Connect 2 Congress: Visual Analytics for Civic Oversight

Peter Kinnaird
Georgia Institute of Technology
85 5th Street NW
Atlanta, GA 30308 USA
peterkinnaird@gatech.edu

Mario Romero
Georgia Institute of Technology
85 5th Street NW
Atlanta, GA 30308 USA
mario@gatech.edu

Gregory Abowd
Georgia Institute of Technology
85 5th Street NW
Atlanta, GA 30308 USA
abowd@gatech.edu

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
Strong representative democracies rely on educated, informed, and active citizenry to provide oversight of the government. We present Connect 2 Congress (C2C), a novel, high temporal-resolution and interactive visualization of legislative behavior. We present the results of focus group and domain expert interviews that demonstrate how different stakeholders use C2C for a variety of investigative activities. The evaluation provided evidence that users are able to support or reject claims made by candidates and conduct free-form, low-cost, exploratory analysis into the legislative behavior of representatives across time periods.

Keywords
Information Visualization, E-Government, Voting Analysis, Poole-Rosenthal, Roll Call Analysis

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors

Introduction
The most critical watchdog mechanism on democracy in the 21st century is the dispersal of information among the general population. Citizens have unprecedented

access to vast amounts of data regarding numerous governmental processes. All democracies rely on an educated electorate. In representative democracies, such as the United States (US), citizens will be unable to make informed voting decisions without information regarding the actions of their representatives. These actions are numerous and diverse, from speeches and fact-finding missions, to interaction with constituents and voting on legislation. In this paper, we focus on roll call votes. Roll call votes take place in the US Congress and require the representative to explicitly select one of three options (Yes, No, Present), which then becomes a matter of public record.

The US Congress consists of two chambers, the House of Representatives and the Senate, with elected officials who represent constituents in particular geographic regions. The House consists of 435 voting members and 6 non-voting members. The Senate consists of 100 voting members, with the Vice President of the US casting a vote only in the event of a tie. Collectively, we refer to Senators and Representatives as representatives, utilizing the proper noun Representatives only to refer to members of the House of Representatives. The 110th Congress refers to the set of representatives who served during 2007 and 2008. New US laws begin as bills, possibly with amendments (collectively, legislation), in Congress. Our database for the 110th Congress contains 14,039 bills, 6,886 amendments, and 2,533 votes.

Given the magnitude and complexity of the roll call voting data, citizens need assistance in order to keep informed on their representatives' voting record. It is labor intensive to discover how a representative voted on a single piece of legislation, let alone on a categorical issue, such as taxes or healthcare. Performing high-level analyses of voting patterns is a challenging and resource-intensive task typically undertaken only by investigative reporters. Historically, citizens have fulfilled their responsibility to stay informed by relying on journalists to report on important or interesting actions taken by their representatives.

Existing voting data sources contain so much data that users are often left without guidance or understanding. Interactive visualization systems, which reduce the overhead for understanding focus on statistical analysis. However, none of these systems allows the user to view changes in the data over relatively short (less than two year) time periods.

We have developed a visualization system, Connect 2 Congress (C2C), which is designed to reduce the level of complexity and time required to discover and comprehend congressional voting patterns. The system provides three essential services to the user: (1) a gateway to browsing tabularized data; (2) a primary source of new data (Poole-Rosenthal scores [23] and other statistical data); and (3) an interactive visualization system to explore the data and the changes exhibited over time.

We present two primary contributions:

- Merging quantitative Political Science practices with techniques from HCI;
- Demonstrating a generalizable technique for making voting records accessible;

Due to space constraints, we have omitted the details of our qualitative evaluation, but retained the results. Our evaluation identified four user communities and demonstrated utility for each by reducing the cost of investigations, such as testing hypotheses or free form, exploratory investigations.

We begin with an analysis of related work in Political Science literature and related sources of information, including other visual analytic systems. We then detail our design and implementation of C2C. Next, we briefly explain our evaluation and results. We then discuss the significant impacts of our system as well as the contributions. We conclude with a recap of our contributions and suggestions for future work.

