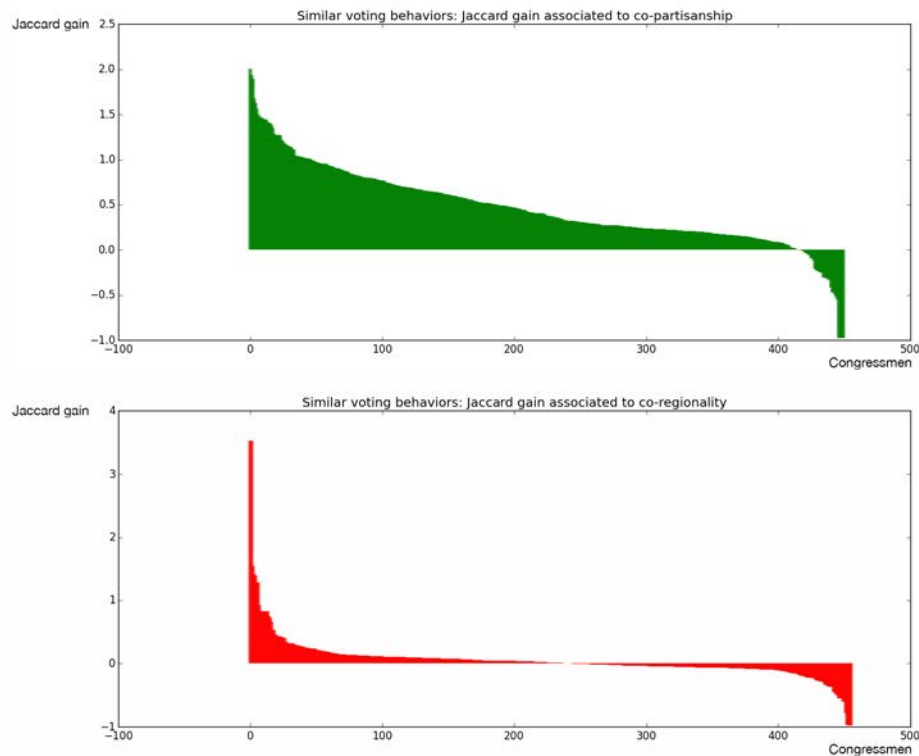
**Figure 3. Similarity gains w.r.t. donations for deputies within the same political party (above) and based on the same federation unit (below)**

The y-axis is the Jaccard Index gain (loss if negative) and the x-axis are the deputies. On average (across all nodes), the Jaccard Index gain in the donation network is 51% (0.51) if deputies represent the same state (co-regionality). Also on average, the Jaccard Index gain is 25% (0.25) if deputies belong to the same political party (co-partisanship). Therefore, in an election that allowed two sorts of donors (corporate donors and regular citizens), co-regionality was a stronger driver of donors similarity. If compared to co-partisanship, which also revealed to be a fairly relevant driver of donors similarity, co-regionality was twice stronger.

Similarly, Figure 4 shows the rank order of similarity gain in the voting network: green for co-partisanship, red for co-regionality. For instance, regarding co-partisanship (green) congressman Chico Alencar (PSOL-RJ) has a similarity gain of 1.94, Mara Gabrilli (PSDB-SP) a gain of 1.01 and Alessandro Molon (PT-RJ) a gain of 0.50. It means that, on average, their voting behaviors are more similar to the behaviors of their respective co-partisans, if compared to their average similarities to all peers. This also means that such identity is stronger for Chico Alencar regarding PSOL, than to Mara Gabrilli regarding PSDB. However, her identity with respect to her party's average behavior is stronger if compared to Alessandro Molon's regarding his party, PT.

Oppositely to what was noted in the donation network, Jaccard Index gain in the voting network is much stronger for co-partisanship if compared to co-regionality. On average, the Jaccard Index gain in the voting network is 46% (0.46) if deputies belong to

156

**Figure 4. Similarity gains w.r.t. votings for deputies within the same political party (above) and based on the same federation unit (below)**

a same political party, versus 5% (0.05) if they represent the same state (co-regionality). On these grounds, one may generally conclude that political parties are almost 10 times more effective in channeling voting behavior, if compared to pure locality. This gives a good measure of the role of political parties in today's legislative – in a general and comparative view, parties do own expressive voting identities, in particular, much stronger than regionality.

