MAPPING WIND ENERGY CONTROVERSIES ONLINE

INTRODUCTION TO METHODS AND DATASETS

Author: Anders Kristian Munk (akm@learning.aau.dk)
Data visualization and design: Anne-Kirstine Bøcher Ellern (www.akellern.com)
Student collaborators: Loretta Møller, Linn Wagner Korsgaard & Line Østereng Jørgensen
November 2014

Wind2050 (www.wind2050.dk) is sponsored by the Danish Council for Strategic Research and led by the Danish Technical University
Principal investigator: Kristian Borch (krbo@dtu.dk)

The Danish Council for Strategic Research

DTU Management Engineering
Department of Management Engineering

AALBORG UNIVERSITY
COPENHAGEN

MAPPING WIND ENERGY CONTROVERSIES ONLINE

Introduction to methods and datasets

As part the Wind2050 project funded by the Danish Council for Strategic Research we have mapped controversies on wind energy as they unfold online. Specifically we have collected two purpose built datasets, a web corpus containing information from 758 wind energy websites in 6 different countries, and a smaller social media corpus containing information from 14 Danish wind energy pages on Facebook. These datasets have been analyzed to answer questions like:

How do wind proponents and opponents organize online?
Who are the central actors?
And what are their matters of concern?

The purpose of this report is to provide a description of the datasets, the methods we used to build them, and suggest possible scenarios of use.

Research shows that public opposition to wind energy is both competent, well informed, and irreducible to narrow NIMBY'ist concerns. It follows that opponents should be recognised and their concerns taken into account in a democratic planning process (e.g. Devine-Wright 2005, Wolsink 2006, 2007, Aitken 2010, Haggett 2011). Studies of other public knowledge controversies confirm the need to take public protest seriously in the pursuit of democratically robust solutions (e.g. Nowotny 2003, Callon et al. 2009). The controversy maps presented here recognise the disagreements as they appear online and provide a visual means of exploring them further.

The report does not represent a final analysis, and does not attempt to take sides or reach settlement. It is meant as a first reference point for anybody who wants to engage with the data, make use of it for their own research projects, collaborate with us, provide commentary or criticism, or inquire into the methods we have used.

Please direct your inquiries about the datasets and the methods to Anders Kristian Munk at the University of Aalborg: akm@learning.aau.dk

Note that the online mapping of wind energy controversies is a subproject under Wind2050, a much wider research collaboration with multiple work packages pursuing multiple research questions. If you have questions about the Wind2050 project as such please direct them to the primary investigator, Kristian Borch, at the Danish Technical University, krbo@dtu.dk, or consult the project webpage, www.wind2050.dk.

Why map controversies online?

There are two overall reasons for taking on this work. The first has to do with controversy mapping in general, the second with the benefits of doing it online.

Controversies, by definition, are situations where the actors disagree. In practice these situations are complicated by the fact that the actors do not often agree on the precise nature of their disagreements. It is rarely the case that one simple question can be posed around which all actors can easily clarify their respective positions. Controversies are complicated by the lack of agreement about what the important questions are to ask and who the authoritative experts would be to answer them.

This is also the case with wind. One could of course ask if wind energy is a good or a bad idea? This would allow actors to be generally for or against it, but a host of additional clarifications would be needed. You might be for wind energy, but against industrial scale wind farms. You might be against wind energy because you question its green credentials, or because you do not recognize the need for a greener energy supply at all. You might be for larger wind turbines because you worry about the health effects of burning fossil fuels, but you could also be against them because you worry about the health effects of low-frequency noise emissions. There are in other words numerous issues around which actors can potentially organise in different ways.

To map a controversy is an attempt to get an overview of these disagreements according to the actors themselves. This is a key point: the method tries to be agnostic about what the important issues, the decisive questions, or the trustworthy knowledge claims should be. It acknowledges the impossibility of taking an a priori stance in such matters without taking a stance in the controversy itself. Instead, controversy mapping works from the pragmatic principle that whatever makes a difference to the actors in the controversy should be charted and given a place on the map relative to its importance.

In some ways this puts controversy mapping in direct opposition to what many actors in a controversy would like to see happen. Having a stake in a controversy is precisely to have a stake in defining the important issues, the decisive questions, and the trustworthy knowledge claims. It is therefore important to stress that controversy mapping is not a method for making decisions or reaching conclusions, but a method for making maps. Maps, by their very nature, can be picked up by actors with different agendas and made to serve different purposes. They can prompt questions, or help you find your way, but they alone cannot provide the answers.

It is the stated ambition of Wind2050 to understand and take seriously the controversies that unfold around wind energy in Denmark and abroad. To do so the project needs to know who the actors are and how they

organise around their respective matters of concern. This is the reason why we are mapping wind energy controversies. The reason why we are doing it online requires a separate set of explanations.

It is generally recognized that discussions on wind energy take place in a highly interconnected network that transcends national borders and offers various channels for information exchange. This is true for both proponents and opponents of wind power. It is also true that much of this networking activity takes place on the open Web. There are websites and social media pages dedicated to various kinds of public wind advocacy work as well as various kinds of public wind protest. These networked online issue spaces make a difference to the controversy on wind energy, they are part of its anatomy, and it has therefore been a key objective to map them.

Whereas the presence of an online issue space where different kinds of protest and advocacy unfold on national and international scales is known from many other controversies, the consistent level of online local engagement that is characteristic for wind is more rare. Individual wind power projects very often have websites or Facebook pages associated with them, made by people who promote them, oppose them, or both. This is interesting because it promises a dataset with higher granularity on the local level. In short it makes it easier to ask questions and explore avenues of analysis on the level of concrete wind energy projects, rather than being restricted to national or international issue organisations speaking on behalf of broader constituencies.

It is important to note that the role of these online issue spaces is already being hypothesized by the actors in the controversy. Some proponents speculate that wind protest is being organised by a small number of issue professionals that draw their inspiration and source their arguments from their international networks. Some wind opponents speculate that there is no such thing as grassroot online wind advocacy work that is not orchestrated and sponsored by the wind industry. This provides a further incentive to map the networks that are subject to such speculations.

In tandem with the online controversy mapping we have also undertaken a series of qualitative, offline case studies. An auxiliary purpose of the online datasets is to test the ability of digital traces to provide complementary insights for traditional case study work on wind energy controversies.

To read more about controversy mapping, see for example:

Beck, G., & Kropp, C. (2011). Infrastructures of risk: a mapping approach towards controversies on risks. Journal of risk research, 14(1), 1-16.

Venturini, T. (2010). Diving in magma: How to explore controversies with actor-network theory. Public understanding of science, 19(3), 258-273.

Whatmore, S. J. (2009). Mapping knowledge controversies: science, democracy and the redistribution of expertise. Progress in Human Geography.

Yaneva, A. (2012). Mapping controversies in architecture. Ashgate Publishing, Ltd.

To read more about digital methods, see for example:

Marres, N. (2004). Tracing the trajectories of issues, and their democratic deficits, on the Web: The case of the Development Gateway and its doubles.Information Technology & People, 17(2), 124-149.

Marres, N., & Rogers, R. (2008). Subsuming the ground: how local realities of the Fergana Valley, the Narmada Dams and the BTC pipeline are put to use on the Web. Economy and Society, 37(2), 251-281.

Rogers, R. (2013). Digital methods. MIT Press.

Venturini, T. (2012). Building on faults: how to represent controversies with digital methods. Public Understanding of Science, 21(7), 796-812.

How is the data archived and published?

We are mapping the controversy as it appeared on the public web in Spring and Summer 2014. All information contained in the two datasets was in other words publically available at the time of harvest. It is nevertheless important to stress that this kind of online public availability does not automatically mean that the information is in the public domain in the sense of having been made explicitly available for public use. This raises ethical questions regarding the storage and usage of the datasets.

We are archiving the information contained in the two corpora for research purposes and will not make this archive available in its raw form. We will however make limited datasets available with the following considerations in mind:

On the web we have harvested the html from 758 websites. We are making public a dataset containing the hyperlink structure between these websites, including their outlinks to neighbouring websites. We are also performing a series of textual queries on the stored information from the websites. We will make the results of these queries public as separate datasets, but we will not publicize the text from the websites in its original form (nor in any other full text format). The web corpus does not allow you to identify statements made by individuals. It can only be queried website by website.

On Facebook we have not applied for membership of any groups, but restricted ourselves to 14 pages which are already open and do not require approval from an administrator. We have anonymized the names of individual users interacting on these pages. We have archived a record of these interactions and we may perform a series of textual queries on this data at a later stage. The results of such queries would be made public as separate datasets. We will again not publicize the text from these pages in its original form (nor in any other full text format).

We strive to adhere to the 2012 recommendations of the Ethical Committee of the Association of Internet Researchers (AoIR), in particular that a case based approach is necessary. As an example it cannot be determined ex ante if the information stored on open Facebook pages is private or public in character. Rather it depends on a concrete evaluation of the interaction between users on a specific page. It is a well established research practice that analysis of news articles, including letters to the editor or discussion threads on online news platforms, do not require informed consent from the authors. Whether or not the interactions on a Facebook page can be said to be of a similarly public nature will vary and require the researcher to make a situated assessment from case to case. We have done that and found that the interactions on the Facebooks pages we are using qualify under such criteria.

It is particularly important to be as open as possible about datasets when they have been collected to map a controversy. As explained above actors in controversies are likely to have strong stakes in how their concerns are being charted and visualized. Opening that process to public scrutiny, eliciting feedback from stakeholders, and enabling symmetrical access to the data is in our opinion a goal worth pursuing.

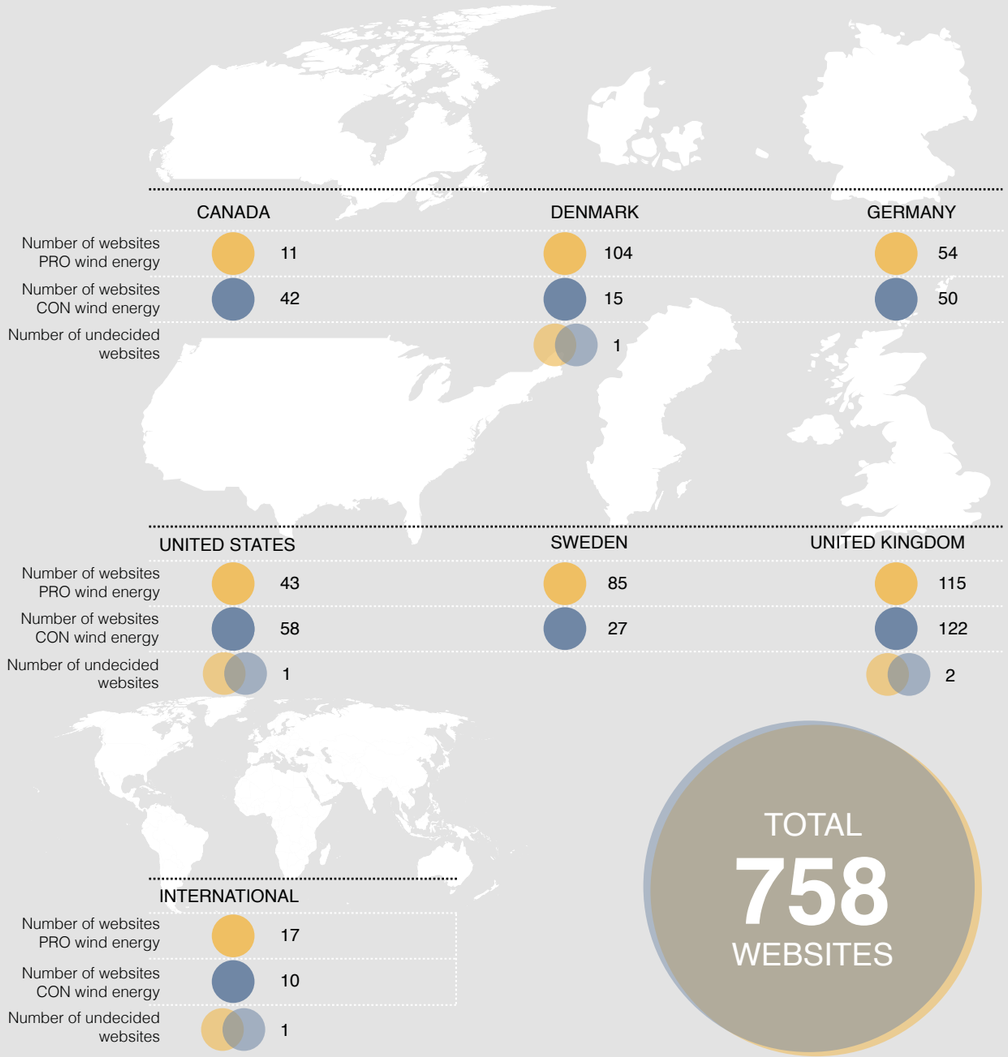
Finally, taking up a recent point made by Neuhaus and Webmoor about agile ethics (2012), we have included our own website (www.wind2050.dk) in the web corpus. We have also made use of open source tools for our data collection process to be as transparent as possible and allow others to redo our datasets in different ways. This report is part of that effort.

---

[1] The recommendations of the AoIR are available here: http://aoir.org/reports/ethics2.pdf

THE WIND2050 WEB CORPUS

The Wind2050 web corpus contains information about 758 wind energy websites from Denmark, Sweden, Germany, Canada, the U.K., and the U.S. (plus a number of international sites). The dataset allows you to query the content of each website and explore its hyperlink connections with other websites. The information stored was publicly available on the websites at the time of collection, but the dataset is not dynamically updated. It represents a snapshot of the wind energy issue space as it appeared on the web in May and June 2014.



Protocol for building the web corpus

We built the web corpus through an iterative process. We began with a seed list of websites that we knew were exclusively dedicated to the issue of wind power. The seed list contained opponents, such as the European Association Against Windfarms (epaw.org), as well as proponents, such as the Danish Wind Turbine Owners' Association (dkvind.dk), from all of the six countries we were interested in (Denmark, Sweden, Germany, Canada, the U.K. and the U.S.). This ensured that our starting points would likely point us in the direction of other wind relevant websites in our countries of interest. It also ensured that our starting points would point us towards websites representing different positions in controversies over wind energy.

Using a web crawler[2] we then harvested all the hyperlinks from the websites on our initial seed list. This allowed us to discover their neighborhood, i.e. other websites receiving hyperlinks from websites on the seed list. The discovered neighbors were qualitatively assessed to decide if they were wind specific, i.e. dedicated to the issue of wind power in a similar manner to the starting points. Based on this assessment we compiled a new list of websites that we fed to the web crawler in order to harvest all hyperlinks and discover yet another and more expansive list of neighboring sites.

We reiterated this process three times to obtain a list of 758 wind specific websites from which we have scraped all hyperlinks and archived all textual content as html.
The qualitative assessment of each website had several purposes, namely:

- To determine if the website was wind specific. This was the criteria used for including a website in the corpus in the first place (see details below)

- To tag the website according to
1. *Its stance on wind energy. We used 'pro', 'con', and 'undecided'.*
2. *Its country of origin (including 'international')*
3. *The placement of the wind project(s). We used 'offshore', 'onshore', 'nearshore', and 'NA' since most websites did not fall clearly into one of the other three*
4. *The type of website. We used 'industry', 'issue', 'government', 'research', etc., but in practice it turned out to be difficult in many cases to make clear distinctions.*

- To get a sense of the different issues at stake in controversies over wind energy. In turn this ongoing work contributed to the construction of the issue dictionary : a list of wind related issues organised as query design with search terms for tracing the presence of these issues across web pages.

