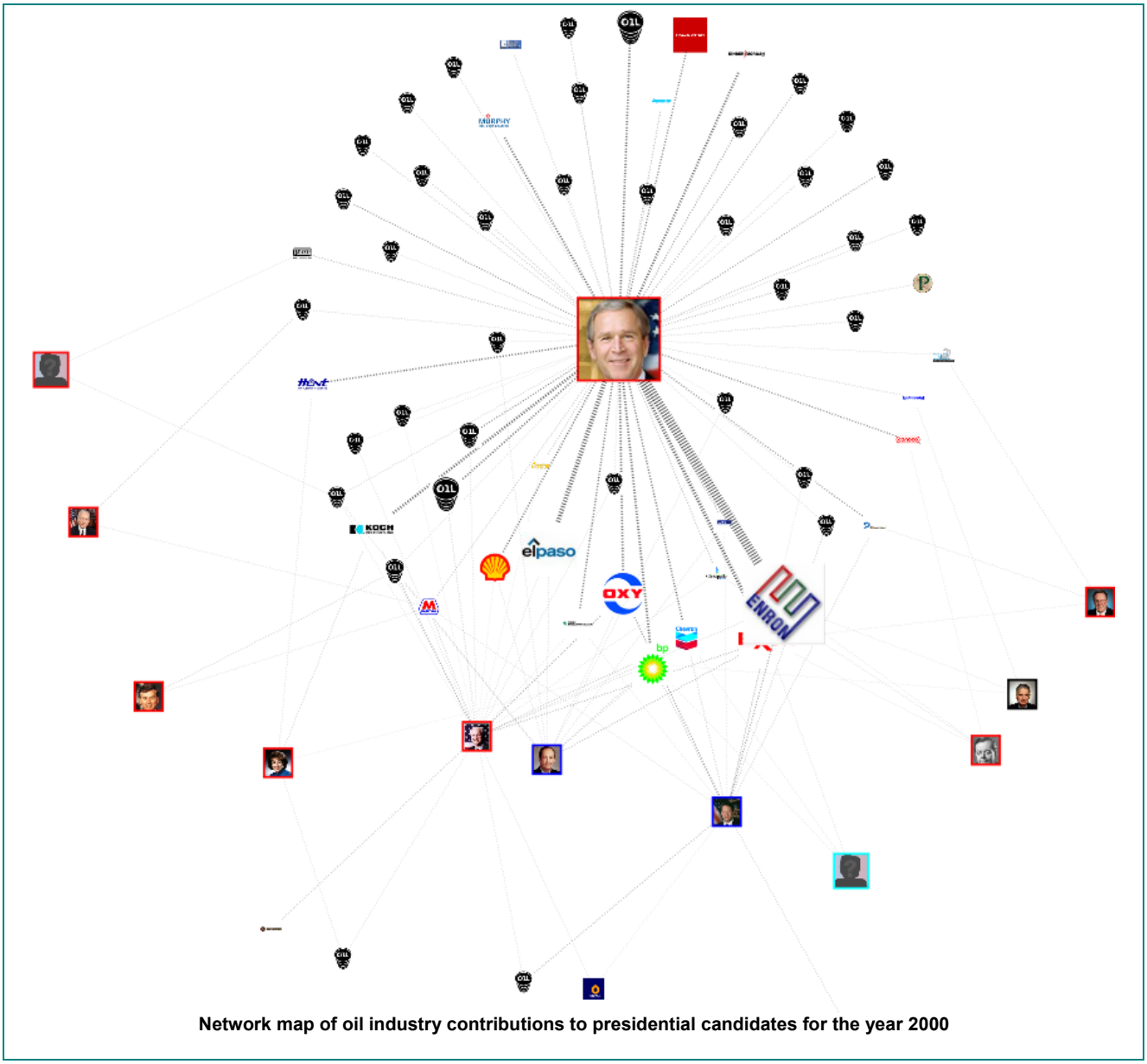


Network map of oil industry contributions to presidential candidates for the year 2000

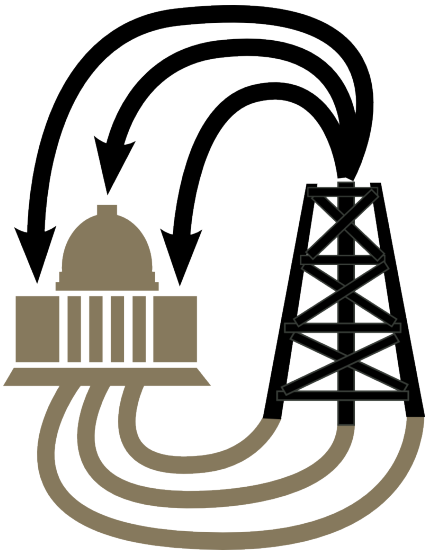
An Interactive Web Application for viewing Oil Industry Contribution Network Data

Skye Bender-deMoll & Greg Michalec

November 2007

OILCHANGE INTERNATIONAL

Overview



In July 2007, Oil Change International (OCI) contracted with Skye Bender-deMoll and Greg Michalec to create a web application that allows users to quickly visualize and examine the contributions to politicians from corporations in the oil and gas industries. OCI's existing website includes features to determine how much oil money elected officials are receiving based on the user's zip code[1] . We want to create a more refined tool that can engage a wide range of people, clearly demonstrate politicians' industry connections and motivate users to take action to "separate oil and state."

