# Charting the Roles and Careers of Ministerial Advisors: A Data-Driven Approach Using NLP - Working Paper IWPP4 Guadalajara

Tom Bellens[1] and Marleen Brans[2]

[1]Public Governance Institute, KU Leuven
[2]Public Governance Institute, KU Leuven

May 2024

**Abstract**

This paper presents a novel methodology for studying ministerial advisors (MAs) by leveraging Natural Language Processing (NLP) techniques to analyze extensive textual data. Focusing on Flemish MAs between 1999 and 2020, we address two primary research questions: the distribution of MAs' roles and the correlation between their career backgrounds and roles. Utilizing Connaughton's typology, which categorizes MAs into experts, partisans, coordinators, and minders, we analyze 18,887 documents related to 341 MAs. By employing classifier entropy, we model the political-neutral axis, and by assessing distances between topic probability vectors, we model the specialist-generalist axis. Our findings reveal a clear distinction between politically active and neutral advisors. However, the specialist-generalist axis requires further refinement. The significant correlation between career data and document partisanship underscores the impact of career paths on advisors' roles. This study highlights the potential of integrating computational techniques with traditional research methods, offering insights into the roles and careers of MAs. Future research should refine the specialist-generalist model and explore broader applications of this methodology.

## 1 Introduction

Ministerial advisors, abbreviated as MAs, have increasingly become a focus of academic research. They play a vital role in the executive triangle, which includes executive politicians and bureaucrats (Bach & Hustedt, 2023). Studies on MAs touch on key themes in political science and public administration, such as politicization (Hustedt & Salomonsen, 2014), the relationship between politics and administration (Connaughton, 2015; Hustedt & Salomonsen, 2017;

Shaw & Eichbaum, 2018), representation (Taflaga & Kerby, 2020), and partitocracy (Walgrave et al., 2004). Despite substantial empirical research on MAs, most studies have been limited to relatively small-N survey research designs. In this paper, we introduce a novel methodology for research on MAs, utilizing a suite of Natural Language Processing (NLP) techniques.

We aim to elaborate on one of the key issues concerning MAs: their varying roles in the ministerial cabinet. A seminal work on this subject is Connaughton (2010) typology, which categorizes MAs into four types: experts, partisans, coordinators, and minder roles. These roles are situated along two axes: one ranging from technical/management to political communication, and another from steering to policy advice. This typology has been validated before in various political contexts through survey research. We use this typology as a case study for our new methodology and aim to answer:

> **RQ1**: How are the roles of MAs distributed in Flanders between 2000 and 2020?"

Another central theme in literature about MAs is their career backgrounds (Alam et al., 2019; Askim et al., 2022; Rouban, 2012). Do advisors mainly come from partisan, private or technocratic backgrounds? Strongly related with Connaughton's typology is then the question of to what extent career background is a predictor of the role an MA assumes in the cabinet. Again, we seek to validate our new methodology by asking the question

> **RQ2**: To what extent do career paths correlate with the different types of advisor we discerned?

In the next section, we elaborate on the theoretical frameworks behind these research questions.

After that, we introduce our new methodology. Despite the extensive empirical research on MAs, there remains a gap in integrating modern computational techniques with traditional survey-based approaches. The advent of big data and advancements in NLP provide an unprecedented opportunity to analyze large volumes of textual data systematically. In this paper, we combine an existing dataset of career paths of 341 Flemish MAs between 1999 and 2020 with a new dataset of 18 887 documents from several sources related to these MAs. For role analysis, we use the latter and introduce two NLP techniques for analyzing this large textual dataset. On the one hand, we model the political-neutral axis through classifier entropy. On the other hand, we model the generalist-specialist axis by comparing distances between topic probability vectors.

Finally, we put these results next to the previously collected career data to see if there is a correlation between these data and the results we obtained from the documents dataset. In discussing the results, we conclude that there are promising results especially on the "political-neutral" axis and that our methodology could be elaborated further in this direction. Regarding the "specialist-generalist" axis, the methodology is yet to be refined after the IPPA workshop.

## 2 Theory: ministerial advisors, their roles and career paths

Before delving into the methodology, it is essential to explore the theory we aim to test. MAs and their roles vary globally. They have a long history and are often most extensive and influential in Napoleonic systems of government, such as those in France, Italy, and Belgium (Brans et al., 2017; Di Mascio & Natalini, 2013; Eymeri-Douzans et al., 2015). Ministerial offices also exist in Scandinavian (Askim et al., 2022; Blach-Ørsten et al., 2020) and Germanic systems, and in recent decades, a process of "cabinetization" has been observed in Westminster systems (Pickering et al., 2023).

Connaughton's seminal work, "Glorified Gofers, Policy Experts or Good Generalists: A Classification of the Roles of the Irish Ministerial Adviser," examines the Irish Fianna Fáil–Progressive Democrat coalition government between 2000 and 2007 but offers insights applicable to other governmental systems beyond the Westminster model. This section will first discuss Connaughton's typology, then explore its applications, and finally outline the specific aspects of her typology that we seek to test.

