An Analysis of Political Views on Blogs
November 3, 2008 Update

Todd Sullivan
todd.sullivan@cs.stanford.edu

1 Introduction

This week we bring our blog, interest, and community FeelScores together into a single FeelScore across time for each class. We explain our aggregate FeelScore method, which is largely influenced by our lack of data for most classes when broken down by date. We end with several expected observations, such as Obama's curve exhibiting a classic lift from the Democratic National Convention with a gradual drop back to normal levels after the event, and highlights for our next update's deeper investigation.

2 FeelScores Across Time

2.1 Blog-based FeelScores

We calculate blog-based FeelScores in a similar way as before. For each blog post, we have the number of positive and negative phrases pertaining to each class (as described in our October 20 update). We use these counts to calculate a blog-based FeelScore for each author over time.

For a given day where the author blogged about a class, the author's positive and negative counts for the class are calculated as an exponential decaying sum of the counts for all blog posts that happened on that day or before that day. For example, the positive count for the day is calculated by the following sum, where the PositiveCount(d) function returns the raw positive count for the day d and diff(a,b) returns the number of days between days a and b:

$$PositiveCount = \sum_{d=Jan.1,\,2008}^{CurDay} PositiveCount(d) \cdot \left(\frac{1}{2}\right)^{diff(CurDay,d)}$$

We use these new positive and negative counts to calculate a FeelScore for each day that the author wrote about a particular class. Since most authors do not blog about each class on a daily basis, we fill out the year's FeelScores by taking the FeelScore that occurred before the given day and closest to it. Thus if an author blogs about Obama once on January 5, the author's blog-based FeelScore for Obama from January 5 through to the end of the year will be whatever FeelScore was calculated from the January 5 blog post.

We create a final blog-based FeelScore curve for each class by, on a given day, averaging the FeelScores for each author that has blogged about the class on or before the day. As one might notice, this gives equal weight to all authors regardless of when they blogged about the class. For example, an author that last blogged about the class 90 days before the current day would have equal weight to an author that blogged about the class on the current day.

While this is somewhat undesirable, it was the only way to generate a non-chaotic curve due to our lack of data. To illustrate this issue, Figure 2.1a shows each candidate's blog-based FeelScore across time if each class' FeelScore for a given day is calculated as an exponential decaying average based on the number of days between the current day and the last day that the author blogged about the class. The average is calculated with a base of

0.9, thus authors blogging on the current day get a weight of 1, authors on the previous day get a weight of 0.9, authors two days away get a weight of 0.81, etc.

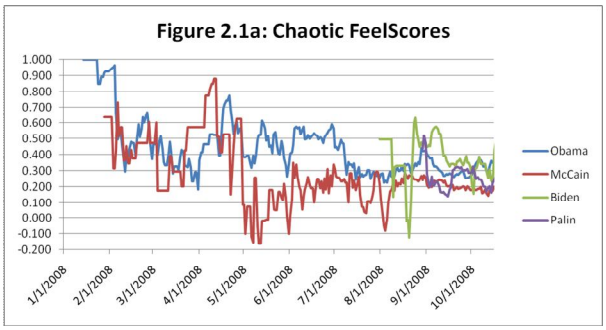


Figure 2.1a is difficult to analyze because of all of the jumps in the curves. This problem is even more severe when only using, say, the last 5 days worth of blog posts. Figure 2.1b shows the distinct number of authors blogging per day about each candidate. While the figure's size does not allow one to easily track the amount of authors talking about a particular candidate, it clearly shows that for the majority of the year every candidate has less than 50 authors making positive or negative statements about them on a given day. The Republican / Democrat / Liberal / Conservative graph follows the same trend of peaking around September 1, which is right in between the Democrat and Republican conventions.

