A Case of Interactional Influence in the United States Supreme Court

Abdul Z. Abdulrahim

May 6, 2019

Contents

1 Introduction

In this paper, I analyse the effect of linguistic influence in the courtroom and how it may have affected the decision in the United States (US) Supreme Court case of *Boumediene v. Bush*. In 2002, Lakhdar Boumediene and five other Algerian natives (the petitioners) were arrested by Bosnian police when US intelligence officers suspected their involvement in a plot to attack the Bosnian US embassy (Oyez, n.d.). The US government had classified these men as enemy combatants in the war on terror and detained them at the Guantanamo Bay Naval Base, located on land leased from Cuba. Boumediene et al. filed a petition for a writ of *habeas corpus*, alleging violations of the US Constitution's Due Process Clause, the common law and international law.

The case was appealed to the US Supreme Court which had to decide whether: (i) the Military Commissions Act of 2006 (MCA) could be interpreted to strip federal courts of jurisdiction over habeas petitions filed by foreign citizens detained at Guantanamo Bay; (ii) the MCA was a violation of the Suspension Clause of the Constitution; (iii) the detainees were entitled not to be deprived of their liberty without due process under the domestic and international laws and conventions; and (iv) the detainees could challenge the adequacy of the judicial review provisions of the MCA (Oyez, n.d.).

As the Supreme Court had earlier held in *Rasul v. Bush* that the habeas statute extends to non-citizen detainees at Guantanamo, the justices heard arguments as to the adequacy of the MCA in protecting these rights and if it was a valid substitute for rights under existing laws and treaties. The oral arguments heard provide insight into the leaning of the justices prior to deciding on the case and the strengths of counsel's arguments in persuading them. While other factors may have influenced the outcome of the case, post-decision analysis of the 5-4 majority in favour of Boumediene and the other detainees indicate that the outcome rested on the persuasion of Justice Kennedy, who voted contrary to his expected ideological stance. For this reason, we analyse the oral argument transcripts using a model of network influence to identify whom and how Justice Kennedy was swayed in favour of the petitioners.

2 Empirical Framework

There have been attempts to identify the influence exerted in interactions between individuals and how this may affect the outcome of said interaction. Blundell, Beck, and Heller

(2012) focuses on turn-taking, leveraging reciprocity to infer social groups and the likelihood of influence being exerted. For example, if Aisha messages Zainab frequently, it increases the probability that Zainab will respond to Aisha in the near future. Knowing that Aisha instantiates most of the conversation, in turn, means she is likely to exert influence on Zainab. In this Bayesian non-parametric model, the time-series data is generative and accounts for the rate of events between clusters and builds on mutually-exciting point processes known as Hawkes processes (Hawkes, 1971; Rasmussen, 2013).

In a follow-up paper, Guo, Blundell, Wallach, and Heller (2015) present an alternative model which focuses on linguistic accommodation between interactions (socio-linguistics). It derives from the basic idea that utterances made can determine the boundary of a conversation, thereby directing the conversation and the conclusions drawn from the interaction (Sanders, 1987). That is, when two people interact, the use of a word by one person can increase the other person's probability of subsequently using that word — capturing dependencies and power relationships inherent in the interactions (West, Turner, & Zhao, 2010). The rest of this paper will use the Bayesian Echo Chamber to model the case of *Boumediene v. Bush.*

2.1 Bayesian Echo Chamber Model

The Bayesian Echo Chamber model specifies a probabilistic generative process for the words that occur in a set of utterances $W(p)_{p=1}^{P}$ made by P people, conditioned on the utterance start times and duration, where $N_{(T)}^{(p)}$ denotes the total number of utterances made by person p over the interval (0, T) and each utterance made by p consists of $L_{n}^{(p)}$ word tokens (Guo et al., 2015).

It draws upon ideas from both dynamic Bayesian language modelling and multivariate Hawkes processes. Each word in an utterance made is drawn from a categorical distribution (with a V-dimensional discrete probability vector ($\phi_n^{(p)}$)) specific to that utterance, and each probability vector is in turn drawn from a Dirichlet distribution ($\alpha^{(p)}, B_n^{(p)}$) that is specific to an individual — where $\alpha^{(p)}$, the concentration parameter, is a positive scalar that determines the variance of the distribution and $B_n^{(p)}$, the base measure, is a V-dimensional discrete probability vector that specifies the mean of the distribution. Given that some of these parameters are not observed, a Markov Chain Monte Carlo (MCMC) algorithm is used to infer these latent influence variables and parameters (see Appendix).

