Replicating 'How Newspapers Reveal Political Power'

Adapted from Ban et al. (2019)

Sean Hambali

February 23, 2024

1 Introduction

Political science, in its essence, is a study of power. However, the empirical measurement of power is an extremely challenging endeavor due to it being a multifaceted concept. Ban et al. (2019) provides a novel strategy to address this challenge; the authors put forward the notion that one can use newspaper coverage to provide a sense of how powerful political actors or offices are.

A measure is only as good as its ability to capture the theoretical concept of interest (Grimmer and Stewart 2013). To further demonstrate the usefulness of their measure, the authors perform validations of their newspaper-based coverage (hereafter referred as "NC") measure by looking at a series of historical reforms and/or political phenomenons, identifying whether the NC measure varies in accordance to their hypothesis.

2 Methods

2.1 Data

The authors use data from newspapers.com, which is an archive that contains text data of newspaper articles from 1690 to 2024. For their purposes, the authors collect a stratified sample of 50% of all articles in the archive, focusing on the period 1877-1977, which comprises the majority of the data in the archive.

Because the OCR results are rather messy and there are instances where letters are not parsed correctly, the authors then use regular expression tools to match the desired terms in the raw text data. Table 1 summarizes the regular expression patterns that were used by

the authors in matching the terms, allowing for non-exact matches in order to capture the OCR parsing errors.

2.2 Measuring Relative Power

The NC measures the *relative* power of an actor compared to her counterpart based on the comparison of the number of newspaper mentions between the two actors, as follows:

$$\hat{Y}_i = \frac{N_i}{N_i + N_{i'}}$$

where $\hat{Y}_i$ denotes the estimated relative power of actor i. N_i and $N_{i'}$ denotes the amount of newspaper mentions for actor i and actor i's counterparts, respectively.

3 Validation Results

First, the authors validate the NC measure by comparing newspaper mentions of congressional committees with that of Stewart and Groseclose (1999), which is derived from member transfer requests. The results are shown by Figure 1 below. The replication result follows closely with the original result, showing the strong correlation between the NC committee ranking and the Groseclose-Stewart ranking.

The authors implement a second validation by comparing NC of party leaders throughout their tenure. They hypothesize that senators gain more power as they are appointed as House Speaker or minority party leader, and therefore should receive more newspaper coverage upon appointment. Figure 2 supports this hypothesis, and I'm able to replicate the original result.

The next validation exercise looks at the variation of the NC measure in settings where city governments implement reforms that limit the executive authority that the city mayor has. Their hypothesis is that the NC of city mayors should decrease relative to the other city offices post-reform, reflecting the decline in their power. Figure 3 shows that the result is consistent with this hypothesis.

As the last validation exercise, the authors look at the 1964 legal reform in Massachusetts that strips out the authority of the Massachusetts Executive Council due to corruption and bribery charges. Their hypothesis is that the NC of the council should decrease as an artifact of their power decline. Figure 4 demonstrates such decline; the replication result is also consistent with the paper's original finding.

Lastly, the authors also provide an application to demonstrate the usefulness of the NC measure in quantifying a political phenomenon. There is a notion that state party committees are no longer as powerful as they once were, and that there has been a transition to candidate-centered politics. The authors use the NC measure to quantify the power of state (or local)

party committees over time and to pinpoint the exact period at which their power starts to decline. Figure 5 shows the results, which suggests a general decline in party power, at least in 9 U.S. states.

4 Replication Details

4.1 Differences from Original Results

Via this replication exercise, I'm able to reproduce the exact same results that were presented in the paper. In their replication folder, the authors have provided thorough codes and instructions, detailing the steps that they perform to pre-process the term count data, and to generate the measures of interest for further visualization. As Figures 1 to 5 have shown, the original and the replication results produce the exact same finding, with minor differences in the aesthetic elements of the visualization.