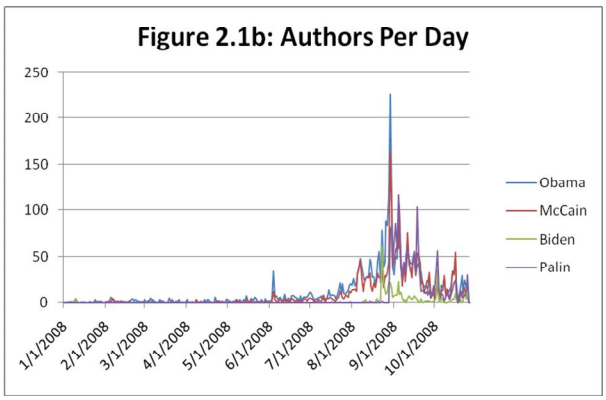
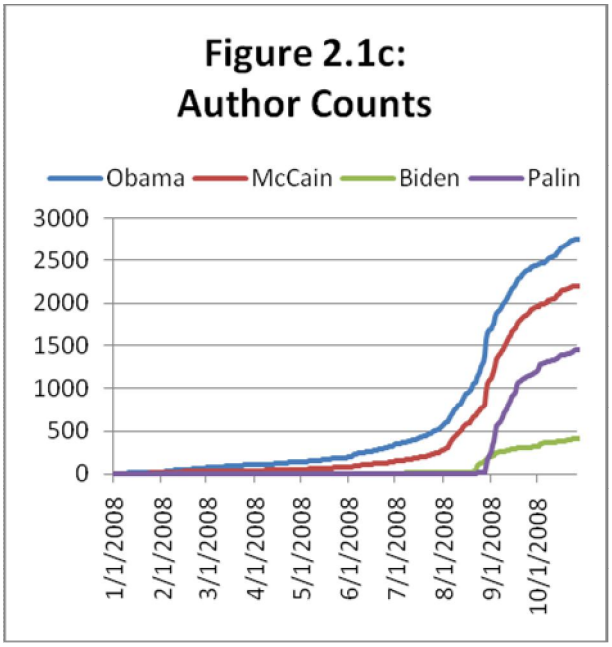
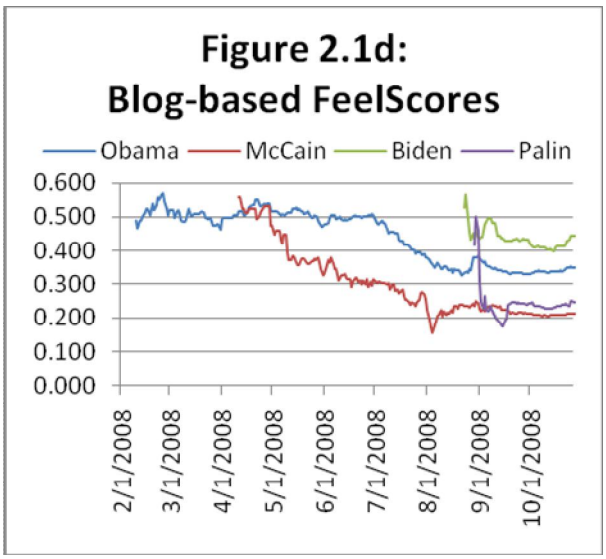


Figure 2.1c shows, for each candidate/day, the number of authors that have a positive or negative opinion about that candidate that was expressed on or before the given day. Figure 2.1c shows that by the time the conventions are occurring (around September 1), each candidate has a few hundred to a few thousand authors that have expressed opinions about them. This will clearly provide more stable FeelScore curves for each class.



Aside from generating non-chaotic FeelScore curves, we believe that our method of giving equal weight to authors regardless of when they last blogged about the class accurately models most people's opinions about the classes. Certainly, some authors are so enthralled by politics that they blog about it on a daily or at least weekly basis. On the other hand, most people will write one or two posts expressing their opinions, and will then never blog about the subject again unless their opinion changes. Our blog-based FeelScore method takes this into account by assuming that an author's opinion about a class does not change over time unless the author explicitly blogs about the class again.