2.2 Linguistic Accommodation in the Court Room

To make real world inferences from interactions, the utterance contents W, start times T, and duration D are observed, while the other latent variables are unobserved, i.e.,

$$\{\Theta = {\phi_n^{(p)}}_{n=1}^{N(p)(T)}, \alpha^{(p)}, \beta^{(p)}, \rho_{q\neq p}^{(qp)}, \tau_L^{(p)}\}_{p=1}^{P}$$

These unobserved parameters are instead quantified via their posterior distribution given W, T, and D, i.e.,

$$P(\Theta|W, T, D) \propto P(W|\Theta, T, D) \cdot P(\Theta)$$

The likelihood term is factorised as a result of the model's independence assumptions. To complete the specification of $P(\Theta)$, gamma priors are placed over the remaining parameters and because the resultant posterior distribution is analytically intractable, posterior samples of $\{\alpha^{(p)}, \beta^{(p)}, \rho_{q\neq p}^{(qp)}, \tau_L^{(p)}\}$ are obtained using a collapsed slice-within-Gibbs sampling algorithm (Guo et al., 2015; Neal, 2003).

3 U.S. Supreme Court Transcripts

3.1 Description & Source

The oral transcript of the arguments presented for the case of *Boumediene v. Bush* are provided by the Supreme Court of the United States (U.S. Surpeme Court, n.d.). These are posted on the U.S. Surpeme Court website on the same day the Court hears an argument. Each oral argument heard by the Court involves up to nine justices, the counsels representing the petitioners and the counsels representing the respondent. The format of each argument is formulaic: the counsel for each party has 30 minutes to present their argument, with those representing the petitioners speaking first. The justices are allowed to interrupt the presentations to make comments and ask questions to clarify or elaborate on the arguments presented.

3.2 Pre-Processing

The time-stamped transcript is converted into XML with the tokens split using the TalkBank library (MacWhinney, 2007). All consecutive utterances by the same person are concatenated, any contributions from people with fewer than two (post-concatenation) utterances are discarded, and the time-stamps are rescaled to the interval (0, T = 100). Lastly, the vocabulary in the dataset is restricted to the 600 most frequent stemmed word types to allow for faster computation and stop words are removed.

4 Interactional Influence in Boumediene v. Bush

4.1 Network Description

From the oral arguments, we derive a network signifying the influence (edge) from one party (node) to another, determined by the linguistic accommodation in a party's utterance given all prior utterances. The expectation is that the influence networks inferred from linguistic accommodation patterns will reveal the influence exerted by the parties involved. Given that the petitioners' counsel presents his arguments first, it is natural to see some influence exerted by them. Similarly, as the justices frequently interrupt to clarify or challenge a counsel's argument, we also expect to see influence exerted by justices who agree or disagree with the points made by counsel.

The justices present for this case were Souter, Roberts, Stevens, Scalia, Kennedy, Breyer and Ginsburg, while the counsels were Waxman (acting for the petitioners, Boumediene et al.) and Clements (acting for the respondent, Bush et al.). The petitioners won the case in a 5-4 majority with Breyer, Ginsburg, Kennedy, Souter and Stevens siding with the petitioners and Alito, Roberts, Scalia and Thomas siding with the respondent.

Table 1 summarises some properties of the nodes and basic centrality measures (which ignores the directions of the edges). We include: (i) the betweenness centrality to detect the amount of influence a party has over the flow of information in the network (serving as a bridges); (ii) the closeness centrality to detect which party is able to spread information efficiently through the network; (iii) the in-degree strength, a sum of incoming edge weights, to indicate influence received from other parties; and (iv) the out-strength, a sum of outgoing edge weights, to indicate influence sent to other parties. It is also worth noting that as neither Justice Alito nor Thomas spoke more than two utterances during the oral arguments, they were not included in the analysis.