Our goal is to give the facts and figures a tangible visual form, handing users a map that they can use to make comparisons. One way of doing this is to represent the politicians as nodes in a network tied by links to their corporate donors. This kind of network map (see cover image) can break apart the aggregate data and help viewers understand the significance and context of contributions by comparing donors' relative positions in the network. (See page 4 for an explanation of network maps.) For example, although the total contributions from the management of Halliburton to George Bush's 2000 campaign are somewhat less than those from BP, Halliburton contributed exclusively to Bush, whereas some BP employees also made contributions to other candidates.

We also provide familiar tabular views of the data, but add interactive navigation so that it is possible to quickly navigate from aggregate totals through lists of companies, all the way down to individual contribution records without leaving the page. Throughout the project, we use the familiar logos of the companies alongside photographs of the politicians to reinforce the sense of "branding" and connection to consumers' gas pump purchases.

The detailed campaign finance data necessary to build this project is available online from the Federal Election Commission (FEC),[2] but requires a great deal of cleaning and integration with data from other sources before it is usable. Executing this kind of complex project requires specialized knowledge and skills and could provide an interesting opportunities for collaboration, whereby several organizations (OCI, the Center for Responsive Politics, Sunlight Foundation, or other transparency and open-government non-profits) can leverage each other's expertise and resources. Although this project specifically targets the oil industry, the tools being developed could be adapted for other areas as well.

Our immediate priority is to first produce timely maps for the 2008 presidential candidates, then expand to members of U.S. Congress, and eventually include data on candidates for state offices. Rough draft versions of many of the features are already operational, although because we are relying on a very conservative set of transactions, the total numbers are quite low. This report gives an explanation of the current project, the issues involved, and our plans for completion.

Project goals

- *Create effective, striking, and numerically accurate visual representations of oil companies' campaign contributions that motivate users to take action*
- *Serve as a test platform and create opportunities for inter-organizational collaboration and sharing of relevant political data*
- *Begin work on a reusable tool-set for network generation and data cleaning that can be deployed and developed by additional organizations*

Site Features

The network mode of representing data is visually striking, but it may not be immediately legible by everyone. While some users are already skilled at reading tabular views, others may find it helpful to think of contributions in terms of a social network; something akin to the popular websites Facebook or MySpace, but the politicians have oil companies as their "friends." We are designing the site to make it easy to move back and forth between various modes of looking at contribution data, allowing both the rapid impressionistic overview delivered by a map, and the detailed comparisons permitted by using a more traditional tabular access point.

1 http://www.priceofoil.org/oilandstate/

2 http://www.fec.gov/disclosure.shtml

Interactive tabular view of data

We've built a basic component that allows us to show lists of candidates with their portraits in a sortable table along with aggregate contribution totals.[3] Clicking on any candidate "drills down" into the candidates contributions by showing a breakdown aggregated by company or Political Action Committee (PAC). The company totals open up further when clicked to show a table of the actual transactions: the contributions from individuals or company PACs. There is also a small link for each company (called "show aliases") that lists all of the alternate spellings and subsidiary organizations attributed to that company in the data-collection process. For each individual contribution there is a "document" icon which opens a new window containing a scanned image of the original FEC filing documents for inspection.

Viewing by campaign

The "campaign view" allows selection of candidate lists by campaign, with drop down menus for selecting House, Senate, or Presidential races and the year of interest.

Viewing by zip code

The "get local" view displays the Congresspeople for the user's zip code, including breakdowns of oil industry contributions by year. Each year has a "show details" link that loads the breakdown table of contributing companies, with the same ability to drill down for individual contributions and FEC links described above. (See expanded table shown in Fig. 1, or test out the mock-up of the page.[4])

The original zip code lookup functionality was built and hosted at DemocracyInAction.org, but for this version we found it easier to do the page generation for the zip code lookup using a newly-developed congressional info API[5] from the Sunlight Foundation. The completed version will also include a candidate name search and the ability to navigate directly to the appropriate network view of the same data.

Viewing by committee

We are still developing a viewing mode that will make it possible to show only the members of select-

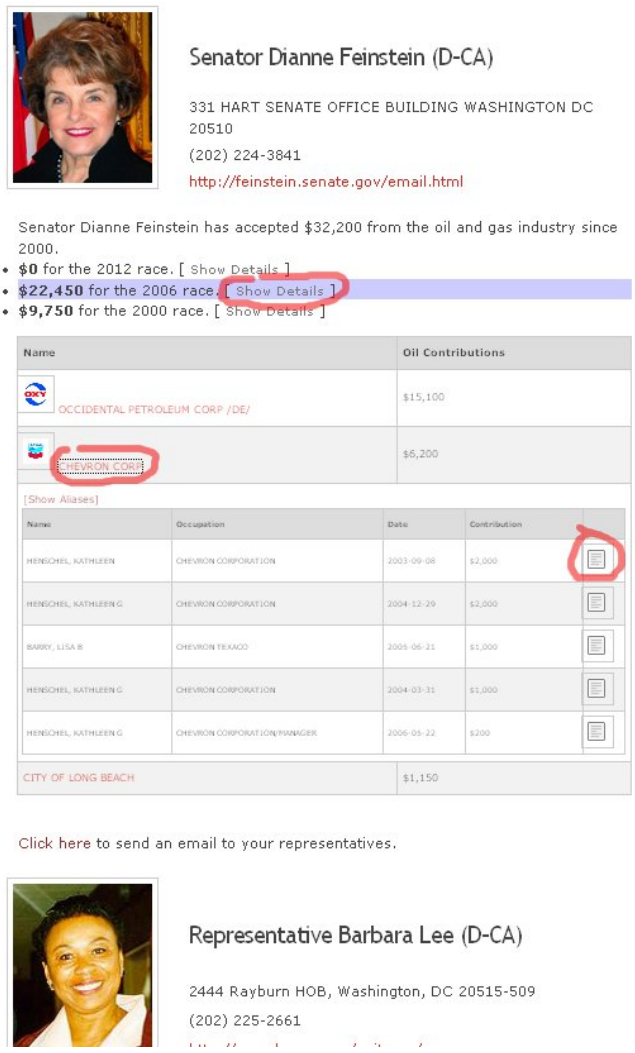


Fig. 1: Portion of view for Oakland zip code "drilled down" to show Feinstein's 2006 receipts from Chevron Corp

ed congressional committees to facilitate comparison between different committees. The network maps for this view will display the committees and companies as nodes so that users can explore the patterns of industry contributions to those members of congress who sit in industry-relevant positions.