Figure 2.1d gives a taste of what our final FeelScore curves will look like. The curves are vastly more manageable than the ones presented in Figure 2.1a. In this figure, each candidate's curve appears once the candidate's average FeelScore includes at least 40 authors.



2.2 Incorporating Interests and Communities

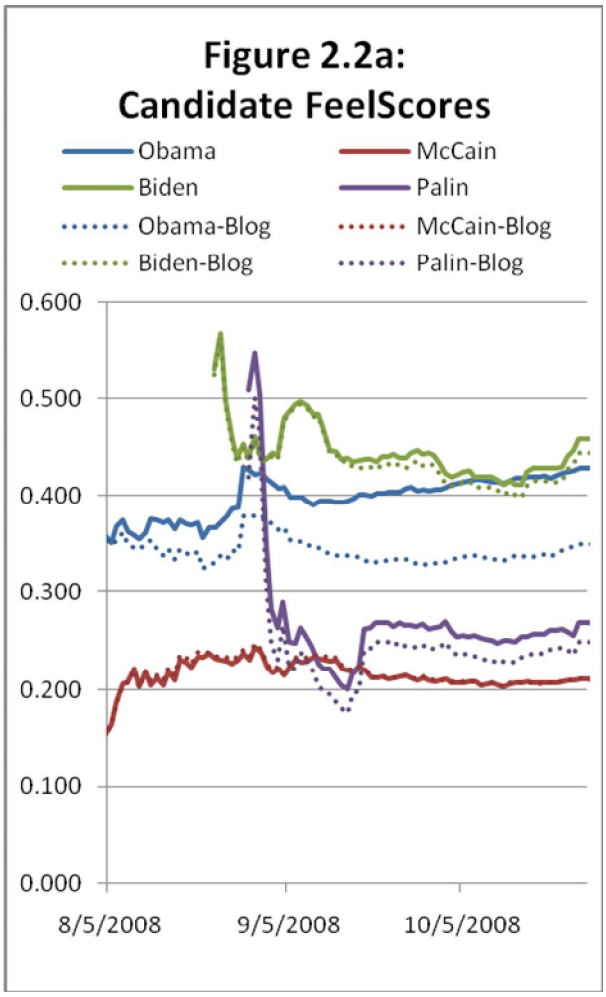
As discussed in our October 27 update, the date associated with each author-interest or author-community pair is the date that the connection was added to the database, and it is not sufficient to graph over time because the crawler bulk inserted connections into the database ever few days. Nevertheless, the dates are still useful for augmenting the blog-based FeelScores.

We added interests and communities by adding 5 positive or 5 negative counts to the day that the connection was added to the database and all days afterwards. Thus if an author lists a positive interest about Obama on April 13, 5 positive counts will be added to all days on or after April 13.

These counts are not used when calculating the blog-based counts as described in Section 2.1. After calculating the blog-based counts over time we have, for each author/candidate pair, a positive and negative blog-based count as well as the interest/community-based count for each day. We calculate the overall day's FeelScore by adding the positive counts together, and the negative counts together, and then performing the normal operation of difference divided by sum.

Since the majority of interests and communities are positive, the incorporation of interests and communities into the overall FeelScore has a general effect of boosting all class' FeelScores. Also, since our interest and community data starts on August 5, the FeelScore curves are the same up to that point.

Figure 2.2a shows the general lift in each candidate's curve brought about by including interests and communities. The solid lines include interests and communites while the dotted lines are the blog-based curves. The interests and communities have the most effect on candidates that had a lot of interests or communities about them. Obama's FeelScore is lifted the most – by about 0.05, while Palin and Biden receive a small boost in scores. Surprisingly, McCain's curve stays almost exactly the same.



The magnitude of the positive lift is mostly a result of interests and communities including authors in the calculation that have not made any blog posts about the candidates. Examining the last day of our current data (October 25) reveals that incorporating interests and communities increased the number of authors with opinions about Obama by 333. The increase for McCain, Biden, and Palin are 41, 10, and 65 respectively. This explains the size of the gap between curves for all candidates but McCain. McCain's curve does not change because of the 41 new authors in McCain's FeelScore calculation, half are from positive interests/communities and half are from negative interests/communities.