From this, we observe that the counsels possess higher betweenness centrality measures, which does not come as a surprise given their role in presenting arguments and responding to inquiries from the justices. They play a vital role in the flow of information within the network. Justice Ginsburg also possessed high betweenness centrality in the network and from reading the oral transcripts, this is explained by the fact that she often clarified and simplified the points made by counsel to the other justices, thus serving as a bridge between the two groups.

Party	**Betweenness**	**Closeness**	**In Strength**	**Out Strength**
Counsel Clement (CLEME)	13	0.39	39.1	39.7
Counsel Waxman (WAXMA)	10	0.41	34.0	43.8
Justice Souter (SOUT)	2	0.33	30.6	26.5
Justice Roberts (ROBE)	1	0.37	35.7	38.7
Justice Stevens (STEV)	0	0.16	23.2	9.4
Justice Scalia (SCAL)	1	0.33	28.4	30.9
Justice Kennedy (KENN)	0	0.30	27.2	28.3
Justice Breyer (BREY)	0	0.37	27.0	25.9
Justice Ginsburg (GINS)	14	0.39	38.6	40.5

Table 1: Centrality Measures for *Bush v. Boumediene*

The closeness centrality measures are less clear in this instance partly because the justices ultimately have to side with either the petitioners or respondent, so they do not tell us much in this particular network. As such, a general measure of how close they are to other members of the network may not be the best representation of their influence on others as by definition we should expect clustering into two distinct groups. However, given the limited scope of this paper, further clustering analysis was not conducted.

A final observation of the in-strength and out-strength is the correlation between the two measures captured in figure 1. That is to say, those who exert more influence receive more influence or vice versa. It perhaps confirms the notion that engaging in debate is the best way to convince others of our views. In figure 2, we graph the network to visualise the influence between the parties in the network. In order to illustrate posterior uncertainty, the network is also drawn with different posterior quantiles (using 25^{th}, 50^{th} (i.e., median), and 75^{th} posterior quantiles) — serving as a visual aid in determining material influences.

These networks look very similar to each other but help identify influences that are more statistically significant. In particular, when looking at the 25^{th} quantile, we see that

