



HAL
open science

Similarity Sampling by Machine Learning A Social Science Experiment with Artificial Intelligence and IPCC Leadership

Tommaso Venturini, Tobias Blanke, Kari de Pryck

► **To cite this version:**

Tommaso Venturini, Tobias Blanke, Kari de Pryck. Similarity Sampling by Machine Learning A Social Science Experiment with Artificial Intelligence and IPCC Leadership. 2021. hal-03464171

HAL Id: hal-03464171

<https://hal.science/hal-03464171>

Preprint submitted on 3 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Similarity Sampling by Machine Learning

A Social Science Experiment with Artificial Intelligence and IPCC Leadership

Tommaso Venturini¹, Tobias Blanke², Kari De Pryck³

1. Centre Internet et Société, CNRS; 2. University of Amsterdam; 3. Sciences Po Paris

Corresponding author:

Tommaso Venturini | tomm.venturini@ gmail.com | +33 (0)6 66 22 92 17
ORCID 0000-0003-0004-5308

Keywords

Network analysis; organisations studies; international organisations; IPCC; sampling; machine learning; artificial intelligence

Statements and Declarations

The authors have no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Abstract

In this paper, we devise a machine learning protocol to tackle a complex sociological task: to create a research sample from a few examples of interest, but in the absence of a clear definition of the target subset. As an example, we create a sample of organisational leaders starting from a list of nominees for the Bureau of the Intergovernmental Panel on Climate Change. The difficulty in this task lies in the impossibility to spell out the characteristics that define leadership in a complex and highly distributed organization like the IPCC. To bypass this lack of explicit definition, we use a series of techniques for anomaly detection to identify IPCC contributors with profiles similar to official Bureau nominees. We find that we can construct a precise (albeit implicit) model of IPCC leadership despite its social and political complexity, and that we can usefully use this model to expand our initial sample.

Introduction

This paper presents an experiment in the use of machine learning to tackle a complex social phenomenon: the identification of informal leaders in a highly distributed international organisation (the Intergovernmental Panel on Climate Change, IPCC). We found that we can build a model capable of an accurate reading of IPCC leadership despite the social and political intricacy of this organisation. Accuracy, however, is not our prime goal. Despite recurring claims that computational methods will revolutionize social sciences [1–5], the rolling out of such methods is still hindered by the mismatch between their functional requirements and the actual conditions of social research, e.g. the lack of large and well-structured datasets [6] and the focus on interpretation [7–9]. For computational methods to realize their full sociological potential, it is therefore crucial to investigate their usage *in vivo*.

This is precisely what we try to do in this paper, by putting ourselves in a situation that is common in social sciences but in many ways opposite to the one generally chosen for machine learning: we deal with a task that is highly disputed among human experts; we rely on a small dataset (the 5,676 individuals who have contributed to the IPCC main reports); we exploit artificial intelligence (AI) techniques available in free packages and compatible with a normal laptop; and, finally, we aim at sociological relevance rather than technical efficiency.

Contemporary AI techniques do not aim at in-depth understanding [10], [11]. Quite the opposite, their success derives largely from their capacity to give up some of interpretability in exchange for better performances [12], [13]. Being deliberately devised as black boxes, it is not surprising these techniques are at odds with the ultimate goal of social research, that is to advance our capacity to make sense of collective life [14]–[16]. This does not mean, however, that AI cannot be helpful. The black boxed nature of machine learning, in particular, can come in handy to perform research tasks whose contours are inherently vague or implicit. In this paper, we exemplify such use of AI for social research by proposing a technique of similarity sampling based on a protocol to teach the machine to recognise individuals with outlying leadership profiles.

Implicit similarity sampling

In order to understand our methodological proposition, it is important to notice that our technique departs deliberately from the conventional approach of random sampling, which aims at selecting items whose features can be generalized to the population it represents [17]. Instead, we propose here a form of *purposive sampling* [18], also known as *non-probabilistic or judgement sampling* [19] because it “involves the pursuit of the kind of person in whom the researcher is interested” [20 p. 137]. This type of sampling is preferred when the objective is to yield “insights and in-depth understanding rather than empirical generalizations” [21 p.1] and when the target of the research is not the whole population but a particularly important subgroup.

In this paper, we are interested in identifying IPCC *informal leaders*. While leadership is a classic topic in International Relations (cf. among others, [22]–[25]), most research tends to focus on *formal leaders* – i.e. individuals who hold specific chairs in international organisations [26] – and struggles to identify “informal leaders” [27] whose influence is palpable but not explicitly defined. In the IPCC, for example, the formal leadership is clearly represented by the Bureau, an organ that is elected at the beginning of each assessment cycle and that coordinates the activities of the organisation. Yet, sitting in the Bureau is not the only way to be influential in the IPCC. In an organisation whose activities are based on voluntary work and whose mandate is to serve as an interface between different national and disciplinary communities, individuals who have been around for a longer time, or played key facilitating roles, or represented powerful countries or disciplines may have considerable sway, whether or not this is officially acknowledged.

The difficulty of identifying IPCC informal leaders exemplifies the difficulty of purposive sampling when the contours of the target subgroup cannot be explicitly defined. In the absence of a clear condition or rule, techniques such as *criterion sampling*, *typical (or deviant) case sampling*, *intensity sampling* are impracticable [18], [28]. Snowball sampling can help extend the

sample to the individuals who are most connected to official leaders, but often (as in our case) no data is available on the connections between individuals.

The failure of classic sampling to cope with the elusiveness of organisational leadership suggests a possible sociological use of machine learning. While we cannot formulate an explicit definition of IPCC informal leadership, we know from fieldwork [29] that the Bureau nominees are generally chosen among the informal leaders of the organisation. As a consequence, we can reasonably assume that informal leaders *resemble* Bureau candidates even though the specific dimensions of this similarity cannot be made explicit. This suggests an avenue for the sociological exploitation of AI techniques, as one of the tasks in which these techniques are most efficient is the extrapolation from a learning dataset. Several techniques originally conceived for prediction (e.g., to anticipate crime outbursts, [30]) or categorization (e.g., to improve marketing campaigns, [31]) can therefore be repurposed for social sciences [32] as methods for *similarity sampling*.