View by key vote

We are exploring techniques that would allow us to compactly represent congressional voting records on key bills that are important to the gas and oil industries, making it possible to visually assess whether the firms are giving differentially to politicians who support their agendas. Maplight.org has done some interesting work in this area using timeline representations[6].

3 http://nopants.primate.net/~unfluence/oilchange/demo_2007_10_01/tables.php

4 http://nopants.primate.net/~unfluence/oilchange/demo_2007_10_01/ priceofoil.php

5 http://sunlightlabs.com/api/

6 http://maplight.org/map/us/bill/40617/default/timeline

Network-Based Visualizations

Tabular data can be very effective, especially if the a viewer wants to make rank comparisons to answer questions such as: "Who are the top five contributors?" However, it is difficult to clearly show relational information in tables. Network views can be helpful if we want to know: "Which companies give to candidates from both parties?" or "Which of the candidates is most well-connected to oil companies?"

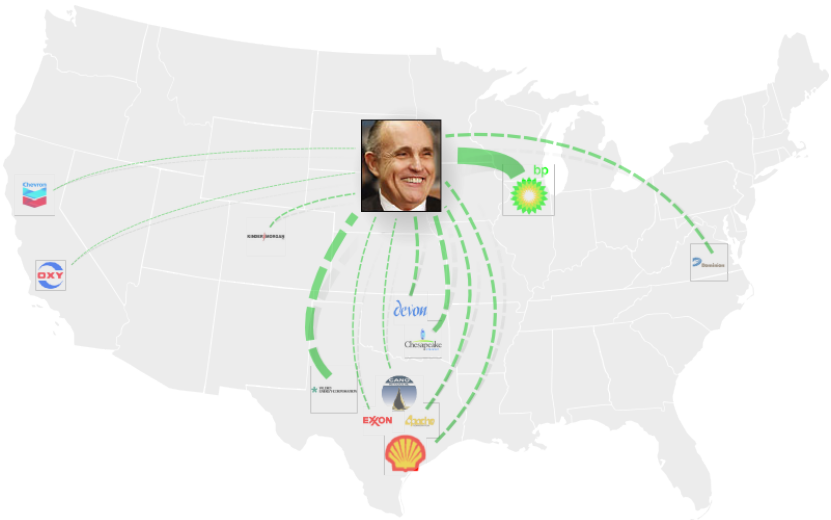


Fig. 2: Hybrid geo/network map showing a single candidate with logos of top contributors positioned at address of headquarters

What is a network map?

A network map (or social graph) is a visual depiction of the pattern of relations (links) between elements (nodes or people). Like geographic maps, network maps use distances on the page to convey information about the proximity of various elements and use size and color to show relative importance and category information. On regular maps, the data points can be positioned easily if there are geographic places associated with the data. For example, we can construct a hybrid map where we geographically position the donors and show contributions to single candidate. In Figure 2, we position company logos at the postal addresses of each corporate headquarters throughout the mainland United States. The relative thickness of the lines represents the relative contribution amounts. Thus we see that there are many oil companies located in Texas and that BP was Giuliani's largest contributor. However, many of these firms are multinational corporations, so merely noting the location of their respective headquarters is not very informative. Furthermore, if we want to include information about multiple candidates this, view will get very hard to read.

Fortunately, we can use network visualization algorithms to "unstick" the nodes from the geographic map and position them as close as possible to their neighbors in the network. The result is a cross between a flow chart (where lines indicate relations between named elements) and an atlas. In a road atlas, elements that are close together on the page will have a short driving distance between them. In a network map, elements will be closer to each other if they are closely connected or have similar patterns of connections.

Comparing networks and table views

If we plot only the contributions to a single candidate and ignore physical geography, we get a "star" network: a candidate surrounded by a constellation of contributors. (Fig. 3) with the size of each logo corresponding to the amounts of that firm's campaign contributions. The star view and the map view both do a good job of demonstrating leading presidential candidate Rudolph Giuliani's contributions from oil companies, but we can present nearly the same information with a tabular view (Fig. 5). The star network view may be somewhat more approachable or intriguing to viewers that are turned off by lots of numbers, but it is not nearly as useful for researchers who are experienced in interpreting tabular data. However, neither the star nor the table views allow direct side-by-side

