'The Ideal Justice': Method Appendix

Johan Lindholm[1], Mattias Derlén[2], and Daniel Naurin[3]

[1]Professor of Law at Umeå University. email: johan.lindholm@umu.se
[2]Professor of Law at Umeå University. email: mattias.derlen@umu.se
[3]Professor of Political Science at Oslo of University and Director of ARENA Centre for European Studies. Email: daniel.naurin@arena.uio.no

October 26, 2021

1 Introduction

This is an appendix to our paper entitled "The Ideal Justice: Who Are Selected to Serve and What Does It Say About Swedish High Courts?" (henceforth 'the paper').[1] The purpose of this appendix is to elaborate on the methods and data used in the paper in order to enhance transparency and to ensure the replicability of our study. Replication data for the study along with the R-script used to reproduce the analysis and to generate the figures in the paper are available at *Harvard Dataverse*[2] and *GitHub*[3].

The appendix particularly elaborates on Sequence Analysis (SA) and how it was used in the main paper to compare the professional background of all Justices who have served on the Swedish Supreme Court (SSC) or the Swedish Supreme Administrative Court (SSAC). As explained in the paper, the vast majority of SSC and SSAC Justices have professional experiences within multiple legal career tracks prior to being appointed Justices. On the one hand each individual Justice's way from law school to serving on one of Sweden's highest courts is unique, but on the other hand some are more similar to each other than other. As explained below, SA helps us bridge this gap to (i) measure the difference in professional background between two Justices and (ii) cluster Justices based on background similarity.

2 Why Sequence Analysis?

In the paper we argue theoretically for distinguishing between six legal career tracks and why and how we believe that a Justice's pre-appointment professional background may be relevant for his or her behavior on a high courts. Whereas

[1]Forthcoming in XX.
[2]`https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/JAD8CI`
[3]`https://github.com/jojolindholm/Ideal-Justice-Replication`

each Justice's professional background is unique some have more similar backgrounds and being able to group together Justices that have 'sufficiently similar' professional backgrounds have great analytical value. However, how does one determine in a non-arbitrary manner which Justices have sufficiently similar backgrounds to belong to the same type (or conversely sufficiently dissimilar backgrounds to belong to different types)?

A methodological concern in this regard is that the vast majority of all SSC and SSAC Justices have, as presented in the paper, professional experiences within multiple legal career tracks prior to being appointed Justices. For example, most Justices have some judicial experience from the lower courts prior to joining the high courts. Should these all be treated as one 'type' of Justice, even though only some of these have exclusively judicial experience and other have e.g. served extensively as politicians or ministry lawyers?

In the paper we employ Sequence Analysis to overcome this methodological problem. Adapted from the biological sciences where it was used to code DNA molecules, SA was subsequently adopted and developed by life course scholars. Compared to analysis based on events as population covariates, SA takes a holistic approach to events, treating the sequence of events as a single conceptual unit (Billari 2005, pp. 273–275). The main value added by a sequence approach is that it allows us to move "from units to context, from attributes to connections, from causes to events" (Abbott 1995, p. 93). Careers provide an excellent example of a phenomenon that can benefits from being studied using the type of holistic approach offered by SA as "we expect both a fair amount of pattern and a fair amount of fluctuation" (Abbott 1990, p. 140). Thus, by using SA and applying to to Justices' professional backgrounds we can capture patterns over entire career paths that risk being missed using an event-based approach. For this we use concretely the tools included in the package `TraMineR` for the programming language `R` (Gabadinho et al. 2011).

3 The Data

Sequence Analysis is based on the ability to represent a time-related phenomenon – in our case the professional path that leads a lawyer to end up on a Swedish high court – as sequences of fixed-length where each unit of observation represents a period of time. "An individual sequence in this perspective is the succession of the observed states for one unit of observation over a given time period." (Gauthier, Bühlmann, and Blanchard 2014, p. 5) To achieve this we must reformat the professional background of Justices in two ways.

First, absolute time cannot be used for our purposes as (i) Justices' were active at different points in time over the almost 250 years studied and (ii) the length of their pre-appointment careers varied in absolute time. We therefore instead use a symbolic time scale where each unit represents 1 percent of a Justices' pre-appointment professional background. However, it is worth pointing out that we lack accurate time information about when in time these transitions took place, especially for early in the Court's history, and consequently exactly how long the judges held each position. In creating the sequences we therefore

assume that the judges held each of his or her prior positions for an equal length of time. While this is not accurate, we treat all judges in the same way and are primarily interested in the shape of career paths and transitions between states, which is why this should have limited impact on our findings.

Second, we must reduce the near-infinite variance in professional background to a finite and manageable number of possible 'states' that a unit can have. In SA the possible values that states can have is referred to as 'an alphabet'. As explained in the paper, we have manually identified 27 types of positions commonly held by Swedish Justices and categorized these positions into what we consider to be six major, pre-appointment legal career tracks: (i) academia, (ii) judicial, (iii) ministerial, (iv) political, (v) other public service, and (vi) private practice. We use these six career tracks in a our alphabet as the possible states of each unit of a sequence.