Background and Related Work

Roll Call Analysis

The Spatial Theory of Voting, put into practice with roll call analysis, has gained considerable traction in the past 30 years. Clinton concisely highlights [8] the value of roll call analysis:

In short, roll call analysis makes conjectures about legislative behavior amenable to quantitative analysis, helping make the study of legislative politics an empirically grounded, cumulative body of scientific knowledge.

The Spatial Theory of Voting can be applied to map a representative's political position to a point on a spectrum [9:2-3]. Poole and Rosenthal determined a method of computing these positions, and showed that it is possible to determine the position of a politician based only on the voting records [23], absent all knowledge of political party, bill content, etc. Poole and Rosenthal's methodology and algorithm ("a scaling procedure that performs parametric unfolding of binary choice data") compares favorably to more modern algorithms [7], such as a Bayesian approach [8].

Poole-Rosenthal scores and other similar scoring methods are sometimes computed by large newspapers or magazines such as *The New York Times* or *The National Journal.* They are typically computed only once every few years and the vote set is carefully selected by the authors. *The National Journal* used similar scores to run cover articles, for instance, in 2004 proclaiming John Kerry the most liberal Senator [18], and in 2008 declaring Barack Obama the most liberal Senator [10].

Raw Data Sources

The Congressional Record maintains the official record of all public actions and votes taken in Congress. This data is publicly available through the Library of Congress's THOMAS website [1]. Users can browse the site or search for specific information regarding bills, amendments, votes, and representatives. The THOMAS system provides a wealth of data to the determined user. Turning that data into information is extremely labor intensive. The Washington Post also provides a voting database [2].

Visual Analytic Systems

A number of other organizations provide reports and metrics for Congressional voting, often attaching their own bias. Project Vote Smart provides a thorough catalog of these ratings [3]. GovTrack [4] also provides visibility to actual votes as well as original analysis in the form of static graphs for each representative. For example, GovTrack created a Political Spectrum, which ran a Singular Value Decomposition (SVD) on cosponsorship data in January 2009 with the 2007 and

2008 data in order to generate a graph displaying each representative and how liberal or conservative the representative is [20].

VoteWorld [17], created by Howard Rosenthal, provides visualizations of representatives' Poole-Rosenthal scores from 1789 to 2000. VoteWorld does not implement filtering and only performs the position computations for each 2 year period. As a result, users are not able to see changes in behavior except from one two-year period to the next.

The SocialAction system visualized common voting groups as edge strengths in a force-directed graph for the Senate alongside limited filtering controls [14]. The system only visualized these patterns at a single point in time. In addition, the applied graphing technique did not yield a useable visualization until the user filtered out a vast amount of data. One significant contribution of SocialAction in this domain, however, was that the combination of visualization and statistics not only enabled discovery, but also assisted with communicating those discoveries to peers.

An earlier version of C2C [13], much like Govtrack's Political Spectrum, provides an SVD analysis, choosing to focus on voting patterns rather than cosponsorship analysis. C2C conducted the analysis in two-week increments, presenting snapshots which enable partisanship analysis as well as pair-wise distance comparisons. The two-week increments are arbitrarily selected both in terms of start and end date, and duration. Additionally, the SVD algorithm provides a poor layout for this kind of data since the scale of the graph continually changes and changes in absolute position are not meaningful.

Connect 2 Congress

We now present our software system, emphasizing its support of the visual information-seeking mantra [19]: "overview first, zoom and filter, then details on demand." C2C is implemented as a web application (Java Applet, PHP, MySQL, and Javascript) capable of running in any modern browser. All of the data in C2C comes from govtrack.us [4].

The Controller

Aside from standard web-based widgets, the primary control scheme is a custom-coded, vertical timeline. The timeline resembles a dynamic query slider [6] functionally, but is designed as a scented widget [22]. We display small, light gray tick marks on all the dates during which voting took place. The tick marks often appear in groups corresponding to five-day business weeks. There are two triangular indicators on the timeline that can be moved to any date. By placing one slider at a starting date and the other on an end date the user specifies a period for analysis, updating all corresponding views. By default, the entire two-year period is selected providing *overview first.*

The Spectrum Visualization

C2C shows two square scatter-plot graphs each representing a political spectrum. On the spectrum, each representative is displayed as a small, semi-transparent circle in the color typically associated with their political party (Democrats as blue, Republicans as red, and Independents as yellow).