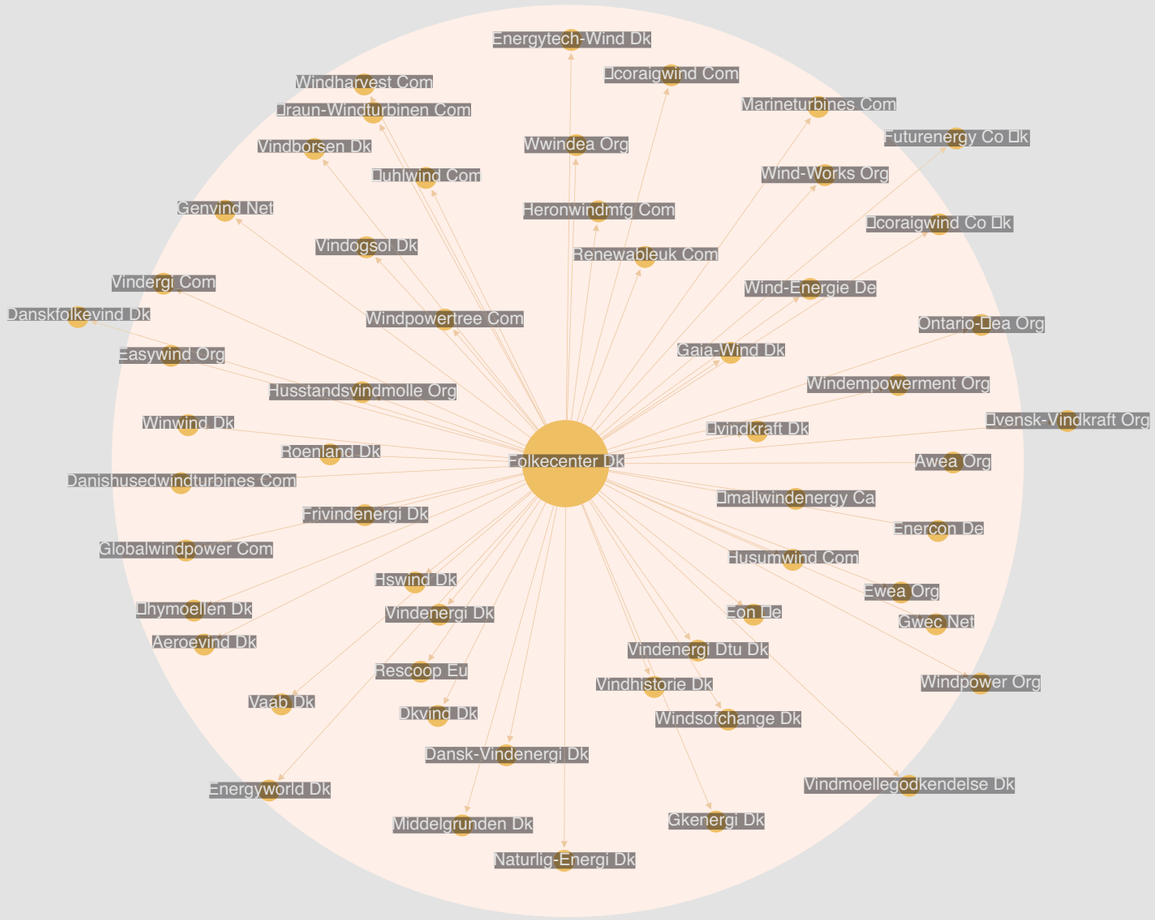
To illustrate some of the challenges faced in the qualitative assessment of websites we can take the Nordic Folkcenter for Renewable Energy (folkecenter.dk) as a case in point. It is relatively trivial to establish that this is a Danish wind energy proponent. It is much less clear if the site can be said to be wind specific. The Folkcenter engages in a range of initiatives as part of its practice to, as it states, "achieve measurable increases in the utilization of renewable energy technologies

2. We used an open source tool called Hyphe (also known as the Hypertext Corpus Initiative) that has been developped by the SciencesPo médialab: http://hyphe.medialab.sciences-po.fr/demo/. The code is available on GitHub: https://github.com/medialab/Hypertext-Corpus-Initiative.

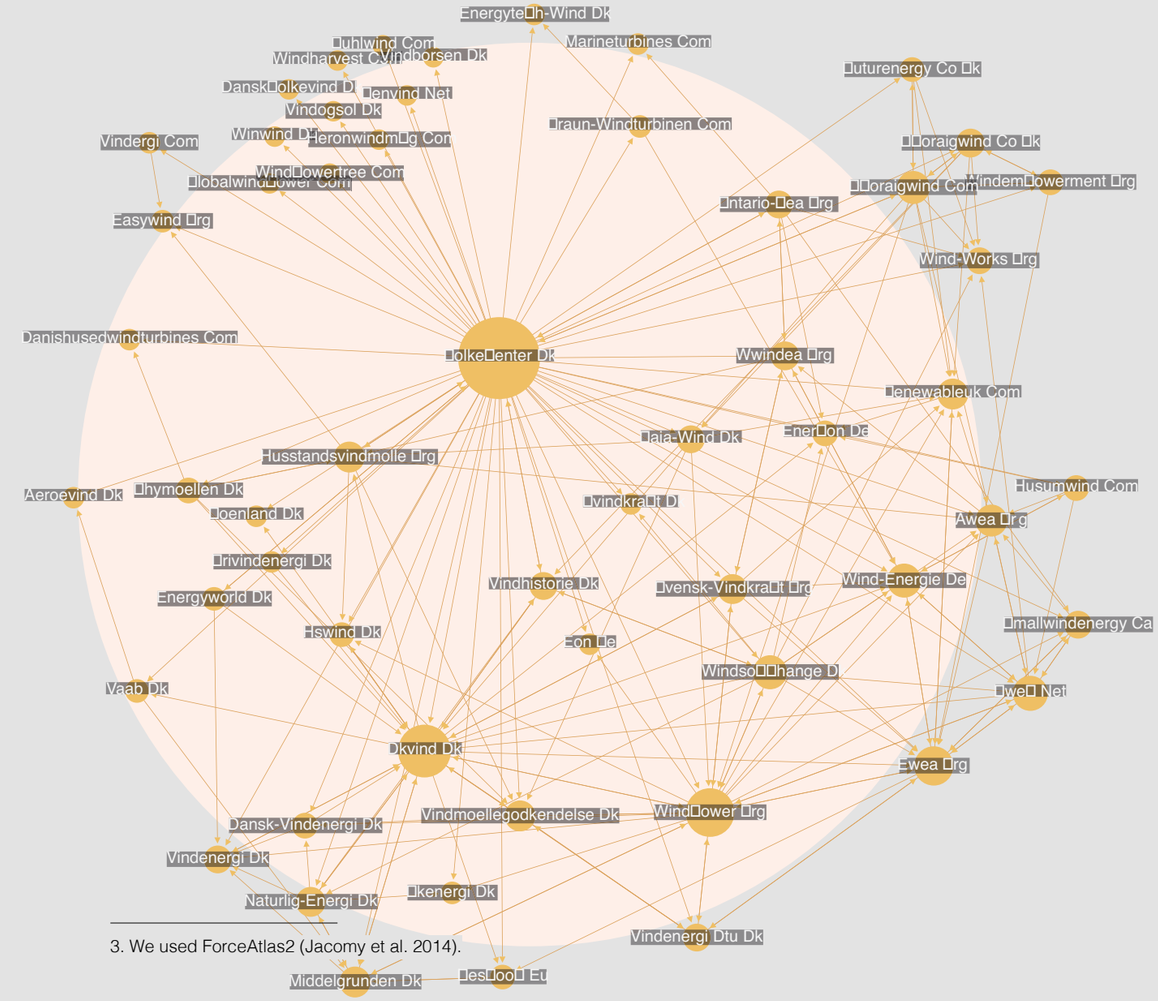
and thereby significant reductions in environmental pollution associated with energy use in Denmark and elsewhere". This is not restricted to wind energy. And yet when you read through the website it becomes clear that wind energy plays a central role in this work. We therefore decided that it made sense to include folkecenter.dk in the corpus. Whereas most websites were much easier to classify, this was an individual decision taken for all websites based on a qualitative assessment.

The analysis of the hyperlinks harvested from folkecenter.dk confirmed the decision to keep it in the corpus. Below is a rendering of the immediate neighborhood of other wind specific sites that receive hyperlinks from folkecenter.dk. It tells the story of a website that is well connected to other proponents of wind energy in Denmark and abroad. It confirms that wind energy is indeed a central concern for the Folkcenter.

To illustrate the iterative process of harvesting hyperlinks from an expanding corpus of discovered wind specific websites we can also look at what happens when the immediate neighbors of folkecenter.dk are crawled using Hyphe. Once we have crawled the neigbouring sites and know how they link to one another we can visualize them as a monopartite



network with directed edges (websites connected by hyperlinks pointing from one to another). Below we have included the hyperlinks between the neighboring websites of folkecenter.dk and spatialized the resulting network with a spring-based algorithm. This allows us to see emerging clusters and explore which websites appear central. It is not surprising that folkecenter.dk, our starting point, is the most connected node, but we can begin to see that other websites, such as The Danish Wind Industry Association (windpower.org), the Danish Wind Turbine Owners' Association (dkwind.dk), or the European Wind Energy Association (ewea.org), are also receiving a relatively high number of links from other websites in the network. We can say that these websites are considered authorities by the emerging cluster of Danish and International wind proponents emerging below and to the right of folkecenter.dk. This makes it particularly interesting to explore what other websites these authorities link out to. The Danish Wind Turbine Owners' Association, for example, maintains a page with links to its members. The websites of these members would then become part of the next iteration of crawls.



3. We used ForceAtlas2 (Jacomy et al. 2014).

The idea behind this iterative method of building the web corpus is to follow the actors and let them point out other interesting websites, instead of, say, delegating that task to a search engine. This does not mean that the web corpus is exhaustive. Far from it - there will always be websites that have either been missed or were not in existence at the time the dataset was built. What we can say is that by following the actors we take advantage of the fact that many wind specific websites curate wind specific link collections, valuable resources when building a dataset for controversy mapping. We could also say that it is a way of imposing as few extraneous delineations on the controversy as possible.

It would have been possible to choose a more automated and structured approach where only the starting points were wind specific, but the resulting web corpus would consist of all websites linked to by either them or their immediate neighbourhood. The only criteria for inclusion in such a corpus would in other words be a crawl distance of 2 sites from the starting points. The main (albeit not only) reason for not doing this and keep the corpus entirely wind specific instead was to enable a better textual analysis of the content of the websites.

Besides harvesting hyperlinks Hyphe also indexes the text from each web page it visits. This makes it possible to run queries to determine the resonance or co-occurrence of different search terms across the corpus. Since we know that all websites are wind specific, it makes it easier to claim that the presence of a search term is indicative of a wind related issue. If the corpus had included more generic websites, like energy.gov, or websites that were specific to tangential issues, like solarenergy.org or wavenergy.dk, it would have made it considerably more difficult to interpret, for instance, a frequent co-occurence of two search terms across the web corpus.

All websites have been crawled to a depth of minimum 2 pages from the homepage.

Exploring the structure and properties of the web corpus

By spatializing the network graph[4] and calculating its modularity[5] it becomes apparent that the web corpus clusters in two overall communities, which can then be subdivided into several smaller clusters. Spatialization basically means that we visualize the network in such a way that websites which are stronger linked to one another also appear closer to one another. In a similar manner the modularity algorithm tries to find the optimal way of subdividing the network in such a way that websites in the same division are relatively more connected to one another than to the rest of the websites in the network.

In the visualization below we have coloured the websites according to their stance on wind energy. On this general level the spatialized network graph displays an almost perfect split between websites that have been tagged as pro wind power (coloured yellow and found to the left) and websites that have been tagged as con wind power (coloured grey and found on the right).

This clear partisanship in the corpus is in itself a noteworthy result. The dataset contains websites from six different countries and websites in general have a tendency to cluster along existing linguistic or administrative boundaries (they typically link to other websites speaking the same language or referring to the same local/national institutions and media outlets). Since our corpus is wind specific and does not contain major media outlets or more generic government sites, some of these effects are mitigated. Yet, as we will explore further below, the degree to which pro and con websites from the same country do not actually link to each other, but to other like minded websites abroad, is still noteworthy.

4. This work was done in Gephi (http://gephi.github.io/). We used the spring based algorithm ForceAtlas2 (Jacomy et al. 2014).
5. We used modularity only as an initial way of exploring the clustering (for a description of the algorithm, see Blondel et al. 2008).

The degree to which pro and con websites link out to their opponents can be discerned from this general overview. It is worth noting that con sites are 12 times more likely to link out to their opponents in the pro cluster than the other way around. There can be several explanations for this. Contemporary controversies around wind energy very often take the form of a wind turbine installation being proposed by a municipality and a developer and subsequently meeting various forms of resistance. It is in other words up to the opponents to try and make their case heard and acquire a public voice. In the overwhelming majority of cases wind energy proponents are not reacting to an established policy or a proposed project, which would explain why they do not link towards it. Wind opponents are in the exact opposite situation, which would explain why they do provide links.

Of course, this does not preclude wind energy proponents from engaging with their opponents, and the lack of outlinks from the pro cluster to the con cluster could also suggest that wind proponents in general do not take much notice, or at least do not pay much official attention, to wind energy protesters. As we shall see in the semantic analysis of the corpus, however, the picture is more complicated than that. It is for example worth noting that noise is still one of the top 5 most talked about issues in the pro cluster.

There is also a marked difference in the types of websites that populate the pro and the con side. Whereas the con side consists almost exclusively of what you might call issue advocates - websites set up with a specific policy agenda in mind - the pro side includes a lot of industry websites as well. These corporate actors generally have a lower tendency to curate link collections on their websites. This may also explain why the pro cluster is less densely clustered than the con side.

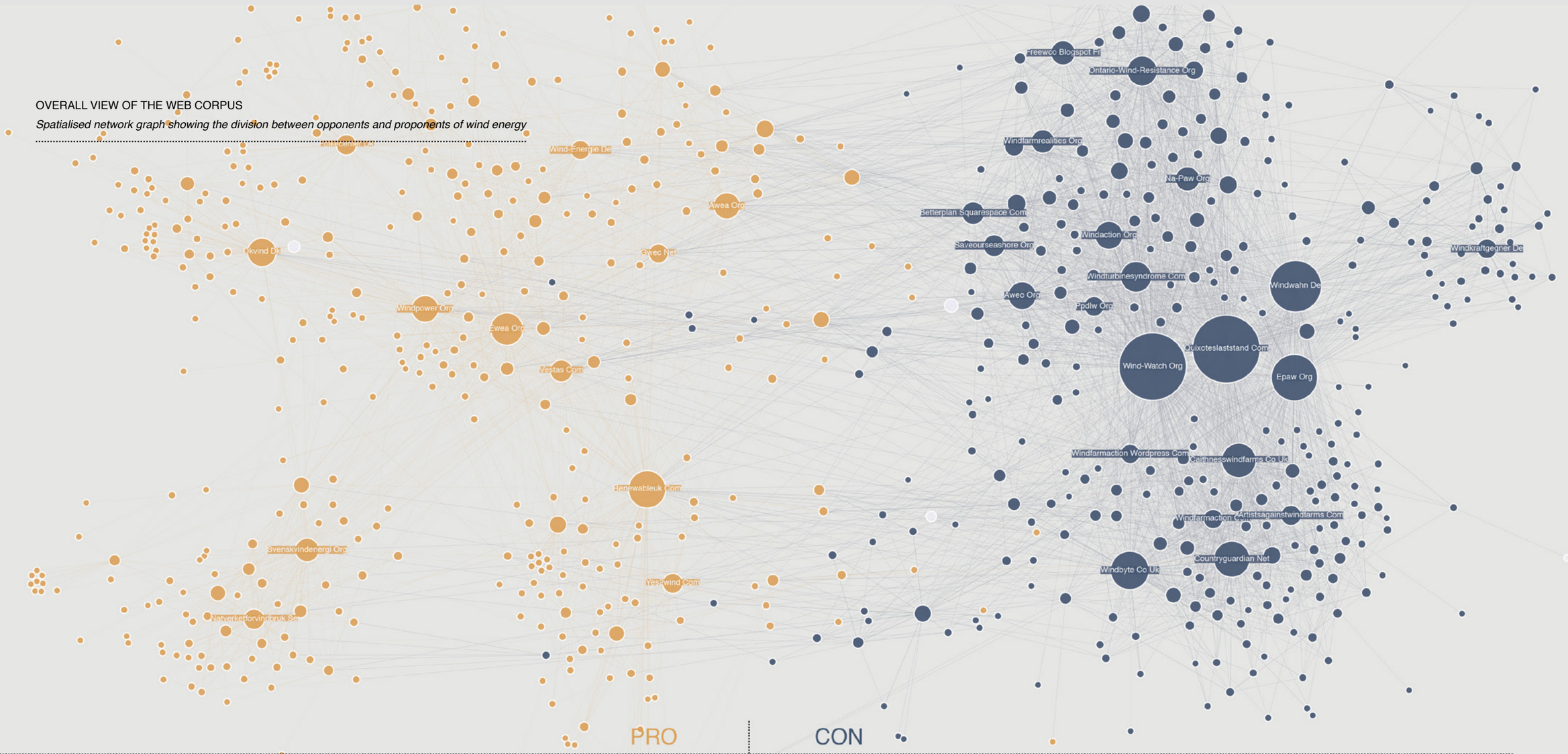
In order to understand the clustering of the web corpus on a more detailed level than the overall separation into pros and cons we experimented with filtering the graph according to the other tags we had attributed to the websites in the qualitative analysis: Country, Type (Government, NGO, Research, etc.), and Placement (Offshore/Onshore). There were no interesting patterns emerging from the Type or the Placement tag. We did not see offshore websites cluster together, for example, nor did we detect an NGO cluster that was separate from, say, government agencies. What we did detect, however, was a very good alignment between the Country tag and the subdivision of the corpus into smaller clusters.