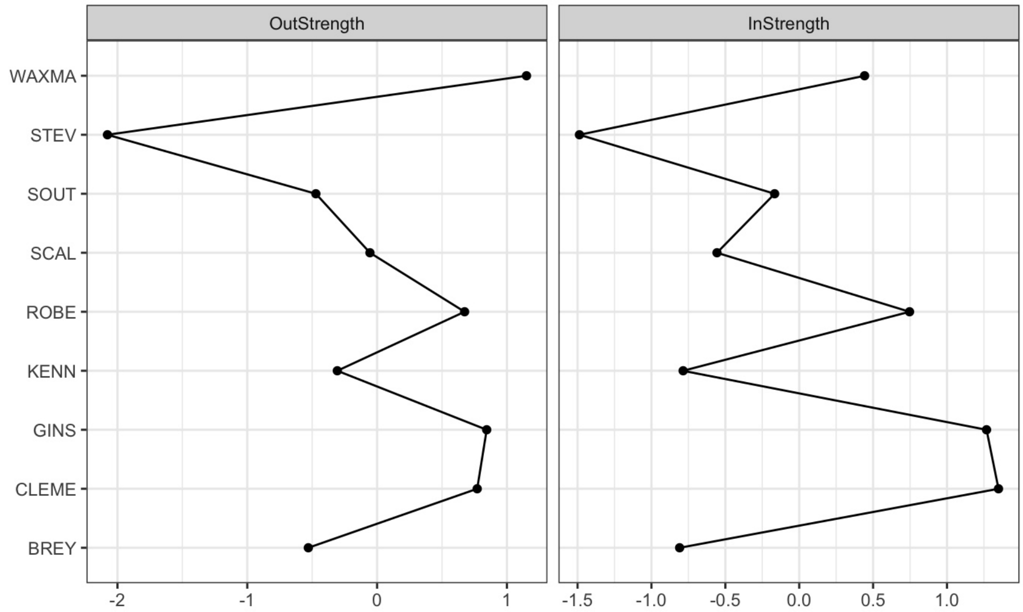


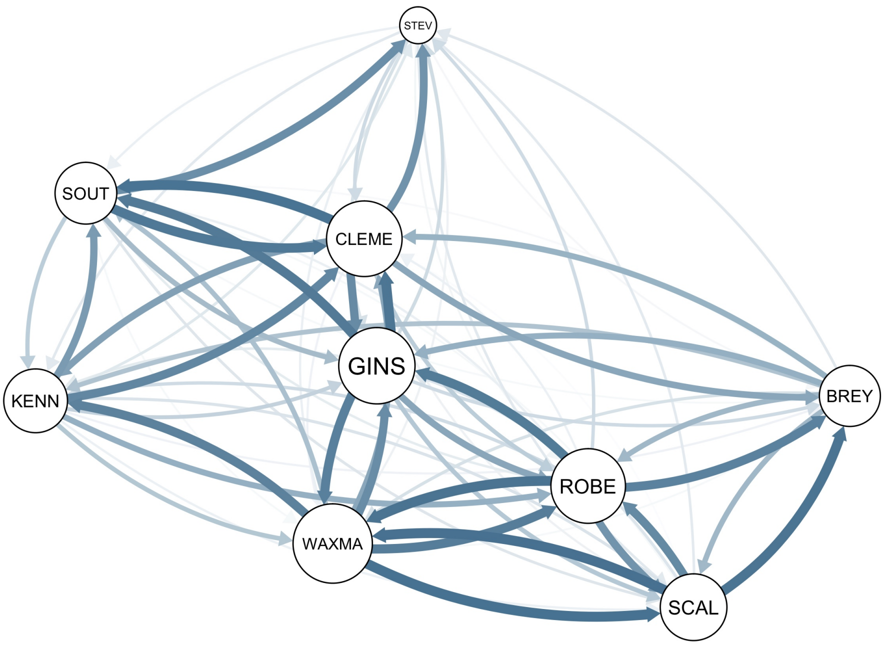
Figure 1: A centrality plot of the in-strength and out-strength to observe correlation between the two

in terms of influence between counsels and justices, Clements had the most influence on Justice Souter, while Waxman had the most influence on Justice Kennedy — who was a crucial vote in the final decision for the petitioners.

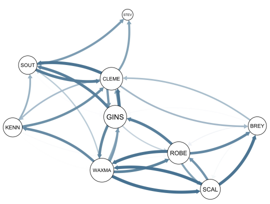
4.2 Implications for the Case

A five-justice majority answered yes to each of these questions outlined at the start of this paper. The opinion, written by Justice Kennedy, stated that the procedures laid out as substitutes were not adequate for a habeas writ, and thus the MCA was an unconstitutional suspension of that writ (Oyez, n.d.). The Court reversed the D.C. Circuit's ruling and found in favour of the detainees. The controversial nature of the case meant that most of the justices voted closer to their political inclinations, i.e. liberal or conservative. This did not come as a surprise to spectators of the case, but what was unexpected was Justice Kennedy's vote in favour of the petitioners as he had ideologically been classed as a conservative (Epstein, Martin, Quinn, & Segal, 2007). He, in essence, was the swing vote. To understand how this may have happened, we look a bit deeper into how he was influenced during the oral arguments as an indication of which direction he was leaning before the vote.

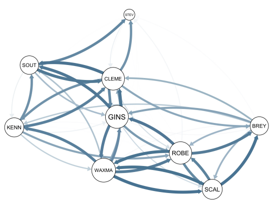
Overall, counsels Clements and Waxman and Justice Ginsburg appear to have the most