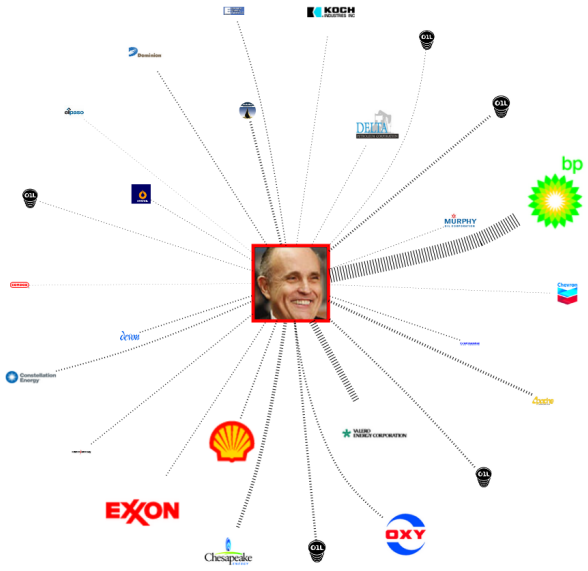


Fig. 3: Network view of company contributions to a single candidate

comparison of the linkages between multiple companies and candidates. Fortunately, the network visualization algorithms allow us to include a large number of nodes and relations — up to a point. Figure 4 shows the network of contributions for several candidates in the 2004 presidential race.

Interactive exploration

Like the tabular view, the network mode can show maps of candidates for the House, Senate, and Presidential races for current and previous years[7]. Because the maps tend to become cluttered when too many relationships are shown, there are adjustable criteria for including candidates, companies, and contributions. Specifying a minimum total for either the candidate or company removes from the map any node receiving (or giving) less than the specified amount. Individual contributions range from very small ($200) to quite large (>$2300), and in some situations hiding the smaller contributions can help focus in on the big

Fig. 5: Tabular view of relations for single candidate, sorted by total contributions

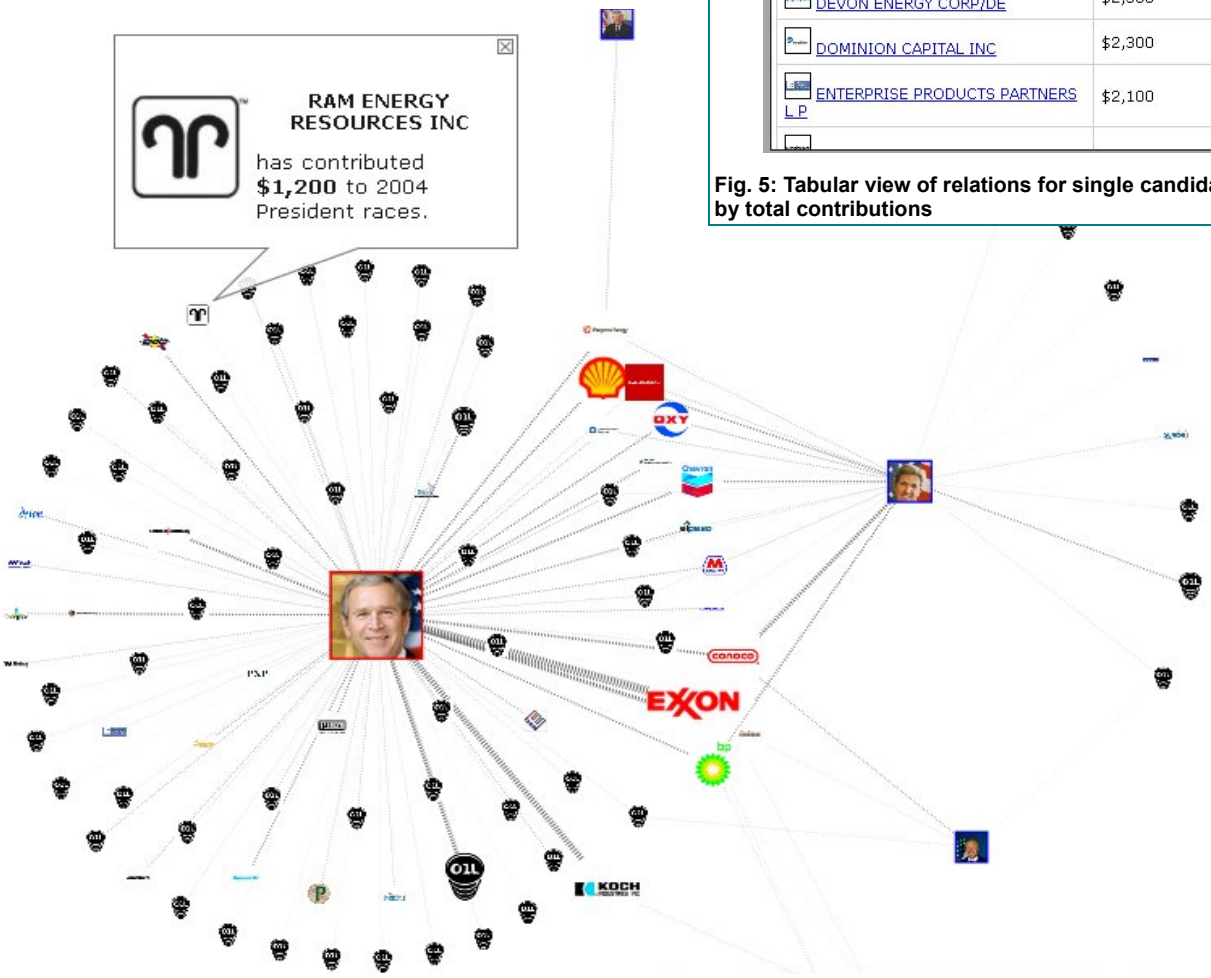


Fig. 4: Network map of contributions to 2004 presidential race, $1000 minimum company total

7 http://nopants.primate.net/~unfluence/oilchange/demo_2007_10_01/

players. This "minimum contribution threshold" can be used to filter out the lower-valued contributions, thus removing the "lightweight" donations and greatly simplifying the graph structure.