There are interesting differences in the national composition of the pro space and the con space. Notably the Danish component of the pro space (24%) is five times bigger than the Danish component of the con space (4.6%). Whereas the reverse is the case for Canada (3% of pro space vs. 13% of the con space), and the German components are more equally balanced (13% vs. 15.4%), the Danish wind industry makes itself felt here. The corpus comprise 104 Danish pro sites, many of which are corporate, and only 15 Danish con sites. This is of course and important feature of the Danish wind debate (jobs and export revenues are at stake) which is necessary to keep in mind when comparing it, for example, with its Canadian counterpart.

On page 13-14 is a sketch of how where websites from different countries are located in the web corpus. Notice that it is necessary to divide each country into a pro version and a con version. this is still the primary association maker in corpus and it now becomes apparent how for example German opponents are far more interlinked with their Canadian or UK compatriots that with German wind proponents. There are of course connections across the pro/con divide. These can for example be clearly seen in the outlinks sent from the Swedish Con cluster towards the Swedish Pro cluster. But again the connections are much stronger with the UK Con cluster.

OVERALL VIEW OF THE WEB CORPUS

Spatialised network graph showing the division between opponents and proponents of wind energy



PRO

NODES IN THE PRO CLUSTER: 429

OUTLINKS TO CON SITES: 24

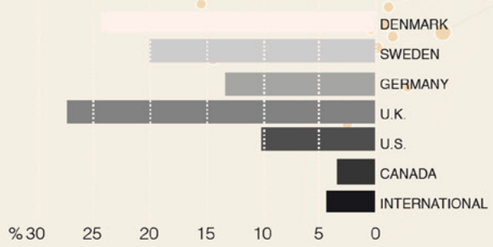
GRAPH DENSITY. LOW

CLUSTERS AS INDICATED BY MODULARITY:

1 2 3 4 5 6

OUTLINKS TO CON CLUSTER AS PERCENTAGE OF MAXIMUM POSSIBLE OUTLINKS: 0.017%

NATIONAL COMPOSITION OF THE PRO CLUSTER

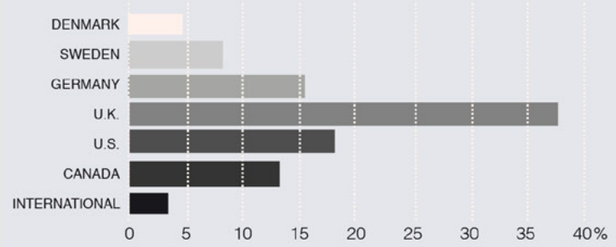


AVERAGELY MOST MENTIONED ISSUES

OFFSHORE, JOBS, ONSHORE, CLIMATE CHANGE, NOISE

CON

NATIONAL COMPOSITION OF THE CON CLUSTER



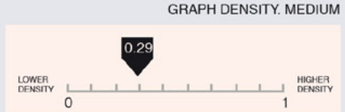
AVERAGELY MOST MENTIONED ISSUES

NOISE, HEALTH, BIRDS, HOUSE PRICES, INFRASOUND

OUTLINKS TO CON SITES: 300

NODES IN THE CLUSTER: 324

GRAPH DENSITY. MEDIUM



CLUSTERS AS INDICATED BY MODULARITY:

1 2 3 4

0.21% OUTLINKS TO CON CLUSTER AS PERCENTAGE OF MAXIMUM POSSIBLE OUTLINKS

DENMARK PRO

GERMANY PRO

US & CANADA PRO

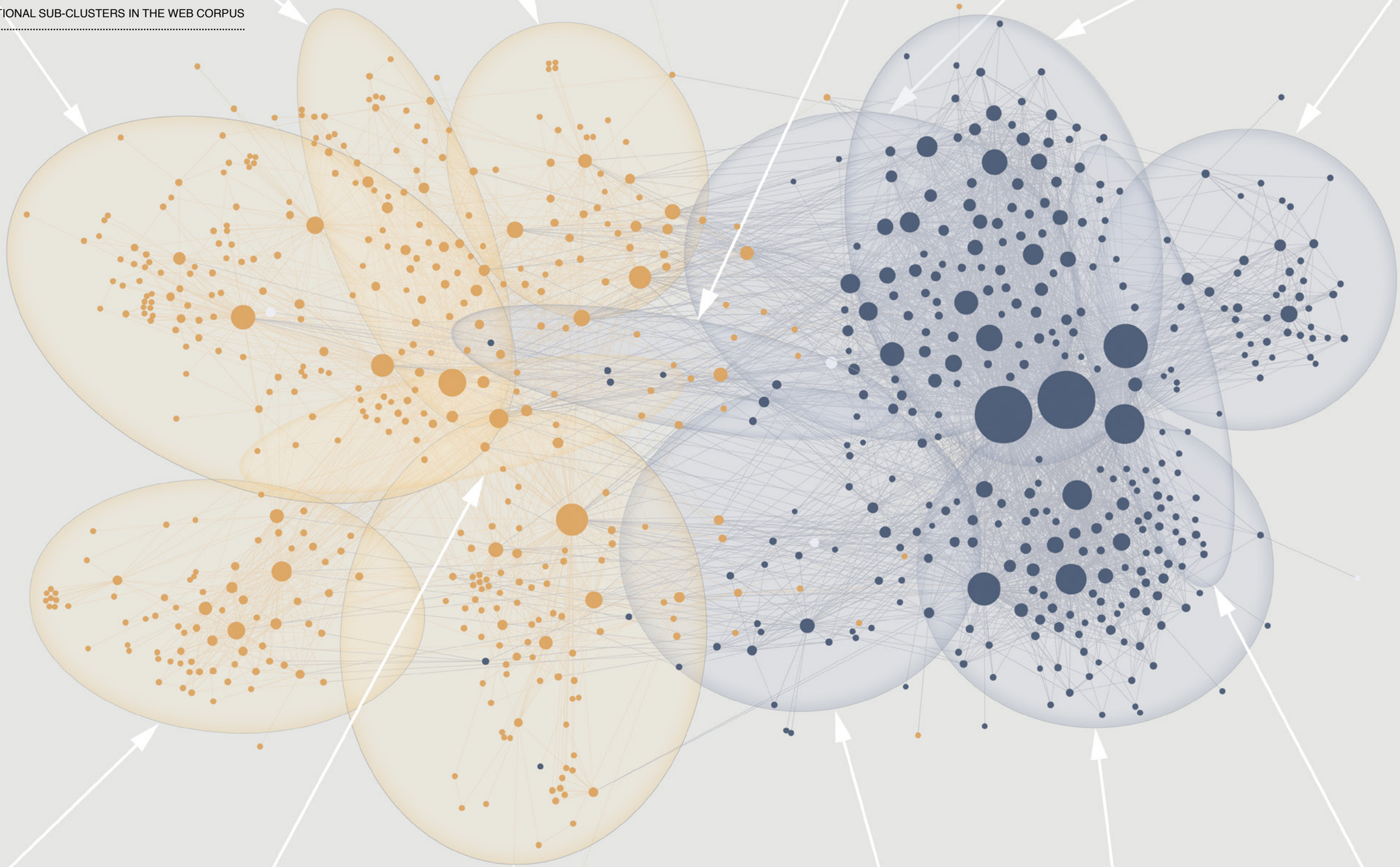
DENMARK CON

US CON

CANADA CON

GERMANY CON

NATIONAL SUB-CLUSTERS IN THE WEB CORPUS



SWEDEN PRO

INTERNATIONAL PRO

UK PRO

SWEDEN CON

UK CON

INTERNATIONAL CON

Based on the realisation that the websites cluster into communities that are primarily for or against wind energy and secondarily nationally oriented, we decided to cut the corpus into 14 sub-clusters for a more detailed analysis: Canada Pro, Denmark Pro, Germany Pro, Sweden Pro, US Pro, UK Pro, International Pro, Canada Con, Denmark Con, Germany Con, Sweden Con, US Con, UK Con, and International Con. We have left out the 5 websites that were tagged as undecided in their stance towards wind power for this part of the analysis.

The clusters are ranked by their density to give a sense of the degree to which the websites interlink. Each cluster has been separately spatialized to obtain 14 new network graphs that can be explored in the same way as the full corpus. The websites have been sized according to the number of links they receive and provide to other websites in the cluster (degree). The names of the most interlinked websites have been retained in each of the network graphs.

Below each network graph it is possible to explore how it connects to other clusters. This can either be done through the general overview of inlinks and outlinks to/ from the cluster (orange/black doughnut chart). While some clusters receive about the same amount of inlinks as they provide outlinks (e.g. Denmark CON or U.S. CON), others (like Sweden Pro or Germany Con) tend to provide a lot more outlinks than they receive inlinks.

It can also be done by exploring where each cluster directs its outlinking activity. If we look at the pro clusters it is worth noticing that while most of them are very exclusively linking out to pro clusters in other countries, a few of them seem to take notice of the con side. This is especially true for the Swedish Pro cluster and to some extent for the UK Pro cluster. If we look at the con clusters it is interesting to note that while they in general link more out to their opponents in the pro cluster, clusters like the Danish Con is exclusively focussed on the Danish Pro, while clusters like the UK Con outlinks to pro sites in several other countries. A possible explanation for this practice could be found in the degree to which protest sites use examples from other countries in their national debates.

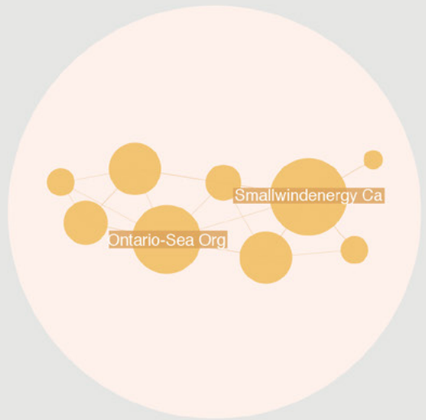
Finally it is possible explore differences in the issues discussed by websites in the respective clusters. We have highlighted the three on average most talked about issues in each cluster. We have also highlighted which issues deviate positively and negatively from the overall most talked about issues in the pro and con spaces. This makes it possible to see, for example, that while noise, infrasound and health are on average the most discussed issues on Danish wind protest websites, this is not just because it reflects a tendency among wind protest websites in general. If you compare it with the average discourse in the con space, infrasound (+15%), noise (+12%), and noise regulation (+11%) are on average more talked about in the Danish Con cluster. Health, on the other hand, is comparatively less talked about (-14%), together with climate change (-12%) and house prices (-11%).

CLUSTERS RANKED BY THEIR DENSITY

1. CANADA PRO

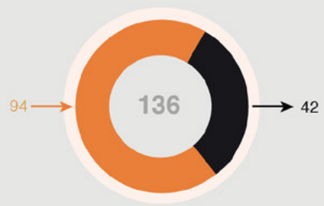
NODES IN THE CLUSTER
11

The most connected websites within the cluster (ranked by degree) are indicated on the graph

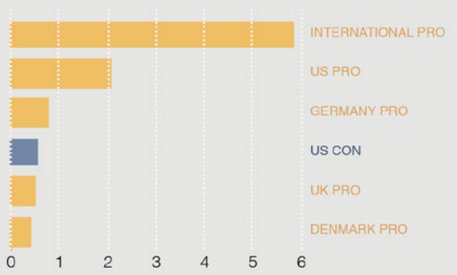


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



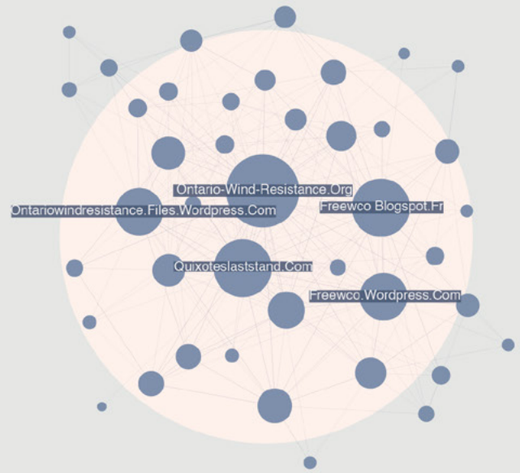
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



2. CANADA CON

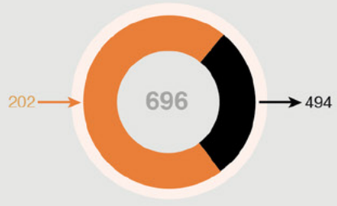
NODES IN THE CLUSTER
42

The most connected websites within the cluster (ranked by degree) are indicated on the graph

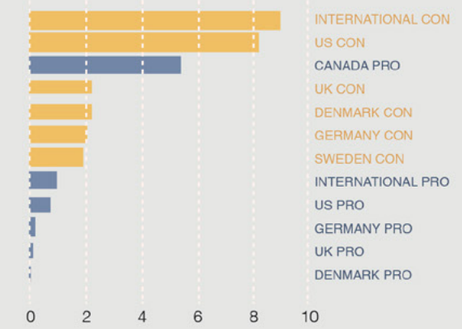


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.