In short, we may conclude that co-regionality drives donations more than co-partisanship, while co-partisanship drives voting behavior much more than co-regionality.

## 4.2. Partisan cohesion – How cohesive are political parties?

We can use the Jaccard Index gain value to identify which congressmen are more dissimilar with respect to his co-partisans. Such analysis indicates that the following deputies not only have greater tendency to vote unlike their co-partisans, but also are more similar to congressmen outside their respective parties: Guilherme Campos (PSD-SP), Giovanni Queiroz (PDT-PA), João Ananias (PC do B-CE), Betinho Rosado (PP-RN), Roberto de Lucena (PV-SP), Silvio Costa (PSC-PE), Danilo Cabral (PSB-PE), Zoinho (PR-RJ), Davi Alves Silva Júnior (PR-MA), Sandra Rosado (PSB-RN). From the voting pattern point of view, these congressmen are not aligned with their respective parties.

A second method that we propose to address partisan cohesion is as follows. For each political party, hereby represented by $C$, we calculate its intraparty average similarity $d_{int}(C)$: the average similarity of the links among party members; the interparty average similarity $d_{ext}(C)$: the average similarity of links between members of party $C$

and deputies who are not a member of $C$; and the overall average similarity across all network links. These definitions have the following equations:

$$d_{ext}(C) = \frac{\sum_{i \in C, j \notin C} w_{ij}}{n_c(n - n_c)} \quad d_{int}(C) = \frac{\sum_{ij \in C} w_{ij}}{n_c(n_c - 1)/2} \quad d(C) = \frac{\sum_{ij} w_{ij}}{n(n - 1)/2}$$

Where $n$ is the number of deputies and $n_c$ is the number of deputies belonging to political party $C$. In theory, cohesive parties should have $d_{int} > d > d_{ext}$, and the wider the gap between $d_{int}$ and $d_{ext}$, the more cohesive is a political party. Finally, we limited this analysis to parties with more than three congressmen.

The values measured for the averages are shown in Figure 5. In order to support the evaluation of our results, Figure 6 shows the number of deputies in each party. The ratio between intraparty similarity and overall average similarity suggests that some political parties are indeed much more cohesive than others. On the other hand, all parties have the interparty similarity lower than the overall average similarity, indicating all parties have some cohesion in comparison to average behavior. The most cohesive party is PSOL, which has the highest $d_{int}/d$ ratio and the second lowest $d/d_{ext}$. Furthermore, contrary to anecdotal evidence, PMDB does have a mild internal cohesion. Also contrary to common knowledge, PV's intraparty similarity is about the same as the average similarity, indicating low cohesion in the voting behavior despite having only 8 deputies. Finally, we stress the fact that the most and least cohesive parties are small ones, but we also find parties with larger sizes with both high and low levels of cohesion, such as PSDB and PT (high cohesion) and PP (low cohesion).
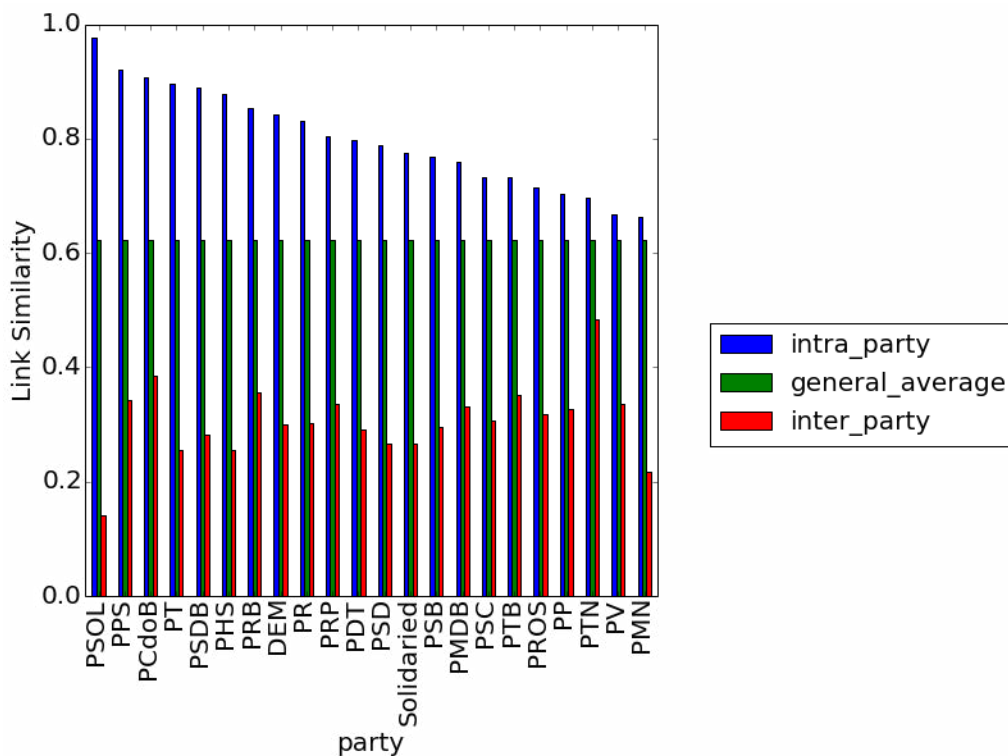