Moving the mouse over any node or link on the network shows a "tooltip" with the name and total dollar amount. This makes it possible for us to avoid cluttering the graph with labels, with the tradeoff that the non-interactive printed versions of the graph are not as useful.

Fig. 6: Popup "bubbles" are shown whenever a node is is clicked

Clicking on a node also brings up a popup bubble with more information (Fig. 6). For the candidate nodes, the bubble includes a "Take Action" link that presents the user with an email form letter to contact their particular Congressional representatives to ask them to stop accepting oil-related contributions. The "Show contributor details" link brings up an interactive table — instantly detailing the link information in an alternate form (Fig. 8).

(Fig. 4). However, the 2008 presidential race still shows very little structure or distinction between candidates. This may be because it is still relatively early in the race. Patterns in funding may show up more clearly after the candidate field has been narrowed by the primary elections and more data has been released.

One difference between the Democratic (blue borders) and Republican (red borders) party candidates is visible in the network map of the 2008 Senate race (Fig. 7). Very few Democratic candidates appear on this map at all, with the notable exception of Mary Landrieu, a Democratic candidate who seems well-integrated into the oil industry funding network, receiving donations from the same major central donors (Exxon, Koch, Pilot, Oxy, BP) as the major Republican candidates. Another initial observation is that Pat Roberts and John William Warner appear to have very similar sets of donors.

Examples

The default settings produce a very clear image (see cover image) showing distinctions between candidates for the 2000 presidential race as well as a reasonably useful image for the 2004 presidential race

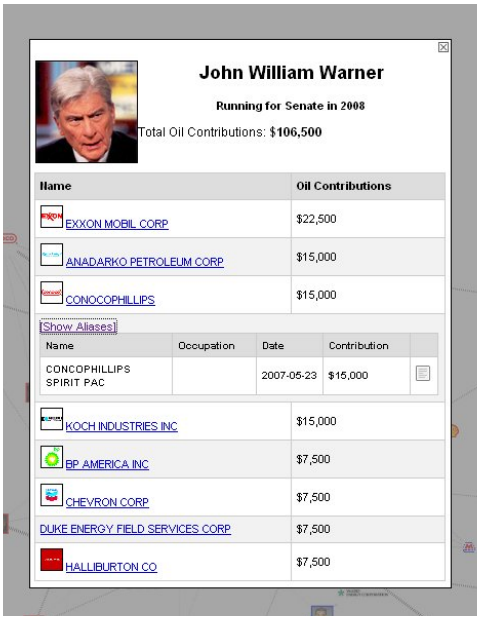


Fig. 8: Clicking on a politician's "Show contributor details" link brings up the interactive table view of the data

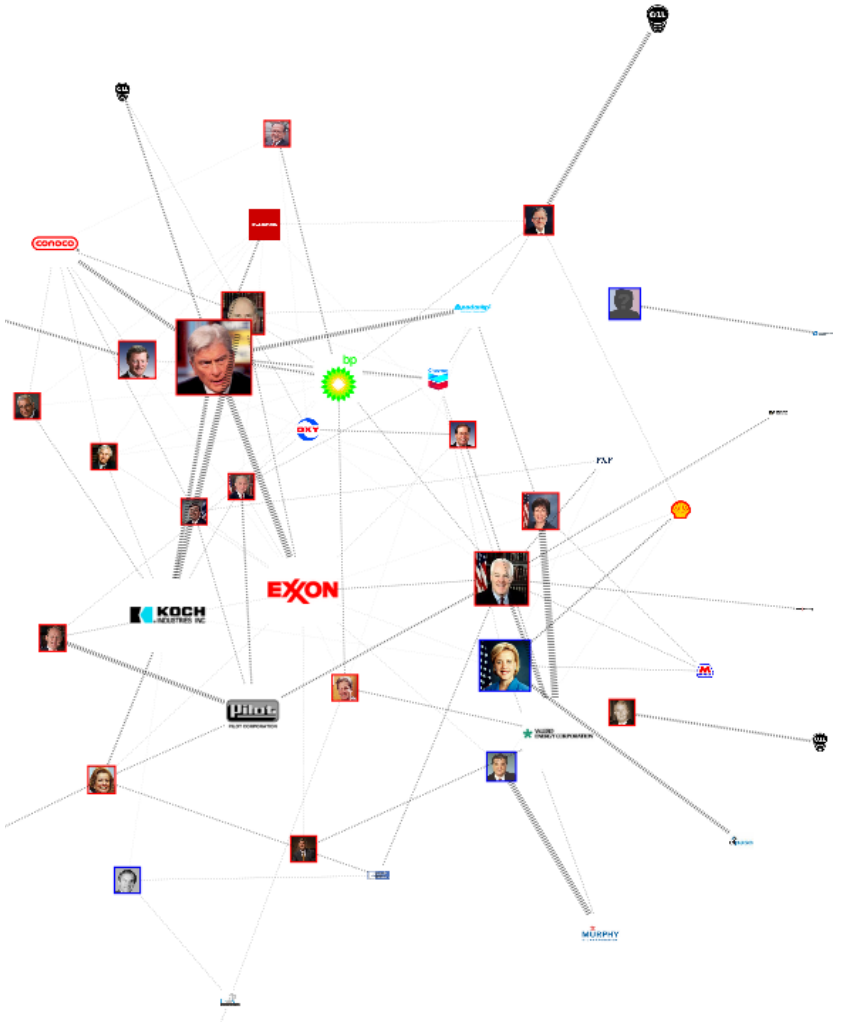


Fig. 7: Contributions in the 2008 Senate race (online version has interactive labels)

Alternate network views.