The IPCC leadership

The case of IPCC leadership is interesting for the key role that the organisation plays in the climate regime [33]. Founded in 1988, the IPCC produces regular assessments of climate change research. Its reports are meant to offer a scientific foundation to the negotiation in the *United Nations Framework Convention on Climate Change* (UNFCCC). For its “efforts to build up and disseminate greater knowledge about man-made climate change”, the IPCC was awarded the *Nobel Peace Prize* in 2007 and has since become a model for several other international expert organisations, notably the *Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (IPBES) and possibly the future *International Panel on Artificial Intelligence* (IPAI). The IPCC is a remarkable organisation for the cooperation it establishes between its authors who write the reports and its national delegates who supervise the assessments work and approve its outcomes. While this dual nature allows the IPCC to serve as an interface between science and politics, the cohabitation of scientists and diplomats is not always easy and tensions may arise in relation to the distribution of power in the organisation [34], and particularly in its Bureau.

Elected by the member states at the beginning of each assessment cycle, the Bureau coordinates and oversees all IPCC activities. The composition of the Bureau (which comprises 34 members in AR6) is the result of complex voting procedures and of intense behind-the-scenes negotiations [29]. The election of the IPCC Bureau is based on a complicated combination of scientific and diplomatic constraints meant to assure that its composition reflects a “balanced geographic representation with due consideration for scientific and technical requirements”, *Principles Governing the IPCC Work*¹, paragraph 5). To guarantee both the geographical equilibrium and the scientific value, the IPCC established complex set of rules detailed in seven-page document the “Procedures for the election of the IPCC Bureau”². Rule 19, for instance, stipulates that governments “should refrain from nominating non-nationals without the consent of the nominee’s national government”. The rule was introduced after a major controversy broke out in

¹ <https://www.ipcc.ch/site/assets/uploads/2018/09/ipcc-principles.pdf>

² <https://archive.ipcc.ch/pdf/ipcc-principles/ipcc-principles-elections-rules.pdf>

2002, when the sitting Chair Robert Watson was nominated by Portugal and New Zealand against the will of his own country (the United States), which ended up supporting the election of the Indian economist Rajendra Pachauri.

Being nominated for election in the IPCC Bureau is an extremely subtle social and political process which signals the influence achieved by someone within the organisation but in ways that cannot be straightforwardly measured or explicitly defined. And this is where machine learning can come in handy, offering a series of techniques to define a sample of informal leaders similar to the Bureau nominees without having to explicitly define the nature of this similarity.

Featurization

The first step of our protocol is the definition of the features that will be used to create a model of IPCC leadership. In this paper, we draw on a database of IPCC authors and delegates that we started to collect in two previous research projects (ANONYMISED)³ and extended and updated since. The database contains the names of all the individuals who have contributed to the first five assessment reports (ARs) of the IPCC (and the members and candidates of the Bureau)⁴. Great effort was invested to disambiguate homonyms and merge different names of the same person, but errors may remain. In total, we have counted 5,676 individual contributors to the IPCC. Our database, furthermore, separates the different roles held by the same individual, thus containing about 18,000 rows, each corresponding to the contribution by a given individual in a given role (delegate, Bureau member, coordinating lead author, lead author, review editor, contributing author). For each of the contributions we also collected the national affiliation declared by the contributor. All the features employed in this machine learning model described in this paper are, more or less directly, extracted from this database.

Engagement features (ENG-)

A first set of features is meant to measure the level of engagement of IPCC contributors and can be extracted directly from the database.

1. The number of plenary sessions to which an individual has participated as a member of a national delegation (ENG-CountSessions)
2. The number of chapters authored by an individual (ENG-CountSignatures)
3. The responsibility roles hold by an individual (ENG-ClaSpmSyrBureau) – i.e. *Contributing Leading Authors* (CLA), author of *Summary for Policymakers* (SPM) or the *Synthesis Report* (SYR), or member of the *Bureau* in an earlier AR.
4. The last AR to which a participant has participated (ENG-LastActive), assuming that individuals active in closer ARs have more chances to be nominated for the Bureau.

³ ANONYMISED

⁴ The database includes authors that have contributed to the IPCC Assessment Reports (AR), and not to the Special Reports (SP).

Networks features (NET-)

While the features above capture the level of engagement in the organisation, they do not consider the nature of the IPCC as a ‘network organisation’. Unlike other international organisations, the IPCC has no permanent organs and no stable employees (except for the dozen people of its Secretariat). The 5,676 individuals who contributed to its assessments did it voluntarily and on the payroll of other institutions. Rather than by hard institutionalization, the IPCC is kept together by the networking activities of its participants. To account for the relational features of IPCC participants, we turn our database into a bipartite network of individuals⁵ and capacities, defined as the 24 different roles in which IPCC members may have contributed to the organisation⁶ (see the table below)

Table 1. temporal, functional and thematic divisions of the IPCC and the roles they form

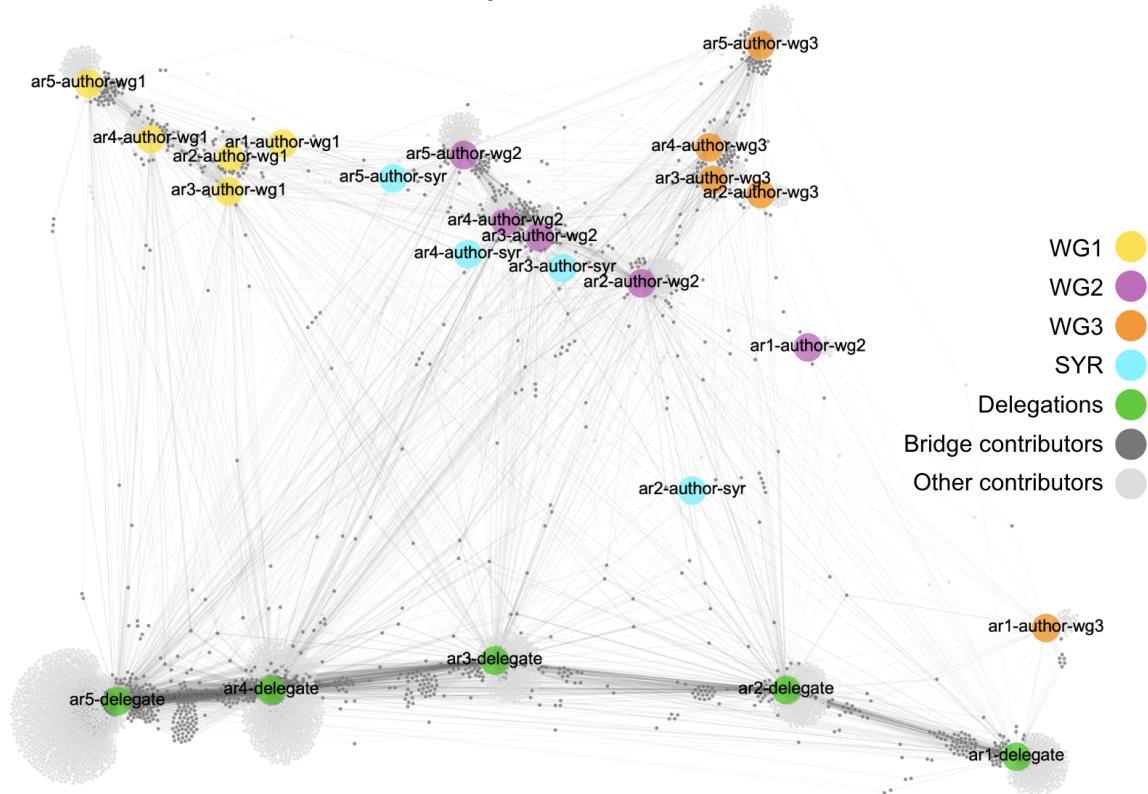
Functional divisions	Thematic divisions	Temporal divisions				
		AR1	AR2	AR3	AR4	AR5
Delegates		ar1-delegate	ar2-delegate	ar3-delegate	ar4-delegate	ar5-delegate
Authors	WGI	ar1-author-wg1	ar2-author-wg1	ar3-author-wg1	ar4-author-wg1	ar5-author-wg1
	WGII	ar1-author-wg2	ar2-author-wg2	ar3-author-wg2	ar4-author-wg2	ar5-author-wg2
	WGIII	ar1-author-wg3	ar2-author-wg3	ar3-author-wg3	ar4-author-wg3	ar5-author-wg3
	SYR	/	ar2-author-syr	ar3-author-syr	ar4-author-syr	ar5-author-syr

