ACCESSING DATA TO VIEW AFFINITIES IN VOTING BY BRAZILIAN SENATE PARLIAMENTARIANS

artigo de revisão

*Ricardo César Gonçalves Sant'Ana**
*Fernando de Assis Rodrigues***

ABSTRACT The increase in citizen participation in the public sphere is directly dependent on information access related to the performance of elected representatives, especially with regard to their decisions in voting during their terms. Information Science can collaborate in this process, proposing and evaluating models of access to this information that can be obtained through data to be made available by the official sites of the Legislative Power in the Federal, State, and Municipal spheres. This article proposes to analyze the process of collecting and usage of data on senator's votes looking for the appropriation of this model for data collection and its use in other spheres. Based on the data, the elaboration of an affinity matrix that allows the identification of the relationship between each of the parliamentarians with the others is analyzed, according to the similarities of the decisions taken in the general open votes. It also analyzes the preparation of initial visualizations and the expansion of the scope of the research through the application of the data obtained in all affinities between parliamentarians and in obtaining an average affinity between parties, allowing new dimensions of analysis to the collected data. The elaboration of the complete matrices of affinity relations between parliamentarians can provide a rich horizon of possibilities for the elaboration of new forms of visualization and analysis, increasing the visibility of parliamentary actions within society.

Keywords: Open Data. Data visualization. Public Transparency.

* Professor in Information Science Graduate Program at São Paulo State University,BrazilORCID https://orcid.org/0000-0003-1387-4519. E-mail: ricardo.santana@unesp.br
** Professor in Information Science Graduate Program at Federal University of Pará, Brazil ORCID https://orcid.org/0000-0001-9634-1202. E-mail: fernando@rodrigues.pro.br

1 INTRODUCTION

Even throughout and within the provided resources from Information Technologies Communication, there are greater difficulties in expanding popular participation in public administrations, yet new ways for the access of data that can favor citizen participation can be sought, as for example in the monitoring of parliamentary votes. This article presents a proposal for the construction of a resource that can visualize these votes, which provide analyzes of their interrelations and a brief description of how available data on the votes of parliamentarians in the Brazilian Senate is available.

A social graph model was adopted in order to show affinities between senators, according to work done by Wilson (2009) on votes taken in the US Senate in which data on senators' votes were used to generate an affinity relationship, showing consistency between Senator Arlen Specter votes in relation to votes by members of the Democratic and Republican parties.

The concept of building social graphs is based on the mathematical definition of graph theory, in which the relationships between elements of a set are the focus of this area. These elements are denominated as vertices and each of their relations is composed of a pair of vertices, being called an edge. The study of these links between elements of a set allows its application in several areas of knowledge (DIESTEL, 2000; WILSON, 1996). Thus, rules are obtained for displaying relationships between two or more objects in a given context, making it necessary to choose the data set to be analyzed and the rule or formula that will define the relationships between the elements of the set, forming the edges. An example of using this theory would be a graph representing part of the World Wide Web where pages are represented as vertices and links as edges (LOCKE, 2000).

Odewahn (2010) has made use of the visualization model proposed by Wilson (2009), considering the premise that the US Senate is usually understood by society as being composed of two very stable coalitions, according to the bipartisan concept. Thus, Odewahn (2010) sought in his proposal to create a graph that could display a broader picture, revealing that there is a factor of variability in the formation of the relations of senators of both parties over time when taking votes during the legislature.

When building his type of chart, Odewahn (2010) had access to primary data (namely raw data) on the votes, that is, he used data on the senators' votes as a source of information to create a graphical representation of the affinities arising from their decisions.

In the work carried out by Odewahn (2010), for the making of a visualization model coming from these data, it was necessary to develop a graph (from the open data provided) that would explore these issues visually to observe the antithesis with the premises taught in the American student colleges, which define the Senate as an inherent and conservative body, in the opposite and literal sense to that of changing opinions among its members. On the one hand the Republicans and on the other (opposite) the Democrats (ODEWAHN, 2010).

Thus it was possible to display, treat and visualize existing data, but in a way that it was possible to extract and analyze "something" hidden, not visible in another approach or in another visualization of the same data - demonstrating these connections between senators (which a priori were hidden in the primary data) rather than viewing only the votes singularly in an ordered list.

The concept of primary data looks to identify data that have not been obtained from transformations or manipulations in such a way that they represent values as close as possible to their origins, that is, they are not the result of sums, groupings analyzes or any other processing. Thus, primary data cannot allow divisions in which parts of this data maintain semantic reference in the context of their application. Therefore, when analyzing the billing of a given project, for example, the total daily sales value is the result of the sum of the notes issued and is not a primary data, unlike the value of a sales note that represents primary data for the context of analysis.

Access to primary data is considered to be one of the key success factors in the construction of alternatives for accessing data, avoiding eventual errors of interpretation, and still allowing varied paths of analysis due to the high degree of freedom of combinations and possible processing that a primary data offers.

> Although the raw data sets at Data.gov do require aggregation and synthesis, at the same time those data come with an important advantage, which is precisely that they can be analyzed, synthesized, and repackaged in any way. In other words, raw data are flexible, and upon making such data available to the public, agencies have explicitly invited the public to put that information to new uses, tailored however users of that information desire (THE WHITE HOUSE, 2009, p. 26).