Moving either slider updates each Spectrum with the representatives' new positions based on their behavior

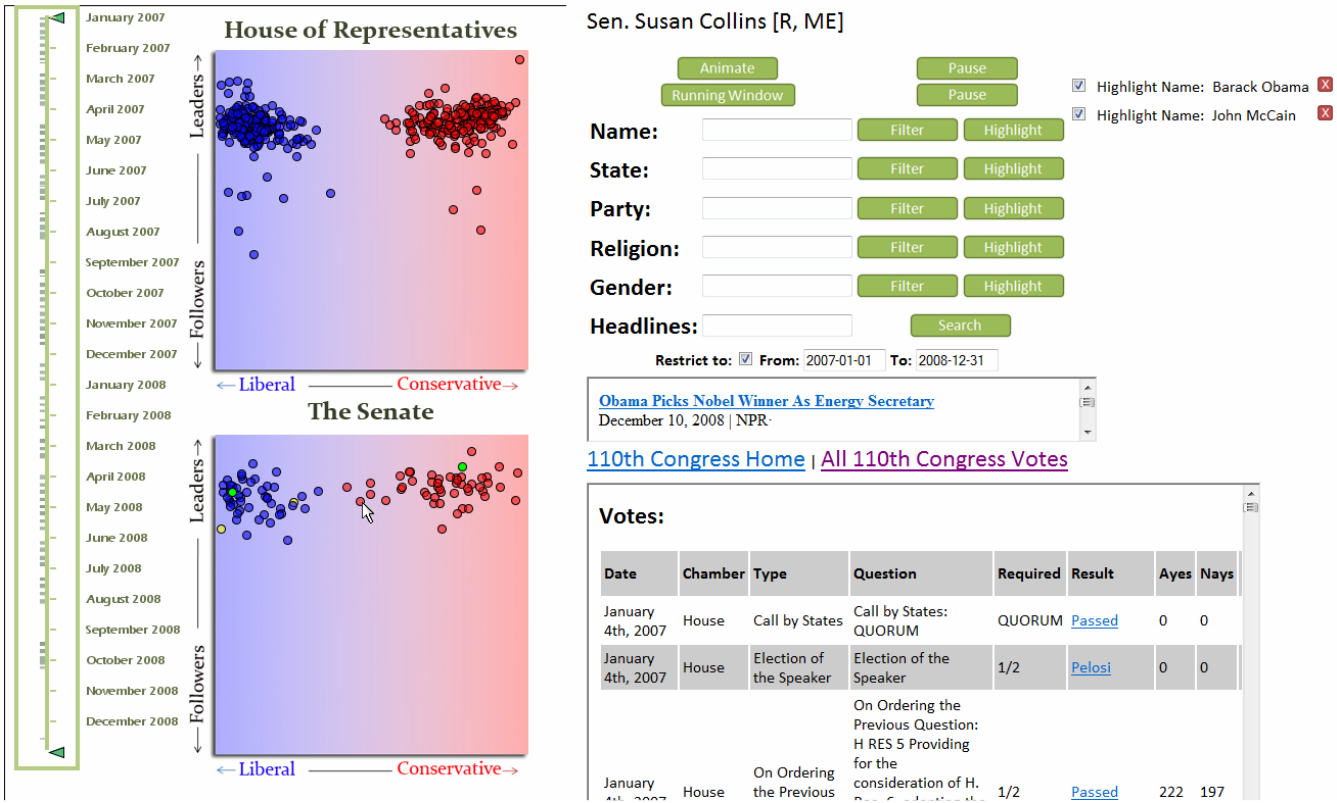

Figure 1: Connect 2 Congress displaying behavior of Congressional representatives exhibited during the entire 110th Congress. Barack Obama and John McCain are highlighted. Senator Susan Collins, Republican of Maine is under the mouse cursor.

during the newly selected period. To enable visual tracking, we animate the shift from one position to the next using linear interpolation. Clicking on a representative's circle loads relevant information in the InfoBox, providing *details on demand.* Mousing over a circle displays the representative's name, party, and state.