The only minor difference was observed on the standard errors in the original and replication regression results (Tables 2 and 3, respectively). For example, using the same data, the standard errors for the first specification of both tables differ (in the tables they look the same due to rounding). The standard error magnitude reported by Table 2 is `0.0239`, while in Table 3 that number is `0.02344`. That said, such minor difference likely arises due to different software being used for implementing the regression (the original paper uses Stata, while this replication exercise uses R), and it bears no significance on the paper's overall results.

4.2 Replication Autopsy

While the paper's main findings have all been replicable, I haven't been able to replicate the process of converting the raw newspaper text data to a cleaned, aggregated term count data. This is because the newspaper data is a proprietary data and thus is not publicly accessible. Therefore, I wasn't able to replicate the regex matching process, or to see if the results are robust to alternative regex specifications.

The replication folder provides both Stata and R codes. The authors mainly use Stata for pre-processing and wrangling the data as well as for implementing the regression analysis. They use R primarily for visualization purposes. In this replication exercise, I try to streamline the process by converting all codes to R. All the scripts that I use for reproducing the paper findings can be found under the `0_scripts > replication` folder of this repository.

When going through their do-files, I notice that there is a minor error in the construction of the analytical variables. Specifically, in the part of the `make_main_items` do-file that generates the variables of interest for producing Figure 3, the authors used `r_mayor` instead of `r_mayor_x` to generate the sum of mayor-council proportions for the city name filter approach. This can be seen in the fifth line of the following code chunk:

```
gen rel_mayor_council_total = r_city_council + r_mayor
gen rel_mayor_council_control_total = r_control_city_council + r_control_mayor
gen rel_mayor_council = r_mayor / rel_mayor_council_total
gen rel_mayor_council_control = r_control_mayor / rel_mayor_council_control_total
gen rel_mayor_council_total_x = r_city_council_x + r_mayor
gen rel_mayor_council_x = r_mayor_x / rel_mayor_council_total_x
```

That said, such minor error does not have any bearing on the replication results, since eventually the city name filter approach isn't used to produce Figure 3.

4.3 Possible Extensions

There are several possible extensions to the methods and analysis presented in Ban et al. (2019). First, further endeavor to replicate the results of this paper would benefit greatly from access to the raw text data. With the availability of the data, one can test whether or not the regex rules employed by the authors have successfully captured the imperfections and possible errors of the OCR parsing results. To the extent that OCR parsing errors are concentrated among earlier newspaper entries, the estimated NC measures may have higher variance in early years. Further, with the availability of the raw text data, others can also test out whether the study conclusions are robust to alternative regex patterns.

Second, I would be interested in exploring the applicability of the NC measure in other news medium, such as news broadcast. Given the limited amount of space that can be covered in a news broadcast (relative to newspapers), these news outlets may focus only on stories with highest newsworthiness values. That is, only the most important subjects/actors are covered by the broadcasts. It will be interesting to explore whether such broadcast-based measure can provide a closer approximation to a person/actor's political power.

Third, I would also be interested in applying the NC measure in the context of developing countries. In Indonesia, for example, on the aftermath of the 1998 financial crisis, the government experienced a massive decentralization of authority from the central government to sub-national government units. The NC measure will be useful in this case to visualize the dramatic decline in power of central executive authorities, such as the President, and at the same time, the relative increase in power of sub-national executive authorities, such as governors, mayors, among others. That said, the applicability of NC in developing countries may be limited by the fact that there is less transparency in governance in their contexts. Since the NC measure cannot measure the influence of actors that remain outside of spotlight, it might provide a less accurate approximation of power in developing countries.