Another important factor in data access is the possibility of these data being processed directly by machines, that is, they must be available in such a way that they can be used without the need for retyping or additional treatments to be directly obtainable. This characteristic allows the interoperability of applications and the use of this data by external researches, regardless of the initial format and specifications linked to the technological platform (BERNERS-LEE, 2009). For example, data that is available in a file that has presentation characteristics, such as a file in

Portable Document Format (PDF) format may not be accessible for direct processing by the machine, since it requires interpretation of the positions in that the values are presented for the semantics to be obtained.

In the year 2007, in the city of Sebastopol in the United States of America (USA), a multidisciplinary working group was put together to develop principles that can conceptualize data as open data. This initiative was named "Open Government Working Group". The analysis of this work was initiated based on the records of US government agencies and resulted in eight principles for assessing whether a given record can be considered open data. These principles attribute that an open data must be complete, primary, temporal (it must be made available in time to be still useful), accessible, processable by machine, non-discriminatory (the user has no need to identify himself), non-owner and license-free (OGD, 2007). Therefore, open data must be primary, but not all primary data is open data. In this article, two characteristics were considered: primary data and data treatable by machines as the main ones by the definition of "data" as open data (OGD, 2007).

This article proposes to apply of the US model to build a visualization model for the Brazilian scenario, but with the objective of analyzing whether the primary data of open voting (in the period from 2007 to 2010) available through the Information and Communication Technologies (ICTs) from the Brazilian Senate to give the data support to affirm the existence or not of a dividing structure of the coalitions between the parties that make up the allied base of the government, the opposition and the independents.

2 OBJECTIVE AND PROCEDURES

The main goal is to present an analysis of the process of collecting and elaborating the visual structure of the data generated by results from the Brazilian Senate bills, through a graph. In this context, we developed a collection of vertices (nodes), edges (edges) and a non-coordinate graph (ODEWAHN, 2010, p. 123) of social relations, built based on a variety of affinities, adapted in the US conceptual model by Odewahn (2010), as explained in the book "Beautiful Visualization", by Steele and Iliinsky.

To obtain the primary data, we used only the data channels offered by the official ICTs of the Brazilian Federal Senate that were used, more precisely the senators' metadata and the primary data of public votes, from 2007 to 2010. With this strategy, we offer a visual proposal to analyze deeper issues, such as the possibility that these synthesized results indicate the existence or not of affinities between parties (and their members) that comprise opposition and government.

2.1 Construction of the social graph

In the paper published by Wilson (2009), the vertices represent senators and each vertex receives a hidden label which when triggered the mouse-over behavior of the browser (action of leaving the mouse pointer on an object) shows the name of the respective senator. The vertices are colored and each color represents a party (red for Republican members, blue for Democrats, green for independent senators, and yellow for lack of this information). The edges represent the connections between two senators when they voted equally more than 65% in the given period.

Odewahn (2010) added the division of members due to their parties in his approach. The graph presented Democrats on the left and Republicans on the right for a better understanding of the links between party members and their opposition group. It was also defined that each biennial legislative session is the basic time unit of the graphs. For example, the 150th session corresponds to the period that started on January 3rd, 1997, and ended on January 3rd, 1999. Its main justification for choosing this unit of time is that the body of the US Senate undergoes changes at each period, considering the election of new senators and the dismissal of others (ODEWAHN, 2010, p. 125).

For Odewahn (2010), three steps are needed for the creation of this visualization model: the acquiring of primary data; the calculation and generation of the affinity matrix, and; the construction of a graph from these data.

2.2 The acquisition of primary data

For the construction of the visualization, it is necessary to obtain two data sets: the first is the set of metadata containing individual information about each senator, with elements

such as his name and party. The second set is composed of primary data from votes made by these senators in a given matter, in the desired period (ODEWAHN, 2010). Metadata is a key factor for the functioning and interoperability of information systems, especially when there is a need to acquire primary data from an external database and-or not visible to the software since it is the metadata itself that guarantees the correct representation of the informational content of a given resource (SANTOS; ALVES, 2009). After defining the metadata that will compose a representation of a data set, it is necessary to choose which metadata format is capable of representing it efficiently; and thus create the basis for the correct conduct of connectivity with the most varied ICTs and tooling applications available. Thus, metadata formats or standards are standardized structures, which, through the set of data-attributes, allow the correct representation of informational content (SANTOS; ALVES, 2009).

The main obstacle in this step was the absence of these primary data in US government websites in a structured manner and that was machine-readable (ODEWAHN, 2010, p. 125). However, civilian projects such as GovTrack have organized and standardized this primary data (about senators and their votes) into machine-readable metadata – and even made them available in accordance with the concept of open data– allowing for the obtaining the same official data as the US government websites, but adding a metadata standard. This modeling allows other software to use these resources through technologies that help interoperability, such as eXtensive Markup Language (XML), JavaScript Object-Notation (JSON), among others. GovTrack has, in its base, structured and complete primary data of retroactive congressional votes until 1991; and structured, but incomplete, before 1991. This civil initiative is in accordance with the eight principles necessary for the denomination "open data" (OGD, 2007).

In the Brazilian Senate website, the data about current senators are found in a document in HyperText Markup Language (HTML) format with simple type metadata embedded, only with the purpose to indexing the page by search engines, through the use of web crawlers (SENSO, 2003, p. 103). Web crawlers are algorithms from search engines that collect data of content or file available on the Internet and store in a database. Thus, the data contained therein about senators are not structured for use facilitated by this kind of applications, generating the need for treatment through an abstraction layer, that is, a conversion algorithm for a certain structured pattern, making data machine-readable, such as an XML format file; and this behavior does not meet one of the eight principles of open data: that these data must be processable by machine (OGD, 2007). This characteristic makes it difficult to use data through technological tools without human intervention.