The edges of the graph in Figure below represent the contribution of each individual. An individual, for instance, is connected to “AR4-author-WG1” if during AR4 he/she has authored one of the chapters of the report of Working Group I. Edges are weighted according to the number of times that each individual has served in the same capacity (for example by authoring two or more chapters in the same AR and WG).

Figure 1. Bipartite network of the IPCC contributors (dark grey) and the capacities in which they have served (coloured according to their functional and thematic division) in the first five assessment reports.

⁵ It is important to remark that, as far as the authors are concerned, our network only includes contributing lead authors, lead authors or review editors (i.e. the authors with roles of coordination and responsibility and who are directed selected by the IPCC Bureau). We have not considered contributing authors who are not selected by the Bureau (but invited by the selected authors), who generally do not participate in the meetings and whose contribution is in most cases limited to specific paragraphs or topics.

⁶ The main role division in the IPCC corresponds to the *functional* separation between the scientists that review the scientific literature on climate change and write the assessment reports (i.e. "the authors") and the diplomats who serve in the national delegations overseeing the work of the organisation (i.e. "the delegates"). The authors are then subdivided in three Working Groups with different *thematic* specialisations: WGI focuses on the physical bases of climate change; WGII on impacts, adaptation and vulnerability and WGIII on mitigation. We also include participation to the writing of the Synthesis Report, which brings together the conclusions of the three WGs. Finally, the work of the IPCC is *temporarily* articulated in Assessment Cycles each lasting several years.



We also transform the bipartite network above in a monopartite network of individuals who have served in the same capacities⁷. From the two networks, we derive the following features (NET-)

1. The number of different capacities occupied by an individual (NET-Degree).
2. The betweenness centrality of an individual, as the number of shortest paths passing through his/her node in the monopartite network (NET-Betweenness)
3. The closeness centrality of an individual, as the sum of shortest-paths distance of his/her node to all the other nodes (NET-Closeness)
4. The eigen-centrality of an individual, as a recursive measure the connectivity of nodes, their neighbours, the neighbours of their neighbours, etc. (NET-Eigen).

Bridgeness features (BR-)

While the network features capture the relational connectedness and centrality of individuals, they do not consider the specificity of the IPCC network. In particular, they do not take into account that some roles are easier to cumulate, while others are more difficult. As Figure 1 suggests, some capacities are more “distant” than others and connecting them is then more valuable in the IPCC. The organisation itself acknowledges this point and identifies a group of

⁷ To generate a monopartite network from our bipartite graphs, we create a matrix in which each couple of individuals is connected by an edge if the Pearson correlations between the list of capacities they have occupied is positive. For all capacities occupied by two contributors x and y , the Pearson correlation divides the sum of their frequencies in x and y minus their respective mean frequencies $m(x)$ and $m(y)$ by the square root of the squares of $(x - m(x))$ and $(y - m(y))$. As illustrated by the figure below (relative to the first five ARs), the distribution of the Pearson correlation for each couple of individuals is highly uneven and this is why we chose zero as

contributors of particular relational importance. The so-called “bridge authors” are authors who deal “with cross-cutting topics across WGs⁸.” In a previous paper (ANONYMIZED), we extend this notion, by calculating three distinct types of bridgeness⁹ as well as their sum:

1. Thematic bridgeness, characterising the individuals, who have participated in different working groups during the same AR (BR-Thematic).
2. Functional bridgeness, characterising the individuals who have been both author and delegate during the same AR (BR-Functional).
3. Temporal bridgeness, characterising the individuals who have served in the same capacity across different ARs (BR-Temporal).
4. The sum of all the bridgeness above (BR-BridgenessSum).

National feature (NAT-)

Finally, as in all international organisations, the composition of the IPCC leadership is decided not only by the individual profiles, but also by the power equilibrium between countries. While IPCC Bureau Election Procedures demand a “balanced geographic representation”, this provision only stipulates that the five regions of the world (according to the classification of the World Meteorological Organisation) should be represented in the Bureau but does not specify which countries should represent each region. As a consequence, the most influential and active countries in the organisation (e.g., the US, the Russian Federation, Japan, Brazil, China, the UK, Canada, Saudi Arabia, India, Germany or France) are also more present in its Bureau.

To account for the national affiliation of IPCC members, we add to our feature list the average percentage of gross domestic product of the country dedicated to research and development between 1996 and 2017 according to the World Bank (NAT-GDPRD)¹⁰. We choose this particular feature for its capacity to capture both the richness of countries and their technical and scientific power.

⁸ https://archive.ipcc.ch/scoping_meeting_ar5/doc10.pdf, p. 10.