The party FeelScores in Figure 2.2b show drastically different curves when incorporating interests and communities. Most notably, the incorporation of interests and communities shoots the Liberal class from the worst FeelScore of around 0.05 to the best of around 0.55. The large changes seen in the Democrat and Liberal classes are due to their dominance in the number of positive interests/communities as well as the number of authors prescribing to those interests/communities.

This dominance can be seen in Table 2.2, which contains the summary counts from our October 27 update. The Liberal class has almost 6 times as many authors prescribing to positive related interests in comparison to the right-leaning classes. The ratio is even greater for communities, with the Liberal author count being 11 to 16 times larger than the Republican or Conservative author counts.

3 The Overall Picture

Figures 3a, 3b, 3c, and 3d on the next page provide an overall view of each class' FeelScore over time. We will thoroughly examine these graphs in the coming weeks. In the meantime, we present them as a hint of results to come, and point out a nice feature related to each party's national convention.

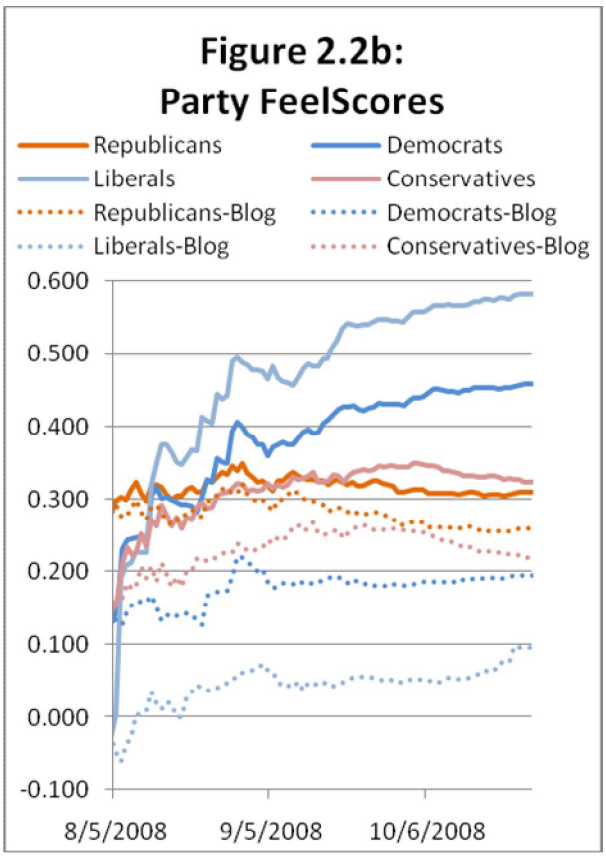


Table 2: Counts by Class

Positive Interests		
Class	Interest Count	Author Count
Republicans	25	118
Democrats	41	373
Liberals	100	605
Conservatives	38	112

Negative Interests		
Republican	15	28
Democrat	2	3
Liberal	12	13
Conservative	10	17

Positive Communities		
Class	Community Count	Author Count
Republicans	8	26
Democrats	19	111
Liberals	45	281
Conservatives	9	17

Negative Communities		
Republicans	0	0
Democrats	0	0
Liberals	3	10
Conservatives	0	0