Using this approach, we are able to create single-dimension, equal-length sequences that represent each Justice's professional experiences over time. This is represented in a matrix with the dimensions of 453 (all Justices) by 100 (symbolic time) with six possible states (career tracks).[4]

4 Distances

The core of Sequence Analysis consists of measuring how similar or different two sequence are, which in SA referred to as their pairwise 'distance'. In our paper we employ for this Optimal Matching (OM) (Kruskal 1983), which is the standard method used for this in SA (see e.g. Ritschard and Studer 2018, p. 2). OM determines how dissimilar two sequences are by calculating the 'cost' of transforming one sequence into another and the transformation is achieved by conducting one of three operations: substituting, inserting or deleting states (Abbott and Tsay 2000, pp. 5–6). The more operations that are required for transformation, the more dissimilar the two sequences are (Gauthier, Bühlmann, and Blanchard 2014).

The transformation cost can be set in multiple ways, including assigning the same cost to all operations and to the transformation from all states to all states or to assign different transformation costs based on a theoretical understanding. We employ an approach of varying transformation costs between states and to assign transformation costs based on how common they are in the data. In other words, we depart from the idea that what transitions occur in reality reflects well what transformation costs should be, i.e. that transformations from one state to another state that are empirically rare should be more costly than transformations that are empirically common.

We achieve this by first calculating the probability of a transition from each state to each other state, where the following unit having the same state as the previous one is the most probable. State transition costs were subsequently set where no state transition has a cost of 0 and transitions between states have some

[4] In the replication data this matrix is stored in the files `judges_seq.csv` and `judges_seq.RData`.

value lower than 2 depending on probability (Gabadinho et al. 2011, pp. 16–17). The transition probabilities and costs respectively are presented on page 4.

Transition probabilities

From state	To State *Academic*	*Judicial*	*Ministerial*	*Political*	*Private practice*	*Public service*
Academic	0.993	0.00286	0.00200	0.000573	0.000859	0.000859
Judicial	0.000214	0.985	0.00922	0.00141	0.000342	0.00406
Ministerial	0.000197	0.0119	0.982	0.00148	0.000295	0.00443
Political	0	0.00459	0.00328	0.986	0	0.00591
Private practice	0.00313	0.00547	0.00235	0.000782	0.987	0.000782
Academic	0.000604	0.0103	0.00664	0.00121	0.000201	0.981

Transition costs

From state	To State *Academic*	*Judicial*	*Ministerial*	*Political*	*Private practice*	*Public service*
Academic	0.000000	1.996923	1.997799	1.999427	1.996013	1.998537
Judicial	1.996923	0.000000	1.978875	1.993998	1.994185	1.985676
Ministerial	1.997799	1.978875	0.000000	1.995244	1.997359	1.988930
Political	1.999427	1.993998	1.995244	0.000000	1.999218	1.992887
Private practice	1.996013	1.994185	1.997359	1.999218	0.000000	1.999017
Academic	1.998537	1.985676	1.988930	1.992887	1.999017	0.000000

Based on these transition costs, the cost of transforming each sequence to each of the other sequences was calculated. As the data includes 453 unique sequences (Justices) this resulted in a matrix with the dimensions 453 by 453 where each value represents the distance between or dissimilarity of two sequences (Justices). This varies between 0 for identical sequence pairs to nearly 200 for the most dissimilar pairs.[5] Figure 1 presents how these pairwise distances are distributed in the data.

5 Clustering

In the paper, we use Sequence Analysis to cluster the Justices. "Cluster analysis is the art of finding groups in data." (Kaufman and Rousseeuw 1995, p. 1) The essence of the clustering problem, generally expressed, is to identify individuals that are similar to each other and different from the rest of the set and on this basis divide them into mutually exclusive groups. In our paper, the relative level of similarity concerns the Justices' professional backgrounds. In this section we seek to elaborate on and validate that process. We used agglomerative hierarchical clustering (Ward 1963), as incorporated in the package `TraMineR` (v. 2.2.2) for `R`, to produce the clustering used in the paper. As the name suggests, this is an approach where clusters are merged hierarchically (Kaufman and Rousseeuw 1995, p. 44).

This is the traditional approach to clustering in SA, but it is not without criticism (Aisenbrey and Fasang 2009, pp. 431–432). One particular concern is that the approach used does not provide the number of clusters (k) which must be selected. It should initially be pointed out that there is no definitive method for deciding the number of cluster. There are a number of clustering quality measurements that can be used as assistance for this purpose, but these did not provide a clear ideal number of cluster for our data (Figure 2).