5 Tables and Figures

Table 1: Regular Expression Patterns

Error Type	Correct Character	OCR	Regex
1:1 Substitution	e	c	[ec]
	v	y	[vy]
	o	c	[oc]
	i	l	[il]
	t	l	[tl]
	b	h	[bh]
	g	y, j, q	[gyjq]
	gf	t	[ft]
	a	u, o	[auo]
1:2 Substitution	m	rm	[m(rn)]

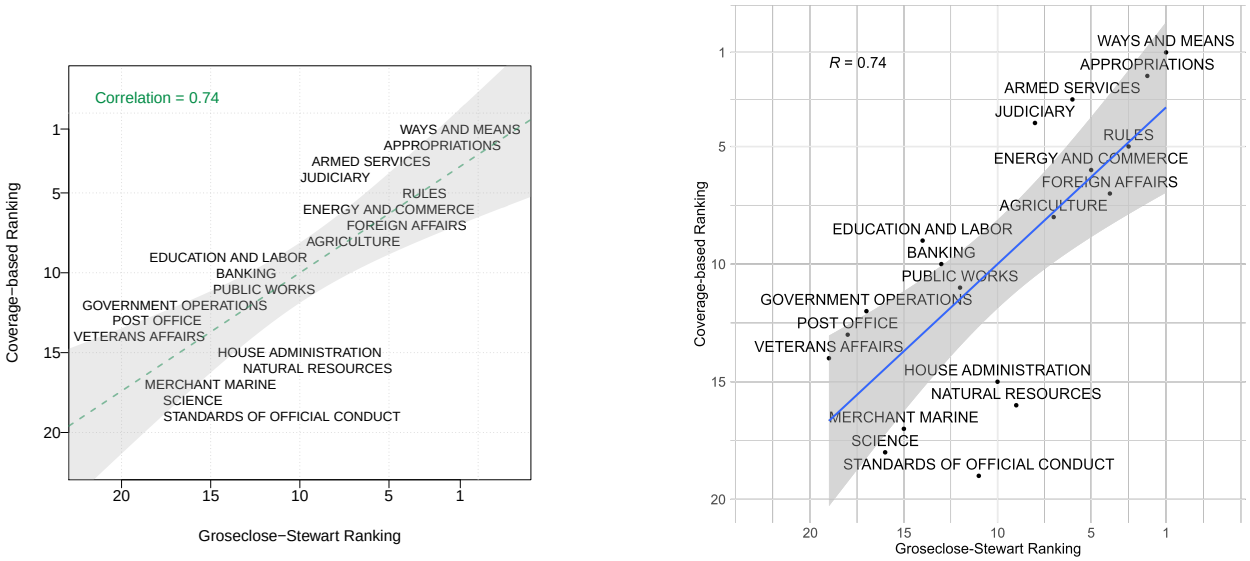
Table 2: Impact of Switch from Mayor-Council to Council-Manager City Government, Original

	All Mentions		Using City Name Filter	
	Relative Coverage of Mayor	Relative Coverage of City Manager	Relative Coverage of Mayor	Relative Coverage of City Manager
Council-Manager Govt Form	-0.18 (0.02)	0.18 (0.02)	-0.25 (0.04)	0.29 (0.03)
N	3540	3540	2376	2376
City Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes

Standard errors, clustered by city, are in parentheses.

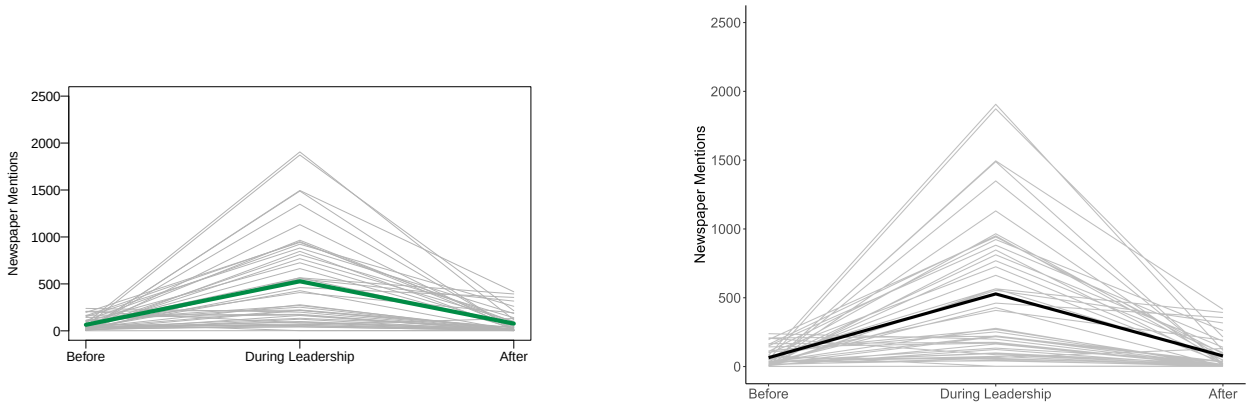
Table 3: Impact of Switch from Mayor-Council to Council-Manager City Government, Replication