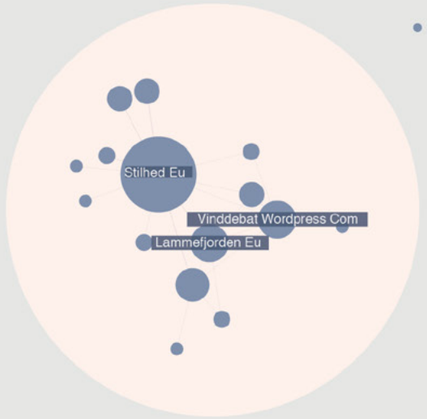


CLUSTERS RANKED BY THEIR DENSITY

3. DENMARK CON

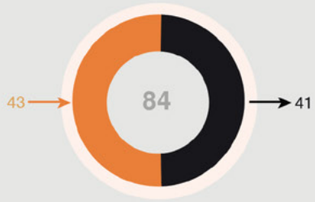
NODES IN THE CLUSTER
15

The most connected websites within the cluster (ranked by degree) are indicated on the graph

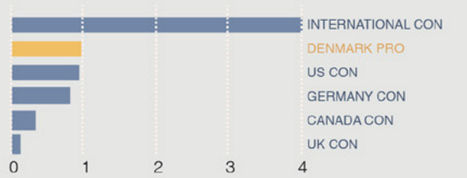


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



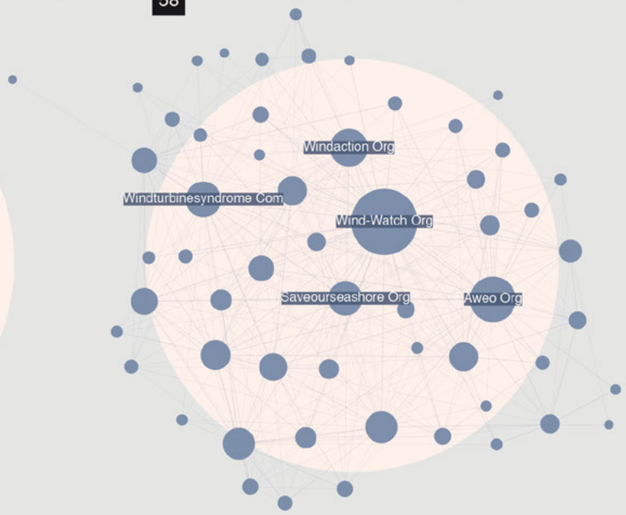
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



4. U.S. CON

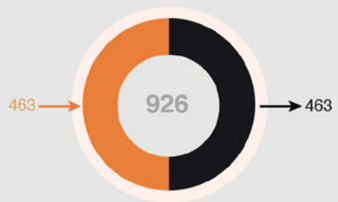
NODES IN THE CLUSTER
58

The most connected websites within the cluster (ranked by degree) are indicated on the graph

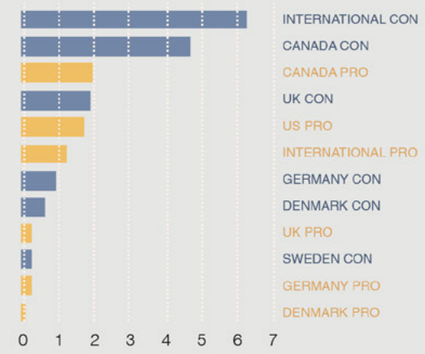


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



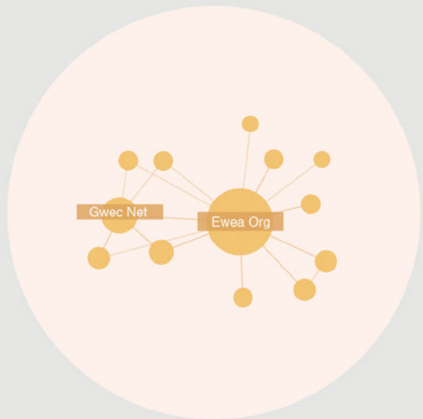
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



5. INTERNATIONAL PRO

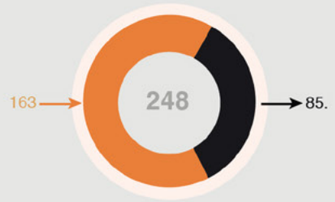
NODES IN THE CLUSTER
17

The most connected websites within the cluster (ranked by degree) are indicated on the graph

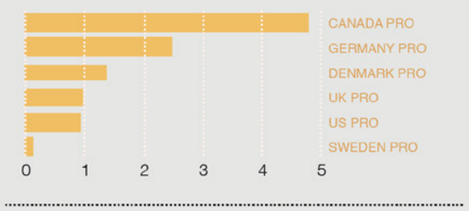


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



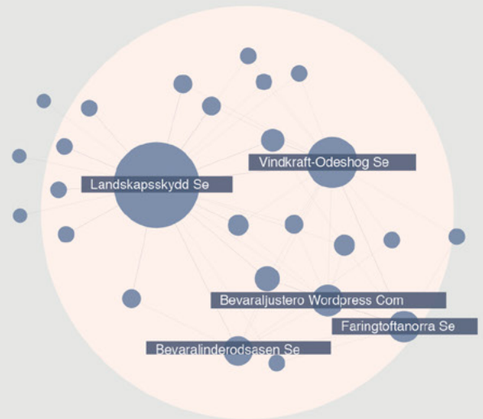
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



6. SWEDEN CON

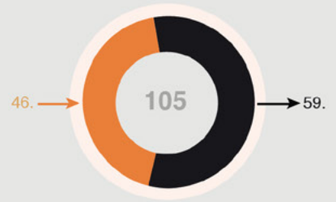
NODES IN THE CLUSTER
27

The most connected websites within the cluster (ranked by degree) are indicated on the graph

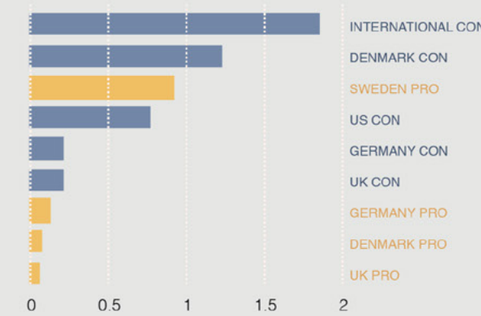


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.

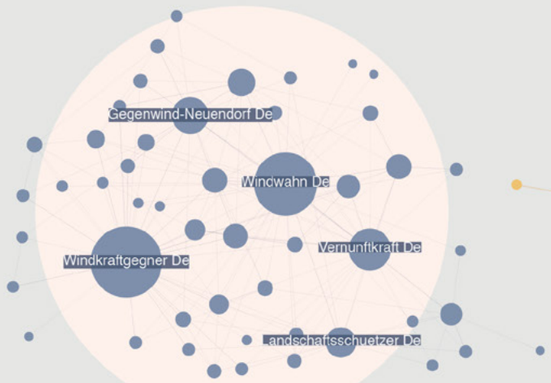


CLUSTERS RANKED BY THEIR DENSITY

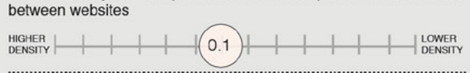
7. GERMANY CON

NODES IN THE CLUSTER
50

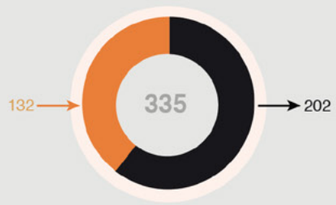
The most connected websites within the cluster (ranked by degree) are indicated on the graph



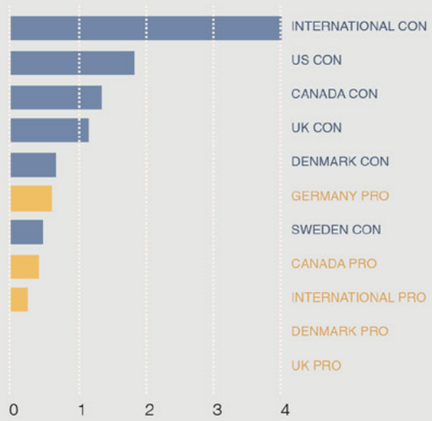
Cluster density: 1 = complete network with all possible connections between websites



Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



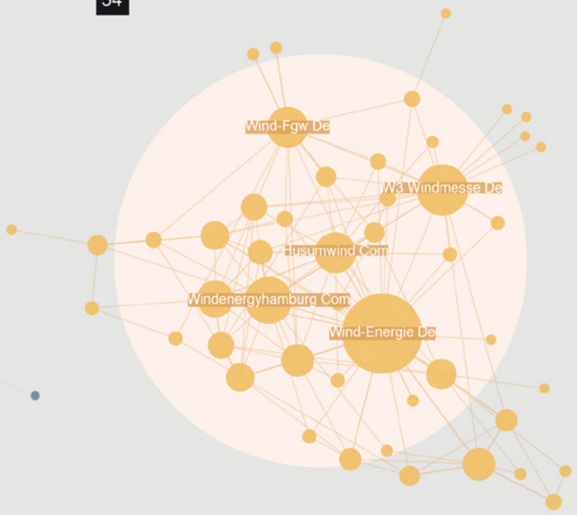
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



8. GERMANY PRO

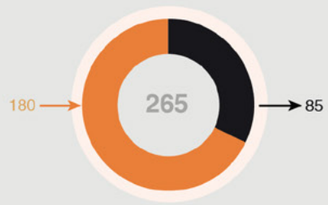
NODES IN THE CLUSTER
54

The most connected websites within the cluster (ranked by degree) are indicated on the graph

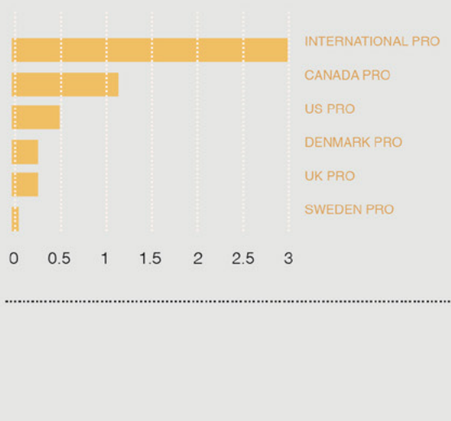


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



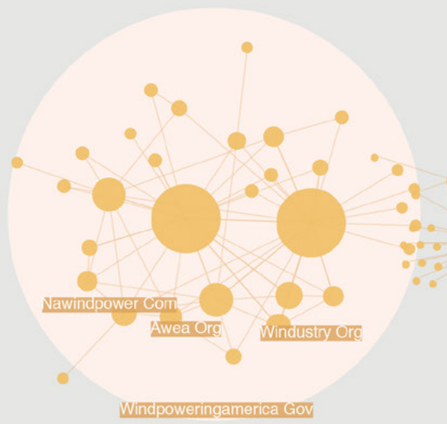
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



9. U.S. PRO

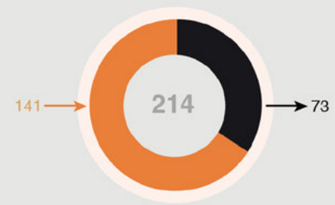
NODES IN THE CLUSTER
43

The most connected websites within the cluster (ranked by degree) are indicated on the graph

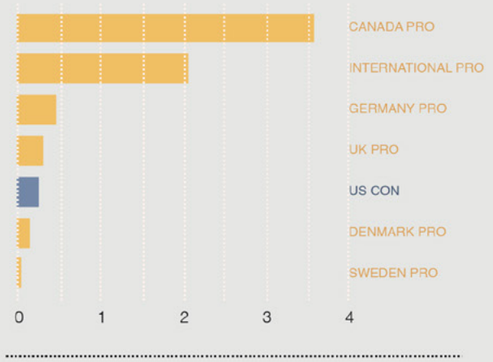


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



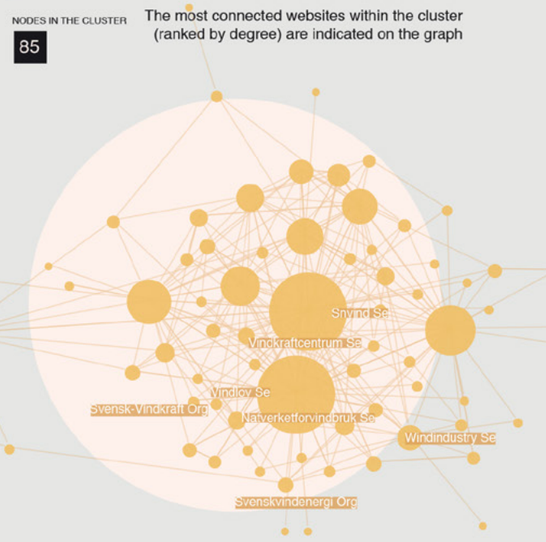
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



10 SWEDEN PRO

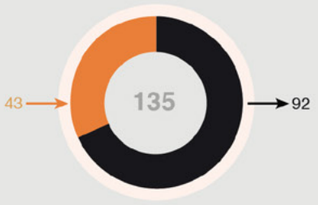
NODES IN THE CLUSTER
85

The most connected websites within the cluster (ranked by degree) are indicated on the graph

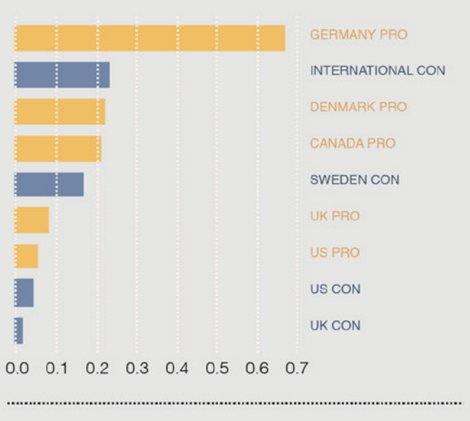


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.

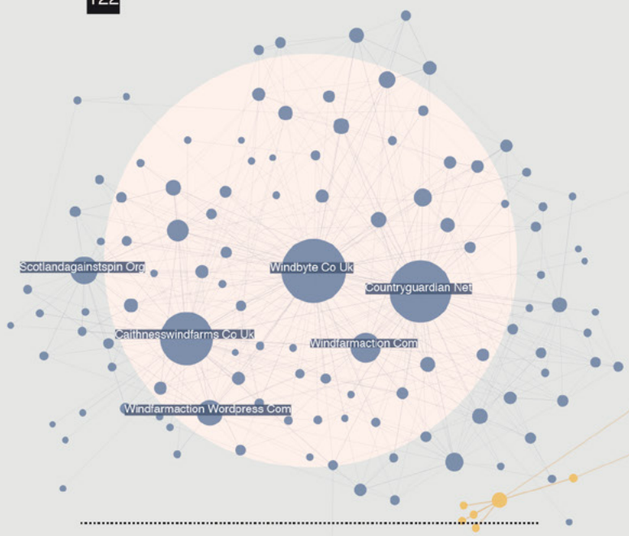


CLUSTERS RANKED BY THEIR DENSITY

11. U.K. CON

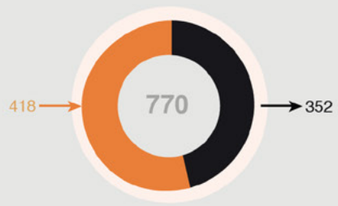
NODES IN THE CLUSTER
122

The most connected websites within the cluster (ranked by degree) are indicated on the graph

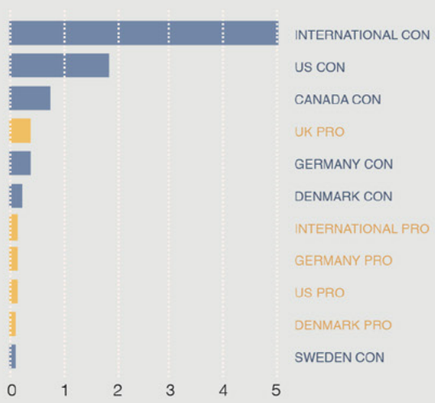


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



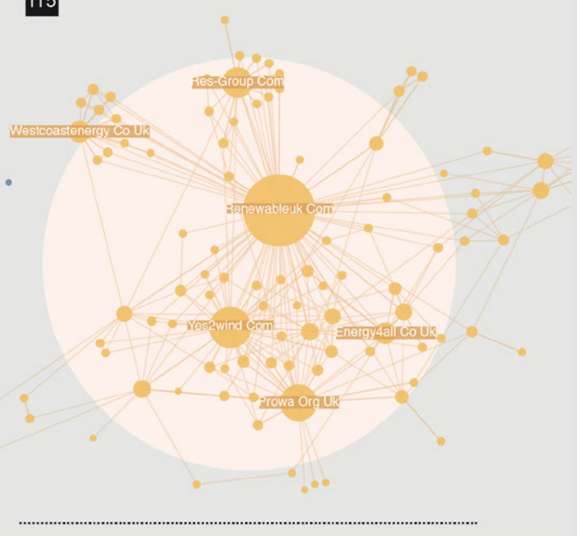
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



