Old Questions, New Techniques: A Research Note on the Computational Identification of Political Elites

Jacqueline Hicks,1 Vincent Traag2 and Ridho Reinanda3

Elite Network Shifts Project, Royal Netherlands Institute of Southeast Asian and Caribbean Studies (KITLV), Reuvenstraat 2, 2311 BE Leiden and Intelligent Systems Lab Amsterdam (ISLA), University of Amsterdam, Science Park 904, 1098 XH Amsterdam

hicks@kitlv.nl
r.reinanda@uva.nl
traag@kitlv.nl

Abstract

This paper presents a new method of identifying a nation's political elite using computational techniques on digitised newspaper articles. It begins by describing the three most widely used methods of identifying political elites: positional, decisional and reputational. It then introduces the “reported elite method”, exploring the kinds of elites it detects and how well it reflects the composition of political elites in our case study of Indonesia. Compared to the other existing methods, we find that our method casts a much wider net when searching for political elites, resulting in many more people from civil society, far fewer formal politicians, and challenging conventional notions of who is a political elite. The method has two major underlying assumptions: (1) the newspapers from which the texts are drawn are free and fairly representative and (2) political power can be inferred from frequent appearance in newspapers alongside other frequently appearing individuals in computational “communities” of political elite.
Keywords

political elites – Indonesia – named entity recognition – community detection

Context: Existing Methods of Identifying Elites and their Implications

It is a question that has endured for centuries: Who rules? While the field of elite studies is made up of many strands – elite recruitment, their social and attitudinal characteristics, elite-mass linkages, and interaction with institutions – this overarching question remains. Are we ruled by a small core of elites with shared backgrounds and overlapping interests? Or are there several distinct elite groups whose mutual competition limits the possibility of domination?

Unlike other areas of political enquiry focusing on institutions or social and economic structures, elite studies seem to promise rich empirical research of relatively more observable “events”. Yet, the conceptual fuzziness of terms such as power, interests and elite has a great influence on the methods used to study elites and, by extension, the answers to this question.

One of the major methodological decisions facing the researcher is how to identify political elites. The three main options are by now well-rehearsed, having been debated in more or less the same way for several decades. First is the positional method, which follows the Weberian definition of elites as those in “command positions” at the top of major bureaucracies (Scott 2003). These typically include cabinet and legislative members, but the addition of top business and military personnel opens the possibility of an industrial-military “power elite” of the type that Mills first described in 1950s North America (Mills 1956).

A researcher’s decision about which elite positions to include will obviously influence the final conclusions about whether a cross-sectoral “power elite” exists. In methodological terms, Hoffmann-Lange points to the absence of clear “guidance on specifying the boundaries of the elite” in the positional method (2007:914), which means that each researcher must decide themselves which institutions and positions to include. Putnam also notes the method’s assumption that power is tied to the resources associated with both the position and its institution, so potentially missing those with more informal

---

sources of power. As he puts it, “power is never perfectly correlated with position” (Putnam 1976:16).

The positional method is, however, one of the most widely used since it is the most practical for researchers (Hoffmann-Lange 2007). This is confirmed by a review of the many national level elite studies that this journal has published over the last ten years. The tendency among these studies is to focus only on parliamentary and/or cabinet level elite, rather than casting a wider societal net to elites within the media, think-tanks, corporations or military. Perhaps this reflects an ideological viewpoint that the latter types of elite do not have sufficient political power to merit inclusion, but it could equally be the case that the identification of these types of elite is just more difficult so that researchers tend to avoid including them. As we will shortly see, this is a gap that could be filled by the reported elite method presented in the next section.

The second main way of identifying elites was conceived as a direct response to the findings of the “power elite” scholars – the decisional method. Its best known proponent, Dahl, called for more empiricist methods of studying a nation’s political elite, arguing that an elite’s power should not be inferred from institutional positions or relationships, but rather observed. A political elite is here defined as a “minority of individuals whose preference regularly prevails in cases of differences in preferences on key political issues” (Dahl 1958:464). They are the active participants in decision-making processes, to be found by following a trail of documents and interviews related to particular decision-making processes.