Based on surveys and interviews concerning advisers' backgrounds, motivations, and specific tasks, Connaughton develops a typology of MAs. She establishes two axes: a "communication" axis ranging from "political" to "technical/neutral" and a "policy implementation" axis ranging from "steering" to "policy advising." Along these axes, Connaughton identifies four types of advisers.

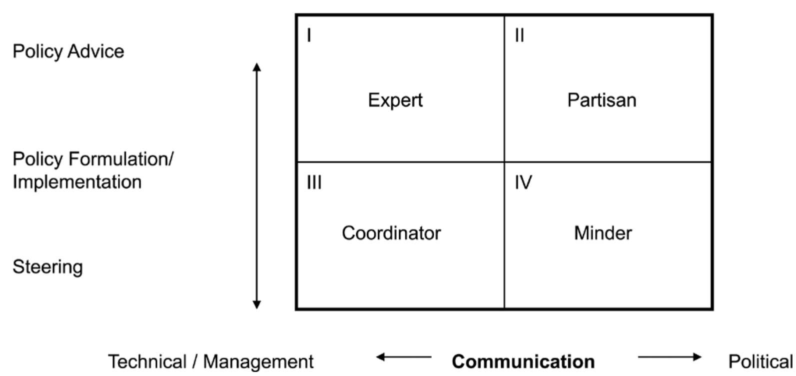


Figure 1: Connaughton (2010)'s typology.

The first type is the **expert** (technical/policy advising), who provides specialized advice and technical knowledge within specific policy sectors. The second type is the **partisan** (political/policy advising), who is politically aligned with the minister's party and offers advice within particular sectors. The third type is the **coordinator** (technical/steering), who, without political affiliation, focuses on ensuring alignment and communication between different government sectors. The fourth type is the **minder** (political/steering), who acts as a trusted aide, defending the minister's political and reputational interests. Connaughton notes that the Irish advisers between 2000 and 2007 fall mostly into the minder category.

Gouglas et al. (2015) empirically test Connaughton's typology in Belgium, Greece, and the European Union. Based on surveys and semi-structured interviews, Gouglas et al. identify different predominant types of advisers in various political contexts. In Belgium, cabinet officials in 2014 were typically a hybrid of minders and coordinators; these politically active individuals assisted ministers politically rather than advocating specific party issues. In Greece, advisers were generally found to be less politically active than their Belgian counterparts, fitting more closely with the pure coordinator type. For advisers in the European Commission, the profile was more complex: both highly skilled generalists with varying levels of political activity and specialists were identified, leading Gouglas et al. to conclude that both hybrid minder/coordinators and specialist types are prevalent in EC cabinets.

Thus, in addressing the first research question, our objective is twofold: to align with existing research for comparison purposes and to ensure that our hypothesis is testable within the framework of our new methodology. To this end, we propose the following axes for testing our hypotheses. First, we have the communication axis, which we modify to the political-neutral axis. In the next section, we explain how our methodology is better suited to this modification. Second, we have the policy formulation/implementation axis, which we name the specialist-generalist axis.

Therefore, the hypotheses we propose for research question 1, in relation to our methodology, are as follows:

> $H_1$: *On the political-neutral axis there is a clear distinction between neutral and politically active advisers.*
>
> $H_2$: *On the specialist-generalist axis there is a clear distinction between specialist and generalist adivsers.*

The works of Connaughton and Gouglas et al. focus explicitly on the roles played by MAs during their tenure in cabinets or ministerial office. However, there is a line connecting this with another substantial body of literature on MAs and their career paths. This overlap is to be expected; the roles MAs assume during their time in cabinet are likely influenced by or at least correlated with their pre-cabinet activities. The careers of MAs are a heavily researched topic in various governance systems and are often studied in the French-language literature under the term "prosopography" (Alam et al., 2019; Brans et al., 2023; Delpu, 2015; Lemercier & Picard, 2011). While this research often implicitly addresses the relationship between career paths and roles in the cabinet, a comprehensive, large-scale systematic study of this relationship is still lacking.

Research on career paths can be broadly divided into two strands. The English-language literature typically presents MAs as partitocratic actors with strong ties to political parties, primarily interested in consolidating political control over policy. In contrast, the French-language literature has historically emphasized the technocratic nature of MAs (Charle, 1984; Mathiot & Sawicki, 1999), portraying them as highly educated elites who seek to technocratize or bureaucratize the political process. Although these perspectives are presented simplistically here and more nuanced accounts exist, it remains clear that both fields of research—on the roles and career paths of MAs —would benefit from a unifying perspective. The methodology proposed in this paper takes initial steps towards this integration. The hypothesis we propose here regarding research question 2 is the following:

> $H_3$: *The distributions determined on the neutral-active axis and the specialist-generalist axis correlate with the career profiles of ministerial advisors.*

# 3 A novel methodological approach

To develop a typology of MA's roles and compare this with their career paths, we follow two distinct tracks. First, we must gather information about the advisers' occupations and devise a methodology to transform this data into a meaningful classification. Second, we need to collect information about their career paths and process it to facilitate a comparison with the derived typology. This section outlines these steps in detail.

We focus on MAs who served in the Flemish government cabinets between 1999 and 2020. This focus is driven not only by the availability of a comprehensive dataset on Flemish cabinet employees but also by the presence of previous similar research the federal Belgian cabinets, which are closely related to the Flemish ones. This context provides an excellent opportunity to apply our new methodology.