12. U.K.PRO

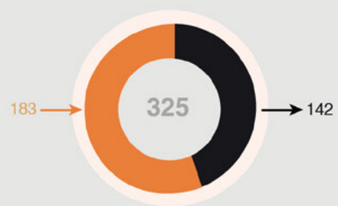
NODES IN THE CLUSTER
115

The most connected websites within the cluster (ranked by degree) are indicated on the graph

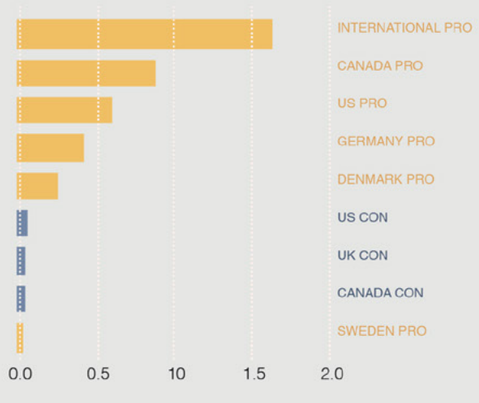


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



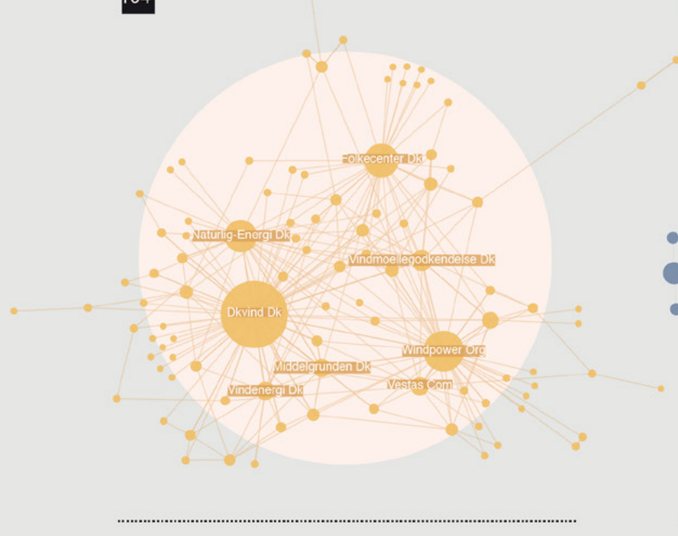
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



13. DENMARK PRO

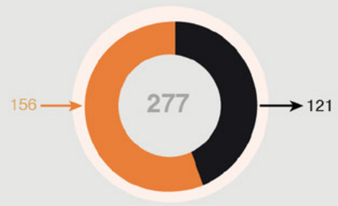
NODES IN THE CLUSTER
104

The most connected websites within the cluster (ranked by degree) are indicated on the graph

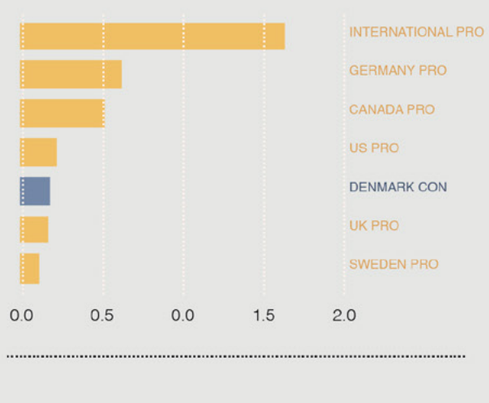


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



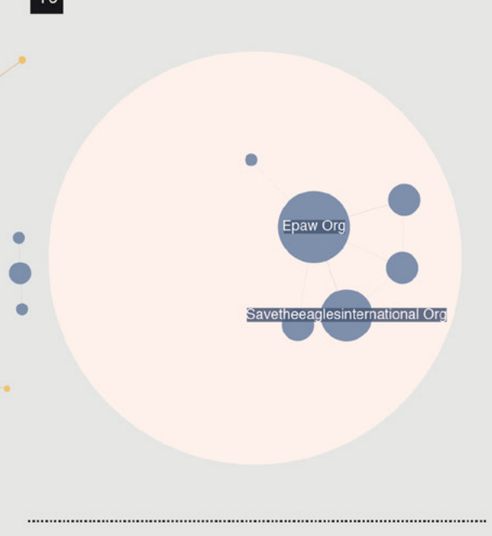
Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



14. INTERNATIONAL CON

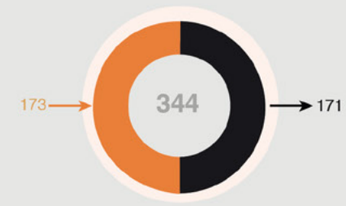
NODES IN THE CLUSTER
10

The most connected websites within the cluster (ranked by degree) are indicated on the graph

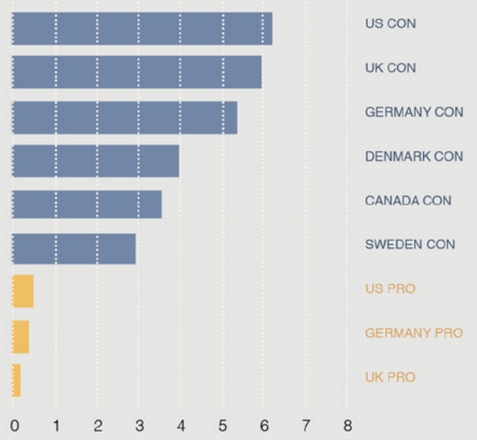


Cluster density: 1 = complete network with all possible connections between websites

Connections with the rest of the web-corpus measured as links received (in-degree) and links provided (out-degree) by the cluster.



Connections with other clusters measured as the percentage of total possibleout-links from this cluster to each of the other clusters.



If we focus on individual websites instead of clusters it also becomes interesting to ask questions about their authority in the network (how important are they considered to be by others?) and how they contribute to the construction of the network (how active are they in providing links to others?).

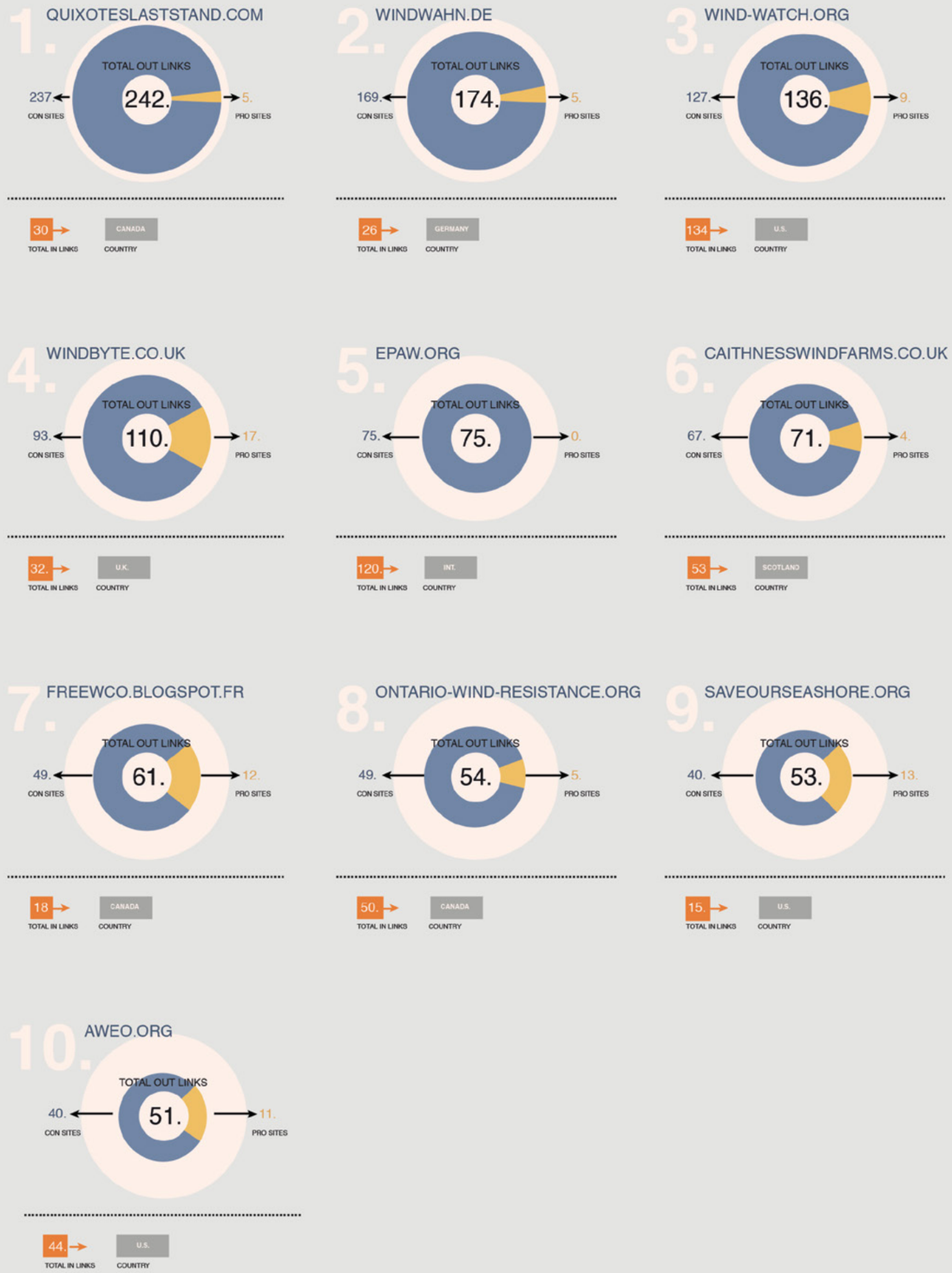
The centrality of a website can be defined in several ways. Here we have restricted ourselves to two of the more basic: hubs and authorities. Hubs are websites that provide a lot of outlinks to other wind specific sites. They drive an issue by pointing a concerned public towards a range of resources that might be of interest or assistance. Authorities, conversely, are websites that receive a lot of links from other wind specific sites. They gather attention and are likely to be considered important for one reason or another. Maybe they are frequently posting news stories that other websites like to share. Maybe they provide a good explanation of a particular set of rules or procedures that other websites like to refer their readers to.

Below we have ranked the 10 most in-linked pro sites and con sites to determine who are the most important authorities in both spaces. Similarly we have ranked the 10 most out-linking pro sites and con sites to determine who are the most important hubs in both spaces. For each of the authorities we show how they receive links from pro and con sites respectively.
For each of the hubs we show how they outlink to con and pro sites respectively.

It is again noteworthy that the top 10 hubs in the pro space (i.e. the most out-linking pro sites) barely ever provide links to websites in the con space. This picture is also confirmed if we look at the top 10 authorities in the con space. They almost exclusively derive their centrality from links received from other con sites. The top 10 authorities in the pro space, on the other hand, receive up to a third of their links from opponents in the con space.

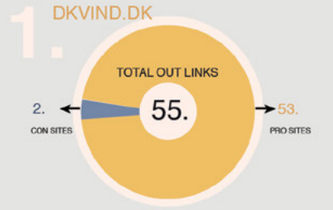
We have also indicated which countries these central websites come from. Whereas one might have expected them to be largely international organisations like the European Platform Against Windfarms (epaw.org) this is actually far from the case. There is just one website tagged international in each of the four top 10s. This might indicate that many national wind energy organisations (pro and con alike) are not exclusively national phenomena but are referred to extensively in other countries as well.