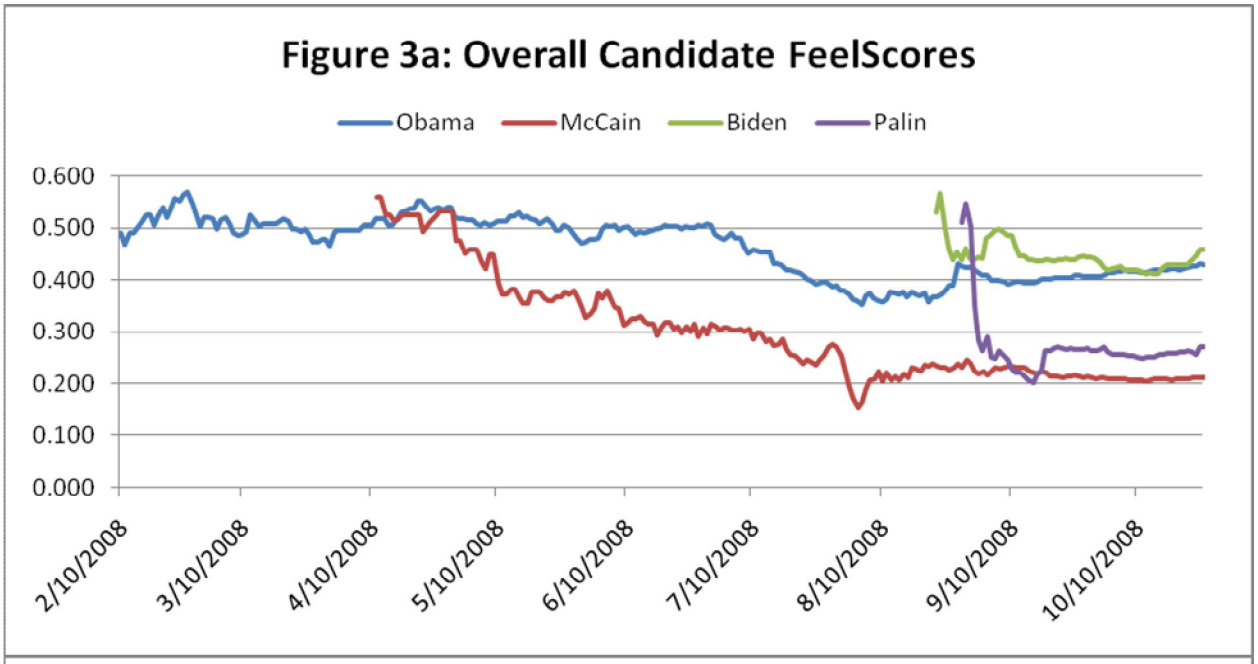


Figure 3a: Overall Candidate FeelScores
Obama
McCain
Biden
Palin
0.600
0.500
0.400
0.300
0.200
0.100
0.000
2/10/2008
3/10/2008
4/10/2008
5/10/2008
6/10/2008
7/10/2008
8/10/2008
9/10/2008
10/10/2008