Ultimately for this study we selected to divide the data into six clusters. The resulting model and the member sequences of each to the six clusters, ordered by end state, is presented in Figure 3. As explained in the article, we find that $k = 6$ provides a clustering that makes theoretical sense and can be usefully applied to explore the research questions. These clusters can be thought of Justice archetypes, i.e. types of Supreme Court Justices who have professional backgrounds that are similar to each other and that distinguish them from Justices belonging to other communities: (1) *the Ministry Judge*, (2) *the Professor*, (3) *the Public Servant Judge*, (4) *the Practising Lawyer*, (5) *the Career Judge*, and (6) *the Politician Judge*.

In Figure 4 we plot the Justice in a two-dimensional space. We applied multi-dimensional scaling to all Justices using each pairs' distance calculated using Optimal Matching (MO) and identified by clusters. This helps illustrate how internally coherent and distinct from each other the clusters are. It reveals that some clusters are clearly distinct from each other, but also that there is

[5] The maximum distance is 200 as there are 100 states that each at most can cost 2 to transform.

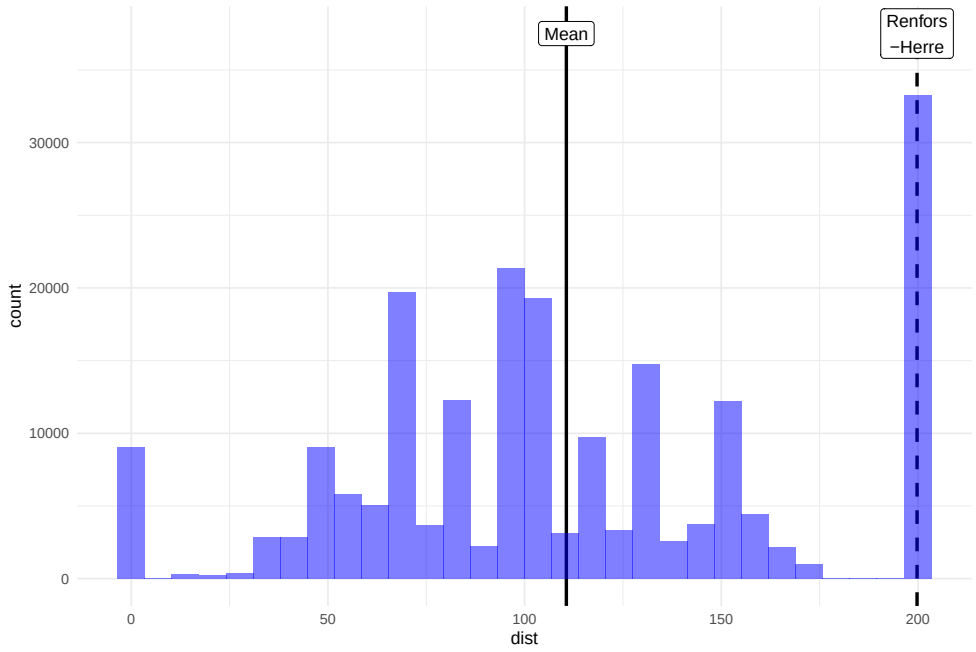

Figure 1: The figures shows how the pairwise distances of all sequences are distributed in the data.

some overlap.

Ultimately, we decided on the number clusters by qualitatively comparing the clustering produced using the range of four to eight clusters. These are presented below in Figures 5–8. In this context, we can compare what changes in the clustering as we deviate from the selected model (k = 6).

The main difference between the chosen model and if remove one cluster (k = 5) is that the *the Politician Judge* and *the Ministerial Judge* merges. In the paper we explain theoretically why we believe that a professional background in politics may shape judicial behavior and how this is different than other types of backgrounds, including ministerial work. The six-cluster model is superior to the five-cluster model in that it incorporates this distinction. The four-cluster model (k =4) is even less useful that it in addition to the merging included in the five-cluster model also merges *the Practising Lawyer* and *the Public Servant Judge*. This is simply nonsensical and strongly indicates that a greater number of clusters are appropriate.

Having more than six clusters is also not valuable. If we add one cluster (k = 7) *the Ministerial Judge* is split into two clusters: one consisting of those who served in the lower courts both before and after serving in the ministries and one consisting of the rest. While this is a thought-worthy distinction, we fail to see any theoretical explanation for how this distinction could be valuable for the purpose of exploring the issues studied in the paper. The eight-cluster model (k = 8) suffers from this feature as well and additionally breaks out from *the Public Servant Judge* Justices who exclusively served in other public office prior