TOP 10 HUBS IN THE CON SPACE

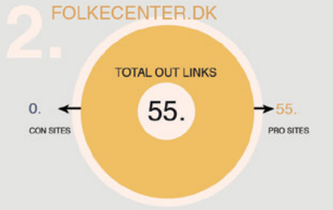


TOP 10 HUBS IN THE PRO SPACE

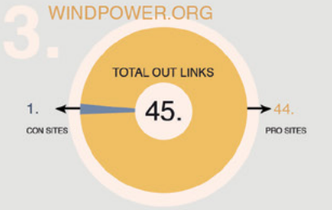
1. DKVIND.DK



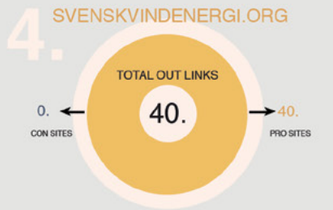
2. FOLKECENTER.DK



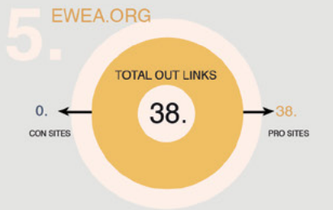
3. WINDPOWER.ORG



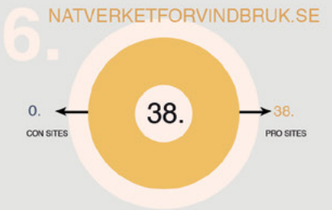
4. SVENSKVINDENERGI.ORG



5. EWEA.ORG

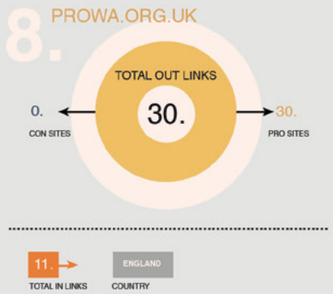


6. NATVERKETFORVINDBRUK.SE

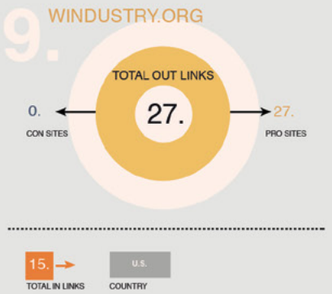


7. WINDINDUSTRY.SE

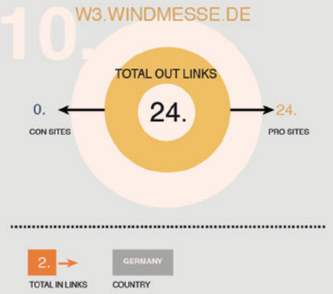
8. PROWA.ORG.UK



9. WINDUSTRY.ORG

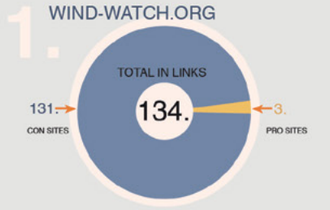


10. W3.WINDMESSE.DE



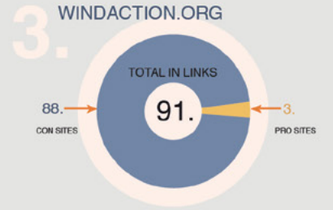
TOP 10 AUTHORITIES IN THE CON SPACE

1. WIND-WATCH.ORG

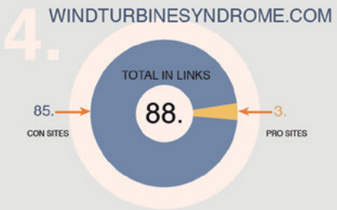


2. EPAW.ORG

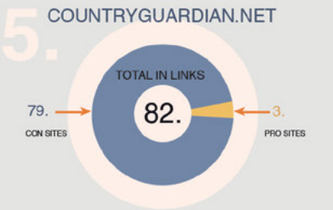
3. WINDACTION.ORG



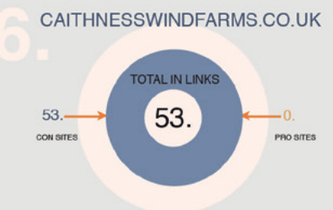
4. WINDTURBINESYNDROME.COM



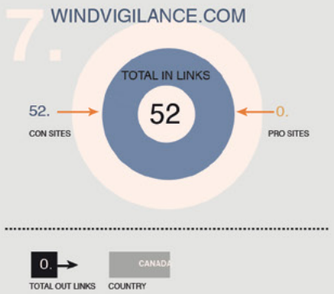
5. COUNTRYGUARDIAN.NET



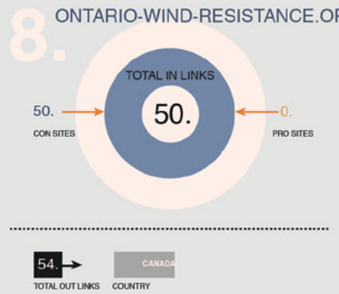
6. CAITHNESSWINDFARMS.CO.UK



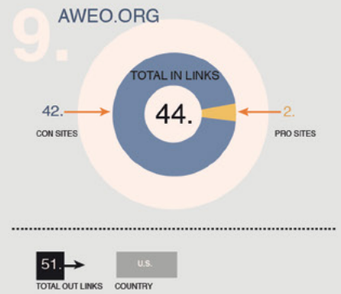
7. WINDVIGILANCE.COM



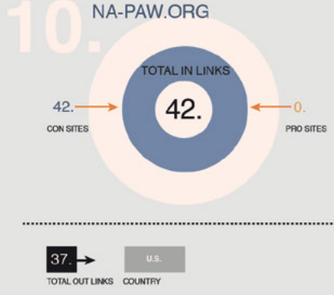
8. ONTARIO-WIND-RESISTANCE.ORG



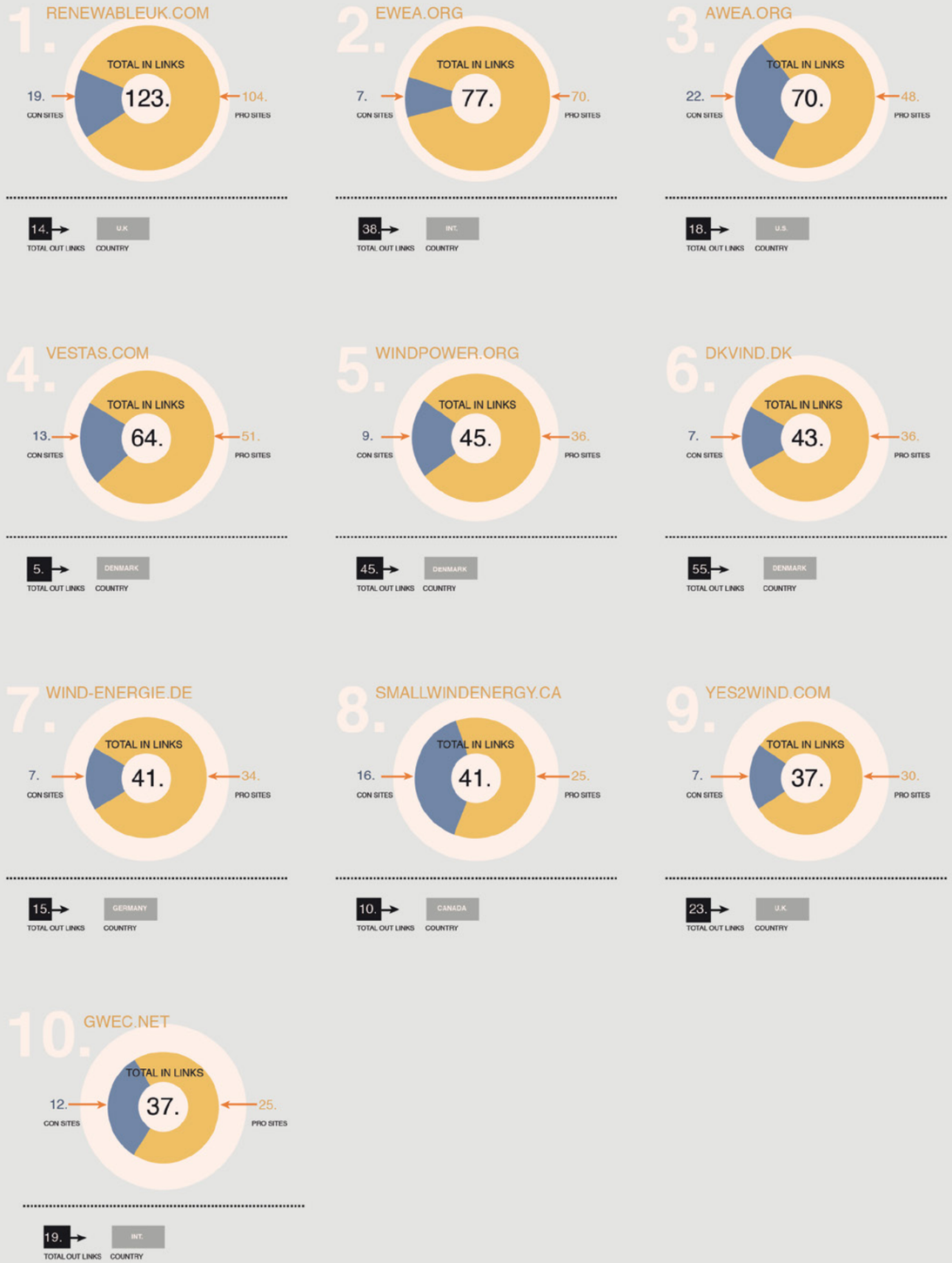
9. AWEO.ORG



10. NA-PAW.ORG



TOP 10 AUTHORITIES IN THE PRO SPACE



Building the issue dictionary

In order to query the text of the web corpus in a systematic manner for resonance of wind related issues we decided to construct an issue dictionary: an evolving set of search terms that are known to be indicative of different aspects of the controversy. The issue dictionary currently contains 47 tried and tested multilingual queries. This is not a definitive list. It can be expanded both in depth and scope, it can be made more detailed, or it can be tailored to specific research questions later on. Here we will present it in its current form and describe the principles behind its construction.

Wind2050 is both a cross-institutional and cross-disciplinary research project that comprises researchers and stakeholders with different interests in wind energy. In order to profit from their prior knowledge we decided to circulate a short questionnaire for the project kick-off meeting. The responses yielded a rough list of issues that might be interesting to detect, but we needed good, discrete search terms to be able to do that in practice. During the kick-off meeting we tested some of the queries on a makeshift pilot web corpus. This was helpful in terms of understanding what type of queries were actually used in the online . More importantly, there was a risk that the issue dictionary would be biased towards the concerns of experts who were overwhelmingly proponents of wind energy.

We therefore decided to take a more experimental and user-driven approach to the construction of the issue dictionary. This work was undertaken by Linn Wagner Korsgaard, Loreta Møller, and Line Østereng Jørgensen as part of their MSc in techno-anthropology at the University of Aalborg (for details, see Møller et al. 2014). Over the course of Spring 2014 they organised a series of workshops with stakeholders both in and outside the project. In the early stages of the process participants were asked to collaborate to produce good keywords to trace issues. In the later stages they were asked to work on mock-ups of what they would consider to be a good issue dictionary.

The conversations prompted by these activities were recorded and transcribed. This yielded valuable insights about the difficulties of translating an issue like “they are destroying my view” or “they are good for Danish export” into discrete, operational search terms.

From the workshops we extracted a longer list of issues and suggested search terms that we began testing on the web corpus. We looked for terms that would both be common enough to be used across a number of websites, but also be precise enough to avoid ambiguities about their meaning. Terms like “view” or “export” are for example present in force across the web corpus, but they are also so generic that it becomes impossible to interpret their presence on a web page as meaning anything in particular. Terms like “shadow flicker” or “migratory birds” are on the other hand quite precise and interpretable when you take into account that they have been found on websites that we already know to be specifically about wind energy.

We translated and tested all queries into English, Danish, Swedish and German to make them compatible with the languages of the entire web corpus.

The list below represents the results of this work. Note that the queries use standard Google search operators. If you want to use them on Google to look for material specifically related to wind energy and one of the issues you can for example add “AND (“wind turbines” OR “wind energy” OR “wind power” OR vindenergi OR vindmølle* OR vindkraft OR windenergie OR windkraft* OR “wind turbinen”)” to the end of a query and paste it to the search box.

The issue dictionary

Issue	Query
Biodiversity	biodiversit* OR "species diversity" OR artenvielfalt OR artendiversität OR artsdiversit* OR "biologiska mångfaldet" OR "biologisk mångfald"
Wildlife	wildlife OR "wild animal" OR "wild animals" OR "animal specie" OR "animal species" OR dyreliv* OR "vilde dyr" OR dyreart* OR djurliv* OR "vilda djur" OR "vilda djuren" OR djurart* OR "wildtier* OR tierart*"truet dyreart"
Endagered species	OR "truede dyrearter" OR "truet dyr" OR "truede dyr" OR "endangered specie" OR "endangered species" OR Utrotningshotad* OR "Bedrohte tiere" OR "Bedrohte arten"
Animal welfare and protection	"animal welfare" OR dyrevelfærd* OR tierschutz* OR djurskydd*
Birds	bird* OR fugl* OR vogel* OR vögel* OR fågel OR fåglar Migrating birds fulgetræk OR trækfulg* OR zugvogel OR vogelzug OR zugvögel OR vögelzug OR "migratory bird" OR "migratory birds" OR "bird migration" OR "bird migrations" OR "migratory bird" OR "migratory birds" OR flyttfågel OR fågelflytt* OR flyttfåglar
Eagles	eagle* OR ørn* OR örn* OR adler*
Bats	"bat deaths" OR "bat fatalities" OR bats OR flagermus* OR chiroptera OR fledertiere* OR flattertiere* OR fladdermöss*
Sea Eagle	havørn* OR "white-tailed eagle" OR "white-tailed eagles" OR "sea eagle" OR "sea eagles" OR "aaliaeetus albicilla" OR seeadler* OR havsörn*
Shadow Flicker	"shadow flicker" OR "shadow flickers" OR "shadow flickering" OR skyggekast* OR rotorskygge* OR "rotor shadow" OR rotorschatten OR "disco-effekt" OR "rörliga skuggor" OR "roterande skugga" OR "roterende skygge" OR schattenwurf* OR "rotating shadow" OR "rotating shadows"
Cultural heritage	kulturarv* OR "cultural heritage" OR "Kulturelles Erbe" OR Kulturerbe*
Nature Conservation	"nature conservation" OR "landscape conservation" OR "landscape protection" OR "nature protection" OR "protected natural" OR naturvård* OR naturskyd* OR landskapsskyd* OR naturbeskyt* OR naturschutz* OR Landschaftschutz*
House prices	ejendomsværdi* OR huspris* OR "house prices" OR "property values" OR "real estate prices" OR Immobilienpreise* OR Immobilienwerte* OR fastighetsvärde*
Setback distance	afstandskrav* OR "set back distance" OR "Minimum Distance" OR "setback distance" OR avståndskrav* OR mindesabstand OR minimumsafstand* OR "abstände zwischen" OR "avstånd till" OR "avstand til"
Noise	lärm* OR noise* OR støj* OR buller* OR vindmøllestøj OR vindkraftsbuller OR windkraftlärm
Noise regulation	"regler for støj" OR "reglerne for støj" OR støjregl* OR støjgrænse* OR *bullergräns* OR bullerregler* OR "noise regulation" OR "noise regulations" OR "noise control" OR lärmbekämpfung OR lärmgrenze*
Low frequency noise	"lavfrekvent støj" OR infrastøj OR infranoise OR "low-frequency noise" OR "lågfrekvent buller" OR infrabuller OR infraschall* OR niederfrequenzschall* OR "lavfrekvent lyd" OR infralyd OR infrasound OR "low-frequency sound" OR infraljud OR "lågfrekvent ljud"
Vibrations	vibration*
Amplitude modulation	amplitudmodul* OR "amplitude modulation" OR Amplitudenmodul* OR amplitudemodul*
Noise gauging	støjmål* OR "sound level meter" OR "noise measurement" OR lärmmessung* OR schallpegelmesser OR bullermät*
Health	krankheit* OR gesundheit* OR health* OR helbred* OR illness* OR sundhed* OR sygdom* OR hälsa* OR sjukdom*
Sleep disorders	søvnløs* OR insomn* OR sleepless* OR "sleep disorder" OR "sleep disorders" OR Sömnlös* OR schlaflos* OR schlafstörung* OR søvnforstyrrelse* OR sömnstörning*
Miscarriages	miscarr* OR aborter* missfall* OR fehlgeburt*
Cardiovascular	cardiovascular* OR hjertekar* OR "hjerte-kar" OR kardiovaskulär* OR "Hjärt- och kärlsjukdom" OR "Hjärt- och kärlsjukdomar" OR "hjerte karsygdomme" OR "hjerte kar" OR "Herz-Kreislauf" OR Kreislauferkrankung*
Wind Turbine Syndrome	vindmøllesyndrom* OR "wind turbine syndrome" OR "windturbine syndrome" OR "wind turbine syndrom" OR vindkraftssyndrom* OR "Wind turbinen syndrom" OR "Windturbinen syndrom" OR Windturbinensyndrom
Headache	headache* OR hovedpine* OR Kopfschmerz* OR huvudvärk*
Energy prices	energiepreis* OR "energy price" OR "energy prices" OR "energy pricing" OR energipris*
Energy crisis	"energy crisis" OR energikris* OR energiekrise*
Security of Supply	forsyningssikkerhed* OR försörjningstrygghet* OR versorgungssicherheit OR "security of supply"
Energy Security	"energy security" OR energisikkerhed* OR Energiesicherheit OR energisäkerhet*
Local ownership	"local ownership" OR "lokalt ejerskab" OR lokaleje* OR "lokalt ägarskap" OR "locally owned" OR "lokale Eigenverantwortung"
Jobs	arbejdspladser* OR jobs OR jobb OR arbetsplatser* OR arbeitsplätz*
Green jobs	"green jobs" OR "grønne arbejdspladser" OR "grønne job" OR "grønne jobs" OR "gröna arbetsplatser" OR "gröna jobb" OR "grüne Arbeitsplätze" OR "grüne jobs"
Climate change	"climate change" OR klimaforandring* OR "global warming" OR "globale opvarmning" OR Klimawandel OR klimatförändring* OR "globala uppvärmningen" OR "globale Erwärmung"
Climate change adaptation	"climate adaptation" OR "adaptation to climate change" OR "climate change adaptation" OR klimatilpasning* OR "Anpassung an den Klimawandel" OR "anpassning till klimatförändringar" OR klimatanpassning* OR Klimaanpassung
Clean tech	"fossil free" OR "clean tech" OR "clean energy" OR fossilfri* OR cleantech

Issue	Query
Climate scepticism	"climate lie" OR "climate scam" OR "climate hoax" OR "green lie" OR "climate fraud" OR "green fraud" OR "green scam" OR "green hoax" OR "global warming hoax" OR "global warming lie" OR "global warming scam" OR climategate OR klimaløgn* OR "klima løgn" OR "climate lie" OR klimalüge OR "klima lüge" OR Klimaschwindel* OR "global warming swindle"
Stray Voltage	"stray voltage" OR "electrical pollution" OR "omstrejfende spænding" OR "vagabonderande ström" OR Streuspannung OR Electrosmog OR "e-smog" OR "elektrisk forurening"
Air Traffic	flytrafik* OR "air traffic" OR flygtrafik* OR luftverkehr* OR lufthavn* OR airport* OR flygplats* OR flughafen*
Industrial wind	"industrial scale" OR "industrial windpower" OR "industrial windturbine" OR "industrial windturbines" OR "industriel vindkraft" OR "industrielle vindmøller" OR "industriell vindkraft" OR "industriellen Maßstab"
Big wind	"big wind"
Community wind	"community wind energy" OR "community wind power" OR "community windpower" OR Bürgerwind* OR samäg* OR fælleseje* OR lokaleje*
Small scale wind	"small scale" OR Husstandsvindmølle* OR småskalig* OR småskala OR husstandsmølle* OR Kleinwindanlagen
Offshore	"off-shore" OR offshore*
Onshore	"on-shore" OR onshore*
Near shore	"near-shore" OR nearshore*

Exploring the web corpus with the issue dictionary

There are several ways of exploring the web corpus with the search terms from the issue dictionary, some of which have been shown earlier in this report. It is of course possible to profile individual web sites based on how they resonate with the queries. As an example we can take one of the protest sites against the Danish National Test Centre for Large Wind Turbines at Østerild (nationalttestcenter.dk). This website mentions noise on 19%, birds on 18%, and jobs on 12% of its pages. It also mentions noise regulations, offshore, biodiversity, eagles, bats, infrasound, health, air traffic, and noise gauging to a lesser extent, but has no resonance with the rest of the issue dictionary.