The Brazilian Senate activity, the votes are only available in digital documents stored in PDF format, where it is not possible to directly access the data, also requiring an abstraction layer for automated use and manual download of the results of the materials for treatment. These results are obtained through a pre-formatted query directly into the browser, with no direct access to primary data without human intervention at the time of this research.

In order to create the graph of visualization of affinities between the votes of senators, it was necessary to identify which senators were part of the collegiate body in the previous year: from 2007 to 2010. These data were found in a similar format to the current collegiate page, or that is, in a document in HTML format and also available for printing. In this document, it was possible to find three elements: the name of the senator, his federative unit, and his party at the time he was elected for that exercise.

For the extraction of these data, again it was necessary to develop an algorithm that captured this document as a whole and transformed the format of the data contained therein, extracting and converting them into a structure that allowed inserting this information in a database, external to the senate site, built specifically for our proposed visualization.

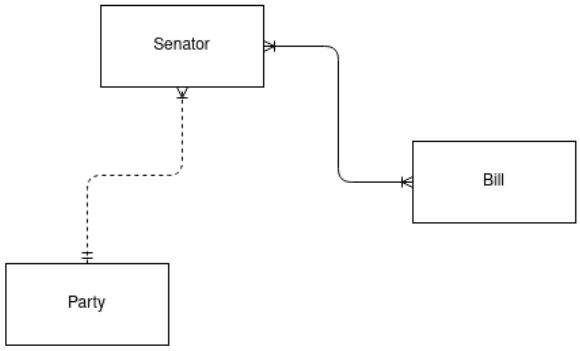
This data collection on the Senate's legislative activities, more specifically on open voting, was carried out using only the data made available on the Senate's official website, without the use of any project kept outside the official government scope.

In this study, the strategy was to acquire data from the Senate web site search about

legislative activities on bills that are already voted. Filtering the type or the kind of bill's proposition was ignored, as well as any other type of filter, except the proposed time unit: bills voted between 2007 and 2010. The volume of documents found, with open votes for this period, was: 34 in 2007, 50 in 2008, 64 in 2009, and 28 in 2010 - totaling 176 articles.

This time period was chosen because it had more recent data and is a closed legislative period at the time of the elaboration of this research: the first semester of the year 2011 (RODRIGUES; SANTANA, 2011). However, this construction model can be applied in any legislative period and exercise, as long as access to parliamentarian's metadata and data on the movement of legislative activities can be obtained.

Figure 1 – Logical representation of primary data and relationships between senators, parties, and voting



Source: Elaborated by the authors

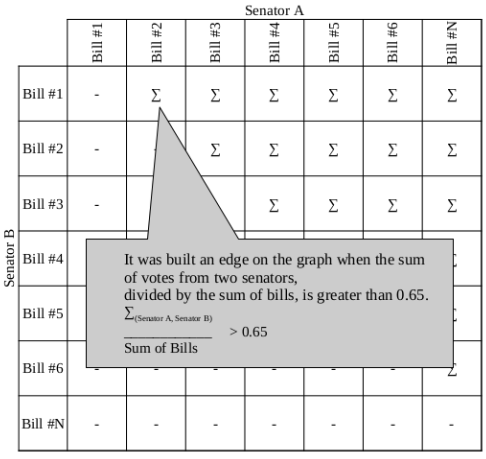
A viable alternative for the use of data obtained from official sources for the generation of graphs and subsequent visualization is to transform them into records in a relational database (ODEWAHN, 2010), preferably instantiated in an open-source database (open source), such as SQLite Database, PostgreSQL or MySQL. However, the bipartisan entity-relationship modeling proposed by Odewahn (2010) does not fit the Brazilian scenario, mainly due to the Brazilian characteristic of multi-party representation. Figure 1 presents the proposed logical model for the primary data necessary for the specific construction of the graph on the affinities between parliamentarians based on the votes in plenary.

2.3 Generation of the affinity matrix

After obtaining the primary data and transforming it into a relational base, the second step is to extract the information about the affinities between senators in the chosen period of time.

These connections can be built by calculating an affinity matrix, fed with the data (votes) of these senators. The affinity matrix (Figure 2) has a structure similar to the concept of rows, columns, and cells in a spreadsheet. Its columns and rows represent all the subjects that had an open vote in a given period, with the column's subjects representing the votes of a senator (e.g. Senator A) and those of the lines representing another parliamentarian. (e.g. Senator B).

Figure 2 – Affinity matrix of votes between two senators



Source: Adapted from Odewahn (2010)

If the meeting of these lines has common values for both, that is, if the meeting of the same matter, the two senators voted equally, a point of affinity is marked. Subsequently, if the sum of the affinity points, divided by the total number of articles, is greater than the 0.65 factor, a factor proposed by Odewahn (2010) from Wilson's article (2009), an edge is created among senators in the graph. The matrix must be executed every two senators until everyone is confronted with each other (ODEWAHN, 2010, p. 124).

In generating the affinities of open votes in the Brazilian Senate, the same algorithm proposed by Odewahn (2010) was used. The treatment of this data can be done with any

programming language, since the technology to create the affinity matrix is irrelevant, provided that the algorithm respects the proposed rule. The affinity matrix can generate thousands of edges in a period of time, depending directly on the volume of materials computed and as the data volume of these affinities is relatively extensive to be calculated on-demand by a personal computer, the results obtained must be recorded in a new database instance.