The neighbor-based layout algorithm isn't the only option for network maps. We can force the layout to arbitrarily position nodes in order to sort them by category. One example is to compare the differences between the parties by placing all the Democratic candidates on one side and the Republicans on the other, allowing the oil companies to land in the middle. (Fig.) This definitely permits comparison of every candidate at the same time and creates a strong "Democrats vs. Republicans" metaphor. However, it is not immediately obvious if it is more helpful than the corresponding network view. This type of view may be more interesting when used to display candidates sorted by their yay/nay votes on a specific bill.

Other kinds of networks

The software framework we have been writing to construct these maps from databases is fairly flexible. It was possible to quickly adapt it to display the "backbone" of the PAC funding network. The network in Figure 10 shows all transactions (including loans, etc.) between the various known oil industry PACs (shown as gray diamonds). We have done some initial work on the interactive version[8] and it seems that it may be worthwhile to include some of the PAC information on the candidate maps.

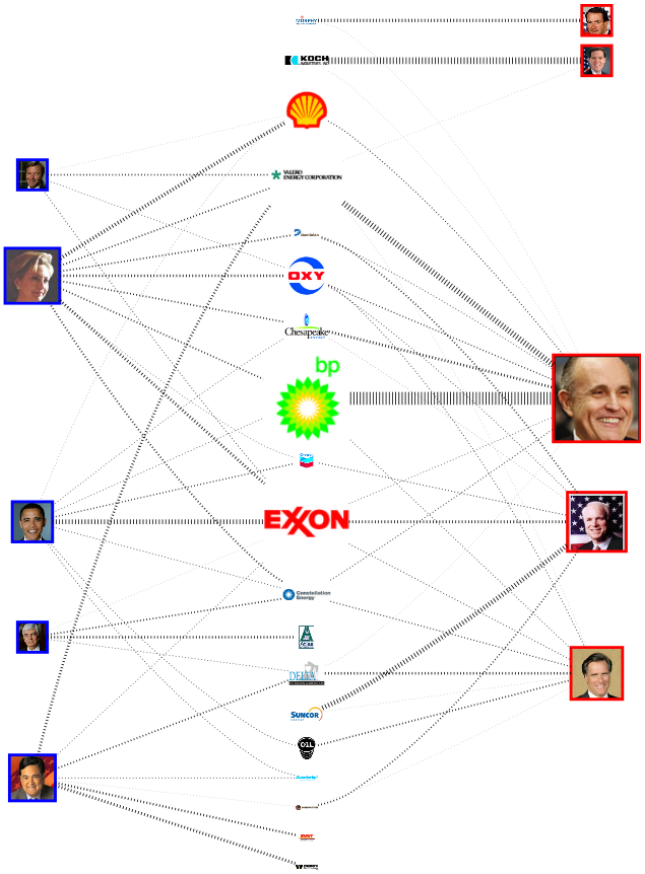


Fig. 9: A alternate layout showing the 2008 candidates sorted horizontally by political party

Fig 10: Transactions between oil-industry PACs (online version has interactive labels)

Data Sources

Candidates and Political Action Committees are required to file reports with the FEC listing all of their donors and transactions. This data is available from the FEC in files giving transactions between committees (filed by both PACs and candidates' committees) and files giving contributions from individuals (as reported by the recipients). Individuals are required to list their occupation, employer, and street address, and are subject to donation maximum limits.

There are several ways in which a company can fund a candidate:

1. The company's PAC can make a contribution to the candidate's committee.
2. CEOs, board members, VPs or other leaders within the company can make personal contributions to the candidate or to their own company's PAC.
3. The company can contribute to an industry PAC or leadership fund that in turn makes a contribution or independent expenditure for a candidate.

The most direct and unambiguous transactions of are the first type. There are a great number of these in Senatorial campaigns (see Table). Most of the money for presidential campaigns, however, comes from the second type: individual contributions.

While arguments can be made that the second type of contributions are simply individuals expressing their

8 http://nopants.primate.net/~unfluence/oilchange/demo_2007_10_01/index.php?setupfile=pacGraphSetup.php

political opinions, the fact that the vast majority of contributions come from company officers makes it is reasonable to assume that their political convictions are aligned with the company. (The converse of this is also possible: that the companies' positions are aligned with those of their leaders, but for our purposes the distinction is not important.) This means that the employer field of the individual contributions can be used as a proxy for giving by the company. However, the entries in this field are not standardized within the FEC databases and must first be matched up with company names before they can be tabulated. We have written software tools to assist with this "normalizing" or "aliasing" process.