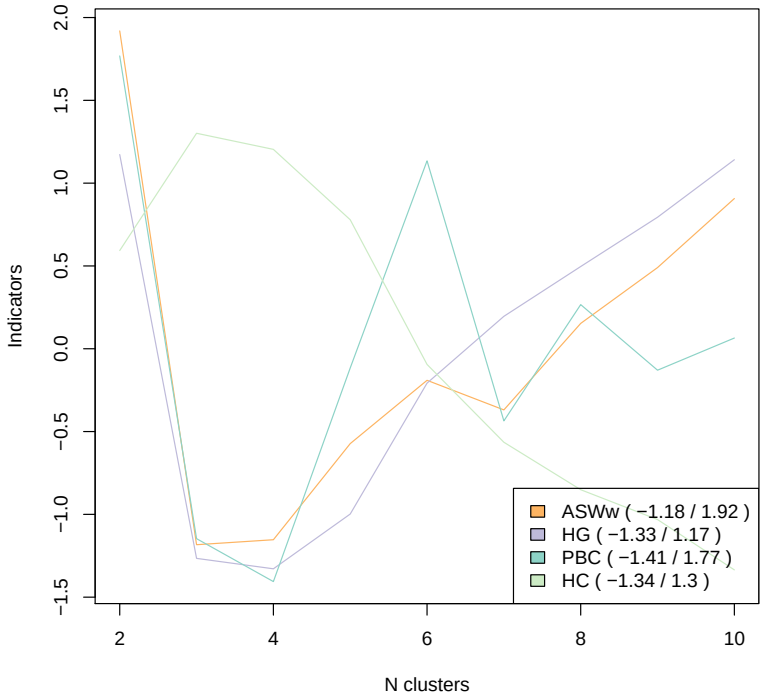

Figure 2: The figure displays four major clustering quality statistics for models using 2–10 clusters: Average Silhouette Width (weighted) (ASWw), Hubert's Gamma (HG), Point Biserial Correlation (PBC), and Hubert's C coefficient (HC).

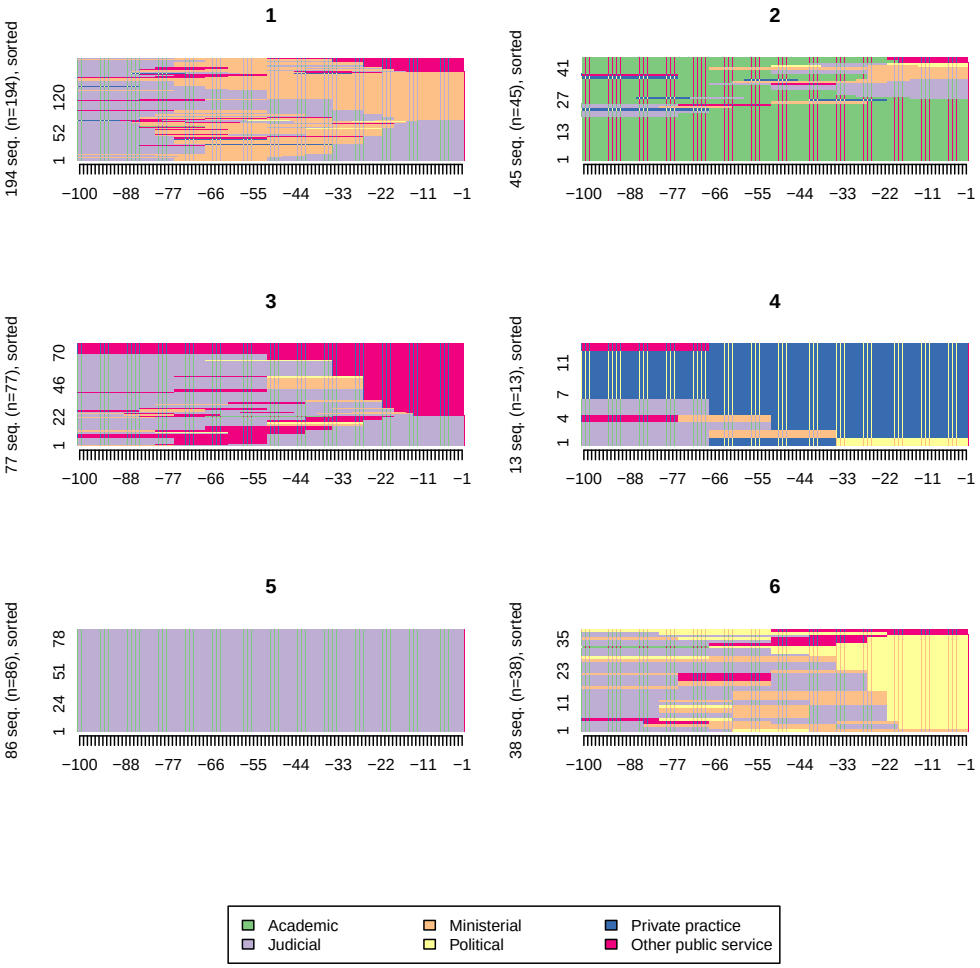

Figure 3: The model used in the paper, clustering of sequences into six clusters (k = 6).

