Similarity Sampling by Machine Learning

A Social Science Experiment with Artificial Intelligence and IPCC Leadership

Tommaso Venturini[1], Tobias Blanke[2], Kari De Pryck[3]
1. Centre Internet et Société, CNRS; 2. University of Amsterdam; 3. Sciences Po Paris

Abstract

In this paper, we devise a machine learning protocol to tackle a complex sociological task: extending a sample of organisational leaders starting from a list of individuals nominated 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 organisation. 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 found that we can build an accurate model of IPCC leadership despite its social and political complexity and that we can usefully use that model to extend our initial sample.

Introduction

This paper presents an experiment in the use of machine learning (ML) to tackle a complex social phenomenon: the identification of informal leaders in a highly distributed 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 its social and political complexity. Accuracy, however, is not our prime goal. In spite of claims that computational methods will revolutionize social sciences (Carley, 1996; Lazer *et al*. 2009; Conte *et al.* 2012; Raghavan, 2014; Alvarez 2016), the rolling out of such methods is often hindered by their mismatch with the conditions of social research, eg. the lack of large and well-structured datasets (Kitchin & Lauriault, 2015) and the focus on interpretation (Woolgar, 1985; Grimmer, 2015; Venturini *et al.*, 2015). For computational methods to realize their sociological potential, it is therefore crucial to investigate their use *in vivo*.

This is why in this paper we put ourselves in the difficult (but common the in social sciences) situation of dealing with a task that is highly disputed (identifying leaders); relying on a relatively small and complex dataset (5,676 IPCC contributors) and standard artificial intelligence (AI) techniques (available in free packages and compatible with normal laptops); and aiming at sociological relevance rather than technical efficiency. Our results show that, even with these constraints, ML can be useful if its ambitions are attuned to its conditions of use. ML techniques have been criticised for their incapacity to deliver in-depth understanding (Andler, 1992; Lecun *et al.*, 2015) and because their performances are often obtained at the cost of interpretability (Jordan & Mitchell, 2015; Cardon *et al.*, 2018). Being deliberately devised a blackboxes, it is not surprising these techniques are at odds with the ultimate goal of social research, i.e. to advance our capacity to make sense of collective life (Burrell, 2016; Mackenzie; 2017; Rudin, 2019). This does not mean, however, that ML cannot be helpful, in particular if used for research tasks otherwise hindered by lack of explicitness. In this paper, we exemplify such use of AI, proposing a technique of similarity sampling based on a protocol to teach the machine to recognise individuals with outlying leadership profiles in a complex, internationally distributed organisation.

Implicit similarity sampling

In order to understand our proposition, it is important to notice that our technique departs deliberately from the conventional approach of random sampling (Teddlie & Yu, 2007). Instead, we propose a form of *purposive sampling* (Patton, 2002), also known as *non-probabilistic or judgement sampling* (Etikan, 2016) because it "involves the pursuit of the kind of person in whom

the researcher is interested" (Thomas, 2013, p. 137). This type of sampling is preferred when the objective is to yield "in-depth understanding rather than empirical generalizations" (Patton, 2015, p.1) and when the target of the research is not the whole population but an 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, Cox, 1969; Schechter, 1987; Thorn, 2012; Reinalda & Verbeek, 2013), most research focus on *formal leaders* holding chairs in international organisations (Tallberg, 2010) and struggles to identify "informal leaders" (Pielstick, 2000) whose influence is palpable but not explicitly defined. In the IPCC, for example, the formal leadership is clearly represented by its 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, 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 over the IPCC, whether or not this is officially acknowledged.