The meticulous demands of this type of research on actual decisions tends to limit the analysis to only one policy domain, making system-wide generalisations difficult. A researcher’s decision about which policy issue(s) to include will therefore greatly influence any more general conclusions about who rules. As Bacharach (1970) pointed out, the decisional method also misses those who are influential but not actively involved in the decision, thereby effectively excluding the possibility of detecting any undocument ed exercise of political power.

The third main method is based on reputation and involves identifying a number of elite from either their official positions or the advice of an expert panel, and then expanding this selection by asking those elites to nominate others. Proving what a rich source of insight this debate has been, it was again conceived in relation to Dahl’s insistence on empiricism. Originally practiced

5 See for example, Best 2007; Best, 2009; Higley 2011; Rodriguez-Teruel 2011.
by Hunter (1953) and nicely conceptualised by Isaac (1997), it assumes that elites maintain power based solely on the relations in which they participate (Isaac 1997). If someone is perceived to have political power by their peers, then it may not perfectly reflect the possession or exercise of such power, but it nevertheless can have similar and lasting effects (Ellis 2006). The unobserved exercise of power is therefore allowed for with this method. Since the respondents that tend to be accessible to researchers are likely to have knowledge of fairly limited regional or policy-specific spheres of influence, the kind of grand cross-sectoral relationships that determine a national “power elite” are likely to be underplayed. It has, however, been used to produce descriptions of power elite-type groupings at the local level under the rubric of “community power research.” Since the method relies on people knowing each other to nominate, it is perhaps not surprising that the resultant elite population tends to have indications of a shared personal background or interests.

All three of these methods have their strengths and weaknesses in uncovering different political elites, as defined by the type of political power they hold. These can be charted according to three different types of power – (1) formal to informal channels of influence (2) national system-wide coverage to policy sector specific (3) potential power associated with office to overt observable power. This is, of course, a highly schematic representation which also obscures the reality of elites who exercise or hold different types of power at the same time or in relation to different objects, but serves the analytical purpose of helping to make connections between the methods and the types of results.

The next section will introduce the “reported elite” method and examine what categories of elite are detected by the method when used in practice on the national case study of Indonesia.

Test: The “Reported Elite” Method and its Results

The basic technique of the reported elite method is the automatic extraction of people's names from digitised newspaper articles. A major sub-task in the field of automatic information extraction, “named entity recognition” involves designing computer code-based methods to automatically recognise, extract and categorise entities such as names, places and organisations. It is not a search task in its usual sense where specific words are manually input to find them in documents. Rather, the computer code, or “algorithm”, finds these words based on linguistic characteristics like word capitalisation and word position in relation to nouns and verbs.
In our test, the process of locating a population of political elites began with Reinanda extracting all named entities from the digitised newspaper articles and then filtering to include just persons. Since we were interested in elites, these persons were again filtered to only include those who appeared in more articles than average.

The next stage was to choose only those persons who frequently co-occur within the space of one sentence with other frequently occurring persons. Traag detected groups of people that frequently co-occur using a technique known as “community detection” with social network analysis software (Traag et al 2011). These steps are illustrated in Figure 1: (a) names are detected in a sentence using named entity recognition (b) a co-occurrence network is created (c) and (d) communities in the network are detected.

The way the persons are grouped in the final image in Figure 1 above represents how often they co-occur together in one sentence. Since there is a tendency for sports celebrities, for example, to be mentioned together in sentences with other sports celebrities, they form one “community” distinct from, say, politicians or entertainment celebrities. As the resolution parameter of the community detection was tightened, increasingly more fine-grained communities became apparent, so that communities reflect politicians from a particular country [Figure 1 (c)] and then, even more fine-grained, they reflect particular issues (e.g. state oil policies), scandals (e.g. corruption cases) or topics (e.g. terrorism) [Figure 1 (d)].

For the purpose of identifying a population of national elites, such community detection effectively allows the researcher to quickly identify and discard the groups of sports celebrities or foreign elites, reducing extremely large quantities of persons down to a more manageable number. For someone interested in seeing which people frequently appear together in connection to certain issues then, a higher resolution would be better.

To further illustrate these techniques and examine the results in greater detail, we tested them on a corpus of digitised newspaper articles about Indonesia, choosing one particular year: 2008. The number of persons who

---

6 The persons appearing Figure 1 (c) and (d) were grouped into communities computationally, but I manually selected a small number of names from each group for illustration.