2.4 Making of the graph of vertices and edges

The third step in the making of the visualization is the conversion of the data being generated by the affinity matrix into a graph, through a technological tool. Among various technological tools available for the construction of graphics, the adopted tool was the open-source software GraphViz which has the native construction of vertices and edges graphs - known as the neato model. The model of vertices and edges follows a similar model to that proposed by Wilson (2009), in which the vertices represent the senators and the edges represent the affinities between them. Thus, for each vertex, there was a numeric label representing a senator in a list (external to the graph) alphabetically ordered.

In building the graph for the Brazilian scenario, it was necessary to obtain an additional element of the primary data: the party representation. Brazil has a multiparty political system and, therefore, the colors of each vertex represents a party that had at least one representative in the 2007-2010 period. The identified parties were: Democratas; PC do B; PDT; PMDB; PP; PR; PRB; PSB; PSC, and also senators without a party affiliation.

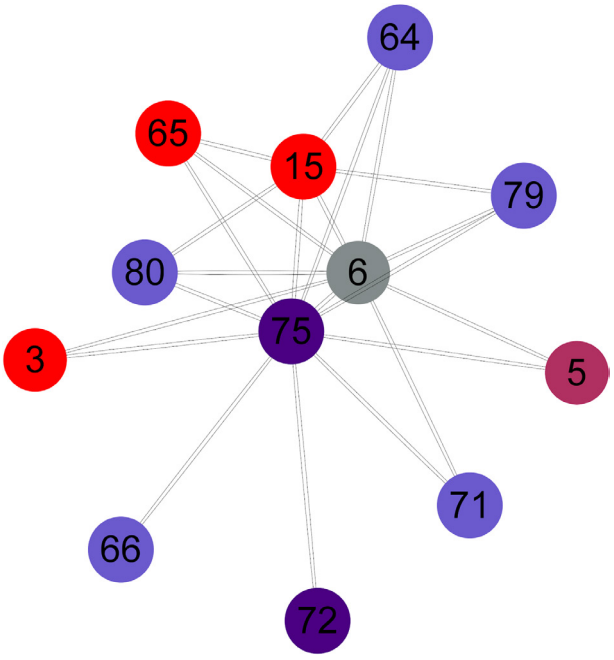
3 RESULTS

After completing the three-step sequence proposed by Odewahn (2010), the image result of the affinities of the open polls between senators, in the period from 2007 to 2010, is shown in Figure 3. The image shows only the members who had an affinity greater than or equal to the factor used (0.65). Senators who do not have an affinity greater than 0.65 with any other member of the Senate are not displayed.

This cluster is formed mainly by the parties PMDB (5 members), PT (3 members), PTB (2 members), DEM (1 member), PSB (1 member) in a total of 12 members with great affinities in their voting, mainly conducted by PTB senators (75) and PSB (6) who have a high-affinity index with PT senators (red vertices), with the opposition (senators at the bottom of the graph) and with PMDB (slate blue vertices), which have senators from the government base and center opposition).

Then, an analysis of smaller and isolated time units within the same mandate of this collegiate of senators is presented, using the same model of construction of the graph, but with matters voted in just one year, with the main objective of isolating important facts in the proposed time period.

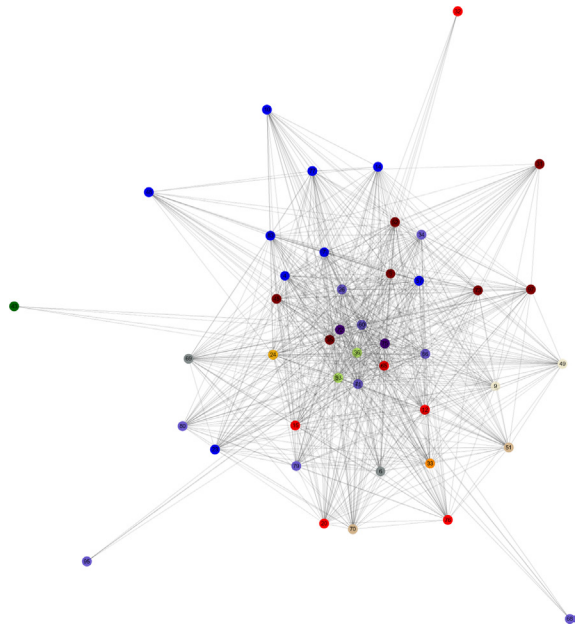
Figure 3 – Graph of vertices and edges representing senators who had affinities between their open votes between 2007 and 2010



Source: Elaborated by the authors

Figure 4 shows the graph of affinities among senators based on the subjects voted only in 2010. Affinities are found between party members: PMDB (9 members), PSDB (9 members), DEM (7 members), PT (6 members), PRB (2 members), PTB (2 members), PDT (2 members), PSB (2 members), PR (2 members), PC do B (1 member), PP (1 member), PSOL (1 member).

Figure 4 – Graph of vertices and edges representing senators who had affinities between their open votes in 2010



Source: Elaborated by the authors

However, Figure 4 shows members of the PSDB and DEM parties with a strong affinity at the top of the graph and four PT members permeated by members of acronyms with fewer representatives in the collegiate, such as the PR, PSB, and PRB; but these have strong links with the government's allied base. Parties such as PSOL (senator number 43, dark green node on the left) maintain a long distance from the allied base and the opposition, despite this party being represented in that mandate and in this graph by only a single senator. This graph makes perceptible a group of members articulated around the votes during 2010.