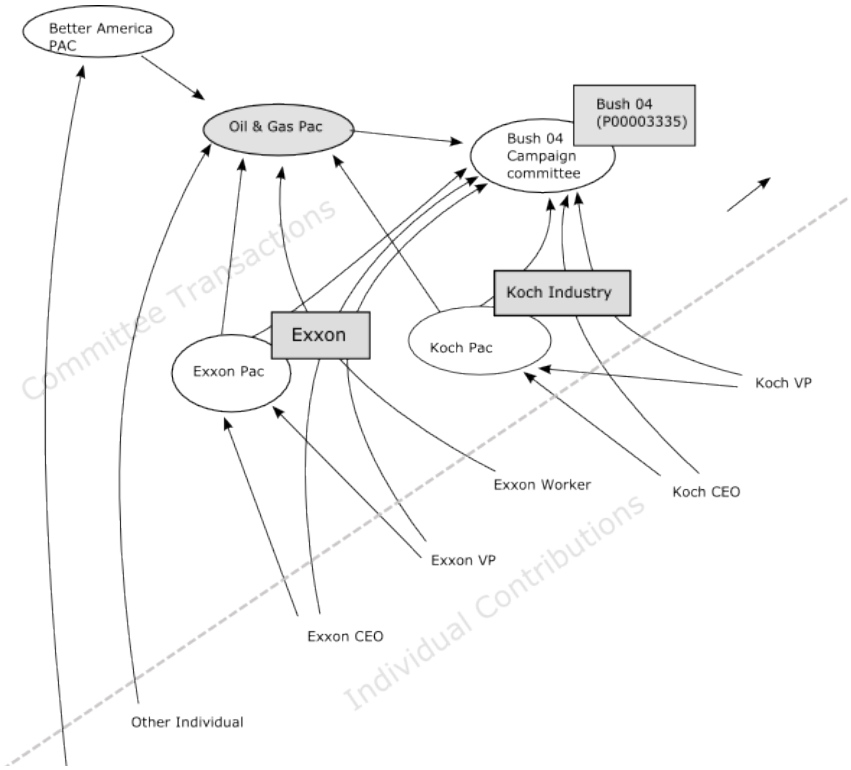


Fig. 11: Diagram of sources of information and types of contributions in FEC data

For the third type of contribution, it is difficult to prove that there is a direct relationship between a company's contribution to an industry PAC and that PAC's contribution to a specific candidate. This means that even though we can include such contributions in overall sector totals and display them as PAC-to-candidate transactions, it is probably not correct to code them as company-candidate transactions unless we can derive an appropriate discounting rule. These transactions are also fairly rare in the U.S. presidential campaigns, since money seems to go to the party committees instead, but they are very common in the Senate races.

	President		House		Senate	
Year	**Total $**	**Count**	**Total $**	**Count**	**Total $**	**Count**
2000	$89,952	37	$8,000	4	$528,810	457
2002			$6,500	1	$573,897	503
2004	$72,000	27	$42,500	6	$469,758	335
2006			$85,730	42	$315,250	211
2008	$26,673	16	$2,500	1	$314,604	108

Table 1: Oil industry (non company-affiliated) PAC contributions

Contribution Limits

We were surprised to learn that it is fairly common for individuals to make contributions to presidential races that exceed the normal limit of $2,300. According to the FEC, as long as the campaigns maintain separate accounting systems, the over-the-limit amounts can be automatically transfered from the candidate's primary fund to the general election funds. The only restriction is that the committee must give the money back to the donor if the candidate doesn't make it past the primary. A number of contributors seem to have done this for Giuliani's 2008 campaign (see for example the top contributors in the online table view for Giuliani, such as Mr. Pickens, CEO of BP Capital[9].)

Another way to get around the individual contribution limits is to have a spouse or family member make donations. The Center for Responsive Politics has stated that it is able to assign many of these contributions by matching last name and address of individuals to determine household contributions. Although the street level address information needed for this is not included in the files available from the FEC, OCI is entering into an agreement to purchase data directly from CRP in order to increase both the sector totals and overall accuracy.

9 http://nopants.primate.net/~unfluence/oilchange/demo_2007_10_01/fetchFECRecord.php?id=27930900227

Transaction types

Another complication in the between-committee transactions is that the FEC's coding scheme is quite complex and apparently ill-enforced. In theory, every transaction involving two committees should be reported twice: once by the contributor, and once by the recipient. The transaction should have corresponding transaction types, so the donor would record a gift 24K and the recipient would report it as 18K (see Table for type definitions). In reality, the totals for these categories diverge wildly, in the above case, the difference is $31.4 million over the last 8 years. Even transfers among committees on the same campaign (categories 24G and 18G) are often not coded into matching categories. Sometimes they are declared an "affiliated transfer" by one committee and a "regular contribution" by the other. This makes deriving a total for the transactions between a pair of committees problematic. If we only count contributions reported by the recipient filer (as recommended by the FEC records office in our conversations), we will miss many contributions. Yet if we were to includ both recipient- and contributor-reports, we would double-count some transactions. Furthermore, the error-prone coding prevents us from writing simple matching scripts to solve the problem. Fortunately, CRP has put a great deal of time into manually reconciling records, so the data purchased from them should be free of many of these problems.

Total	Type Code	Transaction Description
$41,179,782	24K	CONTRIBUTION MADE TO NON-AFFILIATED
$26,801,274	15	CONTRIBUTION
$9,739,134	18K	CONTRIBUTION RECVD FROM REG FILER
$2,236,768	10	[RECEIPT--EXEMPT FROM LIMITS]
$1,273,051	18J	MEMO(FILER'S % OF CONTRB GIVEN TO JT FR)
$1,050,567	24G	TRANSFER OUT AFFILIATED
$494,752	18G	TRANSFER IN AFFILIATED
$166,387	24I	EARMARKED INTERMEDIARY OUT
$141,233	24Z	IN-KIND CONTRIB MADE TO REG. FILER
$127,328	22Y	CONTRIBUTION REF TO INDIVIDUAL
$39,677	24T	EARMARKED INTERMEDIARY TREASURY OUT
$31,643	15E	EARMARKED CONTRIBUTION
$30,335	22Z	CONTRIBUTION REF TO CANDIDATE/ COMMITTEE
$6,000	24N	COMMUN COST AGAINST CANDIDATE (C7)
$5,342	24P	CONTRIB MADE TO POSSIBLE CANDIDATE
$4,484	24E	INDEPENDENT EXPENDITURE FOR
$1,605	15Z	IN-KIND CONTR RECVD FROM REG. FILER
$1,000	15J	MEMO(FILER'S % OF CONTRB GIVEN TO JT FR)
$1,000	24U	CONTRIBUTION MADE TO UNREGISTERED
$750	24R	ELECTION RECOUNT DISBURSEMENT
-$500	15C	CONTRIBUTION FROM CANDIDATE