The difficulty of identifying IPCC informal leaders epitomises 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 (Suri, 2011; Patton, 2002). Snowball sampling can help extend the sample to the individuals related 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 leadership suggests a possible sociological use of ML. While we cannot formulate an explicit definition of IPCC informal leadership, we know from fieldwork (De Pryck, 2018) that its official chairs are generally chosen among its informal leaders. We can thus assume that most informal leaders *resemble* Bureau candidates. This suggests an avenue for the sociological exploitation of AI techniques, as one classic task of these techniques is the extrapolation of similarity from a learning dataset. Techniques introduced for prediction (eg. to anticipate crime outbursts, Benbouzid, 2015) or categorization (eg. to improve marketing campaigns, Syam, A. Sharma, 2018) can therefore be repurposed for social and human sciences (Blanke, 2018) as methods for *similarity sampling*.

The IPCC leadership

The case of the IPCC leadership is interesting for the key role the organisation plays in the climate regime (Aykut & Dahan, 2015). 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 become a model for other international expert organisations, notably the *Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* and possibly the future *International Panel on Artificial Intelligence*. 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 (Hughes & Patterson, 2017), 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 (De Pryck, 2018). The election of the IPCC Bureau rests 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"[1]. To guarantee such an equilibrium, the IPCC has established a complex set "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 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 thus 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. And this is where machine learning can be useful, 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 a priori.

Featurization

The first step of our protocol is the definition of the features to 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 (Venturini *et al.*, 2014)[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 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 in 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 key roles held by an individual (ENG-ClaSpmSyrBureau) as *Contributing Leading Authors* (CLA), author of *Summary for Policymakers* (SPM) or the *Synthesis Report* (SYR), or member of the *Bureau* in earlier ARs.
4. The last AR in which a participant has participated (ENG-LastActive), assuming that individuals active in more recent ARs have greater chances to be nominated for the Bureau.

---

[1] Principles Governing the IPCC Work: 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

[3] The database used in this paper has been initiated thanks to the support of the EU Project EMAPS and the ANR project MEDEA and about half of it completed at the médialab of Sciences Po, by Ian Gray, Nicolas Baya Laffite and Audrey Banneyx, which we greatly thank for their contribution. The database was updated and extended in a doctoral thesis completed by Kari De Pryck.

[4] The database includes authors that have contributed to the IPCC Assessment Reports but not to its Special Reports.

Networks features (NET-)

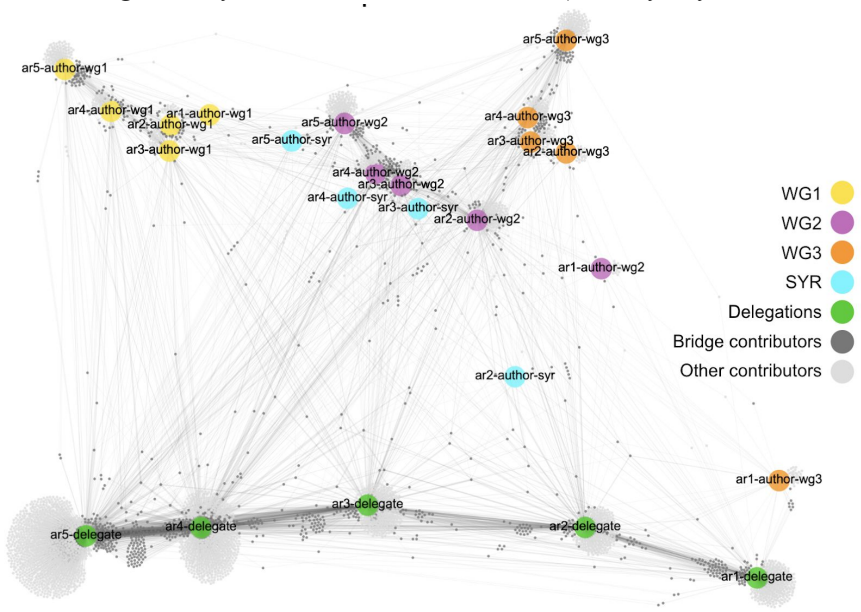
The features above do not take into account the nature of the IPCC as a 'network organisation'. Unlike other international organisations, the IPCC has no permanent organs and no stable employees (except for the dozen people of its Secretariat). Rather than by hard institutionalization, the IPCC is kept together by the networking activities of its participants. To account for the relational features, we turn our database into a bipartite network of individuals[5] and capacities, defined as the 24 different roles in which IPCC members may have contributed to the organisation[6].