The data we use is twofold. We have at our disposition a dataset of all Flemish cabinet officials from 1999 to 2020 and their career paths, using both government sources and LinkedIn profiles. This dataset serves a dual purpose: it helps us establish career profiles and forms the basis for comparison of the career data with the roles typology. From here on we will refer to this dataset as **Dataset I**.

In this paper, we interpret "role" broadly, focusing on the general behavior of cabinet employees. Specifically, we examine whether a cabinet employee tends to behave more politically active or neutrally and whether they function more as a generalist or a specialist. To test these aspects, we utilize the iCandid data hub, a project by FWO and KU Leuven, which integrates several big data sources: all printed press from recent years (BelgaPress), news transcripts from the national broadcaster, Twitter accounts of politically relevant figures and organizations, and parliamentary documents and reports. For each cabinet member in our initial dataset, we query the iCandid data hub and add all relevant media files in full text to the second database. From here on we will refer to this dataset as **Dataset II**.

Dataset I encompasses 341 individuals and 2574 coded career positions (as of the time of writing, with the dataset expected to expand to about 650 individuals after the IPPA workshop). Dataset II contains documents for 276 of these individuals, resulting in a dataset with 18 887 documents distributed among all cabinet employees. In the following sections, we explain how we use these datasets to arrive at answers to the research questions set out earlier.

| **Dataset** | **Content** | **N** |
|---|---|---|
| Dataset I | MAs | 341 |
| | Career positions | 2574 |
| Dataset II | Documents (MA) | 18 887 |
| Dataset III | Politics (MA) | 40 000 |

Table 1: Datasets

### 3.1 Using classification entropy to model the political - neutral axis

The first research question we aim to test concerns the position of cabinet members on the political - neutral axis. We address this question using the iCandid dataset through an innovative yet straightforward approach. Initially, we train (fine-tune) a sequence classifier on a dataset, containing texts associated with various political parties in Flanders. We then apply this classifier to predict the political orientation of texts linked to the MAs. By calculating the entropy over the last layer of the sequence classifier, we assess the political signature of the texts. The principle is that texts with a clear political signature exhibit low entropy, whereas "neutral" texts display high entropy. This method enables us to determine the distribution of texts along the political - neutral axis for the different advisers.

The first step is to compile a dataset to train a sequence classifier that can distinguish between different political parties. For this, we use the iCandid database. We search for the names of the four major political families in Flanders—Christian Democrats ("CD&V"), New Flemish Alliance ("N-VA"), Socialists ("sp.a" / "Vooruit"), and Liberals ("Open-VLD")—and select the 10,000 most relevant media pieces for each. These media pieces are similar in composition to the dataset of texts linked to cabinet officials, comprising full-text newspaper articles, news transcripts, parliamentary documents, and tweets. This is **Dataset III**. Dataset III is then used to train a sequence classifier that attempts to predict a party label based on a piece of text. In principle, a large number of sequence classifiers can be used but we chose in this paper to fine-tune a BERT model on our dataset. This encoder model, with a transformer architecture, is a popular option when it comes to fine-tuning textual data and its weights are readily available online. We chose the standard BERT base uncased model, with 12 transformer block layers (Devlin et al., 2019). Since we have a large amount of data, we think it is justified to fine-tune this large amount of weights without freezing certain layers or other tricks classically used to boost fine-tuning. We have not manually reviewed all 40000 items and thus cannot make any statements about the quality of the data but start from the assumption that “noise” is equally distributed across the different batches and thus does not contribute anything to the classification.

Finally, we use this BERT classifier to classify the texts we have previously collected from the cabinet staff. Instead of focusing on the classification result, we analyze the final distribution of the classification layer. Specifically, we extract the probabilities assigned by the sequence classifier to determine the likelihood that a given text belongs to a particular party. Based on this distribution, we calculate an entropy score.

Entropy is defined as follows:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$$

where $H(X)$ is the entropy, $P(x_i)$ is the probability assigned to class $i$, and $n$ is the number of classes.

The higher the entropy, the more ”uncertain” the model is. Assuming the sequence classifier effectively captures the latent features of the Flemish party political landscape, a higher entropy indicates that the text in question is more ”neutral.” Conversely, a lower entropy suggests that the model is more certain about the partisan / political nature of a given text.

Ultimately, this allows us to test hypothesis 1. Since entropy is a continuous variable we attempt to prove the existence of two groups by fitting a Gaussian mixture model with two components to the distribution of entropy, aggregated by a center measure at the individual level (the entropy scores of all texts belonging to an individual are aggregated). This leads to a minor rewrite of hypothesis 1:

> $H_1$: *The political-neutral axis is a bi-modal distribution, with a clear distinction between neutral and politically active advisers.*

## 3.2 3.2 Specialist - generalist axis: Comparing variances in cosine distances between topic representations

The second research question we test relates to the specialist-generalist axis. Here, we also work with representations built in a sequence classifier, though in a slightly different manner. The basic idea is that the representations from dataset II of a “specialist” should be focused around a well-defined topic, whereas a “generalist” should be associated with a broader range of topics. For each advisor, we calculate a value that situates them on the specialist-generalist axis. This value is derived by examining the distribution of topics covered in their media representations. If an advisor’s texts are concentrated around a narrow set of topics, they are classified as a specialist. Conversely, if their texts span a wide range of topics, they are classified as a generalist. This approach allows us to quantify the extent to which each advisor's media presence aligns with being a specialist or a generalist.