⁹ To quantify the bridging function in the IPCC, we develop a metric called “bipartite-bridgeness” [37], which we define as the summation of the number of indirect connections created by a node weighted by the importance of such connections and by their rarity. More precisely, the bipartite-bridgeness of a node α of partition A is equal to the number of pair of nodes in partition B bridged by α , each weighted by:

- the total number of A-nodes to which the two B-nodes are connected (i.e. the union of their neighbours)
- and the inverse of the number of A-nodes bridging the same pair (i.e. the intersection of their neighbours)

$$BB(n) = \sum_{(i,j)} [|\text{neighbours}(i) \cup \text{neighbours}(j)| / |\text{neighbours}(i) \cap \text{neighbours}(j)|]$$

$$BB(n) = \sum_{(i,j)} [1 / \text{Jaccard}(\text{neighbours}(i), \text{neighbours}(j))]$$

As the intersection of two sets divided by their union is commonly known as their Jaccard coefficient, the bipartite-bridgeness can be defined as the summation of the inverse Jaccard coefficient of the neighbourhoods of all pairs of neighbours of ‘n’. Because the Jaccard coefficient is a measure of similarity between sets, bipartite-bridgeness grows not only with the number of couples bridged but also with their relational diversity.

In the case of our networks of contributors and capacities, the bipartite-bridgeness of an individual is then defined as the summation of all the pair of capacities in which she has served weighted by their importance and the rarity of their connection.

¹⁰ The last four features are extracted from the official data of the World Bank (<https://data.worldbank.org>).

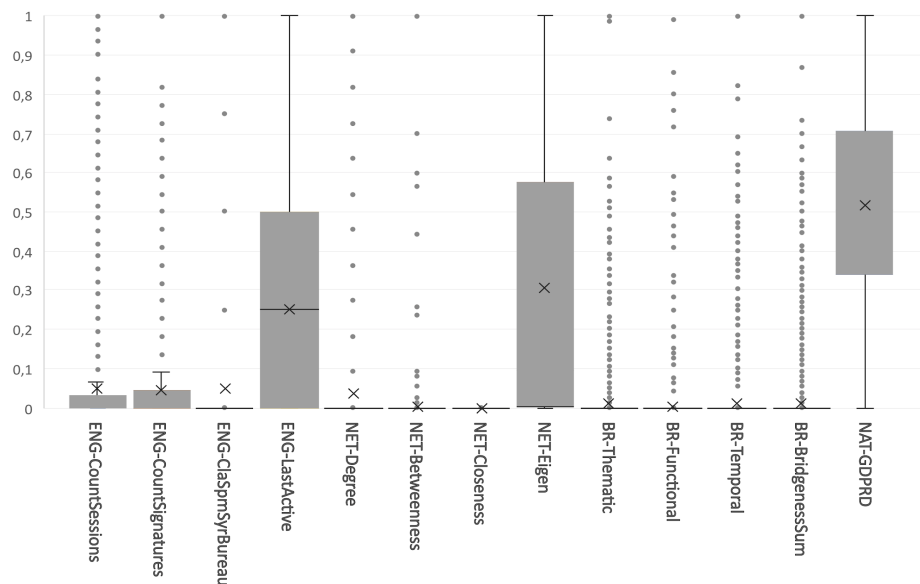
Taking time into account

A final issue in data preparation is related with time. To train our model, we use the list of the Bureau nominees in AR5 and AR6, as well as its elected members in AR4¹¹. Because these individuals have been nominated for the Bureau at a specific moment in time, we have to make sure that our model considered their features at that moment. While considering the AR4 Bureau, therefore, we should only use data from the three earlier reports (AR1, AR2 and AR3). For the AR5 Bureau, we can add the information about AR4 as well. And finally, for the AR6 Bureau, we can consider information about the five previous reports. In practice, this means that all the individual features above (ENG-, NET- and BR-) need to be calculated at three different moments in time and using only the data concerning IPCC participation that occurred before the Bureau election under consideration.

Model training

After having extracted and normalized our 13 features, we train our model to recognise good candidates for the IPCC Bureau. The problem is that out of the 11,742 rows in our dataset, only 148 rows (or 1.3% of the base), correspond to Bureau’s members or candidates. As a consequence, any model that simply assumes that no individual is a Bureau candidate will achieve the stunning accuracy of 98.7%. This “accuracy paradox” prevents us from relying straightforwardly on accuracy as a metric to evaluate and improve our predictive model, thus excluding most basic machine learning approaches. However, as suggested by the figure below, the distribution of values for most of our features is highly uneven and with several outlying values, many of which correspond to Bureau candidates. Individuals nominated to the Bureau seem to have a special profile that sets them apart from all other IPCC contributors and this encourages us to rely on anomaly detection algorithms to identify them and their likes.

Figure 2. Boxplots of the distribution of the values of the 13 features described above.



¹¹ Unfortunately, you could not find the names of the candidates for AR4.

As our main anomaly detection algorithm, we used *isolation forests* as implemented in the open-source H2O platform [35]. This technique is a special case of *random forests*, in which the machine represents the training dataset through a series of tree-like structures with the objective of isolating the members of the target class with as few branches as possible. At each branching, the records of the dataset are divided in two classes according to a threshold on one of the features (e.g., separating the class of individuals who have occupied more than N capacities and then the subclass of those who had been active in the last AR, etc.). Averaging over many trials (hence the name of 'forest'), the machine learns which features are most important to identifying the target class. An isolation forest is an unsupervised version of the random forest, which tries to single out not a target class but each individual item. Being distinctively different from other nodes, outliers tend to be closer to the root and require fewer splits to be isolated. The final outlier score is the average number of splits needed to create a leaf for a given entry across all the trees in the isolation forest. In our case, the lower this number is for an individual, the more likely she or he is to be an anomaly and thus a Bureau candidate.

To improve our results while avoiding overfitting to the training datasets, we first run a grid search with cross-validation to find the optimal hyperparameter range for our isolation forest. Finding the ideal structure is called “hyperparameter tuning” and is a crucial albeit trial-and-error process. In our case, we looked for the overall number of trees in the forest, their maximum depth and the optimal encoding of the capacities. The result is that the best forest has trees of maximum depth 30 (mean depth of ~16 and a minimum depth of 9), with a densely populated forest of up to 10,000 trees. We split our corpus into a training set and a test set with a 75-25 split (taking care that the distribution of bureau candidates is similar in both sets and thus having 37 nominees in the test set). 2,935 entries are thus held back to evaluate how the model reacts to unseen data.

With a default scoring option, our anomaly detection algorithm cannot overcome the accuracy paradox. A straightforward application of the isolation forest method (see table 2) achieves an accuracy of ~0.95 but, while it performs very well in excluding non-Bureau individuals and avoiding false positives (*specificity* = 0.96), it is incapable to identify almost half of the Bureau candidates (*accuracy* = 0.49).