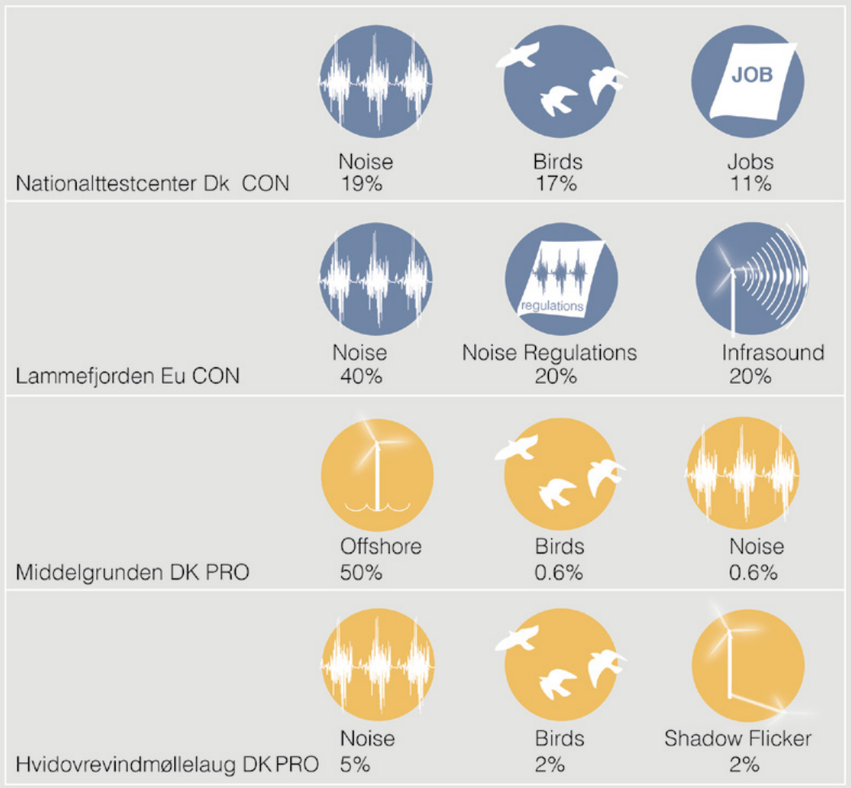
Compare it with other Danish con sites, such as Lammefjorden.eu, and you get a sense of how this information can be used to profile the matters of concern of an actor. On Lammefjorden.eu, for example, a broader range of issues are present, in particular health effects like headaches, cardiovascular disease, or insomnia, and other nuisances relating to the physical proximity of the turbines, such as shadow flicker, vibrations, or setback distance. Compare it with the websites (pro) of two Danish wind turbine cooperatives, middelgrunden.dk and hvidovrevindmollelaug.dk, and you will notice that the issue dictionary overall has lower resonance on these sites, which is perhaps hardly surprising, although noise and birds are mentioned on both and hvidovrevindmollelaug.dk also mentions air traffic and shadow flicker.

This method could be useful to get an initial idea about the scope and composition of the controversy around particular wind turbine projects, depending of course on the availability of websites dedicated specifically to the project in question. It is particularly useful in cases where several websites representing different stakeholders are available around the same wind turbine project.

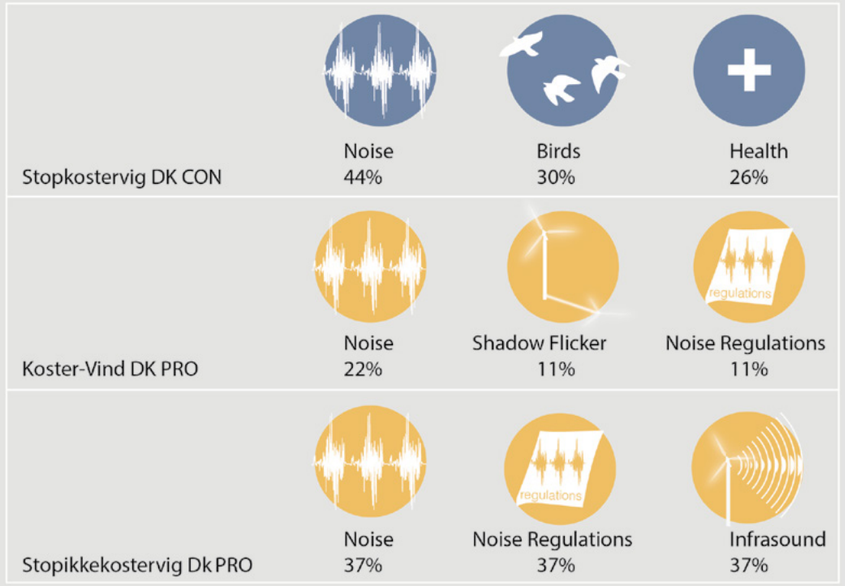
This is for example the case with the Danish Kostervig project. In this instance the project itself has a website, koster-vind.dk, which displays the limited resonance with the issue dictionary that seems to be typical for those kinds of pro websites. There is also a protest group active on the web with stopkostervig.dk, which again displays the typical broader resonance with the issue dictionary for con sites. And finally there is a pro advocacy site set up specifically in response to the protest group. This site, stopikkekostervig.dk, resonates well with the issue dictionary and seems to engage with some of the same issues as the protest website. Besides enabling a better triangulation of what the controversy around Kostervig is about, it also illustrates the importance of considering what kind of pro site you are dealing with. Where as most con sites are advocacy sites whose modus operandi it is to actively raise issues, this is not the case for many of the pro sites.

PROFILING INDIVIDUAL WEBSITES WITH THE ISSUE DICTIONARY

Pages mentioning issue terms as a percentage of the total number of pages on a website



THE KOSTERVIG CASE



Profiling individual websites with the issue dictionary

It is also possible to use the issue dictionary to locate broader discursive regions in the web corpus. Below we show seven sample maps where we have plotted the resonance of a search term onto the network graph of the full corpus. We are essentially using the same method as we did for profiling individual actors, only that we now proceed one issue term at a time and compare its presence across all websites.

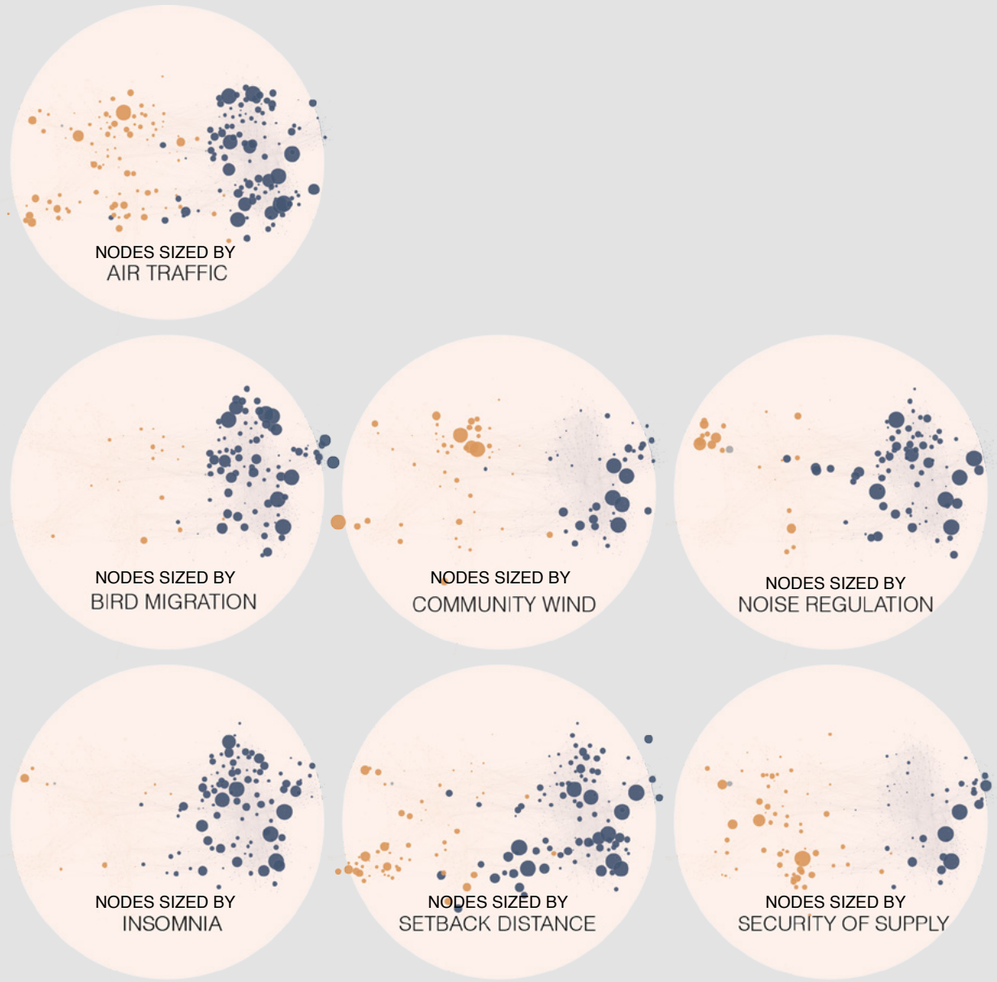
Notice for example that issues like insomnia or bird migration are very exclusive to the con space, where they are prevalent across all the national sub-clusterings. Contrast this with an issue like air traffic which is more or less talked about across the entire web corpus. If we look at an issue like community wind, on the other hand, it is not the overall pro/con divide that springs to mind. It is particularly talked about in the German, Canadian, Swedish, and UK Pro clusters, as well as in the UK and German Con clusters. It is absent from the Swedish Con cluster, however, and relatively absent from the US and Canadian Con clusters. Something similar is the case with security of supply where

both the UK Pro and Con clusters are very active, along with the German Con cluster and the International Pro cluster.

For some issues like setback distance or noise regulation the picture seems to be that they are relatively proliferate across the con space but appear very regional in the pro space. Noise regulations are particularly talked about in the Danish and UK Pro clusters, but almost not at all in the rest of the pro space. In a similar manner setback distance is very talked about in the Swedish Pro cluster, and to some extent in the Danish and UK Pro clusters, but almost not at all in the rest of the pro space.

DISCURSIVE REGIONS IN THE WEB CORPUS

Nodes sized by the number of pages mentioning an issue term as a percentage of the total number of pages on the website



Using the issue dictionary also opens up a different avenue of exploration for the textual data. Instead searching the corpus website by website it is possible to query all the indexed text from all 758 websites for the co-occurrence of search terms. The principle here would be that if two search terms are mentioned in relation with one another somewhere in the corpus we can assume that they have something to do with each other according to the actors in the controversy. We are not able to say where this is the case, but we are able to say something about the relative degree to which two search terms tend to have something to do with one another compared with other search terms. This enables us to profile a search term based on its tendency to occur with, say, the different queries in the issue dictionary.

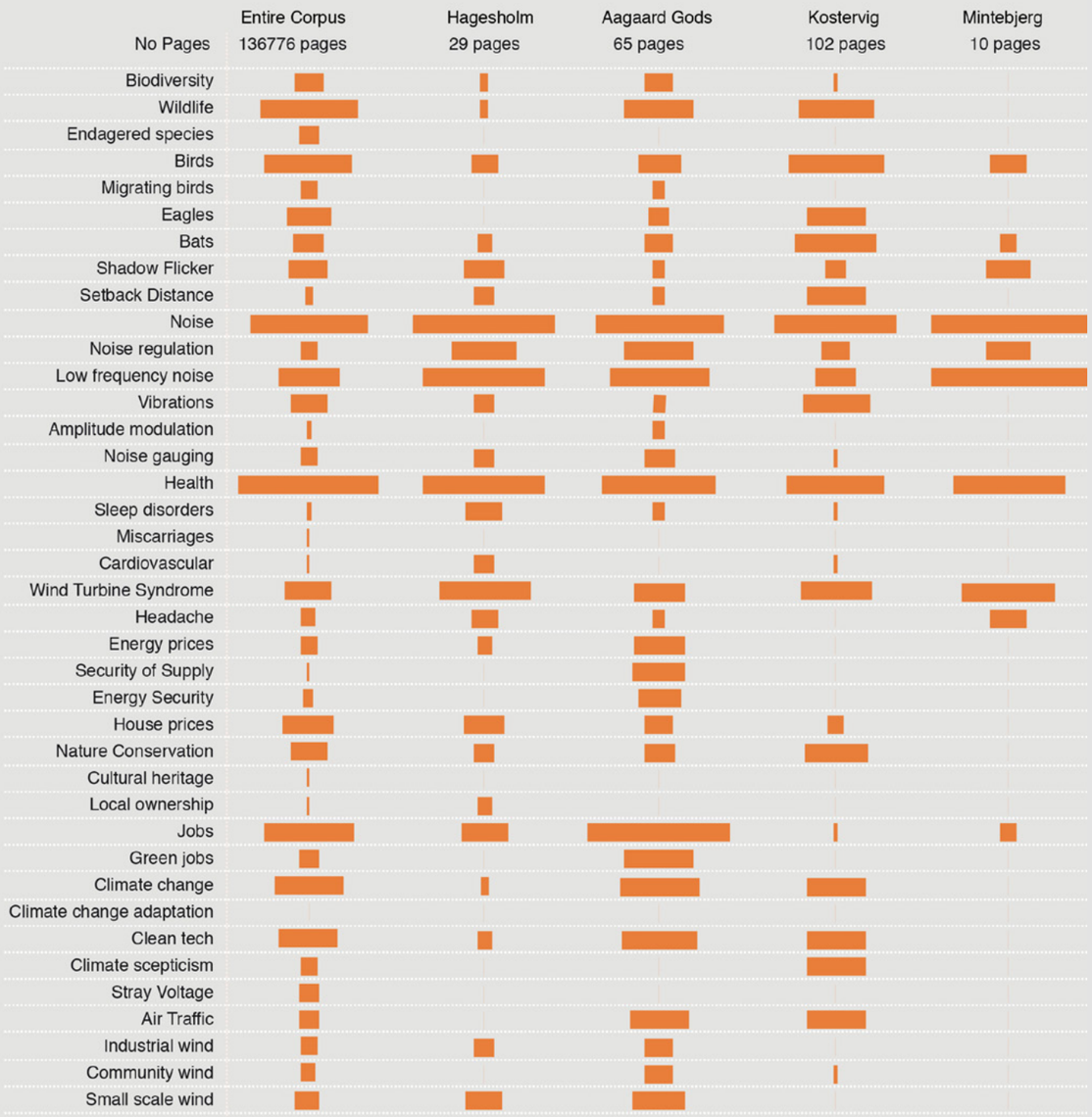
It offers an alternative way forward if you want to use the web corpus to gain knowledge about specific matters of concern that come up in relation to specific wind energy projects. As an example, below we have queried the web corpus for mentions of eight Danish projects, namely Hagesholm, Aagaard Gods, Kostervig, Mintebjerg, Hørslev Solbjerg, Lønborg Hede, Lerchenborg Gods and Brejning, plus two very talked about international projects for reference, namely Caithness (UK) and Amherst (Canada).