The problem to be solved here is a topic modeling problem. For this, we rely on BERTopic, a topic modeling technique that leverages BERT embeddings to create dense representations of documents and then applies clustering algorithms to identify topics (Grootendorst, 2022). The advantage of using BERTopic lies in its ability to capture the semantic nuances of the texts, thanks to BERT's contextual embeddings, and to group similar texts into coherent topics.

Using BERTopic, we first obtain the topic distribution for each document in dataset II. These topic distributions provide a probabilistic representation of how each text relates to the discovered topics. For instance, a text might have a 0.4 probability of belonging to Topic A, a 0.3 probability for Topic B, and so on. The topics and what they represent are not of interest to us.

To quantify whether an advisor is a specialist or a generalist, we aggregate the topic probability distributions of all texts associated with that advisor. This aggregation provides a topic representation for each advisor. To measure the extent of an advisor's specialization, we calculate the average distance between the topic distributions of their texts. This involves computing pairwise distances between all topic distributions for texts related to the advisor and averaging these distances.

An advisor whose average distance between topic distributions is smaller than the average distance in the entire corpus is classified as a specialist. This indicates that the advisor's texts are more concentrated around a few topics, reflecting a focused expertise. Conversely, if an advisor's average distance is greater than or equal to the average distance in the corpus, they are classified as a generalist, indicating a broader range of topics.

For the distance calculation, we employ cosine similarity and Euclidean distance. The Euclidian distance between two topic vectors **a** and **b** is defined as:

$$d(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

Cosine similarity between two topic vectors **a** and **b** is defined as:

$$\text{cosine\_similarity}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$$

## 3.3 Matching the results with career data

Finally, to answer the second research question, the distributions obtained in the previous steps must be compared with the available career data in dataset I. Hypothesis three should be divided into two parts: one where the distribution on the political-neutral axis is compared with information on party political involvement in the career data of the cabinet employees and another where the distribution on the generalist-specialist axis is compared with the information on the occupations of the cabinet employees during their careers.

| **Index** | **Stands for** |
|---|---|
| pol_pos | # political positions |
| pol_dur | duration political positions |
| pol_idx | pol_dur / total duration |
| pub_pos | # public sector positions |
| pub_dur | duration public positions |
| pub_idx | pub_dur / total duration |

Table 2: Career position based indices

For the partisan political data in the career data, constructing a metric based on career positions that situates a given adviser on the political-neutral axis is relatively straightforward. For each career position of an individual, codes are added by a human annotator that indicate whether or not the position is in the political sphere. To be precise, we have five codes: politics in government (adviser or minister positions in the Flemish government), politics in federal government, politics in central party office, politics in parliament, politics on the local level (partly based on Katz and Mair (1990). Based on these codes, a number of metrics can be established for each individual cabinet member. In Table 2, we present these. To arrive at a comparison with the previously obtained distribution on the political-neutral axis based on the texts related to the advisor, we correlate each of the metrics with average entropy determined for a given advisor. We use Spearman's rho correlation for this, because we expect the relationship to be nonlinear but nonetheless monotonically increasing. We expect this correlation to be statistically significant with $p < 0.05$. For redundancy, we perform the same tests on similar indices based on the experience in the public sector.

> $H_{3a}$: *We expect the indices based on the career codes in dataset I to correlate significantly with the entropy distribution on the individual level in dataset II.*

For the generalist-specialist career data, we adopt an approach that closely mirrors the one used for analyzing documents on this axis. Each position in dataset I includes a job title, workplace, and, optionally, a job description. We apply the same BERTopic modeling technique to this career data as we did to the documents. Specifically, we fit a BERTopic model to extract topic probability vectors, then calculate the individual distance versus the average distance between all vectors, using both Euclidean distance and cosine distance. This results in metrics for both the documents and the job descriptions. We label these metrics 'above_avg_cosine' and 'above_avg_euclid'.

Ultimately, we derive labels indicating whether a particular individual has a higher than average inter-topic probability vector distance. We name these variables 'above_avg_cosine' and 'above_avg_euclid' for both the dataset I topic representation and dataset II topic representations. We then test the correlation of these labels using the Phi Coefficient. Additionally, we use Spearman's rho to test the correlation between the average distances per individual. This dual approach allows us to compare the distributions derived from documents and career data systematically.

> $H_{3b}$: *We expect the labels based on the career description in dataset I to correlate significantly with the labels of the documents in dataset II.*

# 4 Results

Regarding the political-neutral axis, the first step was to fine-tune a BERT model. Our fine-tuned model achieved moderate results, with an accuracy of 0.5363 on the test set (with four classes, accuracy by chance would be 0.25). Given our aim was not to use the predictions directly but to assess the overall distribution, we considered this performance sufficient.