Table 2. Confusion matrix (for the test set) showing the results of the isolation forest with default scoring.

	actual “not-bureau”	actual “bureau”	total
predicted “not-bureau”	2775 (true neg)	19 (false neg)	2794
predicted “bureau”	123 (false pos)	18 (true pos)	141
total	2898	37	2935

Our task is therefore to find an approach that maximises the relation of specificity and sensitivity without sacrificing too much accuracy. To counter the accuracy paradox, we thus change our scoring function to over-reward the identification of bureau candidates and over-penalize their missing. The metric we are aiming to maximise is thus the ‘harmonized mean’ of sensitivity and specificity:

$$Score = \frac{Sensitivity * Specificity}{Sensitivity + Specificity}$$

As intended, this approach sacrifices some accuracy (which drops to ~92%), but our confusion matrix results looked more appropriate to the task of predicting bureau candidates. The sensitivity is now improved to 0.65 while specificity is still a strong 0.93.

Table 3. Confusion matrix (for the test set) of the isolation forest with customized scoring function.

	actual “not-bureau”	actual “bureau”	total
predicted “not-bureau”	2683 (true neg)	13 (false neg)	2794
predicted “bureau”	215 (false pos)	24 (true pos)	141
total	2898	37	2935

We can further improve our performance by combining our isolation forest (which identifies anomalous contributors but does not exploit the existing information on who is a bureau candidate and who is not) with a semi-supervised version of autoencoder neural network [36]. Autoencoders aim at reducing the dimensionality of the data, discarding noise by determining the most important latent features of the dataset (or 'deep features') generated as black box combinations of the initial features. Our semi-supervised approach consists in generating our autoencoders using exclusively the data from the 'normal' non-bureau individuals, which will make it easier to detect the bureau candidates as outliers. By building an isolation forest on top of the second hidden layer of the autoencoder and using our custom scoring function, we are able to keep a high accuracy of 92% but at the same time maximise the relationship between sensitivity (0.92) and specificity (0.92). Only 3 bureau candidates are now missed from the test data.

Table 5. Confusion matrix (test set) of the isolation forest with customized scoring and autoencoded features.

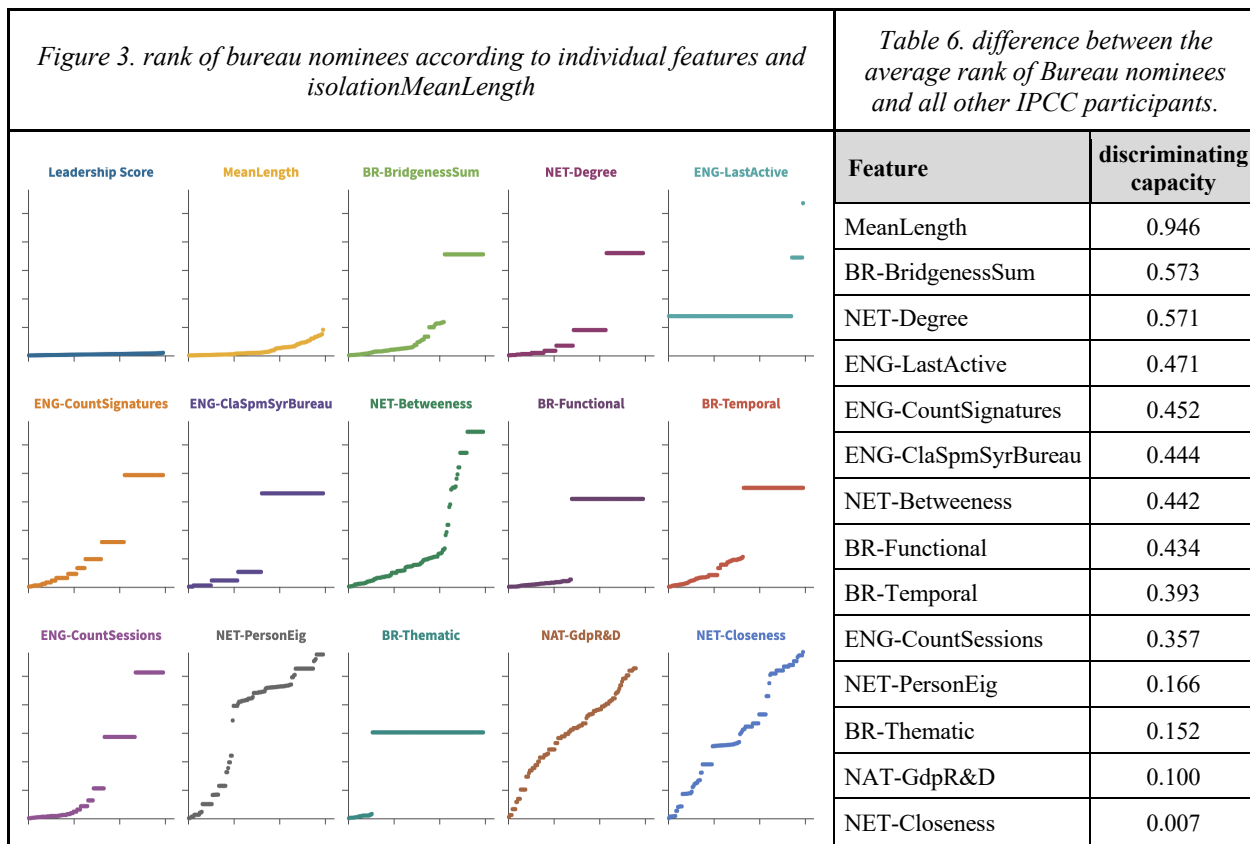
	actual “not-bureau”	actual “bureau”	total
predicted “not-bureau”	2683 (true neg)	3 (false neg)	2794
predicted “bureau”	215 (false pos)	34 (true pos)	141
total	2898	37	2935

Results and discussion

So far, following the conventions of predictive analytics, we presented the results of our model in terms of false positives and false negatives. Our objective, however, has never been to predict which individuals were going to be nominated as IPCC chairs and vice chairs, but to exploit our existing knowledge on the Bureau nominations to identify other individuals with similar 'leadership potential' or, to say it with other words, similar ways to be exceptional in the organisation. We can therefore abandon the binary classifications of the confusion matrices and consider our results according to a continuous measure. For each of the individuals (and for each AR), our algorithm computes the average number of splits needed to isolate the entry across the trees of the isolation forest. This metric, which we call “MeanLength” is our measure of Bureau-likeness.