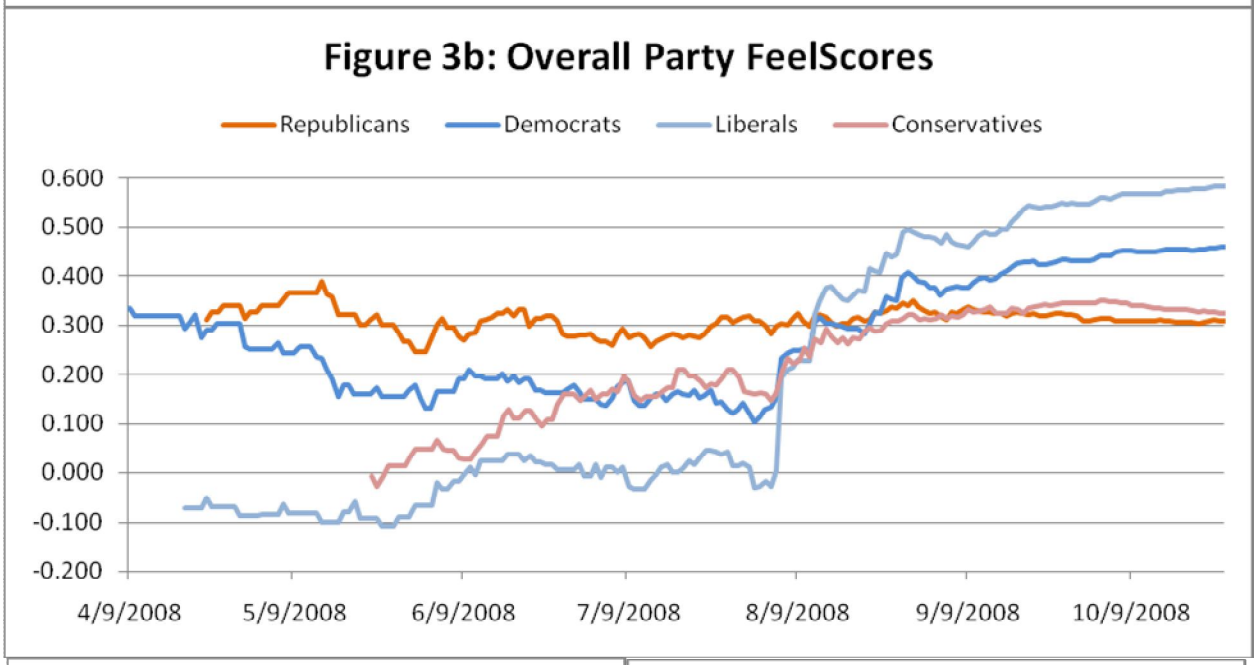


Figure 3b: Overall Party FeelScores
Republicans
Democrats
Liberals
Conservatives
0.600
0.500
0.400
0.300
0.200
0.100
0.000
-0.100
-0.200
4/9/2008
5/9/2008
6/9/2008
7/9/2008
8/9/2008
9/9/2008
10/9/2008

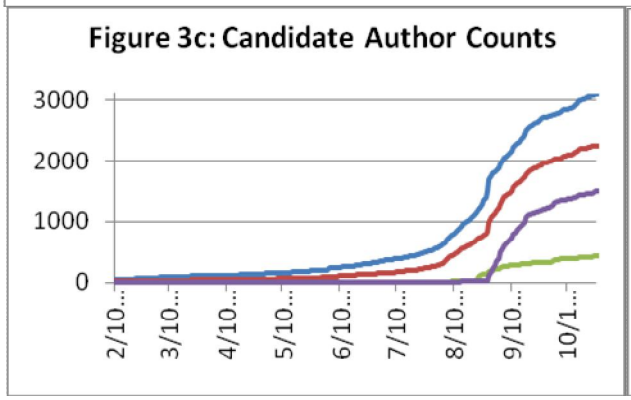


Figure 3c: Candidate Author Counts
3000
2000
1000
0
2/10...
3/10...
4/10...
5/10...
6/10...
7/10...
8/10...
9/10...
10/1...

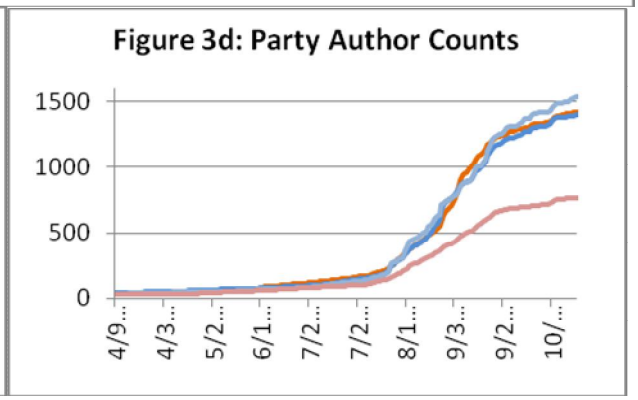
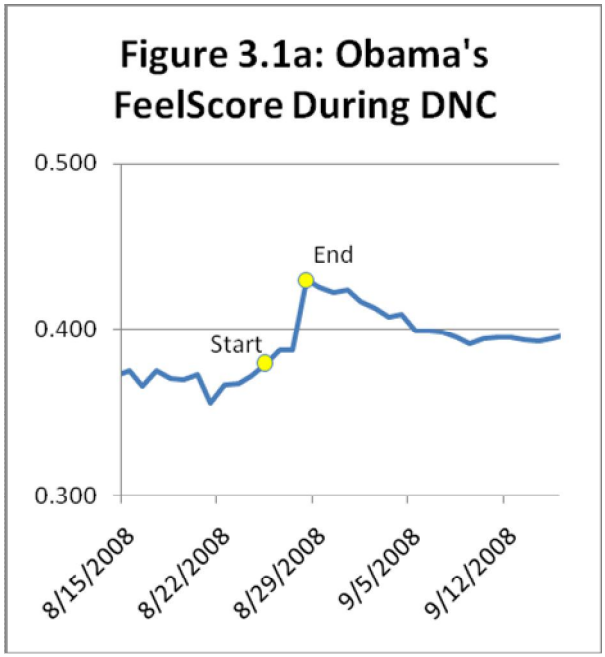