7 We used a collection of newspaper articles from a news service specialising in Indonesian politics called Joyo (http://www.joyonews.org). The number of articles for 2008 was 19,604. Because these articles were originally selected by the Joyo team for their relevance to Indonesian politics, we also successfully tested these methods on the freely available New York Times corpus to check that community detection could still identify groups of sports, entertainment or politicians in an unselected corpus.
FIGURE 1

Steps from named entities to communities.

(c) From left clockwise: Timor Leste; USA; Japan; Malaysia and Thailand. (d) From left clockwise: A mining disaster; state oil policies; terrorism; a corruption trial; a trial on the murder of a human rights activist.
appeared in more than the average number of articles (more than 3) published in 2008 was 5431. After community detection was applied with social network analysis software, I (the primary author) selected only the largest community which contained 2499 persons and then 1500 of those with the highest frequency. I then spent a few days going through that list of 1500 to manually remove errors (places, concepts and organisations = 148), partial names which did not allow the identification of actual people (=263) and those who were obviously not Indonesian elites in 2008 – mostly non-Indonesians, historical or recently deceased figures and reporters (=274). This resulted in a final elite population of 815 persons. If I had wanted to use these results for further research, I would have done a second round of manual filtering to around 500.

Figure 2 shows a visualisation of our final elite population.8

Following the conventions of elite studies, I compared our results against a manually collected list of positional elites for 2008. Two other examples of such comparisons are: Lal (1980), who tested all three methods at community level in India and Hoffmann-Lange (1987), who reflected on the differences between those elite identified by position and by social network analysis based on a survey at national level in West Germany.

As already noted, drawing up a list of positional elite is highly subjective and the types of elites chosen vary widely from study to study. While Mills (1956) focused on military, business and political elites, Lal (1980) also included the leadership of more civil society groups, like trade unions, civic bodies and cultural organisations in addition to the usual ministers, legislative members and business executives.

Hofmann-Lange (1987) included all of the above in her positional lists further adding heads of universities and mass media managers. Such variance is to be expected as different types of political elites are more or less prominent in different countries, but must be balanced with an effort towards systematisation for comparative research.

The positional list of Indonesian elite that I drew up included cabinet ministers and national legislature members; the top 50 richest businesspeople compiled by Globe Asia (December 2008); top departmental bureaucrats

8 To see a high resolution text-searchable and colour-coded by category version of this figure, and an excel file list of our final reported elite population, see F1 and F2 at: https://github.com/Jacky19/Elite-Network-Shifts/tree/master/Old%20Questions%20New%20Techniques%20Additional%20Files. The colour codes are: Yellow = Politician; Blue = Civil Society; Green = Military; Red = Bureaucrat; Purple = Business.
according to scale of seniority (echelon one); top national and regional military commands; head judges of the supreme and constitutional court; regional governors; the leaders of the biggest political parties and the heads of several quasi-governmental organisations. I did not include any civil society actors since in Indonesia they tend to have individual influence rather than institution-based influence, often moving between different organisations in a weakly institutionalised sector. This positional list contained a total of 1178 persons.9

The intention of a comparison is not to measure reported elites against positional elites as if the latter is some sort of gold standard. Rather, I aim to

9 To see an excel file list of the positional elites, see F3 at: https://github.com/Jacky19/Elite-Network-Shifts/tree/master/Old%20Questions%20New%20Techniques%20Additional%20Files.
discuss the different categories of elites that each method uncovers so that researchers can better understand when either method should be chosen, or indeed the ways they complement each other.

The results showed that two hundred and sixty three (22%) of the 1178 positional elites were found in the reported elite population of 815 people. By comparison with the other studies mentioned above, Hoffmann-Lange (1987:44) reported that “about one third” of positional elites could be found in her survey and network analysis-based population, while Lal (1980:34–35) found that 55% of his positional list also appeared in a list using the decisional method and 49% in a reputationally-defined list.

In our test, the type of elites uncovered by each method can be seen below in Figure 3. I categorised the “reported elite” myself manually, although there are some computational methods that can at least partly do this automatically. As Figure 3 shows, the reported elite method finds more than double the proportion of business and military actors as the positional method, and almost half the amount of politicians. It can also be seen that a substantial percentage of the reported elites are from civil society.