To test MeanLength, we compare it to the thirteen individual features described above to show its better capacity to identify 'leader-like' individuals. To gauge such a discriminating capacity,

we rank all the individuals in our dataset according to each of the features and according to the MeanLength and we calculate the average ranking of Bureau candidates and of all other IPCC participants and then simply subtract these two averages. The idea behind this test is that, as Bureau nominees are the people that the organisation itself recognises as its potential leaders, an effective measure of organisational leadership should rank them higher than other participants. In the best-case scenario, in which the 148 candidates are ranked in the first 148 positions, the difference in the average rank between candidates and non-candidates would be equal to 5871. We can use this maximum as a benchmark of the discriminating capability of different metrics. Figure 3 displays the ranking of Bureau nominees for each individual feature and for the MeanLength and table 6 compares their discriminating capacity (as the difference between the average ranking of candidates and the average ranking of non-candidates divided by 5871, the maximum possible difference).

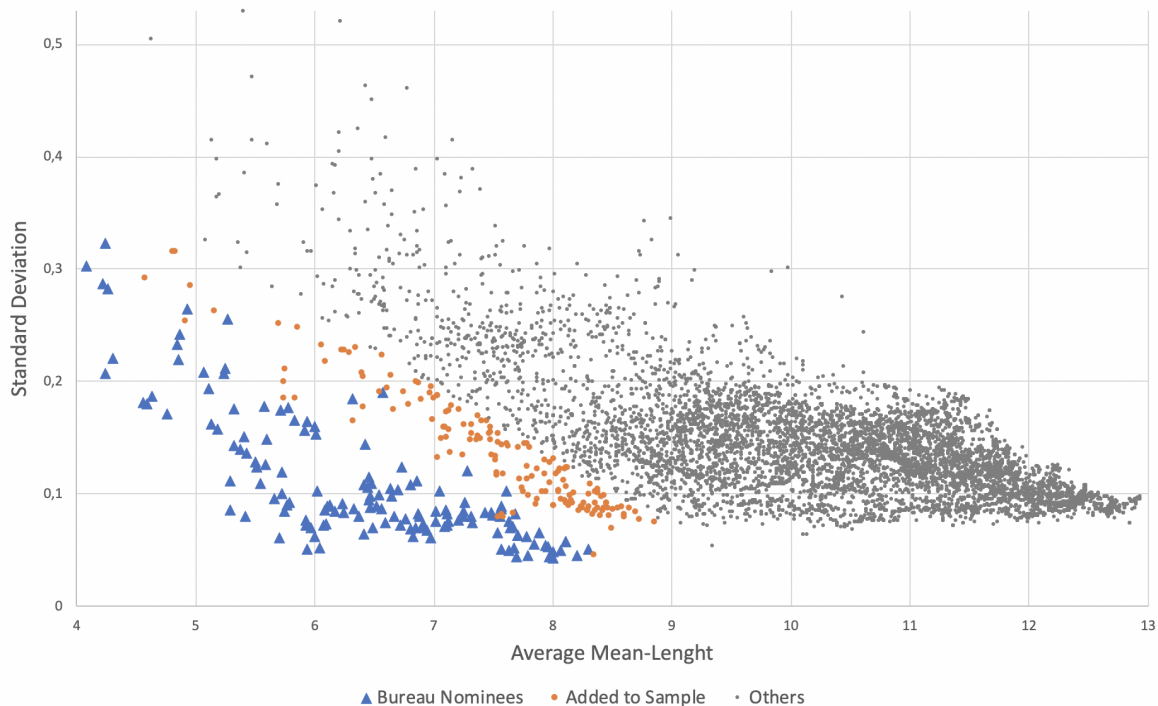


The figure and table above demonstrate that the MeanLength is significantly more effective than all individual features in identifying IPCC Bureau nominees, being almost twice as good as the best features (BR-BridgenessSum and NET-Degree). This suggests that there is a real gain in combining features through machine learning instead of using them in isolation.

To improve our discrimination capacity even more, we can exploit the non-deterministic nature of our protocol – i.e., the fact that, if executed several times, it will produce different models and different final classifications of the individuals in our corpus. Since each of these implementations of our protocol is differently affected by the noise in the data and the chances of the training process, their average results should be more robust than the results of any single

instance. Indeed, running our protocol ten times and averaging the results we obtain an average MeanLength with an even greater discrimination capacity (0.972). But we can do even better. Figure 4 shows that Bureau candidates (the triangles) have not only a lower *average* MeanLength, but also that their evaluation is more stable across different runs of our protocol.

Figure 4. Scatterplot of the average and of standard deviation for the MeanLength of the individuals of our corpus across the 10 implementations of our protocol.



Drawing on this insight, we combine average and standard deviation in a new variable called leadership-score:

$$LS = \text{average of the MeanLength} + (\epsilon * \text{standard deviation of the MeanLength})$$

After exploring all values of ϵ to identify the one (15.5) that maximises the difference in the ranking of bureau candidates and other participants, our new measure achieves an almost perfect discrimination capacity (0.9996). Ranking the individuals of our corpus according to the new measure, all the 107 highest ranked individuals are Bureau candidates, and all the 148 Bureau candidates are ranked among the 184 first individuals (which means that only 36 false positives are necessary to identify all the candidates).

Application to sampling

Given the complexity of the process for the nomination of the IPCC Bureau, obtaining a metric with an almost perfect discriminating capacity is quite a computational feat. By itself, however, this achievement is not particularly useful, as being able to predict Bureau candidates has little use for IPCC actors and little interest for IPCC scholars. To achieve sociological relevance, we still have to demonstrate how our prediction machine can be repurposed into a similarity sampling protocol capable of identifying individuals with special leadership profiles.

Figure 4 shows a clear separation between candidates and non-candidates, but it also shows that some non-candidates have a position that is relatively close to that of Bureau nominees (we highlighted them as orange dots). These are the informal leaders that we should add to the formal ones to complete our sample of IPCC leaders. To identify them, we focus on the current report, AR6. Since our dataset contained 67 candidates for the Bureau of AR6, we decided to use our method to double this sample and identify the 67 other individuals who have not been candidates for AR6 but are ranked the highest for leadership score in that AR. 16 of these informal leaders have been members of the Bureau in one or more of the earlier ARs. We will, therefore, not consider them because they could have added easily to the sample because of their previous role. 51 individuals, however, have never been in the Bureau so their inclusion in the sample is not trivial. Their names are listed in the table below.

Table 7. Informal leaders identified through our similarity sampling protocol.