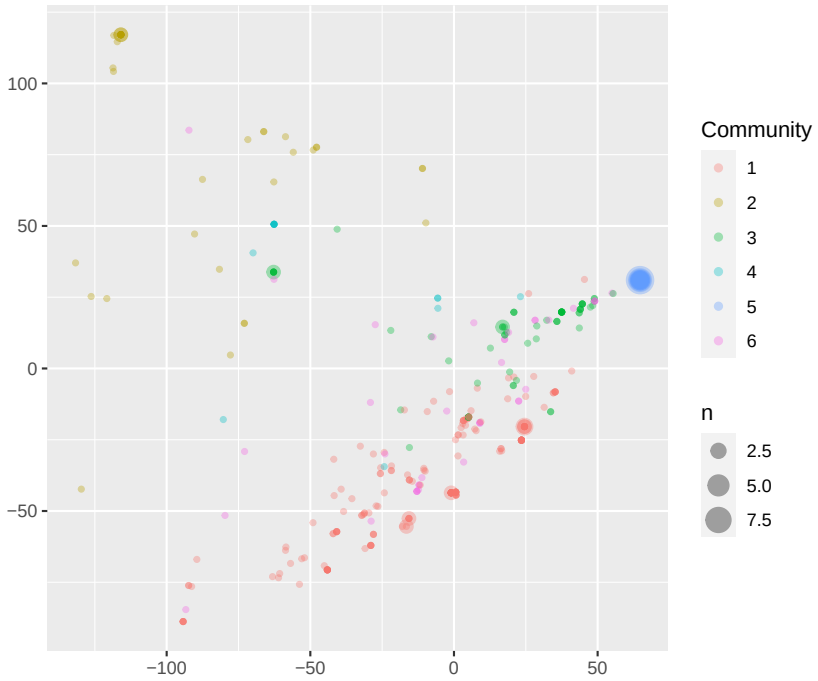

Figure 4: Multi-dimensional scaling applied to all Justices using each pair's distance calculated using Optimal Matching (MO). Size represents number of Justices where these have the same position. Colors illustrate clustering under the six-cluster model used in the main paper.

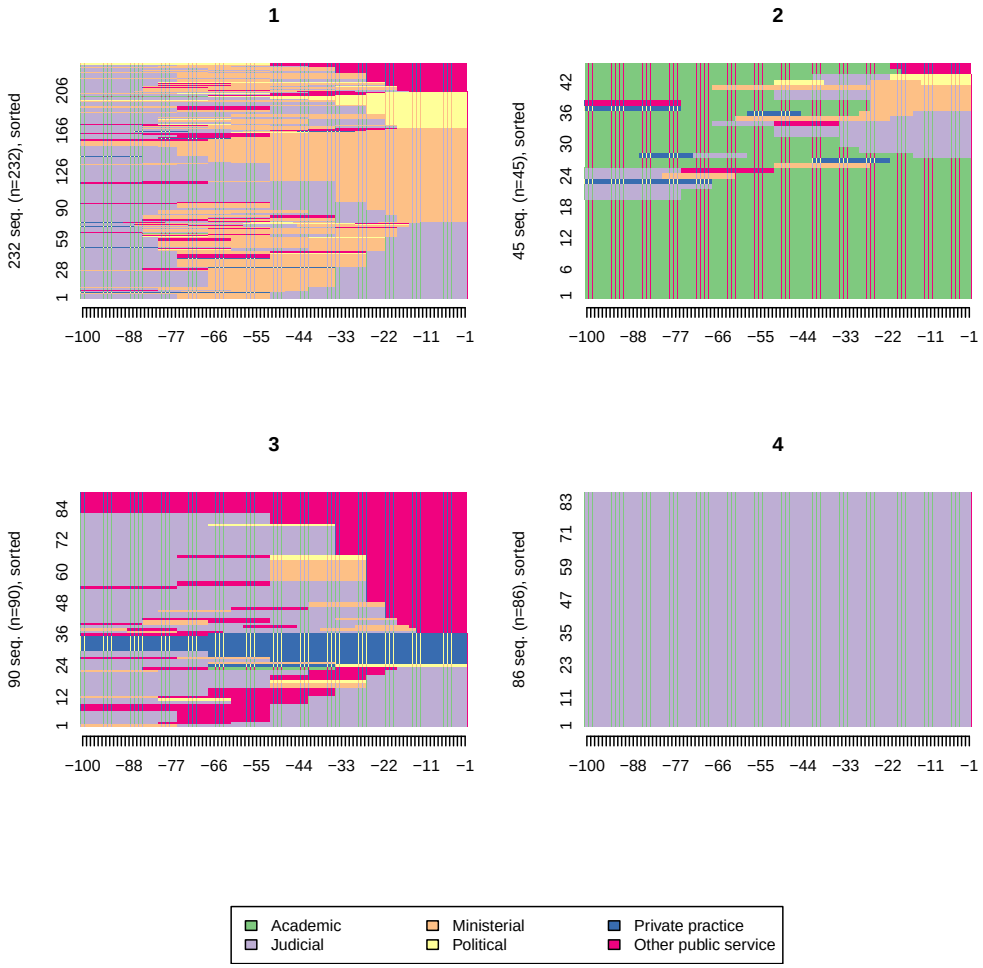

Figure 5: k = 4.

to joining the high courts. This is both a very small group (n = 8), but also we again find no theoretically relevant value of this distinction.