Added to the results obtained is the set of all open votes of senators and the respective affinities calculated among all of them, which allows us to identify a general framework of positioning in relation to the average affinity of senators of each party in relation to the other parties.

This average affinity was obtained by summing up all the votes between each member of parliament in relation to the other members of parliament and dividing this sum by the number of elements calculated. These data, collected in spreadsheets containing a code of the parliamentarian, the party that is affiliated and followed by three other columns containing the code of each of the other parliamentarians, their respective parties and the index calculated between each of them was handled by the consolidation application of affinities between parties, and that allows for the elaboration of correlation tables of affinities between the votes of the representatives of each party, as presented in Tables 1 to 4. For situations in which the party has only one member, in calculating its affinity index within the party itself resulting in an average of 0.00.

Table 1 - Affinity table between Senate parties, in 2007, only by open votes

Party	DEM	PMDB	PT	PSDB	PR	PDT	PTB	PP	PCdoB	PSOL	PRB	PFL	PSB
DEM	0.1194	0.1378	0.1551	0.1396	0.1329	0.1138	0.1051	0.2059	0.1557	0.0606	0.1202	0.0000	0.1488
PMDB	0.1378	0.3269	0.3717	0.2608	0.3101	0.2658	0.2428	0.4468	0.3964	0.1737	0.2675	0.0930	0.3739
PT	0.1551	0.3717	0.4545	0.2484	0.3667	0.3157	0.2849	0.5049	0.4436	0.1520	0.3272	0.0771	0.4330
PSDB	0.1396	0.2608	0.2484	0.3174	0.2261	0.1971	0.1807	0.3592	0.2878	0.2038	0.1670	0.1174	0.2815
PR	0.1329	0.3101	0.3667	0.2261	0.2735	0.1565	0.2353	0.4353	0.3765	0.1235	0.2618	0.0792	0.3490
PDT	0.1138	0.2658	0.3157	0.1971	0.2565	0.2382	0.1904	0.3529	0.3235	0.1176	0.2559	0.0678	0.2745
PTB	0.1051	0.2428	0.2849	0.1807	0.2353	0.1904	0.1607	0.3419	0.2941	0.1140	0.1893	0.0627	0.2745
PP	0.2059	0.4468	0.5049	0.3592	0.4353	0.3529	0.3419	0.0000	0.5294	0.2353	0.3529	0.1176	0.4902
PCdoB	0.1557	0.3964	0.4436	0.2878	0.3765	0.3235	0.2941	0.5294	0.0000	0.1765	0.3088	0.1228	0.4412
PSOL	0.0606	0.1737	0.1520	0.2038	0.1235	0.1176	0.1140	0.2353	0.1765	0.0000	0.0882	0.0969	0.1863
PRB	0.1202	0.2675	0.3272	0.1670	0.2618	0.2559	0.1893	0.3529	0.3088	0.0882	0.2353	0.0199	0.2892
PFL	0.0000	0.0930	0.0771	0.1174	0.0792	0.0678	0.0627	0.1176	0.1228	0.0969	0.0199	0.0984	0.0969
PSB	0.1488	0.3739	0.4330	0.2815	0.3490	0.2745	0.2745	0.4902	0.4412	0.1863	0.2892	0.0969	0.4020

Source: Elaborated by the authors

Table 2 - Affinity table between Senate parties, in 2008, only by open votes

Party	DEM	PMDB	PT	PSDB	PR	PDT	PTB	PP	PCdoB	PSOL	PRB	PSB	PSC	No Affiliation
DEM	0.2122	0.1913	0.2404	0.2385	0.2434	0.1515	0.1929	0.1838	0.3150	0.2650	0.1588	0.2819	0.2463	0.0300
PMDB	0.1913	0.2418	0.3247	0.2187	0.3266	0.1873	0.2561	0.2664	0.3927	0.3427	0.2145	0.3909	0.1955	0.0300
PT	0.2404	0.3247	0.4244	0.2800	0.4342	0.2441	0.3410	0.3369	0.5108	0.4015	0.2877	0.5138	0.2569	0.0462
PSDB	0.2385	0.2187	0.2800	0.2669	0.2808	0.1733	0.2164	0.1862	0.3677	0.3169	0.1692	0.3200	0.2569	0.0262
PR	0.2434	0.3266	0.4342	0.2808	0.4367	0.2575	0.3444	0.3750	0.5050	0.4050	0.2800	0.5200	0.2850	0.0500
PDT	0.1515	0.1873	0.2441	0.1733	0.2575	0.1280	0.1956	0.2133	0.3000	0.2567	0.1633	0.2950	0.1700	0.0200
PTB	0.1929	0.2561	0.3410	0.2164	0.3444	0.1956	0.2589	0.2756	0.4000	0.2956	0.2200	0.4144	0.2111	0.0422
PP	0.1838	0.2664	0.3369	0.2862	0.3750	0.2133	0.2756	0.0000	0.4200	0.3200	0.3600	0.4500	0.2400	0.0200
PCdoB	0.3150	0.3927	0.5108	0.3677	0.5050	0.3000	0.4000	0.4200	0.0000	0.5200	0.2400	0.6100	0.3400	0.0400
PSOL	0.2650	0.3127	0.4015	0.3169	0.4050	0.2567	0.2956	0.3200	0.5200	0.0000	0.2800	0.4600	0.2800	0.0000
PRB	0.1588	0.2145	0.2877	0.1692	0.2800	0.1633	0.2200	0.3600	0.3400	0.2800	0.0000	0.3700	0.2000	0.0200
PSB	0.2819	0.3909	0.5138	0.3200	0.5200	0.2950	0.4144	0.4500	0.6100	0.4600	0.3700	0.6200	0.3400	0.0600
PSC	0.2463	0.1955	0.2569	0.2569	0.2850	0.1700	0.2111	0.2400	0.3400	0.2800	0.2000	0.3400	0.0000	0.0400
No Affiliation	0.0300	0.0300	0.0462	0.0262	0.0500	0.0200	0.0422	0.0200	0.0400	0.0000	0.0200	0.0600	0.0400	0.0000