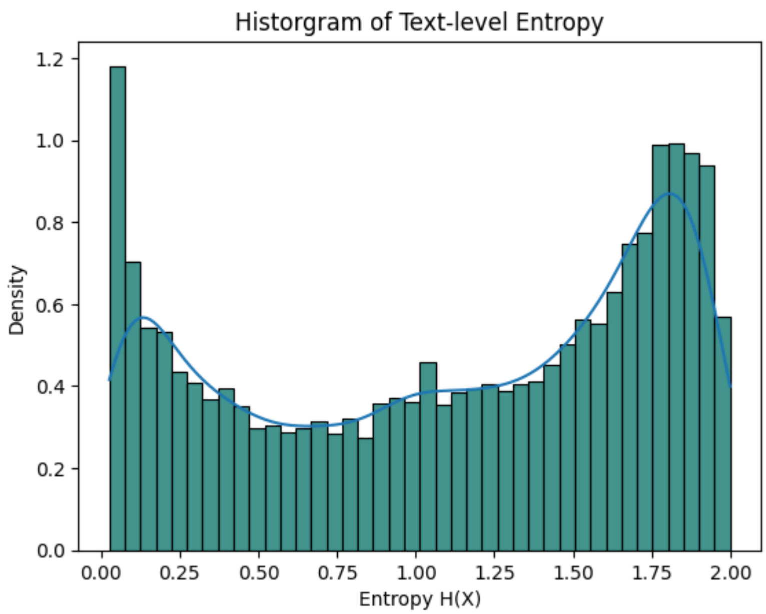


Figure 2: Histogram of text-level entropy

We then used the model to make predictions on our set of texts linked to cabinet members. At the text level, the mean entropy was 1.1115, the median entropy was 1.2250, and the variance was 0.4125. Figure 1 shows the distribution of entropy at the text level. The distribution is clearly bimodal, with peaks near

0 (low entropy, high certainty about partisanship) and near 2 (high entropy, high uncertainty about partisanship). Thus, at the text level, our hypothesis is confirmed. The kernel density estimate also shows two distinct peaks at the extremes of the entropy axis.

At the individual level, the median entropy presents a less clear bimodal distribution. The mean entropy was 1.2308, the median was 1.3512, and the variance was 0.2317. This indicates more variance within individual cabinet employees than between them. The histogram of median entropies at the individual level shows a less distinct bimodal distribution compared to the text level. There is a peak towards the higher end of the entropy spectrum (indicating neutrality), but the density in the lower half is also notable.

A Gaussian mixture model was fitted to this distribution, with the second component (mean=1.5229, variance=0.0564) fitting the data well, suggesting a "neutrality" component. The first component (mean=0.7792, variance=0.1641) had larger variance and a less distinct fit, indicating no strong patterns visually. Nevertheless, it is evident that at the individual level, a neutrality component is well-formed and can serve as a discriminator between politically active and politically neutral individuals. Ultimately, we found that 39.27% of individuals in our dataset fall under the 'political' component, while 60.73% fall under the 'neutral' component.

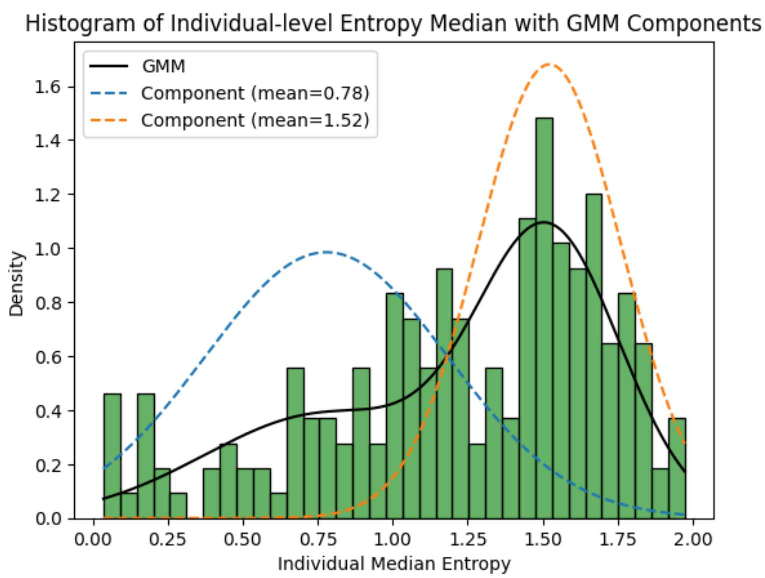


Figure 3: Histogram individual-level entropy

As for the specialist generalist axis, the first task consisted of fitting a BERTopic model to the documents related to cabinet officers. BERTopic identified 208 topics. From the topic model we distilled topic probability vectors.

For graph X, we reduced these to 2 dimensions using t-distributed stochastic neighbor embedding (t-SNE). What is striking about this graph is that although (as expected) the topic probability vectors cluster well there is one topic that 1) does not cluster as well and 2) "absorbs" a large portion of the documents. Indeed, topic -1 absorbs 9151 of the 18887 documents. This is a problem to be solved after the IPPA workshop.