6 Diversity over Time

As described in Section 4 above, Optimal Matching provides us with a measurement, in SA referred to as a 'distance', that captures how dissimilar two Justices' professional backgrounds are on a scale between 0 (identical) and 200 (maximum dissimilarity). Each such value describes the dissimilarity of a pair of Justices but it obviously can also be used to measure the relative degree of background dissimilarity of a group of Justices. In the paper we used these values to measure background diversity over time on the SSC and SSAC respectively.

Our approach in this regard was straightforward. First, for each year they have been in existence we identified which Justices served on the SSC and SSAC.

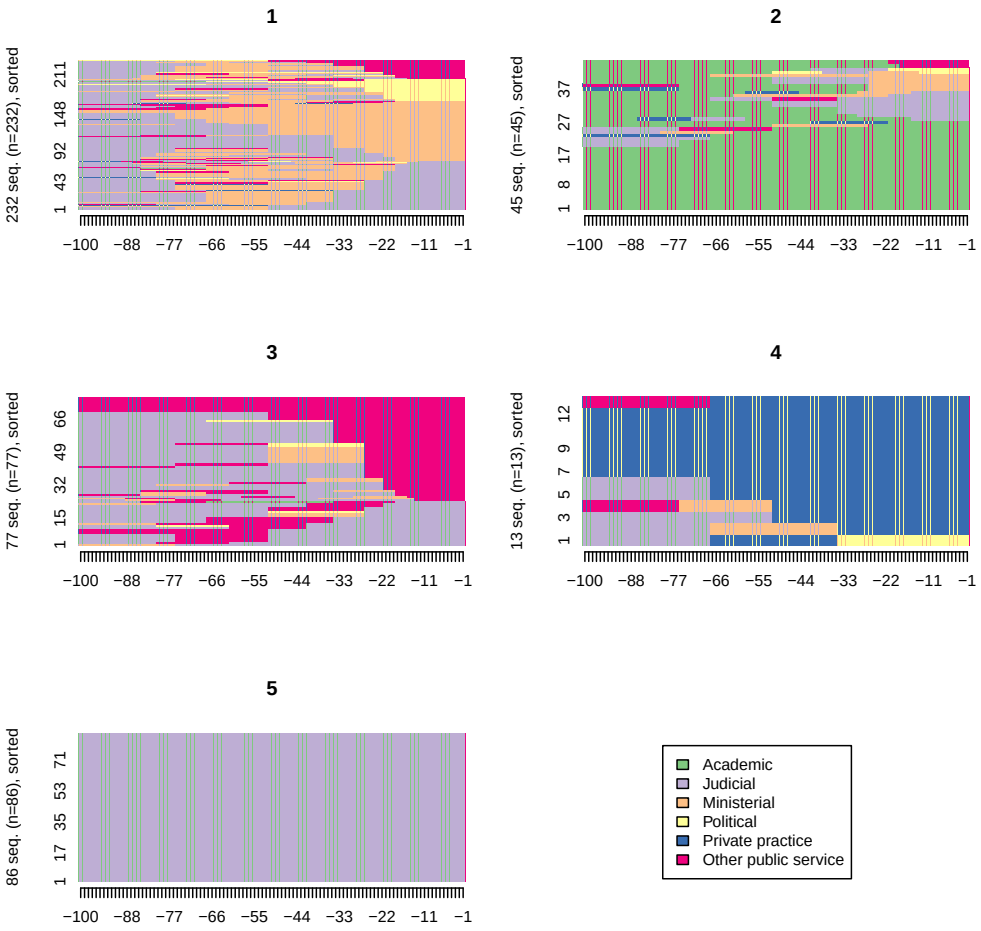
1
2
3
4
5
232 seq. (n=232), sorted
45 seq. (n=45), sorted
77 seq. (n=77), sorted
13 seq. (n=13), sorted
86 seq. (n=86), sorted
−100 −88 −77 −66 −55 −44 −33 −22 −11 −1
Academic
Judicial
Ministerial
Political
Private practice
Other public service

Figure 6: k = 5.