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 fig. 1 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 (eg. by authoring two 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 our network only includes contributing lead authors, lead authors or review editors (i.e. the authors with roles of coordination and responsibility). We have not considered contributing 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 and write the assessment reports (authors) and the diplomats who oversee the work of the organisation (delegates). The authors are subdivided into 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 in 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.

We finally transform the bipartite network above into a monopartite network[7] and derive the following features

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 of 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 accumulate, while others are more difficult. As fig. 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 contributors of particular relational importance. The so-called "bridge authors" are authors who deal "with cross-cutting topics across WGs[8]." In a previous paper (Venturini, 2020), we extend this notion, by calculating three distinct types of bridgeness[9] as well as their sum:

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

---

[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)).

[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", which we define as the summation of the number of indirect connections created by a node and 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)

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

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

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 pairs of capacities, in which she has served, weighted by their importance and the rarity.

National feature (NAT-)

Finally, as typical to international organisations, the composition of the IPCC leadership is decided by the power equilibrium between countries. While the election procedures demand a "balanced geographic representation", this provision only stipulates that the five regions of the world (according to the WMO classification) 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 (eg. the United States, the Russian Federation, Japan, Brazil, China, the United Kingdom, 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 features 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). We choose this feature for its capacity to capture both the richness of countries and their technical and scientific power.

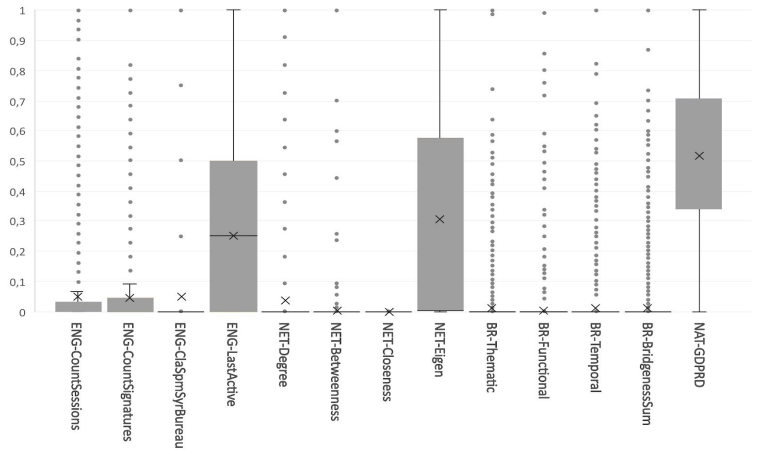
Taking time into account

A final issue in data preparation is related with time. In order to train our model, we use the list of the Bureau nominees in AR5 and AR6, as well as its elected members in AR4.[10] Because these individuals have been nominated for the Bureau at a specific moment in time, we have to make sure that our model considers 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 all five previous reports. In practice, this means that all the individual features above need to be calculated at different moments in time by considering only the data concerning events that occurred before each of the Bureau elections under consideration.

Model training

After having extracted and normalized 13 features, we train our model to recognise the Bureau candidates. The problem is that, out of the 11,742 rows in our dataset, only 148 (or 1.3%), correspond to Bureau's candidates. As a consequence, a model that simply assumes that no individual is a candidate is 98.7% accurate. This "accuracy paradox" prevents us from relying on accuracy to evaluate and improve our model. However, as suggested by the fig. 2, the distribution of features is highly uneven and with several outliers, many of which correspond to Bureau candidates. Bureau nominees seem to have a special profile that sets them apart from other IPCC contributors, which encourages us to rely on anomaly detection to identify them and their likes.