Source: Elaborated by the authors

Table 3 - Affinity table between Senate parties, in 2009, only by open votes

Party	DEM	PMDB	PT	PSDB	PR	PDT	PTB	PP	PCdoB	PSOL	PRB	PSB	PSC	PV	No Affiliation
DEM	0.3116	0.2698	0.2988	0.3087	0.3069	0.2156	0.3503	0.3192	0.3690	0.2031	0.2885	0.4057	0.1953	0.0703	0.0273
PMDB	0.2698	0.2287	0.2658	0.2620	0.2783	0.1826	0.3089	0.2633	0.2352	0.1594	0.2645	0.3539	0.1563	0.0555	0.0219
PT	0.2988	0.2658	0.2951	0.2928	0.3014	0.2193	0.3499	0.3041	0.3017	0.1791	0.3053	0.4153	0.1875	0.0709	0.0331
PSDB	0.3087	0.2620	0.2928	0.3046	0.2958	0.1982	0.3423	0.2896	0.2563	0.1563	0.2823	0.3828	0.1885	0.0458	0.0250
PR	0.3069	0.2783	0.3014	0.2958	0.3021	0.1908	0.3354	0.2461	0.2227	0.1719	0.2969	0.3848	0.1172	0.0508	0.0234
PDT	0.2156	0.1826	0.2193	0.1982	0.1908	0.1548	0.2575	0.2545	0.2522	0.1518	0.2176	0.3080	0.1741	0.0759	0.0212
PTB	0.3503	0.3089	0.3499	0.3623	0.3354	0.2575	0.4107	0.3828	0.3535	0.2129	0.3848	0.4775	0.2602	0.0742	0.0273
PP	0.3192	0.2633	0.3041	0.2896	0.2461	0.2545	0.3828	0.0000	0.3750	0.2813	0.2891	0.4531	0.2969	0.0938	0.0313
PCdoB	0.2690	0.2352	0.3017	0.2563	0.2227	0.2522	0.3535	0.3750	0.0000	0.1719	0.3047	0.4219	0.2813	0.0781	0.0391
PSOL	0.2031	0.1594	0.1791	0.1563	0.1719	0.1518	0.2129	0.2813	0.1719	0.0000	0.1563	0.2734	0.1406	0.0938	0.0313
PRB	0.2885	0.2645	0.3053	0.2823	0.2969	0.2176	0.3949	0.2891	0.3047	0.1563	0.2969	0.4102	0.1328	0.0391	0.0352
PSB	0.4057	0.3539	0.4153	0.3828	0.3848	0.3080	0.4775	0.4531	0.4219	0.2734	0.4102	0.5000	0.2969	0.1250	0.0469
PSC	0.1953	0.1563	0.1875	0.1885	0.1172	0.1741	0.2302	0.2969	0.2813	0.1406	0.1328	0.2969	0.0000	0.0938	0.0000
PV	0.0703	0.0555	0.0709	0.0458	0.0508	0.0759	0.0742	0.0938	0.0781	0.0938	0.0391	0.1250	0.0938	0.0000	0.0156
No Affiliation	0.0273	0.0219	0.0331	0.0250	0.0234	0.0212	0.0273	0.0313	0.0391	0.0313	0.0352	0.0469	0.0000	0.0156	0.0156