The position of a representative's circle is the result of two independent, mathematical analyses, one for each axis. We calculated Poole-Rosenthal scores using a program provided by Keith T. Poole called W-NOMINATE [15]. For each unique pair of selectable dates we run W-NOMINATE on all roll call votes cast during that period. We used Poole-Rosenthal scores for the horizontal positions of representatives.

Although Poole-Rosenthal scores are widely accepted among political scientists, we expected some of our users to be skeptical, not of our computation of these scores, but of their applicability and appropriateness. With that in mind, we developed a simpler calculation for the vertical positions that could be easily explained to the users.

LEADERS-FOLLOWERS ANALYSIS

We developed an original approach for statistical analysis of bill sponsorship and cosponsorship. Most bills have a single sponsor attached to them. This sponsor is generally the person who authored the bill. Bills can have many cosponsors who join later.

For each bill, we compute a representative's leadership on the following basis:

- Sponsors receive 1 point.
- Each cosponsor receives a fraction of 1 point, depending on how far along the bill was in the process when they joined as a cosponsor.

For cosponsors, we look at the entire duration that the bill had any updates or actions taken on it and we assign points accordingly. For example, if the period between a bill's introduction and the last action taken on it is 30 days and a representative joins as a cosponsor after 10 days, the representative receives 0.666 points.

To compute a person's position for a selected period we take the average of the representative's scores for all bills the person sponsored or cosponsored during the given period. The result is a value between 0 and 1 for each unique pair of dates on which voting took place for each representative. We scale these values to the height of the spectrum. Therefore, sponsoring or cosponsoring numerous bills yields a high score, placing the representative near the top of the spectrum. We label the axis with "Leaders" and "Followers" for simplicity.

Filter and Highlight

The default view of the system retains all pertinent data over the largest possible date range. Although we do not implement traditional zooming functionality, we have implemented filtering and highlighting tools to support user queries, a type of semantic *zoom and filter*. The user is able to enter a search for name, state, party, religion, and gender. The user may then select either filter or highlight buttons.

Filter queries remove all representatives from the spectrums that do not match. For example, filtering on party with the term "Democrat" will remove all non-democrats from the display. Highlight functions turn the representatives who match a highlight query bright green (for visibility), but they leave other representatives within the view for context.

NewsBox

On selection of a period we display the top headlines as returned by National Public Radio (NPR)'s search API [5] in an iframe, which functions as a mini-browser. We selected NPR for three primary reasons: it is a source provided by the government; the availability of historic data which fully extends through 2007; it is both easy to use and free of charge.

We also support various headline searches. Users are able to perform keyword searches and specify dates on which to search. The intended use of these searches is to investigate possible causes of interesting behavior, as well as to discover potentially interesting dates on which to seek unusual behaviors. For example, a user may remember "the increase in gas prices following a hurricane which hit Texas sometime in 2007 or 2008." Using the headline search, the user can determine the

date on which this occurred and examine behaviors in that period. Alternatively, noticing a sudden shift in behavior, the user might discover that the shift was temporally correlated with the collapse of Lehman Brothers financial group.

InfoBox

We also present various data primarily as text and tables in an iframe. For example, a simple list of representatives in the 110th Congress along with their birthdays, religions, gender, state, district, and a link to their YouTube page. Clicking on a representative's name loads an alternative page into the InfoBox with information about that representative such as the committees they have served on and the bills they have sponsored.

Animations

We also enable two different modes of animation for different types of analysis and discovery. We name them Growing Window and Running Window. Before describing the functionality, it is necessary to explain a feature of the Poole-Rosenthal scores that became obvious through use.

When looking at more than approximately 8-12 months of data, most representatives settle into a location from which they hardly move. Therefore, the changes in position in a given 2 or 3 month period are much more dramatic than those when shifting from a 19 month view to a 20 month view.

Both types of views can still yield valuable results, however. By looking at slight shifts over a long period, users may be able to spot trends in both individual and group behavior. On the other hand, by examining short durations at different periods, users can evaluate the reactions of representatives to specific events. As an example, one might wish to investigate the reaction of representatives to the financial crisis using a short duration window. Another investigation may be whether an individual began shifting toward more liberal, conservative, or moderate positions in the run-up to an election, using a larger duration.