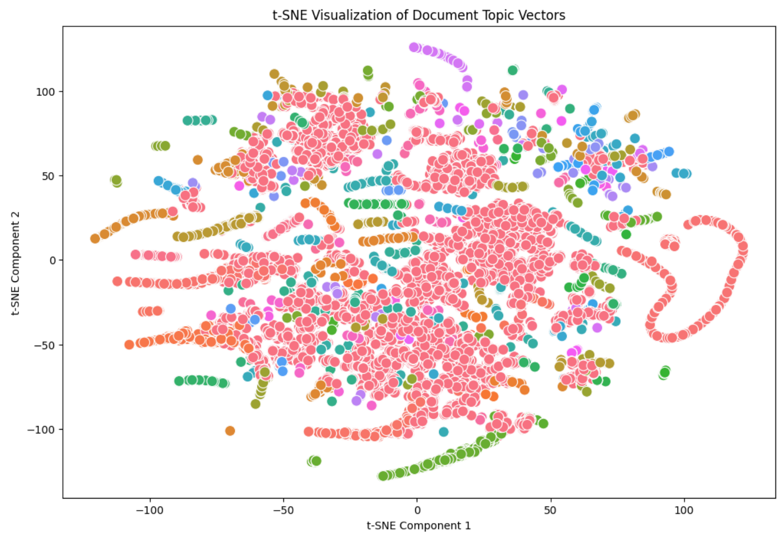


Figure 4: t-SNE visualization of Document Topic Vectors

Then, based on both euclidean distance and cosine distance, the average distance between all topic probability vectors was calculated and then within different individuals. In this way, we arrive at two distributions: 0.3247 of the individuals had a greater average distance between topic probability vectors than the average distance between all topic probability vectors when the distance metric was euclidian distance. For cosine distance, that number was 0.2177. These are relatively small values: this would mean that between 22 and 32% of the cabinet member is a generalist based on the documents related to this cabinet member. However, it can be confirmed that, again, there are two groups within cabinet employees, with a clear distinction between specialists and generalists. At the time of the IPPA workshop, these results do still need to be taken with a grain of salt, given the possible error in the topic model.

The testing of career data proceeded in two parts, as previously indicated. First, we focused on the relationship between career data and the associated documents. Our dataset contains 2,574 coded positions. On average, the proportion of political-related jobs over total jobs for an individual is 0.4068, with an average of 2.7595 positions per individual, covering approximately 11.61 years of their career. For public sector positions, the average proportion is 0.2115, with an average of 1.3372 positions per individual, spanning about 7.45 years of their career.

We examined the relationship between the political index (pol_idx) and the average entropy for a cabinet advisor. Using Spearman's Rho, we found a significant ($p < 0.05$) correlation of -0.1516. For the public sector index (pub_idx), the correlation was 0.1329 ($p = 0.0640$). This indicates a weak but significant correlation between the documents related to the cabinet member and their career data. Specifically, the higher the pol_idx, the lower the mean entropy of the political-neutral classifier. In other words, the more political-related career positions a cabinet member has, the more politically-oriented their associated documents tend to be.

Finally, career data was also tested on the specialist-generalist axis. All descriptions of the positions were fit into BERTopic, which recognized 66 topics in the 2574 positions. As can be seen in graph X, there is a better distribution of topic probability vectors here than the one for the documents,with good clustering and a more even distribution (although there is still one topic that takes up 770 of the 2574 positions)

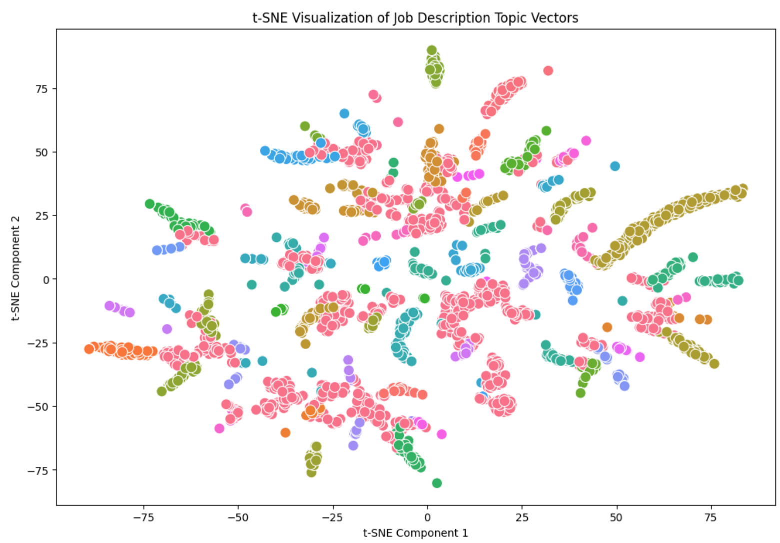


Figure 5: t-SNE visualization of Position Description Topic Vectors

In the career data, we observe a slightly different distribution along the specialist-generalist axis. Calculated using Euclidean distance, we find that 0.4908 of individuals have a greater average distance between their topic probability vectors than the average distance between all topic probability vectors. Using cosine distance, this proportion is 0.3727. Therefore, based on career data, we estimate that the proportion of generalists ranges between 37% and 49%.