**Figure 5. Average link similarities for each political party with more than three congressmen**
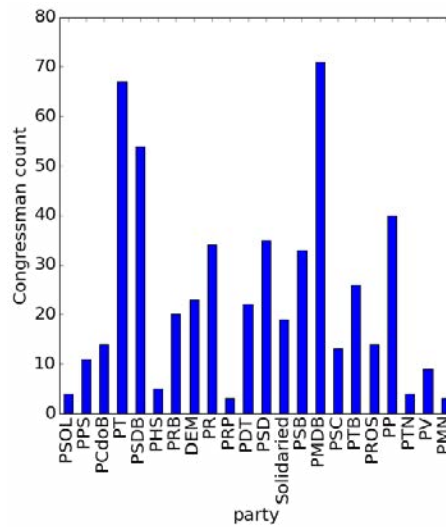
**Figure 6. Number of deputies in each party with more than three congressmen**

### 4.3. Influence prediction – If we use a list of top influencing congressmen as ground truth, could we retrieve that same list by mining voting network's topological features?

The third question is based on the hypothesis that topological characteristics of a network can often be used to identify the most influential nodes, in our case members in the House of Representatives. To address this question we use the yearly report "Heads of Congress", published by the Inter-Union Department for Parliamentary Advisory (DIAP) [Departamento Intersindical de Assessoria Parlamentar 2014], [Departamento Intersindical de Assessoria Parlamentar 2015], and considered an official source for the list of top influencing deputies. DIAP's reports are often referred by major Brazilian newspapers when conveying political analysis, and thus it is a reasonable source for ground truth.

We consider node centrality as the topological feature used to identify top influential nodes. Nodes are sorted based on the following centrality metrics: Strength, Betweenness, Closeness, Katz, and PageRank. These metrics were chosen not only because they are the most common ones, but mainly because they capture different types of centrality. The Strength is defined as the sum of weights on links connected to a node. Both the Betweenness and the Closeness were normalized. The PageRank is used with a 0.85 damping coefficient and the Katz has $\alpha = 0.1$ and $\beta = 1$. Finally, all the metrics were calculated considering the Jaccard Index for the voting network as weights.

An important modification was made on the network. The link weight definition previously used in this paper was a similarity index, i.e. the higher its value, the more similar are the nodes connected by that link. However, the use of those centrality methods require link weights to be defined as a dissimilarity index, i.e. the lower its value, the more similar are the nodes. Therefore, we calculated this new link weight as being the complement of the old link weight, i.e. $weight_{dissimilarity} = 1 - weight_{similarity}$.