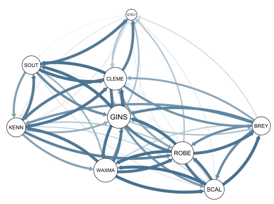
(a) Mean Influence Network



(b) 25^{th} Quantile



(c) 50^{th} Quantile



(d) 75^{th} Quantile

Figure 2: Linguistic Influence Network for *Boumediene v. Bush*

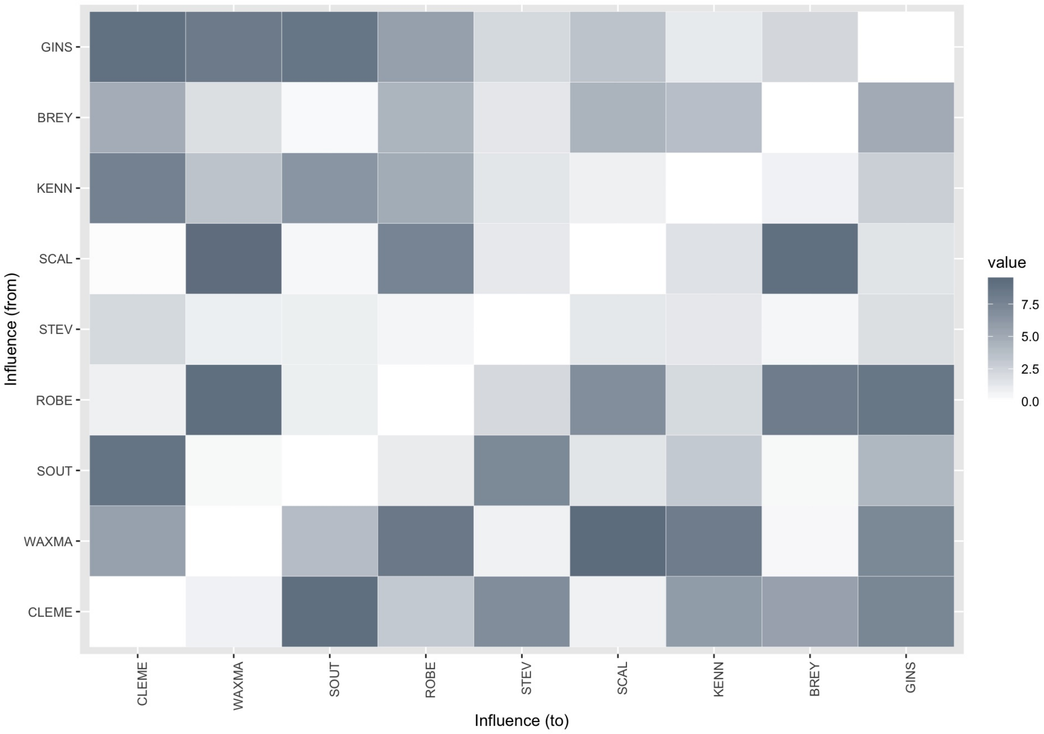
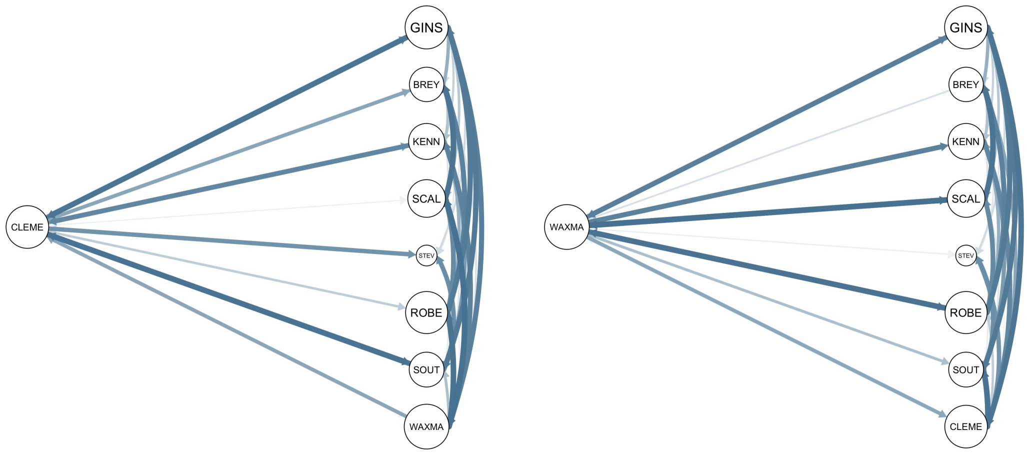


Figure 3: Influence Matrix for *Boumediene v. Bush*

influence on other parties. Given the high betweenness centralities they possessed, it may have been expected that they steered the conversation in our linguistic accommodation model. Delving a bit deeper, the influence matrix in figure 3, which maps the in-degrees and out-degrees between the parties, shows us that Justice Kennedy received the most influence from the counsels Clements and Waxman. However, Waxman (the petitioners' counsel) had significantly more of an effect on Kennedy in the linguistic accommodation model. It is important to note that because a similar level of influence is observed from Waxman to Justices Roberts and Scalia (see figures 3 and 4), who sided with the respondent, this perhaps does not give us a full picture and requires some analysis of the transcripts to understand the dynamics fully.