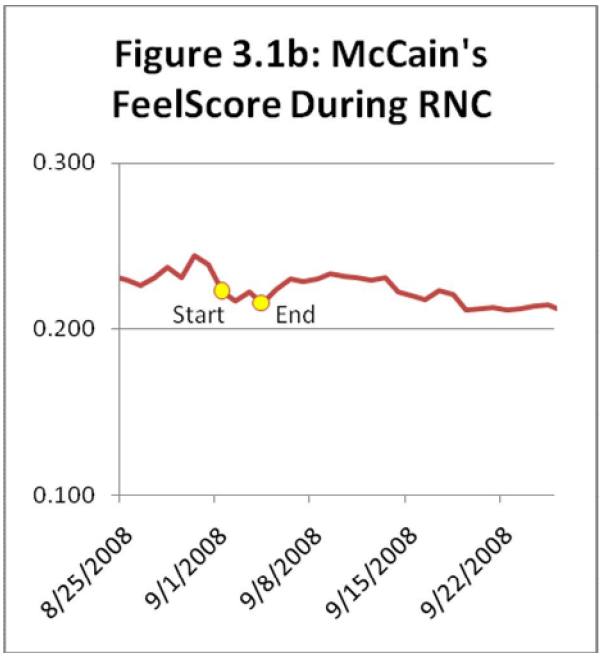
Figure 3d: Party Author Counts
1500
1000
500
0
4/9...
4/3...
5/2...
6/1...
7/2...
7/2...
8/1...
9/3...
9/2...
10/...

3.1 National Conventions

Figure 3.1a shows a close-up of Obama's FeelScore curve during the Democratic National Convention. The yellow dots mark the beginning and ending of the convention. Obama's FeelScore exhibits a clear jump during the convention, with a gradual tapering off back to normal levels afterwards. This general behavior is often noted in the press as occurring each presidential election around each party's national convention.



McCain's curve does not exhibit the same profound trend, as shown in Figure 3.1b. While his curve does seem to have a slight lift at the end of the convention that slowly drops, it is not nearly as drastic as Obama's. This may be in part due to the fact that the Republicans limited the extravagance and overall size of their convention because of Hurricane Gustav. Hurricane Gustav was a category three storm that caused an estimated 1.9 million people to evacuate the Louisiana coast. Due to the size of the hurricane, and especially in light of the result from Hurricane Katrina a few years ago, the Republicans did not want to be seen as throwing a party while an entire state is trashed by a hurricane. This resulted in many cancelled speeches and less media coverage than usual.



4 Future Work

In the coming weeks we will more closely examine the FeelScore curves and attempt to explain why various rises and falls occur. We will also compare FeelScore curves between networks to see which networks seem to be more affected by important events such as the debates. Additionally, we will pair the curves with polls and the actual results of the election on November 4.

Two particular topics that we will explore are:

- Why does Biden's curve rise to a plateau and then fall back to normal levels around September 10? Is this significant or just a result of too few data points?
- Why does Palin's curve rise shortly after September 10 and stay relatively even afterwards?