GROWING WINDOW

In order to use this mode, the user first selects a pair of dates using the indicators, just as in all other interaction with the system. On clicking the "Animate" button, the top-most will remain fixed, while the bottom one will slowly tick to subsequent date selections. At each tick, all displayed representatives' are repositioned according to the behavior displayed over the newly selected period.

Growing Window mode provides historical context used to evaluate the importance of the more dramatic shifts that might be observed with the running window approach. By taking into account an increasing number of votes, the user can determine, for example, whether a short but substantial change in behavior has an important impact on the representative's overall evolutionary trend.

RUNNING WINDOW

The alternative animation mode is to keep a running window. The running window moves both indicators forward at the same time, yielding positions based on an approximately consistent quantity of votes (the actual number of votes from one "frame" to the next could differ based on the number of votes that took place on a given day in Congress).

The running window is most useful for examining behaviors over relatively short durations. For example, if the user is curious about positional or leadership shifts surrounding the Virginia Tech massacre, they might select a period of approximately 2 weeks prior to the massacre, and then activate a two-week running window. The user would undoubtedly notice that immediately after the massacre, representatives from Virginia jumped noticeably, indicating that they took a greater role in sponsoring or cosponsoring legislation.

Evaluation

We conducted a formative evaluation consisting of a focus group and interviews with domain experts to validate our design approach and determine the utility of the system. We identified four primary communities who might be interested in understanding and examining congressional voting records. These are social scientists, political strategists and politicians, journalists, and informed voters.

Results and Discussion

We will now present the results of our evaluation and impacts and contributions of our system based on user community.

Social Scientists

Social Scientists are often interested in models of group behavior. C2C serves as a primary source for Poole-Rosenthal scores, one model of legislative group and individual behavior. We have computed a very computationally intensive algorithm at a never-before seen temporal resolution. Using C2C, social scientists and historians visually explored and analyzed both individual and group behavior on a microscopic scale, relative to prior work.

Political Strategists and Politicians

C2C enables political strategists and politicians to spot trends and changes in behavior at both the individual and group level, which can be used to attack opponents and defend their own behavior. C2C also allowed politicians and strategists to evaluate the effectiveness of strategies, particularly party-unity or spread.

Journalists

C2C provides a primary source of Poole-Rosenthal scores for journalists. Journalists can use C2C to see trends and behavior as well as support, verify, or reject claims. Journalists can also see changes or abnormalities, leading them towards deeper investigations. The political journalist we interviewed appeared frustrated with both the complexity and overhead associated with roll call analysis and the lack of available tools. He was also able to describe six stories that he could imagine writing and publishing after briefly viewing a Running Window and then carefully examining the end state.

Informed Voters

Although C2C is probably too sophisticated a tool for laypeople to fully utilize, it could be made publicly available to the citizenry to use. Informed voters suggested that they might use C2C to find unbiased information about candidates' actions and to conduct their own inquiries into actions which may not otherwise receive much press. One participant from this community said:

When you look at the people who are editing the news …They're only going to make issue of the things they want you to know about. This [C2C] gives you the option of making issue with the other things.

Conclusions and Future work

We designed and implemented Connect 2 Congress (C2C), which computes and visualizes a model of legislative behavior [23] at a high level of temporal granularity, enabling users to understand, experience, and explore voting behavior interactively. We have provided considerable guidance, through a thorough description and considerable background research, to other researchers wishing to replicate the approach of C2C, perhaps for other voting domains.

When we first began working on C2C, it was unclear whether the temporal component of the visualization would be meaningful or helpful. In this yet-unexplored area, it was possible that the legislators would exhibit virtually identical behavior in most or all time frames. We discovered that the behaviors of this social network are extremely dynamic when examined at high temporal granularities less than approximately 1 year.

We have utilized existing mathematical models from Political Science and applied techniques from the HCI community to visualize and interact with the models. The result is an easy-to-use system which enables users to understand voting records, test claims, perform comparisons, uncover interesting trends and behaviors, and, ultimately, improve representative accountability.