Few studies empirically test multiple methods, and of those even fewer categorise the different types of elites found by each. Hoffmann-Lange (1987) finds that her survey and network method uncovered around double the relative proportion of politicians, half of business and a tenth of the military compared to her positional list. The remaining categories – civil servants, and various sectors of civil society – were around the same. Higley’s study of Australian national elites started with a positional list who were then asked to nominate others that were considered influential and not already in the sample. He notes that very few extra elites were nominated, leading him to conclude a substantial overlap between the two methods. The elites missed out by the positional method in that study were mainly judges and second tier civil servants (Higley 1979:70).

A second test compared the contents of a book which lists the 100 most influential Indonesians in 2008 (Ali 2008) with our population of reported elite. After removing eight of those 100 who were sports and entertainment figures, it was found that 70 (76%) of the 92 remaining “most influential” were found among the reported elites.

A third test of the reported elite method compared the results with some of the qualitative literature on policy influentials in Indonesia. Although there were no civil society actors in the positional list, even if I had included 50 people, it would still have only been around 4% of the total number of positional elites. This is a very big difference when compared to the number and proportion
of civil society actors found in the reported method. This is partly an artefact of the nature of the media which seeks comment from commentators as well as actors, but also what we would expect in the Indonesian context where, although patchy and deeply problematic, civil society influence on policy has grown substantially in recent years (see for example, Rosser et al 2005; Ito 2011; Maclaren et al 2011).

The relatively low proportion of positional politicians found in our network also reflects the Indonesian situation. When broken down further into sub-categories, these results showed that relatively few of the total number of legislative members on the positional list made it into our network (14%) compared to the other types of politicians – cabinet members (100%) and regional governors (66%). Again, this could be expected in the Indonesian context where it is known that the rank and file of legislative members are almost completely marginalised from decision-making by a select few party leaders (Sherlock 2003).
In sum, the degree of overlap with a positional list (22%), and the ability to capture a large percentage of an external list of 100 of Indonesia's most influential people (76%), shows that the reported elite method has promise. In our test, it captured many more civil society actors and many less politicians than the positional method, but a brief review of the qualitative literature on Indonesian policy influentials gives some explanation about why this may be an accurate reflection of the reality on the ground.

The next section will now return to consider the reported elite method in terms of the kinds of assumptions and implications for understanding patterns of political power that were laid out in the first section.

Discussion: Assumptions and Implications of the “Reported Method”

Although the sources and techniques outlined here are not new, this is the first time they have been put together in such a way and within the context of elite studies. Newspaper articles are, of course, extensively used by social scientists to better understand all aspects of elite behaviour, particularly by researchers using the decisional method of elite identification and analysis. As a political scientist, my understanding of a country’s politics is substantially informed by reading newspapers, and just as in my qualitative research, any use of newspapers for computational techniques must consider the degree to which they are both free and representative.

Although comparatively still new, extracting entities, measuring their co-occurrence and performing network analysis to find significant clusters of co-occurrence have also been used before, but not for the task of identifying elites. In recent decades, social network analysis methods have been profitably used to study elites to both locate them and elaborate on their relations. These are different from the reported elite method presented here because they generally draw information from surveys and/or interviews, basing their “links” (relationships) on the elites’ answers to questions about their “interaction partners” – those they communicate with (Laumann and Pappi 1976; Moore 1979; Higley and Moore 1981; Hoffmann-Lange 1987). These other studies all begin with a list of elites based on the positional method, even if their main purpose is to further refine those lists with network analysis of the information gathered in the surveys.

The reported elite method does not share the decisional method’s insistence on the empirical observation of the exercise of power, but rather infers power. However, instead of inferring that power is tied to the resources of an
institution (as positional) or the perception of power among their peers (as reputational), the reported method infers power from their appearance in the media. But it is more than just frequently appearing in newspaper articles, it is people who co-occur more frequently with others in a cluster or “community” than with those outside of it. The value of this co-occurrence strategy is that it effectively builds a kind of virtual elite terrain that people must appear on if they are to be included in the reported elite population. Even someone who appears a hundred times in the newspaper articles will not be included unless they appear inside one of these communities.