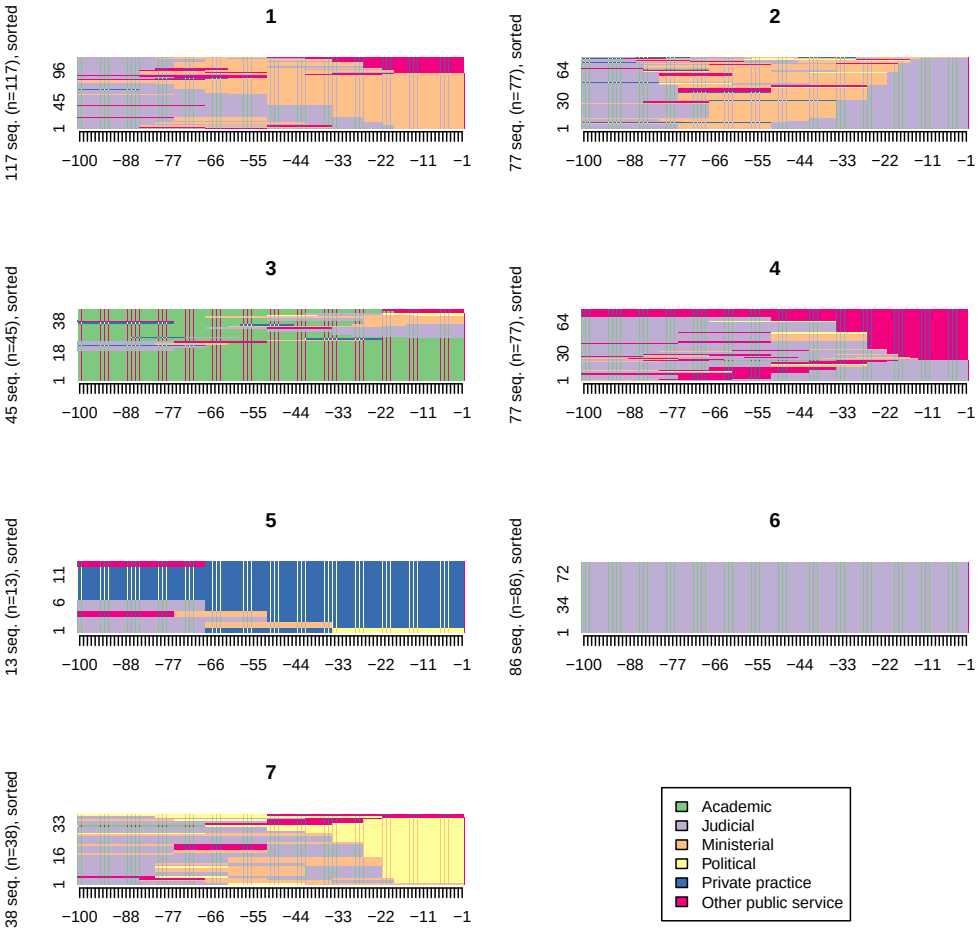
1
2
3
4
5
6
7
117 seq. (n=117), sorted
77 seq. (n=77), sorted
45 seq. (n=45), sorted
77 seq. (n=77), sorted
13 seq. (n=13), sorted
86 seq. (n=86), sorted
38 seq. (n=38), sorted
−100 −88 −77 −66 −55 −44 −33 −22 −11 −1
Academic
Judicial
Ministerial
Political
Private practice
Other public service

Figure 7: k = 7.

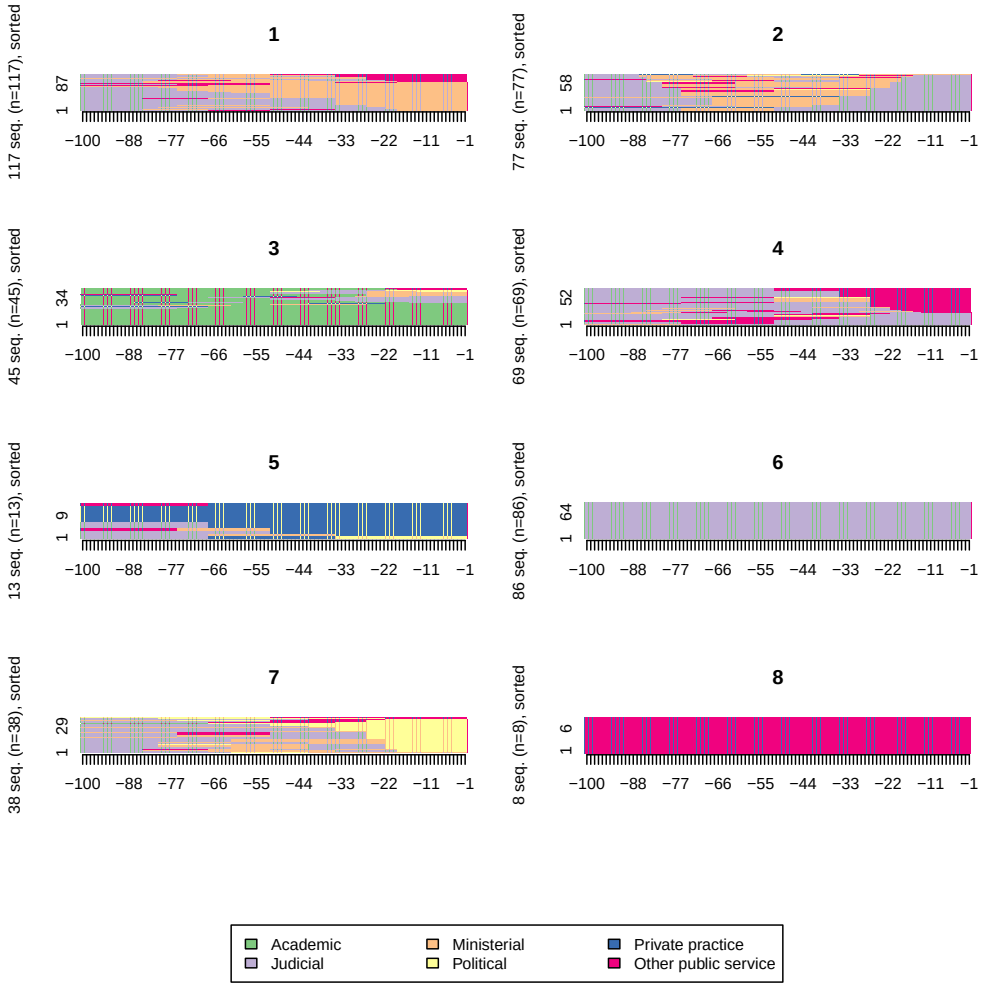
1
2
3
4
5
6
7
8
117 seq. (n=117), sorted
77 seq. (n=77), sorted
45 seq. (n=45), sorted
69 seq. (n=69), sorted
13 seq. (n=13), sorted
86 seq. (n=86), sorted
38 seq. (n=38), sorted
8 seq. (n=8), sorted
−100 −88 −77 −66 −55 −44 −33 −22 −11 −1
Academic
Judicial
Ministerial
Political
Private practice
Other public service