In the transcripts, we see that interactions between Justice Roberts and Scalia and counsel Waxman were more interrogatory — and at times almost hostile. Whereas, the interactions with Justice Kennedy appeared to prompt Waxman to elucidate his points and were perhaps less hostile. At one point Justice Kennedy quips at Waxman: "you're not heartened by the prospect that the detainees could apply to the Cuban courts, which would then hand the process to the Commanding General at Guantanamo?" to which the room



(a) Flow Diagram for Counsel Clement (b) Flow Diagram for Counsel Waxman

Figure 4: *Boumediene v. Bush* linguistic influence network as a flow diagram showing how each counsel is connected to all other nodes

laughs. Our model rightly picked up all of these interactions as potential influence, but while these interactions might have meant that Justices Roberts and Scalia accommodated Waxman's vocabulary, it did not translate into meaningful influence in the final vote for the case. What this tells us is that the model, on its own, can be misleading for inferring actual influence in the courtroom. What it does pick up are the engagements between parties that may lead to actual influence. For the case of *Boumediene v. Bush*, the interaction that appeared to matter in the end was between Waxman and Justice Kennedy.

4.3 Robustness & Limitations

In Guo et al. (2015), the Bayesian Echo Chamber model is tested on various data sets and compared with other influence networks inferred from Blundell et al. (2012) to evaluate its validity and accuracy. The results show that by focusing on linguistic accommodation patterns, a substantively meaningful influence network is arguably derived. They also assess the predictive probability of held-out data (perplexity), a standard metric for evaluating statistical language models.

For all data sets tested, the Bayesian Echo Chamber out-performed both the unigram language model and the dynamic topic model (Guo et al., 2015). While this may give us statistical confidence, given the linguistic influence inferences are in non-statistical domains, it should also be evaluated against criteria derived from the domain studied. Most judicial scholars that have studied the behaviours of justices confront major hurdles in pinpointing their ideologies or leanings as judges are ideally trained to be devoid of any bias. In this case, although we do not intend to identify the bias of a judge, we still attempt to see the inferred influence on their ideology or opinion.

As such, this model still faces similar hurdles to judicial scholars who study bias; and that is causality. To say the oral arguments could have been a necessary condition in swaying Justice Kennedy's view may be plausible (though challenging to infer with certainty), but it cannot be said to be a sufficient condition as we are unable to peek behind the fortress that is his mind. The nature of this case — being politically charged and liberty infringing in nature — means other factors such as personal and social views that go beyond the law could have also affected the decision made.

In the majority opinion, the Court found that no historical habeas case offered by either side was directly relevant and, instead, turned to the fundamental principles underlying the purpose of habeas corpus: 'to allow the courts to act as a check against the abuse of executive power'. It emphasised the need to "protect against the cyclical abuses of the writ by the executive and legislative branches" and noted that the freedom from unlawful restraint was a fundamental precept of liberty (Oyez, n.d.). Therefore, one could argue that an overwhelming influence on the decision made was the gravity of the lengthy imprisonment of the detainees without trial. Lakhdar Boumediene, for example, had been detained for over six years without an indication of what he was to be charged with.

5 Conclusion

The Bayesian Echo Chamber is a generative model for discovering latent influence networks via linguistic accommodation patterns. Guo et al. demonstrated that the model could recover known influence patterns from various datasets. By comparing it to the influence networks inferred from that of Blundell et al.'s turn-taking model they show its ability to infer meaningful insights from linguistic accommodation patterns. In this paper, we take this further to analyse how this pattern might indicate the leaning of a justice in a case during oral arguments, and we see similar meaningful insights. This is however cautioned by the fact the ability to make a causal claim from such analysis is limited by other external factors such as legal precedents, the political climate and perhaps even bias.