Afterward, we consider the top 55 members (about 10% of the House's population) in the centrality ranks, resulting in five separate lists eligible to being tested against

DIAP's Heads of Congress. The quality of each centrality metric is evaluated by its precision in identifying 55 names from the DIAP's report. It is worth noticing that DIAP's original report contains 100 congressmen, but only 55 are deputies. Also, we abstained from using deeper information contained in the report, such as the type of influence exerted by each congressman.

The distributions for four centrality metrics in the voting network may be found in Figure 7. As can be seen, the Strength distribution is similar to PageRank's. This result was not unexpected as it is widely reported in the literature that, for undirected networks, the PageRank for nodes are proportional to their degrees [Grolmusz 2012]. More interestingly, the distribution for Betweenness was highly concentrated on a few nodes whose Betweenness value was zero. These nodes are deputies with the highest levels of abstention. A possible direction for further improvements could be to recreate the voting network with a minimum threshold for attendance, therefore decreasing this disproportional concentration of shortest paths on a few nodes.
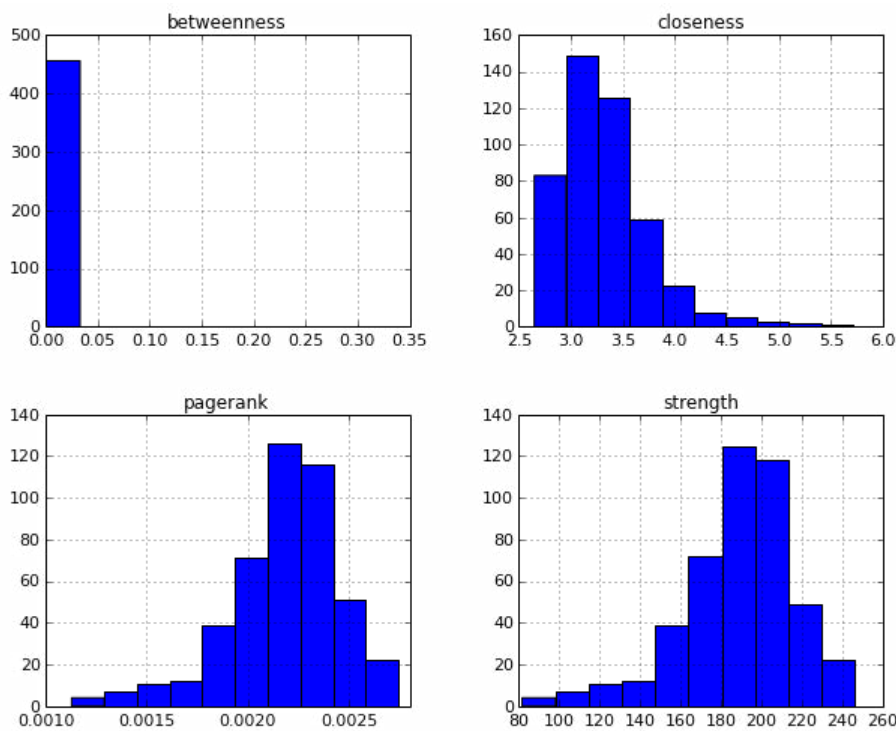


**Figure 7. Histogram of centrality metrics for the nodes on the voting network**

Finally, Table 2 indicates the precision of the five ranks, highlighting the best approach (underlined). The best performance is assigned to the top PageRank (and Strenght) centrality, which manages to identify 27% of DIAP's list. Such result surpasses a naive benchmark by 2.5 times, as random selection would only identify 11% on average. Nevertheless, this performance still is relatively low. A possible explanation is the high density of the network causing nodes to become similar while it should not be the case. Further work must be done in order to verify this hypothesis.