Table 2: Totals for oil related transactions by FEC type code, years 2000-2008

Oil companies used

For the images in this report and the current online demos, we have currently entered and tagged ~250 oil industry companies that appear as employers. These companies were selected from lists of top-producing and most-profitable companies, as well as searching through lists of (~1000+) publicly traded companies in the SEC's Standard Industry Classification (SIC) lists for categories 2911 (Petroleum Refining) and 1311 (Crude Petroleum and Natural Gas). This is clearly not an exhaustive list, but at a minimum it should be well representative of the universe of companies, heavily selected towards those that are most important and publicly traded.

Aliasing and grouping company names

Because many of the companies have complex and rapidly-changing ownership relations, it is not trivial to determine how the data should be aggregated in order to group transactions from subsidiaries and affiliates. In situations where one company has purchased another outright, some of the previous name information is available via the lookup we do in the SEC data. Otherwise we have consulted company websites when possible to determine appropriate ownership (circa 2008). Luckily, the aliasing system we are using makes it very simple to adjust the hierarchy and reprocess the data as needed.

In addition to the useful tables of company aliases, our current database contains 11 million rows of data, with mappings to matching ID numbers in other election related data sources. At a minimum, the data we have successfully coded to date can be considered a representative sample of the federal campaign data that can be directly attributed to campaign filings, even if the overall totals are much less impressive than those the CRP gives. However, as this level of data coverage is not sufficient for the final project, OCI has made the decision to purchase the detailed dis-aggregate data from CRP rather than make an additional significant investment of time and research into a fully independent system for federal data.

State-Level Data

The political presence of the oil and gas industry

reaches much deeper than the national political races. The same corporations make campaign contributions and lobby for favorable legislation at the state level as well. However, there are no nationwide mandatory campaign finance disclosure requirements in races for state offices, and no official centralized repository of data. Instead, each state has its own laws, deadlines, and reporting systems. Fortunately, the non-profit National Institute on Money in State Politics[10] (NIMSP) has stepped forward to partially fill the gap.

NIMSP collects and categorizes all the data available from each state's offices, in some cases laboriously transcribing it from paper records. Embracing the potential of modern internet collaboration, NIMSP provides access to much of their data through a web API, making it possible for others to write applications that can query their contribution databases. For example, the unfluence.net[11] website uses NIMSP's API, along with earlier versions of the technologies being developed for this project, to produce network maps of state level data.

A future phase of this project will involve working with NIMSP to integrate the state level data into our project. This will require extending our current normalization systems to deal with the different naming and categorization conventions used within their databases. The end result will be a system capable of producing comprehensive profiles of companies' political contributions across multiple spheres of activity--and placing that information in the hands of ordinary voters and consumers.

Projected Completion Dates

Our current plan is to have the main features shown in the demo version available to the public on OCI's website by the end of 2007. This will include comprehensive data purchased from CRP and information on the 2000 to 2008 federal races. We hope to have the state-level data integrated and available in the first quarter of 2008.

This project is being built on a shoe-string budget using open-source tools and components. (For the technically curious: this includes MySQL, PHP, JavaScript and AJAX libraries, with the network layout provided by GraphViz.) The source code for this project will eventually be made public domain, with the hope of encouraging other organizations to contribute development resources to improve the tools and adapt them for additional areas of open-government and transparency research.

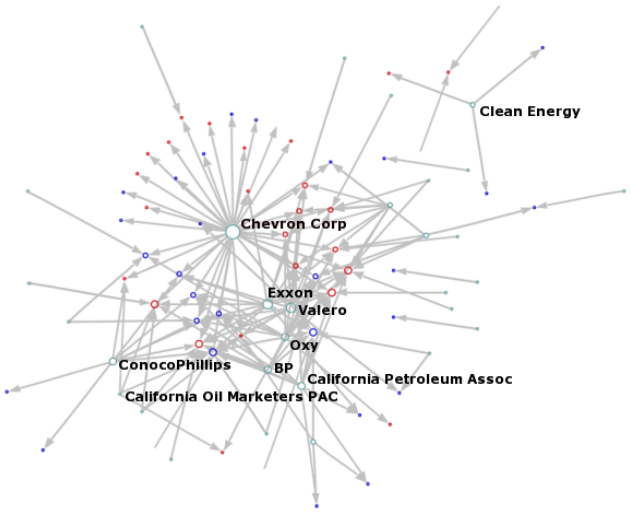


Fig. 12: Image of Oil & Gas contributors to 2006 candidates for California State Assembly (with $1000 minimum contribution) generated by unfluence.net

Maintaining the site, keeping the data current, and developing additional sophisticated analysis features will require continued support. OCI is seeking additional sources of funding and relationships for collaboration.

Authors

Oil Change International campaigns to expose the true costs of oil and facilitate the coming transition towards clean energy by identifying and overcoming political barriers.

Greg Michalec[12] and **Skye Bender-deMoll**[13] are both independent consultants with experience in internet software development, network visualization research, and a desire to work on innovate progressive projects.

10 http://www.followthemoney.org/

11 http://unfluence.primate.net/unfluence.html?StateSelect=CA&YearSelect=2006&Office=R01&Interest=33&valueMin=1000

12 http://greg.primate.net

13 http://skyeome.net