Our system description and related work provide sufficient guidance for other researchers who may wish to build similar systems for deployment or evaluation. C2C is currently running and usable with data from 2007 and 2008 [12]. The underlying statistical modeling that we employed can also be utilized for other voting systems and contexts such as state governments, parliamentary systems or the United Nations General Assembly [16].

From a technical perspective, there are a number of ways that C2C could improve. Improving the speed and efficiency of our analytic algorithms could allow real-time computation, enabling both simple categorical analysis (how does my Senator score on Gun Control?) and complex categorical analysis (how do female representatives, independently analyzed, score on Abortion and Gay Marriage combined?). Like Many Eyes [11] and the Baby Name Generator [21], we anticipate C2C would benefit from a framework enabling social analysis. We are currently working to implement an automatically updating version of C2C.

The affordability and accessibility of blogging software has given individuals a voice; C2C gives them a story to tell and a persuasive way of telling it. By reducing the overhead required to perform complex investigations into Congressional behavior, C2C improves transparency between citizens and their representatives.

References

[1] THOMAS (Library of Congress). http://thomas.loc.gov/.

[2] Congressional votes database http://projects.washingtonpost.com/congress/about/.

[3] Interest Group Ratings. *Project Vote Smart*. http://www.votesmart.org/official_five_categories.php?dist=issue_rating_category.php.

[4] GovTrack: Source Data. *Govtrack.us*. http://www.govtrack.us/developers/data.xpd.

[5] Query Generator. *National Public Radio*. http://www.npr.org/api/queryGenerator.php.

[6] Ahlberg, et al. Dynamic queries for information exploration: an implementation and evaluation. *ACM SIGCHI,* (1992), 619-626.

[7] Carroll, R., et al. Comparing Nominate and Ideal: Points of Difference and Monte Carlo Tests. *Legislative Studies Quarterly 24*, 4 (2009).

[8] Clinton, J., et al. The Statistical Analysis of Roll Call Data. *American Poli. Sci. Review 98*, (2004).

[9] Enelow, J. and Hinich, M. The Spatial Theory of Voting: An Introduction. In Cambridge University Press, 1984, 2-3.

[10] Friel, B., et al. Obama: Most Liberal Senator in 2007. *National Journal*, 2008.

11] IBM. Many Eyes. *Many Eyes*. http://manyeyes.alphaworks.ibm.com/manyeyes/page/About.html.

[12] Kinnaird, P. Connect 2 Congress. http://www.connect2congress.com.

[13] Kinnaird, P., et al. Poster: Connect to Congress. *IEEE Information Visualization*, (2009).

[14] Perer, A. and Shneiderman, B. Integrating Statistics and Visualization for Exploratory Power: From Long-Term Case Studies to Design Guidelines. *IEEE Comput. Graph. Appl. 29*, 3 (2009), 39-51.

[15] Poole, K.T. W-NOMINATE Program Page. http://voteview.com/w-nominate.htm.

[16] Poole, K.T. *Spatial models of parliamentary voting*. Cambridge University Press, New York, NY :, 2005.

[17] Rosenthal, H. JVoteWorld. *VoteWorld*. http://voteworld.berkeley.edu/jvw/JAnim.html.

[18] Schneider, W. 2007 Vote Ratings. *The National Journal*. http://www.nationaljournal.com/voteratings.

[19] Shneiderman, B. The eyes have it: a task by data type taxonomy for information visualizations. *Visual Languages, 1996. Proceedings., IEEE Symposium on*, (1996), 336-343.

[20] Tauberer, J. GovTrack: The Political Spectrum. http://www.govtrack.us/congress/spectrum.xpd.

[21] Wattenberg, M. Baby names, visualization, and social data analysis. *INFOVIS 2005. IEEE Symposium on*, (2005), 1-7.

[22] Willett, W., et al, M. Scented Widgets: Improving Navigation Cues with Embedded Visualizations. *IEEE TVCG 13*, 6 (2007), 1129-1136.

[23] Wiseman, J., et al. The Whole "Poole-Rosenthal Scores" thing: A Primer for APD. *American Poli. Sci. Association*, (2004).