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



[10] 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 (Candel et al., 2016). 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 according to a threshold on one of the features (eg. 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 run a grid search with cross-validation to find the optimal hyperparameters for our isolation forest. The result is that the best forest has trees of mean depth of ~16. We also split our corpus into a training and a validation set with a 75-25 split (with similar candidates distribution). 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 for the validation results) 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 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 change our scoring function to reward the identification of bureau candidates and penalize missing them. The metric we are aiming to maximise is 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 showing the results 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 (Baldi, 2012). 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 blackbox 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 later on 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 validation data according to Table 5.

Table 5. Confusion matrix of the isolation forest with auto-encoded features and customized scoring function.

	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 exceptionality'. We can therefore move on from 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. We call this metric, which is central to isolation forests, "MeanLength".

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

Figure 3. rank of bureau nominees according to individual features and isolationMeanLength

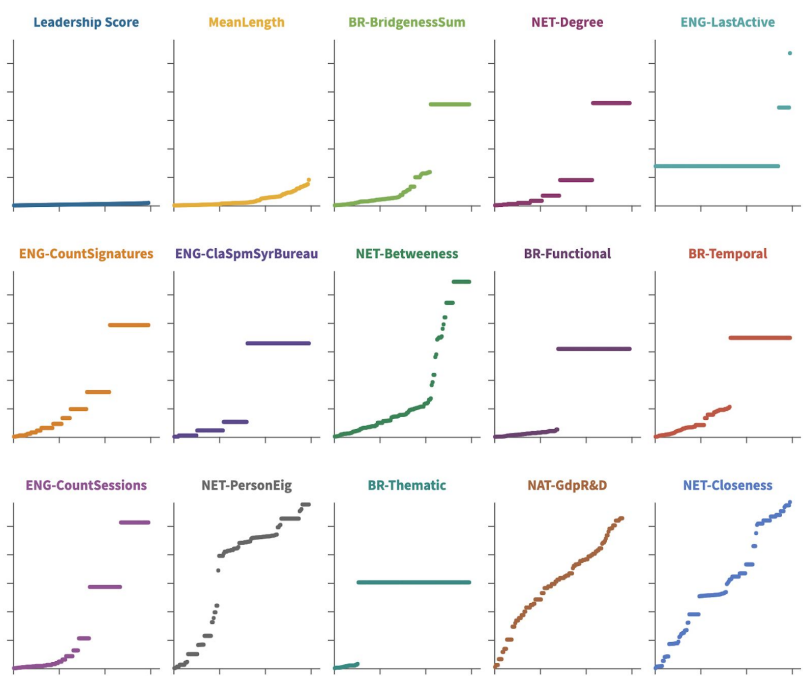


Table 6. difference between the average rank of Bureau nominees and all other IPCC participants.

Feature	discriminating capacity
MeanLength	0.946
BR-BridgenessSum	0.573
NET-Degree	0.571
ENG-LastActive	0.471
ENG-CountSignatures	0.452
ENG-ClaSpmSyrBureau	0.444
NET-Betweenness	0.442
BR-Functional	0.434
BR-Temporal	0.393
ENG-CountSessions	0.357
NET-PersonEig	0.166
BR-Thematic	0.152
NAT-GdpR&D	0.100
NET-Closeness	0.007

The figure and table above demonstrate that the MeanLength is significantly more effective than individual features in identifying 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 – the fact that, if executed several times, it will produce different models and different final classifications. Since each of these implementations of our protocol is differently affected by the noise in the data and the randomness of the training process, their average results should be more robust than that of any single instance. Indeed, running our protocol ten times and averaging the results, we obtain an average MeanLength with a greater discrimination capacity (0.972).

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.