This distribution differs from the one derived from the document data, and there is no significant correlation between the two distributions. For the Euclidean distance indicators between document and career data, we obtain a Phi coefficient of -0.0187 (Pearson correlation $p = 0.7589$). For the cosine distance indicators, the Phi coefficient is 0.0742 (Pearson correlation $p = 0.2236$). Similarly, experiments using Spearman's Rho to compare mean distances within individuals resulted in a Spearman's Rho of -0.0353 ($p = 0.5625$) for Euclidean distance metrics and 0.0736 ($p = 0.2274$) for cosine distance metrics.

# 5 Discussion

The results, though based on about half of the final dataset, show a clear trend. The developed methodology seems effective for the political-neutral axis but is less convincing for the specialist-generalist axis. Therefore, we discuss the results using this distinction.

The approach of modeling the political-neutral axis using classification entropy appears to work well. As hypothesis H1 suggests, we indeed find a bimodal distribution of classification entropy across texts. When these texts are distributed across advisors, the bimodality becomes less distinct, which is expected: "neutral" texts can be written about politically active figures. A neutral person is expected never to be associated with a political party, but once some documents show a clear partisan political color, the neutral label no longer applies. The median entropy scores for the neutral Gaussian distribution are very high, indicating minimal partisan content in these documents. For the political distribution, the scores are more widely distributed across the spectrum, reflecting the mix of partisan and neutral texts.

The distribution of approximately 40% politically active and 60% politically neutral advisors slightly contradicts expectations based on previous literature on Belgium (such as Gouglas et al.) but can be explained in two ways. First, as discussed earlier, someone is declared politically active only if there is reasonable evidence of political activity. However, this does not mean that all individuals labeled as neutral are genuinely so; it simply means no evidence to the contrary was found. Second, our focus is on the Flemish government rather than the federal Belgian government and in a different era than that of Gouglas et al. Literature suggests that the party's influence, including on the profile of cabinet officials, is declining, so these figures should not be surprising.

Regarding the relationship between the documents and career data on the political-neutral axis, the findings are significant and confirm hypothesis 3a. There is a significant relationship between the partisanship of documents and career data, in the expected directions: for the pol_idx, the relationship is negative (more political career positions correspond to less ambiguity about the partisan nature of texts), and for the pub_idx, it is positive. This indicates a correlation between an advisor's actions, behavior, and career path. The implications for further research are interesting: both theoretically and empirically, more integration can be sought between the fields studying the role of MAs and the career paths of advisors.

Second, we address the results related to the specialist-generalist axis, which present more ambiguities. Firstly, we question the relevance of the topic probability vectors to the documents. Given that nearly 50% of the texts were assigned to one of the 208 topics, the tuning of BERTopic to our specific case must be critically assessed. Consequently, the resulting distribution should be viewed cautiously: it identifies two groups, with 22% to 32% of cabinet members classified as generalists. Theoretically, we would expect this proportion to be much higher. However, H2 is not invalidated, as two distinct groups were identified based on the documents. Considering that one topic was assigned to half of the texts, we anticipate a slightly higher proportion in later phases (post-IPPA): if documents are assigned to more varied topics, there will be greater variation, and thus more generalists. The main challenge here is to find optimal hyperparameters for BERTopic.

Given these issues, interpreting the correlation with career data is also challenging. The non-significant correlation could be due to the problems with BERTopic or a genuine lack of correlation. If the latter, we must consider whether this is due to our methodology (perhaps the documents do not capture whether an individual is a specialist or generalist) or the theory itself. It is possible that there is no relationship between a person's career path and their behavior as a specialist or generalist. Before drawing definitive conclusions, further tuning of BERTopic is necessary.

Overall, this study has produced significant results for at least one axis and has paved the way for further research with large datasets to explore the relationship between MA's roles and their careers.

# 6 Conclusion

This study aimed to address gaps in the research on MAs by employing a novel methodology that leverages Natural Language Processing (NLP) techniques to analyze extensive textual data. By doing so, we sought to validate Bernadette Connaughton's typology of MAs' roles and examine the relationship between these roles and the advisors' career trajectories.

Our results demonstrate that the methodology is effective in differentiating between politically active and neutral advisors, confirming the presence of distinct categories as proposed by Connaughton. This success underscores the utility of NLP tools in revealing underlying political orientations within large datasets of textual information.

The analysis of the specialist-generalist axis was less definitive, suggesting areas for improvement in the application of BERTopic for topic modeling. The discrepancies observed highlight the need for further refinement of the model to better capture the nuances of the specialist-generalist distinction.

The exploration of career data and its correlation with advisor roles provided partial support for our hypotheses. The relationship between political career paths and document partisanship was significant, indicating that career backgrounds do influence the political nature of an advisor's role. However, the expected correlation between career paths and the specialist-generalist classification was not observed, prompting further investigation into both the methodology and underlying theory.

Overall, this study demonstrates the potential of integrating NLP techniques with traditional research approaches to enhance our understanding of MA's roles. The findings encourage continued development and application of computational methods in political science and public administration research. This approach promises to yield richer, more nuanced insights into the dynamics of advisory roles and their evolution within government structures, thereby contributing to the academic discourse on MAs. By conducting these analyses at a more detailed level, new research opportunities can emerge. Comparisons over time, across countries, and between governance systems can be relatively easily carried out, provided the necessary data is available.