References

Blundell, C., Beck, J., & Heller, K. A. (2012). Modelling reciprocating relationships with hawkes processes. In *Advances in neural information processing systems* (pp. 2600–2608).

Epstein, L., Martin, A. D., Quinn, K. M., & Segal, J. A. (2007). Ideological drift among supreme court justices: Who, when, and how important. *Nw. UL Rev.*, *101*, 1483.

Guo, F., Blundell, C., Wallach, H., & Heller, K. (2015). The bayesian echo chamber: Modeling social influence via linguistic accommodation. In *Artificial intelligence and statistics* (pp. 315–323).

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90.

MacWhinney, B. (2007). The talkbank project. In *Creating and digitizing language corpora* (pp. 163–180). Springer.

Neal, R. M. (2003). Slice sampling. *The annals of statistics*, *31*(3), 705–767.

Oyez. (n.d.). *Boumediene v. bush.* Retrieved from `https://www.oyez.org/cases/2007/06-1195`, accessed on 2019-04-09.

Rasmussen, J. G. (2013). Bayesian inference for hawkes processes. *Methodology and Computing in Applied Probability*, *15*(3), 623–642.

Sanders, R. E. (1987). *Cognitive foundations of calculated speech: Controlling understandings in conversation and persuasion.* Suny Press.

U.S. Surpeme Court. (n.d.). *Oral argument transcripts.* Retrieved from `https://www.supremecourt.gov/oral_arguments/argument_transcript/2002`, accessed on 2019-03-28.

West, R. L., Turner, L. H., & Zhao, G. (2010). *Introducing communication theory: Analysis and application* (Vol. 2). McGraw-Hill New York, NY.

A Appendix: Supplementary Charts

A.1 Total Influence within the Influence Network

The total influence sent and received by each participant is shown for the model in figure 5. The error bars represent the posterior standard deviation. This is a graphical representation of the in-strengths and out-strengths seen in table 1.

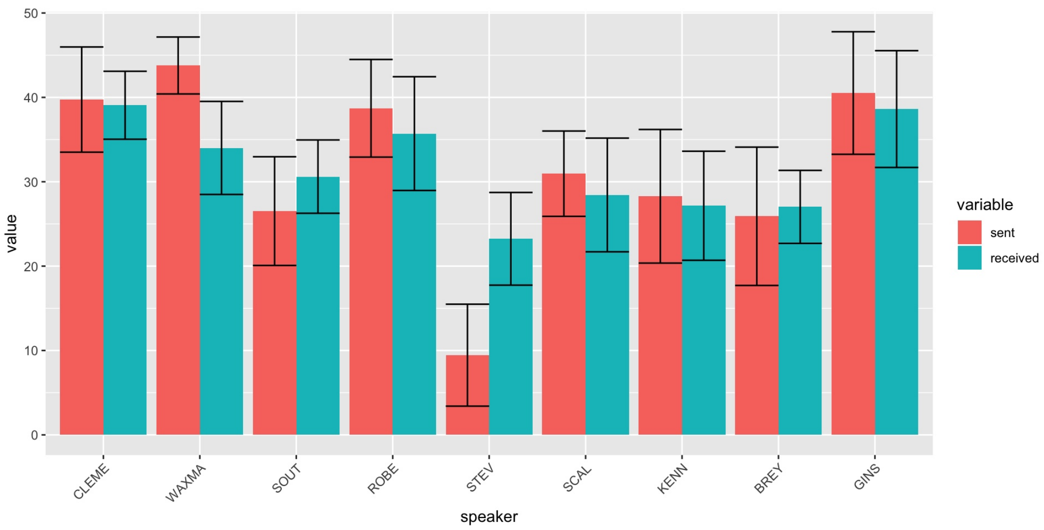


Figure 5: Total Influence Chart for *Boumediene v. Bush*

A.2 Markov Chain Monte Carlo (MCMC) Modelling

MCMC is a class of techniques for sampling from a probability distribution and can be used to estimate the distribution of parameters given a set of observations (Neal, 2003). MCMC allows us to draw samples from the distribution of the latent variables or parameters that are unobserved — or cannot be computed. It can be used to sample from the posterior distribution over parameters, i.e. to compute the distribution over the parameters given a set of observations and a prior belief.