	Rel. Coverage of Mayor	Rel. Coverage of City Manager	Rel. Coverage of Mayor	Rel. Coverage of City Manager
	All Mentions		Using City Name Filter	
	(1)	(2)	(3)	(4)
Council-manager government form	-0.18*** (0.02)	0.18*** (0.02)	-0.25*** (0.04)	0.29*** (0.03)
Observations	3,540	3,540	2,376	2,376
R^2	0.76	0.70	0.70	0.68
Within R^2	0.15	0.29	0.11	0.26
Year fixed effects	✓	✓	✓	✓
City fixed effects	✓	✓	✓	✓



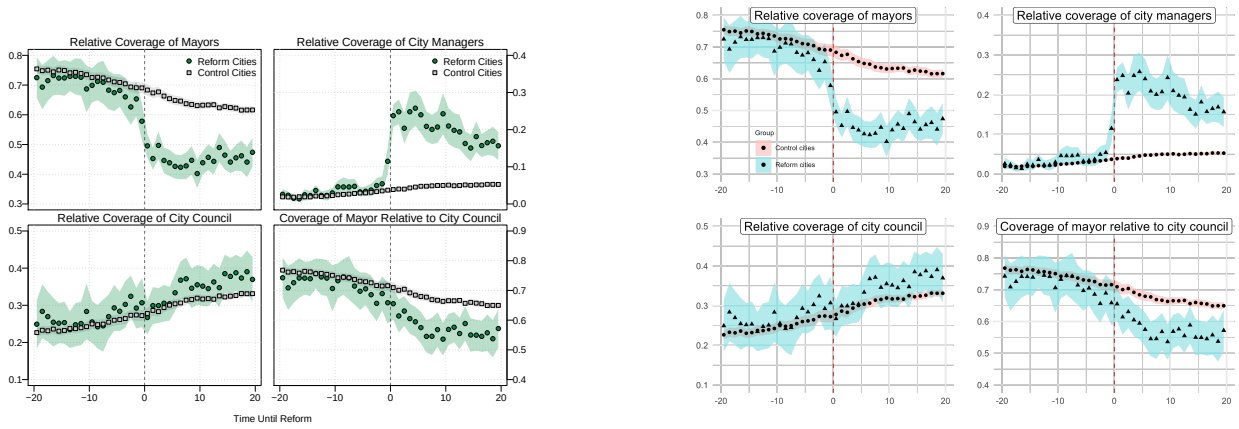
(a) Original (b) Replication

Figure 1: Ranking Congressional Committees



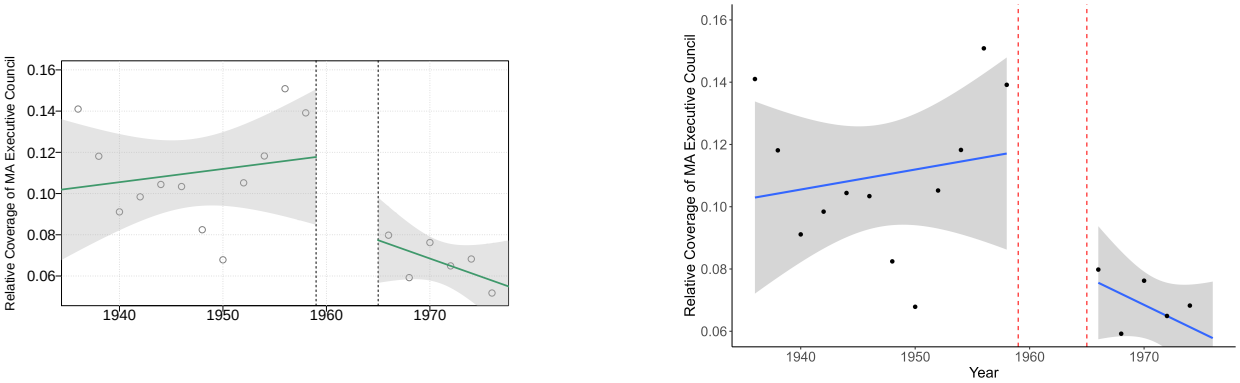
(a) Original (b) Replication

Figure 2: Coverage of Speakers Before, During and After Leadership



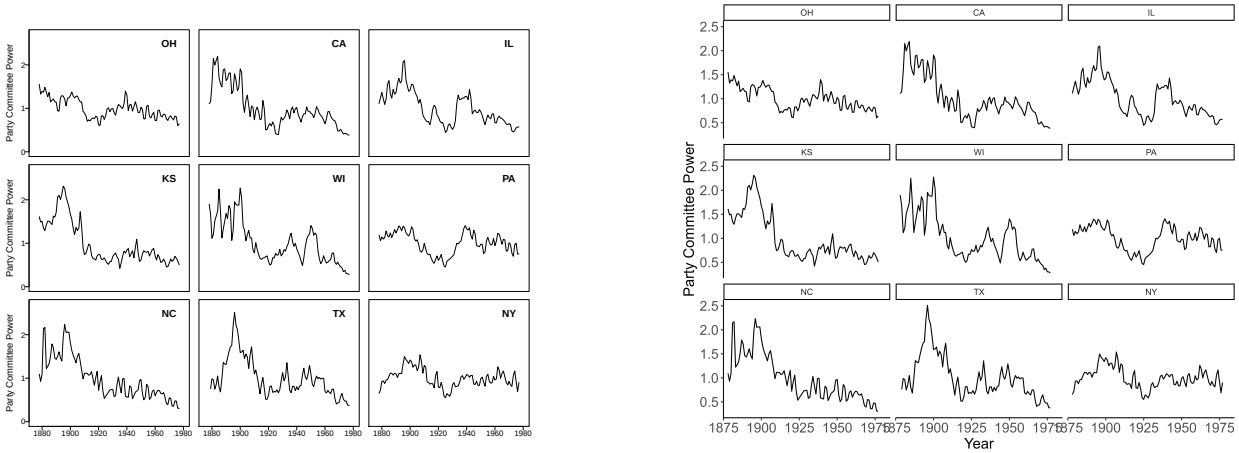
(a) Original (b) Replication

Figure 3: Coverage of City Offices, Before and After Reforms



(a) Original (b) Replication

Figure 4: Coverage of Massachusetts Executive Council



(a) Original (b) Replication

Figure 5: State Party Committee Power Across 9 U.S. States

References

Ban, Pamela, Alexander Fouirnaies, Andrew B Hall, and James M Snyder. 2019. "How Newspapers Reveal Political Power." *Political Science Research and Methods* 7 (04): 661–78. https://doi.org/10.1017/psrm.2017.43.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97. https://doi.org/10.1093/pan/mps028.

Stewart, Charles, and Tim Groseclose. 1999. "The Value of Committee Seats in the United States Senate, 1947-91." *American Journal of Political Science* 43 (3): 963–73. https://doi.org/10.2307/2991842.