## References


Alam, T., Le Mazier, J., & Sawicki, F. (2019). Quarante ans d'enquêtes prosopographiques sur les membres de cabinets ministériels: quelles problématiques pour quels résultats? *Revue francaise d'administration publique*, (4), 751–769.

Askim, J., Karlsen, R., & Kolltveit, K. (2022). Populists in Government: Normal or Exceptional? *Government and Opposition*, *57*(4). https://doi.org/10.1017/gov.2021.30

Bach, T., & Hustedt, T. (2023). Policy-making in the executive triangle: a comparative perspective on ministers, advisers, and civil servants. *Handbook on Ministerial and Political Advisers*, 338.

Blach-Ørsten, M., Mayerhöffer, E., & Willig, I. (2020). From government office to private PR: Career patterns of special ministerial advisers and the privatization of politics. *The International Journal of Press/Politics*, *25*(2), 301–319.

Brans, M., de Visscher, C., Gouglas, A., & Jaspers, S. (2017). Political control and bureaucratic expertise: Policy analysis by ministerial cabinet members. In *Policy analysis in belgium* (pp. 57–78). Policy Press.

Brans, M., Meert Arthur, Moens Pieter, & Squevin Pierre. (2023). Studying ministerial advisers' careers and profiles: the prosopographic method. In R. Shaw (Ed.), *Handbook on ministerial and political advisers* (pp. 155–172). Edward Elgar Publishing.

Charle, C. (1984). René Rémond, Aline Coutrot, Isabel Boussard et al., Quarante ans de cabinets ministériels, Paris, Presses de la Fondation Nationale des Sciences Politiques, 1982, 276 p. *Annales. Histoire, Sciences Sociales*, *39*(3). https://doi.org/10.1017/s0395264900080306

Connaughton, B. (2010). 'Glorified gofers, policy experts or good generalists': A classification of the roles of the Irish ministerial adviser. *Irish Political Studies*, *25*(3), 347–369.

Connaughton, B. (2015). Navigating the borderlines of politics and administration: Reflections on the role of ministerial advisers. *International Journal of Public Administration*, *38*(1), 37–45.

Delpu, P.-M. (2015). La prosopographie, une ressource pour l'histoire sociale. *Hypothèses*, *18*(1). https://doi.org/10.3917/hyp.141.0263

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*.

Di Mascio, F., & Natalini, A. (2013). Analysing the role of ministerial cabinets in Italy: Legacy and temporality in the study of administrative reforms. *International Review of Administrative Sciences*, *79*(2). https://doi.org/10.1177/0020852313477972

Eymeri-Douzans, J.-M., Bioy, X., & Mouton, S. (2015). Le règne des entourages: Cabinets et conseillers de l'executif. *(No Title)*.

Gouglas, A., Brans, M., & Jaspers, S. (2015). Political Advisers and Policy Making in Ministerial Cabinet Systems: the case of Belgium, Greece & the European Commission. *International Conference on Public Policy, Date: 2015/07/01-2015/07/04, Location: Milan*.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Hustedt, T., & Salomonsen, H. H. (2017). Political control of coordination? The roles of ministerial advisers in government coordination in Denmark and Sweden. *Public Administration*, *95*(2), 393–406.

Hustedt, T., & Salomonsen, H. H. (2014). Ensuring political responsiveness: politicization mechanisms in ministerial bureaucracies. *http://dx.doi.org.kuleuven.e-bronnen.be/10.1177/0020852314533449*, *80*(4), 746–765. https://doi.org/10.1177/0020852314533449

Katz, R. S., & Mair, P. (1990). *Three faces of party organization: Adaptation and change*. Department of Government, Victoria University of Manchester.

Lemercier, C., & Picard, E. (2011). Quelle approche prosopographique? *Hal-Shs*, *18*.

Mathiot, P., & Sawicki, F. (1999). LES MEMBRES DES CABINETS MINISTÉRIELS SOCIALISTES EN FRANCE (1981-1993) RECRUTEMENT ET RECONVERSION: PREMIÈRE PARTIE CARACTÉRISTIQUES SOCIALES ET FILIÈRES DE RECRUTEMENT. *Revue française de science politique*, 3–29.

Pickering, H., Bellens, T., & Brans, M. (2023). The prime minister's chief of staff: Comparing profiles and trends in Westminster democracies, 1990–2021. *Governance*.

Rouban, L. (2012). L'état à l'épreuve du libéralisme: Les entourages du pouvoir exécutif de 1974 à 2012. *Revue Francaise d'Administration Publique*, *142*(2). https://doi.org/10.3917/rfap.142.0467

Shaw, R., & Eichbaum, C. (2018). *Ministers, minders and mandarins: An international study of relationships at the executive summit of parliamentary democracies*. Edward Elgar Publishing.

Taflaga, M., & Kerby, M. (2020). Who Does What Work in a Ministerial Office: Politically Appointed Staff and the Descriptive Representation of Women in Australian Political Offices, 1979–2010. *Political Studies*, *68*(2). https://doi.org/10.1177/0032321719853459

Walgrave, S., Caals, T., Suetens, M., & De Swert, K. (2004). *Ministerial cabinets and partitocracy: A career pattern study of ministerial cabinet members in Belgium*. Citeseer.