Source: Elaborated by the authors

Table 4 - Affinity table between Senate parties, in 2010, only by open votes

Party	DEM	PMDB	PT	PSDB	PR	PDT	PTB	PP	PCdoB	PSOL	PRB	PSB	PSC	PV
DEM	0.3565	0.3209	0.3855	0.3763	0.4023	0.3300	0.3170	0.5312	0.4933	0.3996	0.4687	0.4888	0.3437	0.1562
PMDB	0.3209	0.3045	0.3873	0.3339	0.3929	0.3152	0.3048	0.5018	0.4929	0.3875	0.4688	0.4759	0.3000	0.1518
PT	0.3855	0.3873	0.4659	0.4047	0.4786	0.3857	0.3719	0.6179	0.6036	0.4571	0.5875	0.5982	0.3393	0.2071
PSDB	0.3763	0.3339	0.4074	0.4054	0.4169	0.3490	0.3281	0.5670	0.5156	0.3750	0.5033	0.5190	0.3147	0.1965
PR	0.4023	0.3929	0.4786	0.4169	0.4583	0.3854	0.3763	0.6161	0.6161	0.5089	0.5759	0.5938	0.3929	0.1875
PDT	0.3300	0.3152	0.3857	0.3490	0.3854	0.2833	0.3053	0.5060	0.4762	0.3750	0.4673	0.4732	0.3095	0.1667
PTB	0.3170	0.3048	0.3719	0.3281	0.3763	0.3053	0.2670	0.4949	0.4745	0.3827	0.4515	0.4566	0.3061	0.1327
PP	0.5312	0.5018	0.6179	0.5670	0.6161	0.5060	0.4949	0.0000	0.7857	0.6071	0.7679	0.7857	0.5000	0.2500
PCdoB	0.4933	0.4929	0.6036	0.5156	0.6161	0.4762	0.4745	0.7857	0.0000	0.6071	0.7500	0.7500	0.4286	0.2143
PSOL	0.3996	0.3875	0.4571	0.3750	0.5089	0.3750	0.3824	0.6071	0.6071	0.0000	0.5536	0.5536	0.4643	0.0714
PRB	0.4687	0.4688	0.5875	0.5033	0.5759	0.4673	0.4515	0.7679	0.7500	0.5536	0.6786	0.7321	0.4286	0.2500
PSB	0.4888	0.4759	0.5982	0.5190	0.5938	0.4732	0.4566	0.7857	0.7500	0.5536	0.7321	0.7143	0.4286	0.2500
PSC	0.3437	0.3000	0.3393	0.3147	0.3929	0.3095	0.3061	0.5000	0.4286	0.4643	0.4286	0.4286	0.0000	0.0714
PV	0.1562	0.1518	0.2071	0.1964	0.1875	0.1667	0.1327	0.2500	0.2143	0.0714	0.2500	0.2500	0.0714	0.0000

Source: Elaborated by the authors

The occurrence of an affinity index equal to or greater than 0.65 in the case of relations between the parties is considerably lower since it takes into account the participation of all its affiliated parliamentarians. In the collection carried out in 2007, the highest affinity index found was 0.5294, among members of the PC do B and PP. In 2008, the greatest affinity was found in the relationship between the members of the PSB and PSOL and between the PSB and the PR (0.5200). Referring to 2009, once again the affinity in the votes between PSB members (internal) was the highest index (0.5000) and, between parties, it was 0.4775, between the PTB and the PSB. Only in the matrix for the year 2010, affinity indices greater than 0.65 were calculated, and these occurrences are highlighted in Table 4.

The identification of the greatest affinity between the parties perceived in 2010 can also be highlighted by obtaining the average of the affinities calculated in each of the years, whereas in 2007 the average was 0.2331, in 2008 it was 0.2584, in 2009 the calculation indicated the lowest index 0.2141 and in 2010 the highest index, 0.4175.

Table 5 - Affinity table between the Workers' Party (PT) and other parties, only by open votes, from 2007 to 2010

Year	Party																
	DEM	PMDB	PT	PSDB	PR	PDT	PTB	PP	PCdoB	PSOL	PRB	PFL	PSB	PSC	PV	No Affiliation	Average
2007	0.1551	0.3717	0.4545	0.2484	0.3667	0.3157	0.2849	0.5049	0.4436	0.1520	0.3272	0.0771	0.4330	0.0000	0.0000	0.0000	0.2584
2008	0.2404	0.3247	0.4244	0.2800	0.4342	0.2441	0.3410	0.3369	0.5108	0.4015	0.2877	0.0000	0.5138	0.2569	0.0000	0.0462	0.2902
2009	0.2988	0.2658	0.2951	0.2928	0.3014	0.2193	0.3499	0.3041	0.3017	0.1791	0.3053	0.0000	0.4153	0.1875	0.0709	0.0331	0.2388
2010	0.3855	0.3873	0.4659	0.4074	0.4786	0.3857	0.3719	0.6179	0.6036	0.4571	0.5875	0.0000	0.5982	0.3393	0.2071	0.0000	0.3933
Average	0.2700	0.3374	0.4100	0.3072	0.3952	0.2912	0.3369	0.4410	0.4649	0.2974	0.3769	0.0193	0.4901	0.1959	0.0695	0.0198	0.2952

Source: Elaborated by the authors

Another analysis that can be proposed is the relationship of one party with the others, and in this case, we opted for the Workers' Party for being the main party of the executive government at the moment, so that a matrix of affinities between the parliamentarians of each party can be established over the years, thus allowing perception of the fluctuations in the affinities of each party and, also each year, as shown in Table 5.

It is perceived that when analyzed as a whole, one can have a clearer perception of aspects such as party cohesion, that is, the index of affinity between the decisions taken by the members of the same party and also the proximity that each of the parties presents among themselves.

Based on such data, one can elaborate a variety of differing conclusions, and the consistency between what is presented in the speeches as a line of conduct for decision-making and the reality of the votes of each of the parliamentarians and their respective parties can be monitored.

4 CONCLUSIONS

The description of the process for obtaining data on open votes shows that the scattered availability of primary data for votes in the Brazilian Senate reduces the alternatives for building views of parliamentary votes; as well as the lack of machine-readable structures and formats of these data on the official website, that makes difficult to retrieve and process information. The eight principles of open data have not been considered in the presentation of this data by the Senate so far, which would facilitate the diversified use of this information, enabling the construction of applications for these data by society itself.

Analyzes can be constructed from the access to data and the elaboration of visualizations of these data, such as those obtained on each of the votes of the parliamentarians, such as the identification of party clusters, the real position of each representative concerning the others and the parties, as well as the consistency of each of our representatives in voting during their terms.