Name	Country	Ran_LS	NET Degree	BR Temporal	BT Thematic	BR Function	BR Sum	ENG Sessions	ENG Signatur.	ENG-ClaSpmS yrBureau
Cramer, Wolfgang P.	Germany	120	7	816	194	2061	3071	3	14	3
Ishitani, Hisashi	Japan	137	3	0	0	236	236	1	5	2
Jorgensen, Anne Mettek	Denmark	138	5	3824	0	0	3824	25	0	0
Mostefa-Kara, M.K.	Algeria	141	2	0	0	1523	1523	2	1	1
Kashiwagi, Takao	Japan	146	3	0	0	236	236	1	4	2
Friedlingstein, Pierre	France	156	5	121	228	2485	2834	3	9	2
Titus, J.	USA	164	2	0	307	0	307	0	2	1
Melillo, Jerry	USA	165	4	61	173	0	234	1	5	2
Oquist, Mats	Sweden	166	2	71	0	0	71	0	4	2
Perrin, Dominique	Belgium	168	3	878	0	0	878	4	0	0
Clini, Corrado	Italy	171	5	3824	0	0	3824	6	0	0
Taniguchi, Tomihori	Japan	172	3	0	143	1140	1283	3	3	1
Melnikov, P.I.	USSR	174	1	0	0	0	0	0	1	1
Morand Francis, Pascale	Switzerland	176	3	84	0	310	394	6	1	0
Ososkova, Tatyana	Uzbekistan	178	5	1271	0	255	1526	15	1	0
Brown, Sandra	USA	186	2	71	0	0	71	0	8	2
Kerem, A.	Israel	191.5	1	0	0	0	0	1	0	0
Morgenstern, Richard	Germany	194	2	1485	0	0	1485	6	0	0
Mahrenholz, R.	Germany	197	3	878	0	0	878	3	0	0
Banuri, Tariq	USA	200	4	97	122	1046	1265	1	5	3
Egbare, Awadi Abi	Togo	202	4	1271	0	0	1271	24	0	0
Abrol, I.	India	203	1	0	0	0	0	0	2	1
Braima, Tommy	Sierra Leone	207	3	878	0	0	878	5	0	0
Kauppi, P.E.	Finland	211	2	0	0	0	0	0	4	2
Abuleif, Khalid	Saudi Arabia	226	5	3824	0	0	3824	9	0	0
Penman, Jim	UK	232	5	3824	0	0	3824	13	0	0
Ada, Fahim	Afghanistan	233	1	0	0	0	0	1	0	0
Shah, A.	USA	238	1	0	0	0	0	0	1	0
Styrikovich, M.	Russia	243	1	0	0	0	0	0	1	1
Pittock, A. Barrie	Australia	245	4	32	194	865	1091	1	13	3
Jaffe, S.	Israel	249	1	0	0	0	0	2	0	0
Gilbert, J.	NZ	250	2	0	0	310	310	1	2	1
Madhava Sarma, K.	India	251	1	0	0	0	0	0	1	1
Albritton, Daniel L.	USA	254	4	50	156	741	947	1	5	3
Solomon, Allen M.	USA	255	1	0	0	0	0	0	5	2
Tokioka, Tatsushi	Japan	260	5	1141	0	1220	2361	3	2	0
Prentice, I. Colin	Germany	261	2	50	0	0	50	0	7	2
Xie, Shaoxiong	China	264	2	0	0	310	310	1	1	1
Callander, Bruce	UK	268	4	84	173	1077	1334	6	2	1

Benndorf, Rosemarie	Germany	269	3	389	0	0	389	3	0	0
Spradley, Julian R.	USA	275	3	84	0	310	394	4	1	0
Beetham, R.	UK	276	2	0	0	310	310	1	1	1
Barrell, Susan Lesley	Australia	278	4	1271	0	0	1271	8	0	0
Ivanov, Teodor	Bulgaria	279	3	878	0	0	878	13	0	0
Harasawa, Hideo	Japan	281	6	648	0	1051	1699	7	7	2
Oppeneau, Jean Claude	France	282	2	0	0	310	310	3	1	1
Tilley, John	Australia	285	2	0	0	310	310	1	1	1
Sathaye, Jayant	USA	288	5	209	143	0	352	0	12	3
Yokobori, Keiichi	Japan	292	3	84	0	310	394	5	1	1
Luxmoore, R.J.	USA	303	1	0	0	0	0	0	4	1
Hulme, Mike	UK	308	2	0	350	0	350	0	8	2

Our analysis stops at this list. However, since the very objective of our protocol is to build an extended sample of organisational leaders, this result is where a sociological investigation of IPCC leadership could begin.

Conclusions

In this paper we presented an experiment with machine learning in which we purposely put ourselves in a situation that, while decidedly more difficult than most AI experiments, is closest the actual research conditions of the social sciences: we measured ourselves against an elusive phenomenon (IPCC informal leadership) and worked with small data and limited computational resources. The results show that, even with these constraints, it is possible to train an efficient model. By using a combination of autoencoders and isolation forests, we developed a surprisingly precise measure of leadership potential. If we had tried to guess the 67 nominees of AR6 Bureau by taking the 67 individuals ranking the highest according to our isolation MeanLength, we would have predicted 60 of them (quite a stunning result, considering that our model was choosing among 5.559 different individuals).

As social scientists, however, we could not content ourselves with such a result, for prediction is rarely a goal in our disciplines and certainly not as the IPCC leadership is concerned. To make our protocol sociologically relevant, we then repurpose our model shifting its use from prediction to sampling. In this new use, we are not trying to guess which individuals would be nominated for the IPCC Bureau, but we use our knowledge about these nominations to find other individuals with a similar profile and thus devise a protocol for sampling people with a leadership profile. Yes, our “leadership-score” metric is a black box because it is based on an opaque machine learning algorithm that does not provide us with the exact recipe employed to compute it. Yet, opacity is not necessarily a drawback. Because there is no clear or unique definition of what makes an IPCC leader, there is no way to develop a transparent algorithm for leadership sampling. Still, as we tried to show, this does not mean that it is impossible to devise such a sampling *if the transparency condition is relinquished*.

The situation in which we carried out our experiment, we believe, is not exceptional. It is common for social scientists to be interested in the subset of a population (e.g., the elite of an international organisation) that they identify by some prototypical examples (e.g., the official leaders of the same organisation) rather than by an exact definition. Because we were particularly interested in leaders, who, by definition, occupy an outlying position in their organisation, we relied on techniques for anomaly detection, but other techniques of machine

learning can be employed in cases in which the sample to be extended does not have the same character of exceptionality.