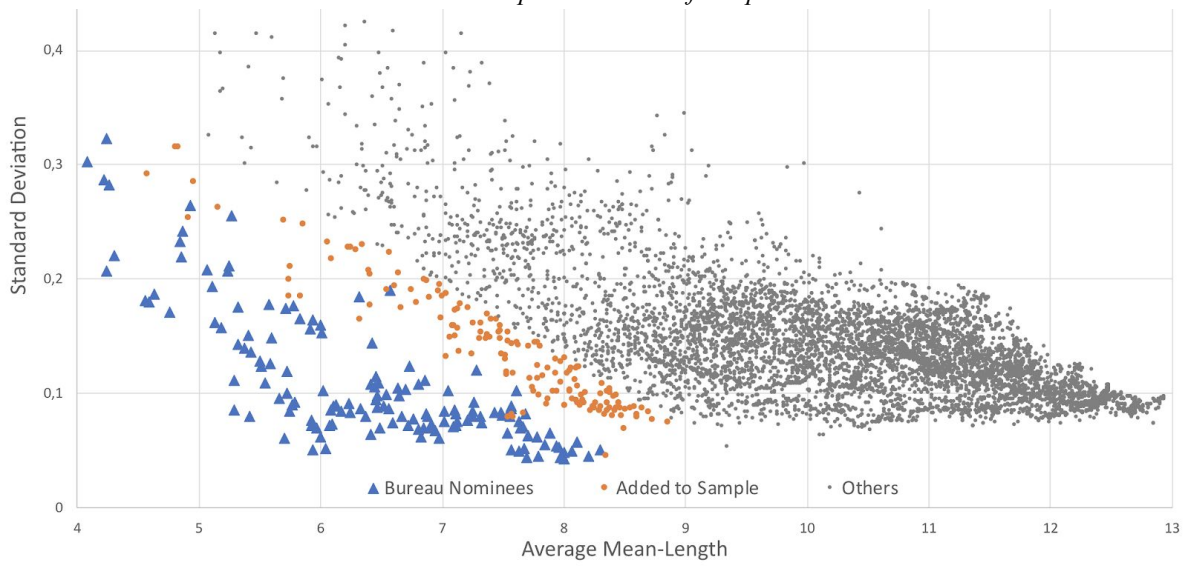


Fig. 4 shows that Bureau candidates (the triangles) have not only a lower *average* MeanLength, but also that their evaluation is more stable across different implementations of our protocol. Drawing on this insight, we combine average and standard deviation of MeanLength in a new variable we call leadership-score (LS):

*LS = average of the MeanLengh + (ε * standard deviation of the MeanLengh)*

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 such a strong discriminating capacity makes us confident that our protocol has potential. 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 sampling protocol capable of singling out individuals with special leadership profiles.

Fig. 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 in orange). 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, which have not been candidates for AR6 but are rankest highest according to 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 Signatures	ENG-ClaSpmSyrBureau
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, as we have fulfilled our objective to build an extended sample of formal and informal leaders which could become the basis of a sociological investigation of IPCC leadership. 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 machine learning can be of use in the traditional task of sampling in social sciences. It is especially useful in situations in which the research is hindered by a lack of an explicit definition of the sampling group.

Conclusions

In this paper, we presented an experiment in which we purposely put ourselves in a situation that, while decidedly more difficult than most machine learning experiments, is close to actual research conditions of the social sciences: we measured ourselves against an elusive phenomenon (IPCC informal leadership) and worked with heterogeneous data and limited computational resources. The results show that, even in these difficult conditions, it is possible to train an efficient model. By using a combination of autoencoders and isolation forests, we developed a strong 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 (an excellent 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 as such is rarely a goal in our disciplines and certainly not as the IPCC leadership is concerned. To make our protocol sociologically relevant, we repurposed 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 device a protocol for sampling people with a leadership profile. While our "leadership-score" metric is a blackbox because it is based on a complex machine learning algorithm that does not provide us with the exact recipe employed to compute it, this 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.

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 that they identify by a few prototypical examples rather than by an exact definition. In these circumstances, the impossibility to formulate an explicit sampling criterion can be sidestep by extending the sample of prototypical examples relying on machine learning and its capacity to find resemblance in a multi-dimensional space.

References

Alvarez, Michael R., ed. 2016. *Computational Social Sciences*. Cambridge: University Press.

Andler, Daniel. 1992. "From Paleo to Neo Connectionism." In *New Perspectives on Cybernetics*, ed. G. van der Vijver. Kluwer: Dordrecht, 125–46.