Fundamentally then, the principles of co-occurrence and frequency dictate who will be captured using this method. Like all computational tools that analyse texts, the information extraction and analysis is based on a kind of naïve word counting, divorced from context or meaning. But because there is some merit to the idea that some frequently co-occurring people in newspapers hold some kind of power, in the context of identifying a population of political elite, counting words can be revealing.

The focus on frequency means that the method results in all sorts of individuals who would not ordinarily be considered elites in the sense of wielding political power – convicted terrorists or a president’s wife, go-betweens in corruption cases and famous former politicians who have just died. But looking through the results forced me to consider whether some high profile terrorists should be recognised as political elites, for example, or a president’s spouse – they both have the potential to exercise considerable influence on the political process under some circumstances. The computational methods alone cannot judge which politicians’ spouses, terrorists or economic commentators are politically influential enough to be considered a political elite, but a researcher can.

A common misconception about computational techniques is that they aim to replace a researcher’s own judgement, but in my view they are better understood as “knowledge discovery tools”, a kind of second generation search engine. It is valuable enough to see lists of frequently co-occurring people in news stories to merit a few days of manual filtering, particularly for anyone interested in knowing who is most reported in relation to a particular issue. The reported elite method can also be used quite profitably in conjunction with the other methods rather than as a replacement. Indeed, exploring the overlap between the results of the different methods can itself reveal something of the nature of a particular political system.10

10 Putnam (1976: 18), for example, posits that a convergence of formal and informal types of power represents a stable political system as oppose to one in flux. In methodological
Compared to the other methods then, the reported elite method casts a much wider net since the boundaries of an elite population are not presumed at the outset by the researcher making decisions about which institutions and positions to include (as in the positional method) nor which policy domains (as in the decisional and reputational methods). Rather, it opens the possibility of finding other kinds of political elite – those from business, military, judiciary and civil society – as well as the usual legislative and cabinet members. It can also cover multiple policy domains and so has a better chance of allowing generalisations across national systems, and can account for informal types of power, unfettered from the institutions which are pre-judged by the researcher to be politically significant. Since journalists are generally obliged to obtain quotes from officials in a news story, it could also be argued that formal power-holders are likely to be fairly well represented in a reported elite population. On the other hand, there will always be a handful of powerful elites who make great efforts to stay outside the media glare, which this method would be unable to capture.

In terms of practicality, the automatic nature of the method sounds compelling, as if pressing a few buttons will complete the task. At the time of writing, however, these techniques still require some knowledge of coding, even when using the off-the-shelf named entity recognition software such as the most widely used one produced by Stanford.\footnote{http://nlp.stanford.edu/software/CRF-NER.shtml} Social network analysis software such as Gephi can be used without any coding skills, but there is a certain art to community detection that must also be learned. In addition to the “data cleaning” – the manual filtering of results discussed above – a good deal of underlying data preparation is also needed at the beginning of the process. The digitised texts must conform to stringent technical specifications and disambiguating people’s names\footnote{Often a single person will be referred to in a corpus by several different names. For example, “Susilo Bambang Yudhoyono” can also be referred to as “President Yudhoyono” or “SBY”. The process of disambiguation ensures that all these different ways of referring to one person are grouped together, effectively recognising them as the same person. The amount of manual work involved in this task depends on the availability of “knowledge sources”, such as Wikipedia, for the different types of elites.} is a laborious task which is still only partially accomplished by automatic means. Nevertheless, like the positional method, the reported elite method can be undertaken relatively efficiently compared to
the field interviews needed to find reputational elites or the intensive archival research needed for the decisional method.

Perhaps, though, what makes the reported elite method most promising is its potential to act as a foundational structure upon which further developments in field of automatic extraction can be built. As presented here, I have not tried to infer the relative importance of the automatically extracted persons from their centrality in the network as I find it difficult to justify based on co-occurrence alone. But it is already possible to go some way towards automatically adding biographical information about the elites by automatically linking them to their Wikipedia pages or automatically extracting some simple relations from the text, such as family or work relations over time (Reinanda et al 2013; Reinanda and De Rijke 2014). This would ultimately enable a new generation of large scale, complex analyses of our elites and their shared interests, answering not only the new questions that are frequently called for in the field of digital humanities, but also the old ones which have been around for centuries.

References