Beyond the details of our protocol, the crucial point of our demonstration was to show that the black box nature of AI does not disqualify its use in social sciences but can be exploited as a workaround in situations in which the research is hindered by a lack of explicit definition. In these circumstances, it may be desirable to extend one sample by similarity, *even if we are unable to characterise precisely the nature of such similarity*. This is a situation in which machine learning can come in handy for its capacity to find resemblance in a multi-dimensional space.

References

- [1] K. M. Carley, "Artificial intelligence within sociology," *Sociol. Methods Res.*, vol. 25, no. 1, pp. 3–30, 1996.
- [2] D. Lazer *et al.*, "Computational social science," *Science (80-)*, vol. 323, no. 5915, pp. 721–3, Feb. 2009.
- [3] R. Conte *et al.*, "Manifesto of computational social science," *Eur. Phys. J. Spec. Top.*, vol. 214, no. 1, pp. 325–346, 2012.
- [4] P. Raghavan, "It's time to scale the science in the social sciences," *Big Data Soc.*, vol. 1, no. 1, pp. 1–4, 2014.
- [5] M. Alvarez, *Computational Social Science*. Cambridge: University Press, 2016.
- [6] R. Kitchin and T. P. Lauriault, "Small data in the era of big data," *GeoJournal*, vol. 80, no. 4, pp. 463–475, 2015.
- [7] S. Woolgar, "Why not a Sociology of Machine? The Case of Sociology and Artificial Intelligence," *Sociology*, vol. 19, no. 4, pp. 557–572, 1985.
- [8] J. Grimmer, "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together," *PS Polit. Sci. Polit.*, vol. 48, no. 01, pp. 80–83, Jan. 2015.
- [9] T. Venturini, P. Jensen, and B. Latour, "Fill in the Gap: A New Alliance for Social and Natural Sciences," *J. Artif. Soc. Soc. Simul.*, vol. 18, no. 2, p. 11, 2015.
- [10] D. Andler, "From paleo to neo connectionism," in *New Perspectives on Cybernetics*, G. van der Vijver, Ed. Kluwer: Dordrecht, 1992, pp. 125–146.
- [11] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [12] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," vol. 349, no. 6245, 2015.
- [13] D. Cardon, J.-P. Cointet, and A. Mazières, "Neurons spike back," *Réseaux*, vol. n° 211, no. 5, pp. 173–220, 2018.
- [14] J. Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms," *Big Data Soc.*, vol. 3, no. 1, p. 205395171562251, Jan. 2016.
- [15] A. Mackenzie, *Machine Learners: Archaeology of a Data Practice*. Cambridge Mass.: MIT Press, 2017.
- [16] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.
- [17] C. Teddlie and F. Yu, "Mixed Methods Sampling," *J. Mix. Methods Res.*, vol. 1, no. 1, pp. 77–100, 2007.
- [18] M. Q. Patton, *Qualitative Research and Evaluation Methods (3rd edition)*. Thousand Oaks: Sage, 2002.
- [19] I. Etikan, "Comparison of Convenience Sampling and Purposive Sampling," *Am. J. Theor. Appl. Stat.*, vol. 5, no. 1, p. 1, 2016.
- [20] Gary and Thomas, "How to Do Your Research Project," 2013. [Online]. Available:

- <http://www.amazon.com/How-Your-Research-Project-Education/dp/1446258874>. [Accessed: 14-Feb-2016].
- [21] M. Q. Patton, "Sampling, Qualitative (Purposeful)," *Blackwell Encycl. Sociol.*, pp. 1–3, 2015.
 - [22] R. W. Cox, "The Executive Head: An Essay on Leadership in International Organization," *Int. Organ.*, vol. 23, no. 2, pp. 205–230, 1969.
 - [23] M. G. Schechter, "Leadership in international organizations: Systemic, organizational and personality factors," *Rev. Int. Stud.*, vol. 13, no. 3, pp. 197–220, 1987.
 - [24] I. M. Thorn, "Leadership in International Organizations: Global Leadership Competencies," *Psychol. J.*, vol. 15, no. 3, pp. 158–163, 2012.
 - [25] B. Reinalda and B. Verbeek, "Leadership of International Organizations," in *The Oxford Handbook of Political Leadership*, Oxford: University Press, 2013.
 - [26] J. Tallberg, "The power of the chair: Formal leadership in international cooperation," *Int. Stud. Q.*, vol. 54, no. 1, pp. 241–265, 2010.
 - [27] C. D. Pielstick, "Formal vs. Informal Leading: A Comparative Analysis," *J. Leadersh. Stud.*, vol. 7, no. 3, pp. 99–114, 2000.
 - [28] H. Suri, "Available from Deakin Research Online : Purposeful Sampling in Qualitative Research Synthesis," *Qual. Res. J.*, vol. 11, no. 2, pp. 63–75, 2011.
 - [29] K. De Pryck, "Expertise under Controversy : The Case of the Intergovernmental Panel on Climate Change (IPCC)," Sciences Po Paris and University of Geneva, 2018.
 - [30] B. Benbouzid, "From situational crime prevention to predictive policing. Sociology of an ignored controversy," *Champ pénal/Penal field*, vol. VII, no. July, pp. 1–19, 2015.
 - [31] N. Syam and A. Sharma, "Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice," *Ind. Mark. Manag.*, vol. 69, no. November 2017, pp. 135–146, 2018.
 - [32] T. Blanke, "Predicting the Past," *Digit. Humanit. Q.*, vol. 12, no. 2, pp. 1–21, 2018.
 - [33] S. C. Aykut and A. Dahan, *Gouverner le climat ?* Paris: Presses de Sciences Po, 2015.
 - [34] H. R. Hughes and M. Paterson, "Narrowing the Climate Field: The Symbolic Power of Authors in the IPCC's Assessment of Mitigation," *Rev. Policy Res.*, vol. 34, no. 6, pp. 744–766, Nov. 2017.
 - [35] A. Candel, V. Parmar, E. LeDell, and A. Arora, "Deep learning with H2O," *Amazonaws.com*, 2016. [Online]. Available: <http://h2o-release.s3.amazonaws.com/h2o/rel-turnbull/2/docs-website/h2o-docs/booklets/DeepLearningBooklet.pdf>.
 - [36] P. Baldi, "Autoencoders, Unsupervised Learning, and Deep Architectures," in *Workshop on Unsupervised and Transfer Learning (27)*, 2012, pp. 37–50.
 - [37] T. Venturini, K. De Pryck, and R. Ackland, "Bridging in Network Organisations the Case of International Panel on Climate Change (IPCC)," *Soc. Networks*, no. Under review, 2021.