Antonakis, John et al. 2004. "Methods for Studying Leadership." In *The Nature of Leadership*, eds. John Antonakis and Anna T. Cianciolo. Thousand Oaks: Sage, 49–70.

Aykut, Stefan C., and Amy Dahan. 2015. *Gouverner Le Climat ?* Paris: Presses de Sciences Po.

Baldi, Pierre. 2012. "Autoencoders, Unsupervised Learning, and Deep Architectures." In *Workshop on Unsupervised and Transfer Learning (27)*, , 37–50.

Benbouzid, Bilel. 2015. "From Situational Crime Prevention to Predictive Policing. Sociology of an Ignored Controversy." *Champ pénal/Penal field,* VII(July): 1–19.

Blanke, Tobias. 2018. "Predicting the Past." *Digital Humanities Quarterly* 12(2): 1–21.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3(1): 205395171562251. http://journals.sagepub.com/doi/10.1177/2053951715622512.

Candel, A., V. Parmar, E. LeDell, and A. Arora. 2016. "Deep Learning with H2O." *Amazonaws.com*. http://h2o-release.s3.amazonaws.com/h2o/rel-turnbull/2/docs-website/h2o-docs/booklets/DeepLearningBooklet.pdf.

Cardon, Dominique, Jean-Philippe Cointet, and Antoine Mazières. 2018. "La Revanche Des Neurones." *Réseaux* 211(5): 173. http://www.cairn.info/revue-reseaux-2018-5-page-173.htm.

Cardon, Dominique, Jean-Philippe Cointet, and Antoine Mazières. 2018. "Neurons Spike Back." *Réseaux* n° 211(5): 173–220. http://www.cairn.info/revue-reseaux-2018-5-page-173.htm?ref=doi.

Carley, Kathleen M. 1996. "Artificial Intelligence within Sociology." *Sociological Methods and Research* 25(1): 3–30.

Choi, H., and H.R. Varian. 2012. "Predicting the Present with Google Trends." *Economic Record* 88(1): 2–9. http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf (December 3, 2010).

Conte, R. et al. 2012. “Manifesto of Computational Social Science.” *European Physical Journal: Special Topics* 214(1): 325–46.

Cox, Robert W. 1969. “The Executive Head: An Essay on Leadership in International Organization.” *International Organization* 23(2): 205–30.

De Pryck, Kari. 2018. “Expertise under Controversy : The Case of the Intergovernmental Panel on Climate Change (IPCC).” Sciences Po Paris and University of Geneva.

Etikan, Ilker. 2016. “Comparison of Convenience Sampling and Purposive Sampling.” *American Journal of Theoretical and Applied Statistics* 5(1): 1.

Gary, and Thomas. 2013. “How to Do Your Research Project.” http://www.amazon.com/How-Your-Research-Project-Education/dp/1446258874 (February 14, 2016).

Ginsberg, Jeremy et al. 2008. “Detecting Influenza Epidemics Using Search Engine Query Data.” *Nature* 457(7232): 1012–1014. http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature07634.html?lang=en (December 3, 2010).

Grimmer, Justin. 2015. “We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together.” *PS: Political Science & Politics* 48(01): 80–83. http://www.journals.cambridge.org/abstract_S1049096514001784.

Hughes, Hannah Rachel, and Matthew Paterson. 2017. “Narrowing the Climate Field: The Symbolic Power of Authors in the IPCC’s Assessment of Mitigation.” *Review of Policy Research* 34(6): 744–66. http://doi.wiley.com/10.1111/ropr.12255.

Jordan, M I, and T M Mitchell. 2015. “Machine Learning: Trends, Perspectives, and Prospects.” 349(6245).

Kitchin, Rob, and Tracey P. Lauriault. 2015. “Small Data in the Era of Big Data.” *GeoJournal* 80(4): 463–75. http://dx.doi.org/10.1007/s10708-014-9601-7.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. “Big Data. The Parable of Google Flu: Traps in Big Data Analysis.” *Science (New York, N.Y.)* 343(6176): 1203–5. http://www.ncbi.nlm.nih.gov/pubmed/24626916.