**Table 2. Performance of centrality ranks w.r.t. precision@55 on DIAP's list**

| Metric | Precision@55 |
|---|---|
| Random | 0.11 |
| Strength | <u>0.27</u> |
| Betweenness | 0.18 |
| Closeness | 0.13 |
| Katz | 0.07 |
| PageRank | <u>0.27</u> |

## 5. Conclusions and Future Work

In this paper, we consolidate data from different public sources concerning donations to and votes by Brazilian congressmen. Donations refer to the election of 2014 and votes refer to legal bills in an 11 month agenda: from January 2015 to November 2015. Using this data we construct and analyze two networks where nodes are congressmen: the donation and voting networks. These relationships are fundamental to the political process, making the study of both networks a relevant subject for research in Network Science.

Furthermore, this paper finds evidence of homophily regarding co-partisanship and co-regionality in both networks. However, homophily patterns differ greatly between these networks, revealing the richness of the phenomena comprised in the original tripartite network. It is worth highlighting that we find quantitative evidence on the importance of parties in voting behavior, being almost 10 times stronger than co-regionality. Moreover, using the voting network we measured the degree of party cohesion and polarization with respect to the rest of the House of Representatives, with results that partially match our domain intuition, but also reveal new insights (e.g., the least cohesive parties). Finally, recognizing the most influential deputies using these networks and an external ground truth proved to be a difficult task that requires further study. A few promising directions for future work would be to test other similarity metrics besides the Jaccard Index, or to extend the dataset with other information such as congressmen participation in special commissions.

Also, due to the Federal Supreme Court's ruling, the landscape of donors is expected to change dramatically in the 2018 elections. That said, it is appealing to compare the networks uncovered by this work, to the ones that are going to arise after the prohibition of corporate donors. For example, homophily for co-regionality could increase significantly since corporate donors are more likely to be transversal to states than regular citizens.

## References

Dal Maso, C., Pompa, G., Puliga, M., Riotta, G., and Chessa, A. (2014). Voting behavior, coalitions and government strength through a complex network analysis. *PloS one*, 9(12):e116046.

Departamento Intersindical de Assessoria Parlamentar (2014). Radiografia do Novo Congresso: Legislatura 2015-2019. Technical report, Departamento Intersindical de Assessoria Parlamentar, Brasilia.

Departamento Intersindical de Assessoria Parlamentar (2015). Os Cabeças do Congresso Nacional : uma pesquisa sobre os 100 parlamentares mais influente. Technical report, Departamento Intersindical de Assessoria Parlamentar, Brasilia.

Etter, V., Herzen, J., Grossglauser, M., and Thiran, P. (2014). Mining democracy. In *Proceedings of the second ACM conference on Online social networks*, pages 1–12. ACM.

Fowler, J. H. (2006). Connecting the congress: A study of cosponsorship networks. *Political Analysis*, 14(4):456–487.

Grolmusz, V. (2012). A note on the pagerank of undirected graphs. *CoRR*, abs/1205.1960.

Lazer, D. (2011). Networks in political science: Back to the future. *PS: Political Science & Politics*, 44(01):61–68.

Madan, A., Farrahi, K., Gatica-Perez, D., and Pentland, A. S. (2011). Pervasive sensing to model political opinions in face-to-face networks. In *Pervasive Computing*, pages 214–231. Springer.

Peng, T.-Q., Liu, M., Wu, Y., and Liu, S. (2014). Follower-followee network, communication networks, and vote agreement of the us members of congress. *Communication Research*, page 0093650214559601.

Porter, M. (2006). A network analysis of committees in the united states house of representatives. In *APS Meeting Abstracts*, volume 1, page 33006.

BBC (2015). STF decide que doação de empresas a campanhas é inconstitucional; veja em que países isso já é proibido.

Folha de São Paulo (2014). As 10 empresas que mais doaram em 2014 ajudam a eleger 70% da Câmara.

O Globo (2015). Nova ferramenta digital revela doadores das ultimas campanhas eleitorais.

Zhang, Y., Friend, A., Traud, A. L., Porter, M. A., Fowler, J. H., and Mucha, P. J. (2008). Community structure in congressional cosponsorship networks. *Physica A: Statistical Mechanics and its Applications*, 387(7):1705–1712.