Figure 8: k = 8.

Second, from the pairwise distance matrix described above we identified the pairwise distances between all Justices that served at the same time. Third, we calculated the mean value, with standard deviations, for all those values, which we use as a measurement of background diversity on the high court concerned at that time.

References

Abbott, Andrew (1990). "Conceptions of time and events in social science methods: Causal and narrative approaches". In: *Historical Methods* 23.4, pp. 140–150. ISSN: 19401906. DOI: 10.1080/01615440.1990.10594204.

— (1995). "Sequence Analysis: New Methods for Old Ideas". In: *Annual Review of Sociology* 21.1, pp. 93–113. ISSN: 03600572. DOI: 10.1146/annurev.soc.21.1.93.

Abbott, Andrew and Angela Tsay (2000). "Sequence Analysis and Optimal Matching Methods in Sociology". In: *Sociological Methods and Research* 29.1, pp. 3–33.

Aisenbrey, Silke and Anette E. Fasang (2009). "New life for old ideas: The "second wave" of sequence analysis bringing the "course" back into the life course". In: *Sociological Methods and Research* 38.3, pp. 420–462. ISSN: 15528294. DOI: 10.1177/0049124109357532.

Billari, Francesco C. (2005). "Life course analysis: two (complementary) cultures? Some reflections with examples from the analysis of the transition to adulthood". In: *Advances in Life Course Research* 10, pp. 261–281. ISSN: 10402608. DOI: 10.1016/S1040-2608(05)10010-0.

Gabadinho, Alexis et al. (2011). "Analyzing and visualizing state sequences in R with TraMineR". In: *Journal of Statistical Software* 40.4, pp. 1–37. ISSN: 15487660. DOI: 10.18637/jss.v040.i04.

Gauthier, Jacques-Antoine, Felix Bühlmann, and Philippe Blanchard (2014). "Introduction: Sequence Analysis in 2014". In: *Advances in Sequence Analysis: Theory, Method, Applications*. Dordrecht: Springer, pp. 1–17. ISBN: 9783319049694. DOI: 10.1007/978-3-319-04969-4_1.

Kaufman, Leonard and Peter J. Rousseeuw (1995). *Finding Groups in Data - An Introduction to Cluster Analysis*. Hoboken: John Wiley and Sons. ISBN: 9781626239777.

Kruskal, Joseph B. (1983). "An Overview of Sequence Comparison: Time Warps, String Edits, and Macromolecules". In: *SIAM Review* 25.2, pp. 201–237.

Ritschard, Gilbert and Matthias Studer (2018). "Sequence Analysis: Where Are We, Where Are We Going?" In: *Sequence Analysis and Related Approaches: Innovative Methods and Applications*. Ed. by Gilbert Ritschard and Matthias Studer. Cham: Springer Open, pp. 1–11. ISBN: 9783319954202. DOI: 10.1007/978-3-319-95420-2_1.

Ward, Joe H. (1963). "Hierarchical Grouping to Optimize an Objective Function". In: *Journal of the American Statistical Association* 58.301, pp. 236–244.