Lazer, David et al. 2009. “Computational Social Science.” *Science (New York, N.Y.)* 323(5915): 721–23. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2745217&tool=pmcentrez&rendertype=abstract.

Lecun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep Learning.” *Nature* 521(7553): 436–44.

Mackenzie, Adrian. 2017. *Machine Learners: Archaeology of a Data Practice*. Cambridge Mass.: MIT Press.

Olah, Chris et al. 2018. “The Building Blocks of Interpretability.” *Distill*. https://distill.pub/2018/building-blocks/.

Patton, Michael Quinn. 2015. “Sampling, Qualitative (Purposeful).” *The Blackwell Encyclopedia of Sociology*: 1–3.

Patton, Michael Quinn. 2002. *Qualitative Research and Evaluation Methods (3rd Edition)*. Thousand Oaks: Sage.

Pielstick, C. Dean. 2000. “Formal vs. Informal Leading: A Comparative Analysis.” *Journal of Leadership Studies* 7(3): 99–114.

Raghavan, Prabhakar. 2014. “It’s Time to Scale the Science in the Social Sciences.” *Big Data and Society* 1(1): 1–4.

Reinalda, Bob, and Bertjan Verbeek. 2013. “Leadership of International Organizations.” In *The Oxford Handbook of Political Leadership*, Oxford: University Press. http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199653881.001.0001/oxfordhb-9780199653881-e-039.

Rudin, Cynthia. 2019. “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.” *Nature Machine Intelligence* 1(5): 206–15. http://dx.doi.org/10.1038/s42256-019-0048-x.

Schechter, Michael G. 1987. “Leadership in International Organizations: Systemic, Organizational and Personality Factors.” *Review of International Studies* 13(3): 197–220.

Sebastiani, Fabrizio. 2002. “Machine Learning in Automated Text Categorization.” *ACM Computing Surveys* 34(1): 1–47.

Suri, Harsh. 2011. “Available from Deakin Research Online : Purposeful Sampling in Qualitative Research Synthesis.” *Qualitative Research Journal* 11(2): 63–75.

Syam, Niladri, and Arun Sharma. 2018. “Waiting for a Sales Renaissance in the Fourth Industrial Revolution: Machine Learning and Artificial Intelligence in Sales Research and Practice.” *Industrial Marketing Management* 69(November 2017): 135–46. https://doi.org/10.1016/j.indmarman.2017.12.019.

Tallberg, Jonas. 2010. “The Power of the Chair: Formal Leadership in International Cooperation.” *International Studies Quarterly* 54(1): 241–65.

Teddlie, Charles, and Fen Yu. 2007. “Mixed Methods Sampling.” *Journal of Mixed Methods Research* 1(1): 77–100.

Thorn, I. Marlene. 2012. “Leadership in International Organizations: Global Leadership Competencies.” *Psychologist-Manager Journal* 15(3): 158–63.

Venturini, Tommaso, Kari De Pryck, and Robert Ackland. 2020. “Bridging in Network Organisations the Case of International Panel on Climate Change (IPCC).” Forthcoming. www.tommasoventurini.it/wp/wp-content/uploads/2020/07/Venturini_etal_BridgingNetworkOrg.pdf

Venturini, Tommaso, Pablo Jensen, and Bruno Latour. 2015. “Fill in the Gap: A New Alliance for Social and Natural Sciences.” *Journal of Artificial Societies and Social Simulation* 18(2): 11. http://jasss.soc.surrey.ac.uk/18/2/11.html.

Venturini, Tommaso et al. 2014. “Climaps by Emaps in 2 Pages (A Summary for Policy Makers and Busy People in General).” *Social Science Research Network* (ID 2532946). http://papers.ssrn.com/abstract=2532946 (February 18, 2015).

Woolgar, Steven. 1985. “Why Not a Sociology of Machine? The Case of Sociology and Artificial Intelligence.” *Sociology* 19(4): 557–72.