We cannot associate any party cluster in the 2007-2010 period from the data collected, nor between allied parties and opponents. It is also not possible to affirm the existence of groups of senators within parties. There is no standard from the data collected that identifies the existence of two opposing groups (opposition and government) in the Brazilian senate. It is considered that the other senators, who participated in the collegiate between 2007-2010 and are absent in the graph due to lack of affinity with all collegiate in more than 65%, had a dispersed posture in the open votes. that is, they are never part of a fixed group of senators.

The graph of the year 2010 (Figure 4) shows the amount of associations increases due to the articulation of the opposition and government in favor of the approval (or veto) of matters. However, it is important to note that, in 2010, the number of voting materials captured on the Senate website was the least significant compared to other years in the chosen period - 28 articles. However, this period also includes open voting on socially important matters, such as the ineligibility of candidates seeking to protect public administrative probity; the creation of the Social Fund; the onerous assignment to PETROBRAS activities in the research and mining of oil, natural gas, and fluid hydrocarbons; the extension of benefits to the Manaus Free Trade Zone; the timeless permanence of the Poverty Combat and Eradication Fund; the new wording that prevents the use of the magistrates' retirement as a disciplinary measure and the permission to lose their position; and the increase to one hundred and eighty days in the duration of the period of maternity leave, many of them with strong popular appeal.

Data access and the many possibilities for building visualization for this kind of data are the key elements in building citizen participation and for achieving a more transparent relationship between the government and society.

VISUALIZAÇÃO DE AFINIDADES ENTRE PARLAMENTARES MEDIANTE DADOS DE VOTAÇÕES NO SENADO BRASILEIRO

RESUMO *A ampliação da participação cidadã na esfera pública depende diretamente do acesso as informações relativas a atuação dos representantes eleitos, principalmente no que diz respeito a suas decisões nas votações durante seus mandatos. A Ciência da Informação pode colaborar neste processo, propondo e avaliando modelos de acesso a estas informações que podem ser obtidos através de dados a serem disponibilizados pelos sítios oficiais do Poder Legislativo nas esferas Federal, Estadual e Municipal. Propõe-se neste artigo a análise do processo de coleta e uso de dados sobre votações de senadores com vistas a apropriação deste modelo para coleta e uso de dados nas demais esferas. A partir dos dados é analisada a elaboração de uma matriz de afinidades que permita identificar a relação entre cada um dos parlamentares com os demais, em função das similaridades das decisões tomadas no conjunto das votações abertas. É analisa também a elaboração de visualizações iniciais e a ampliação do escopo da pesquisa através da aplicação dos dados obtidos em todas as afinidades entre os parlamentares e na obtenção de uma afinidade média entre partidos, permitindo novas dimensões de análise aos dados coletados. A elaboração das matrizes completas das relações de afinidades entre os parlamentares, pode propiciar um horizonte rico de possibilidades para elaboração de novas formas de visualização e análise, ampliando a visibilidade das ações parlamentares junto a sociedade.*

Palavras-chave*: Dados Abertos. Visualização de dados. Transparência Pública.*

Artigo recebido em 26/06/2012 e aceito para publicação em 17/03/2013

REFERENCES

BERNERS-LEE, T. On the next web. **TED Conference 2009.** Available at: http://www.ted.com/index.php/talks/tim_berners_lee_on_the_next_web.html. Accessed on: 28 Sep. 2011.

DIESTEL, R. **Graph theory.** New York, USA: Springer- Verlag. 2000.

LOCKE, S. **Graph Theory.** Department of Mathematical Sciences at Florida Atlantic University: Boca Raton, FL, USA. 2000.

ODEWAHN, A. Visualizing the U.S. Senate Social Graph (1991-2009). In: STEELE, J; ILIINSK N. **Beautiful Visualization: looking at data through the eyes of experts.** O'Reilly: Sebastopol, CA, USA, 2010.

OGD - OPEN GOVERNMENT DATA. **8 Principles of Open Government Data.** 2007. Available at: <https://public.resource.org/8_principles.html>. Accessed on: 22 Sep. 2011.

RODRIGUES, F. A.; SANTANA, R. C. G. Visualizando grafos sociais do Senado brasileiro (2007-2010) através de ferramentas de Tecnologia da Informação e Comunicação. In: ENCONTRO INTERNACIONAL DE INFORMAÇÃO, CONHECIMENTO E AÇÃO (EIICA), 7., 2011, Marília, SP, Brazil. **Resumos...** Marília, Brazil : UNESP, 2011. p. 27-28.

SANTOS, P. L. V. A. da C.; ALVES, R. C. V. Metadados e web semântica para estruturação da web 2.0 e web 3.0. **DataGramaZero**, v. 10, n. 6, Nov./Dec. 2009.

SENSO, J. A.; ROSA PIÑERO, A. de la. El concepto de metadato. Algo más que descripción de recursos eletrónicos. **Ciência da Informação**, Brasília (Brazil), v. 32, n. 2, p. 95-106, May/Aug. 2003.

THE WHITE HOUSE. **The Obama Administration's Commitment to Open Government: status report.** Available at:

https://obamawhitehouse.archives.gov/sites/default/files/opengov_report.pdf. Accessed on: 29 Jan. 2011.

WILSON, R. **Introduction to graph theory**. 4. ed. Harlow, UK: Longman. 1996.

WILSON, C. The Senate Social Network. Accessed on, Washington (USA), Apr. 2009. Available at: http://www.slate.com/id/2217204. AAccessed on : 01 Jun. 2011.