Our MCMC sampling was conducted with 4000 samples and 500 burn-ins, and the observed priors were the utterance contents W, start times T, and duration D.

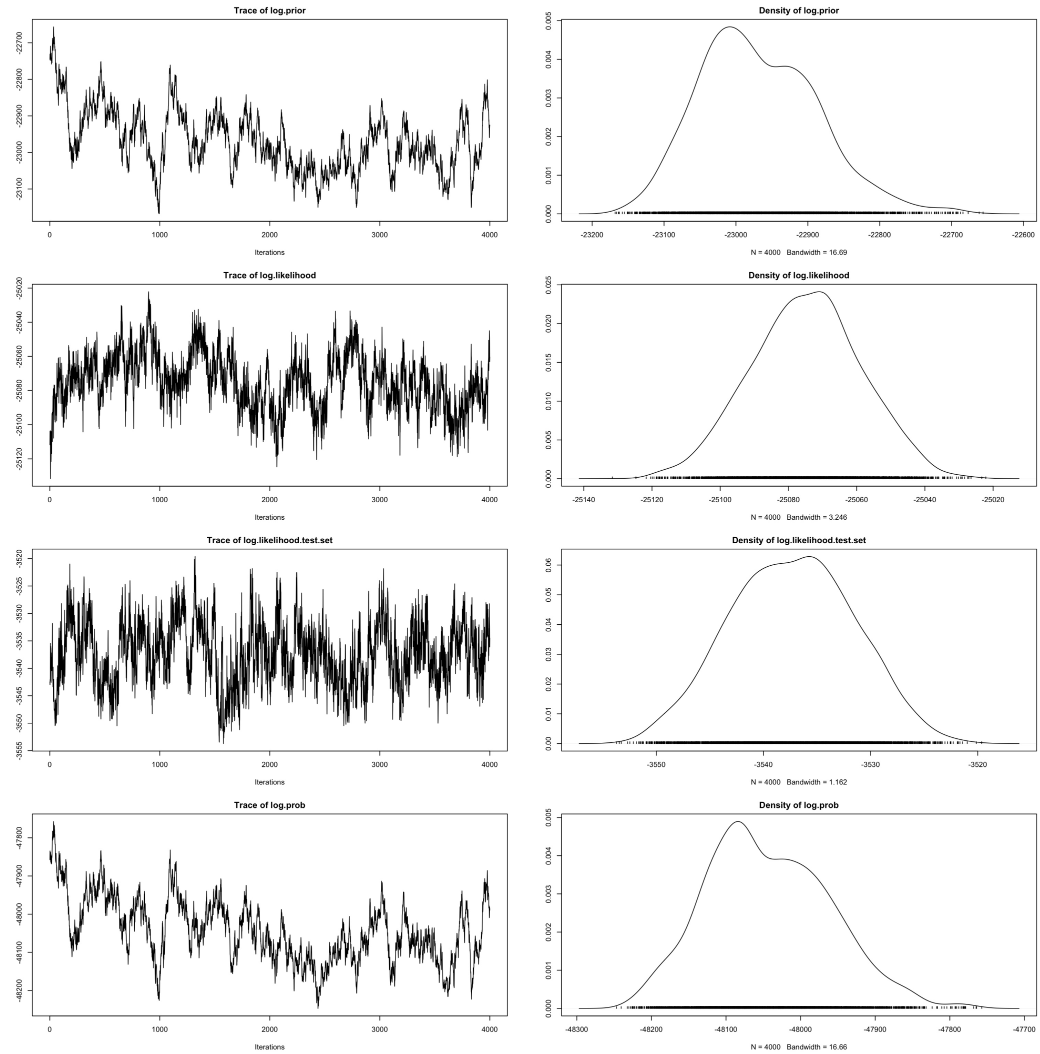


Figure 6: Plot of Markov Chain Monte Carlo Trace and Density of log-Likelihoods for *Boumediene v. Bush*