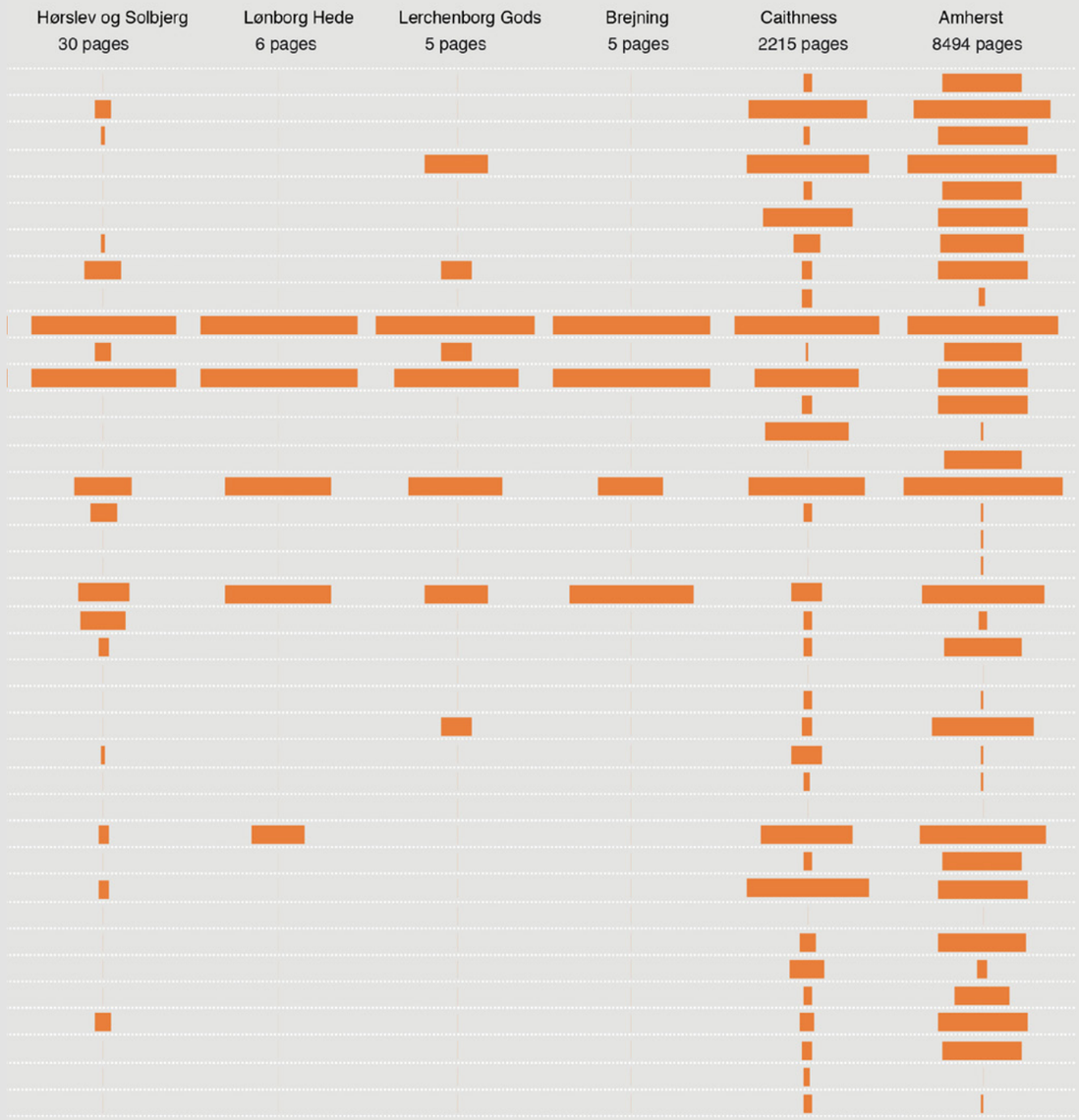
In the row immediately under the projects we show the number of webpages (individual pages on a website) where they are mentioned. In the subsequent rows we have queried the issue dictionaries to map the extent to which each issue is mentioned on the same pages as each of the projects. The result is ten vertical semantic profiles for each of the projects.

Danish projects are quite consistently mentioned together with issues like noise and low frequency noise. This is also true for the international projects, but here you have a significant presence of issues like birds or wildlife which is typically not found in relation to the Danish projects (with the exception of Kostervig). As the only Danish project, Aagaard Gods has been mentioned in relation to issues like security of supply and energy security. This way of cross counting the web corpus with the issue dictionary can be tailored to other enquiries. The current cross count provides an illustration of the method.

CROSS COUNTING ISSUE TERMS IN THE WEB CORPUS

How are different wind turbine projects associated with different issues?





THE WIND2050 SOCIAL MEDIA CORPUS

The social media corpus is a dataset containing information from 14 Danish wind energy pages on Facebook. It is a smaller and geographically more discrete dataset than the web corpus, but it is also a dataset that allows for more analytical detail. It consists of timestamped posts and comments by pages and users, as well as a range of metadata provided by the Facebook API. Metadata includes likes, comments, and engagement statistics, as well as categorizations like type (photo, status, link, video, etc.), sex, or user locale. The dataset has been harvested using Netvizz (Rieder 2013, see also: https://apps.facebook.com/netvizz/) and spans a period from October 2009 to August 2014.

We have restricted ourselves to Facebook pages even though there are also several Danish wind energy Facebook groups. In contrast to Facebook groups which can be closed and require approval of membership by a group admin or member, pages are essentially public profiles. Facebook describes pages as being "for businesses, brands and organizations to share their stories and connect with people" but the extent to which this is the case in practice depends on a concrete evaluation of the actual interaction that takes place on the pages in question.

The Facebook pages that comprise the Wind2050 social media corpus are typically set up to pursue a specific policy agenda and/or to raise public awareness about a specific issue. At least this is true for the 12 pages advocating against wind energy in one way or another, and mostly in the form of local protest against a concrete project. A qualitative analysis of the interactions of these pages reveal that they function as information sharing fora where protesters voice their opinions and post links that they consider relevant for their agenda. It is our assessment that they mobilise and organise public protest, and does it in a way that is comparable in its 'publicness' to writing letters to the editor or taking part in other forms of public debate.

The 2 pro pages have significantly less interaction and seem to be curated more with professional/strategic communication in mind than advocacy and debate. One of them keeps a photographic diary of the construction process of the turbine project in question, the other provides updates on the planning process of another project. In practice they are sometimes engaged by critical users who take the opportunity to ask questions about the project or the process.

The dataset that we are archiving has been anonymized. It will still be possible to see posts and comments by identifiable users if you visit the Facebook pages in question. Here, however, users are free to retrospectively edit or delete their comments.

The social media corpus contains data from 3965 posts by pages and users from the following 14 pages on Facebook:

CON PRO

- Nej tak til flere Vindmøller i Slæggerup
- Nej tak til kæmpe-vindmøller mellem Hørselv, Herskind og Skovby
- Nej tak til kæmpevindmøller i Hørslev og Solbjerg
- Nej tak til kæmpevindmøller i Ørsted, Viby sj.
- Nej til urentable vindmøller på land, nej til "symbolpolitik"
- Vindmøller/Sundhed/Viborg
- Nollund vindmøller - Giv din mening
- Imod Megavindmøller i Thy
- Gruppe imod kæmpe vindmøller i Slagelse
- Gruppe mod vindmøller i Manna Kær
- Borgere mod Kæmpevindmøller i Vejen Kommune
- Møllegruppen Rynkeby

- Krogstrup Enge Vindkraft
- Kastrup Tiset Enge Vindmøller

Exploring the social media corpus

The social media corpus offers a range of analytical avenues of exploration that are not available with websites. Despite being smaller, and restricted to a Danish context, it allows for temporal analysis and can be examined on the micro level of interactions between individual actors in the controversy (here represented as anonymized Facebook user profiles).

The level of detail also means that many of the questions that can be pursued with this dataset are of a qualitative nature. In general, and if available, we would recommend visiting the Facebook pages relating to a particular turbine project before building a case study around it or engaging its actors in other ways. For the most part these pages are rich in information and provide a good grasp of the matters of concern that are central to their users.

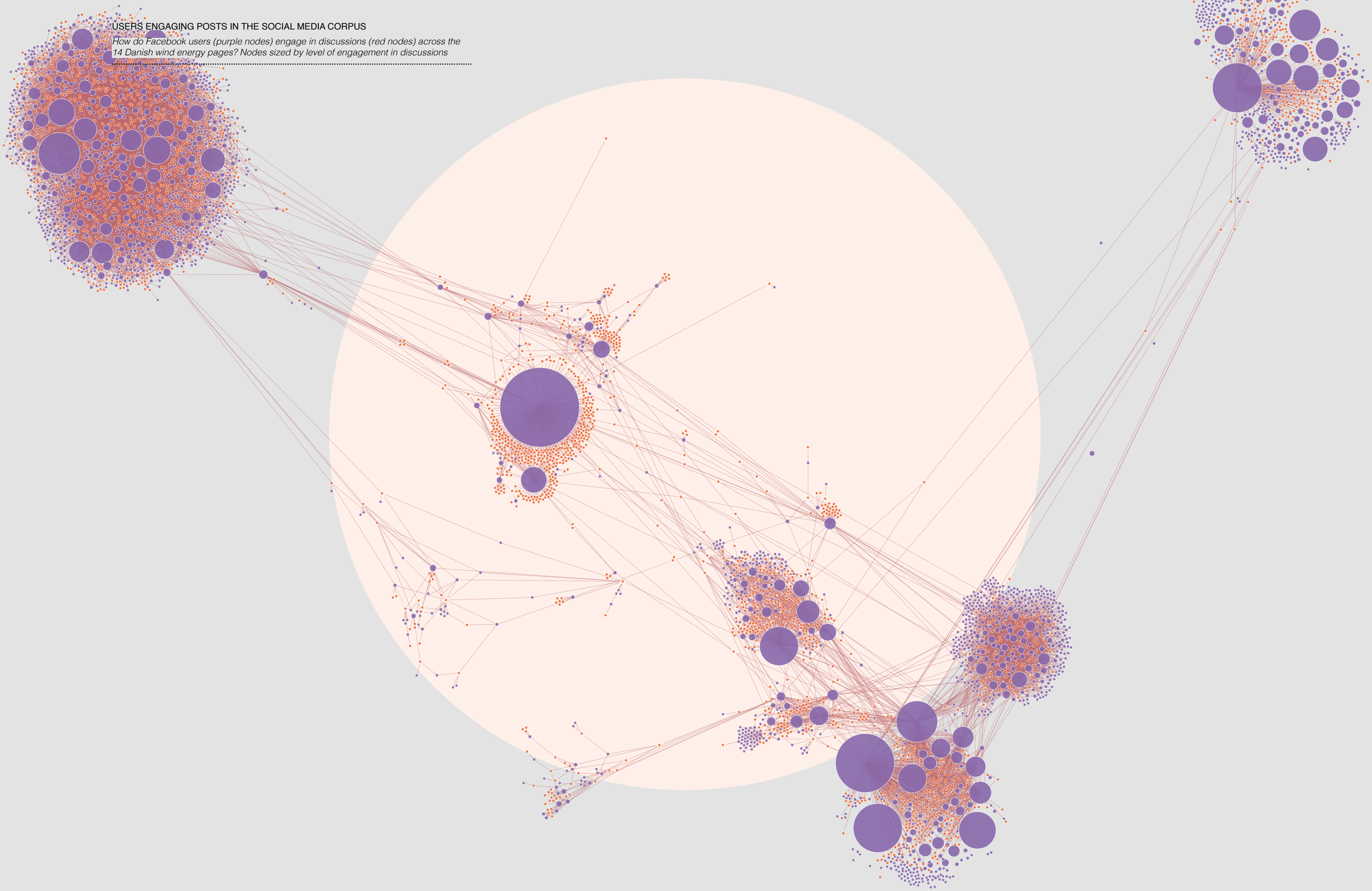
One of the ways in which the dataset can be of help in such qualitative inquiries is by directing the researcher's attention, filtering the data to focus on particular time intervals or keywords, or prioritising it through activity metrics like comments or engagement.

It is however also possible to make more structural inquiries with the corpus. Below we have mapped the how users (purple nodes) engage in posts (red nodes) by commenting or liking them (which will generate a link between the user and the post). The more a user engage in the same post the stronger becomes the link between them. The users have been sized by their level of engagement. The clustering corresponds to the 14 pages in the corpus, which means that users tend to engage mainly on one page rather than moving between pages. There are occasional bridges between the clusters when a user has been engaging posts on two or more of the pages. The most active users, however, typically restrict their engagement to one page.

This is potentially interesting because 11 out of 12 con pages on Facebook are dedicated to a specific local controversy. This would indicate that wind protest on Facebook tends to be genuinely local rather than driven by a group of national issue advocates making themselves heard across the different discussions.

USERS ENGAGING POSTS IN THE SOCIAL MEDIA CORPUS

How do Facebook users (purple nodes) engage in discussions (red nodes) across the 14 Danish wind energy pages? Nodes sized by level of engagement in discussions

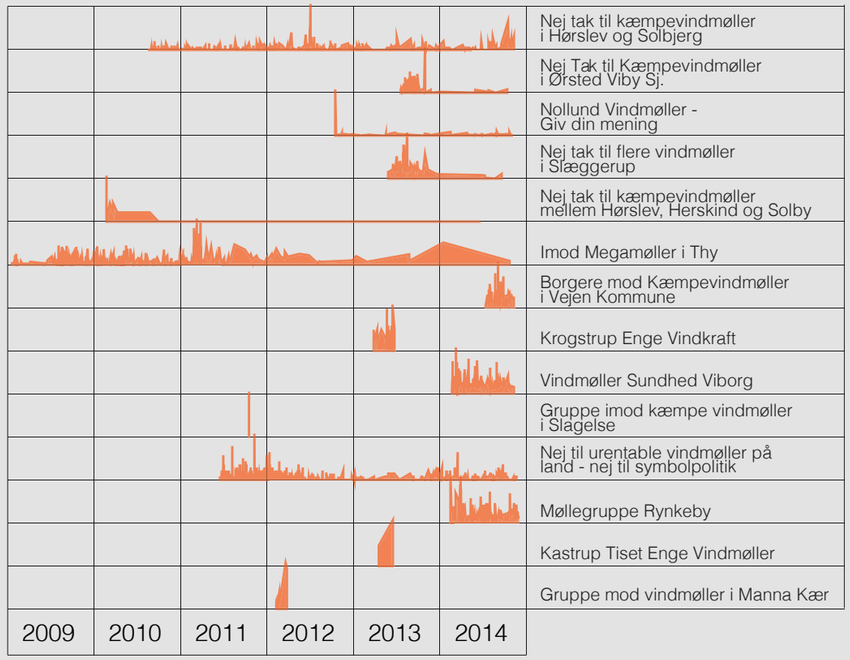


In contrast to the web corpus, which provides a synchronic snapshot, the social media corpus offers the opportunity of exploring discussions in time. There are numerous ways of doing so since all posts are time stamped and one might imagine asking questions about the popularity of different sources or different topics over time. Below we compare the engagement of users on the different pages of the corpus over time. Engagement of a post on Facebook is defined as “the percentage of people who saw a post that liked, shared, clicked or commented on it”. It provides an overall perspective on the changes in activity on a page.

Some pages, such as Nej tak til kæmpevindmøller i Hørslev og Solbjerg, Imod Megamøller i Thy, or Nej tak til urentable vindmøller på land - nej tak til symbolpolitik, have been consitstently active over a long period of time and continue to be active. Others, such as Borgere mod Kæmpevindmøller i Vejen Kommune, Vindmøller Sundhed Viborg, or Møllegruppe Rynkeby, have only recently become active. Others still have become inactive, either gradually or abruptly. The latter seem to be the case with the two pro pages, Krogstrup Enge Vindkraft and Kastrup Tiset Enge Vindmøller.

ACTIVITY OVER TIME IN THE SOCIAL MEDIA CORPUS

Measured as the engagement of posts on the 14 Danish wind energy pages



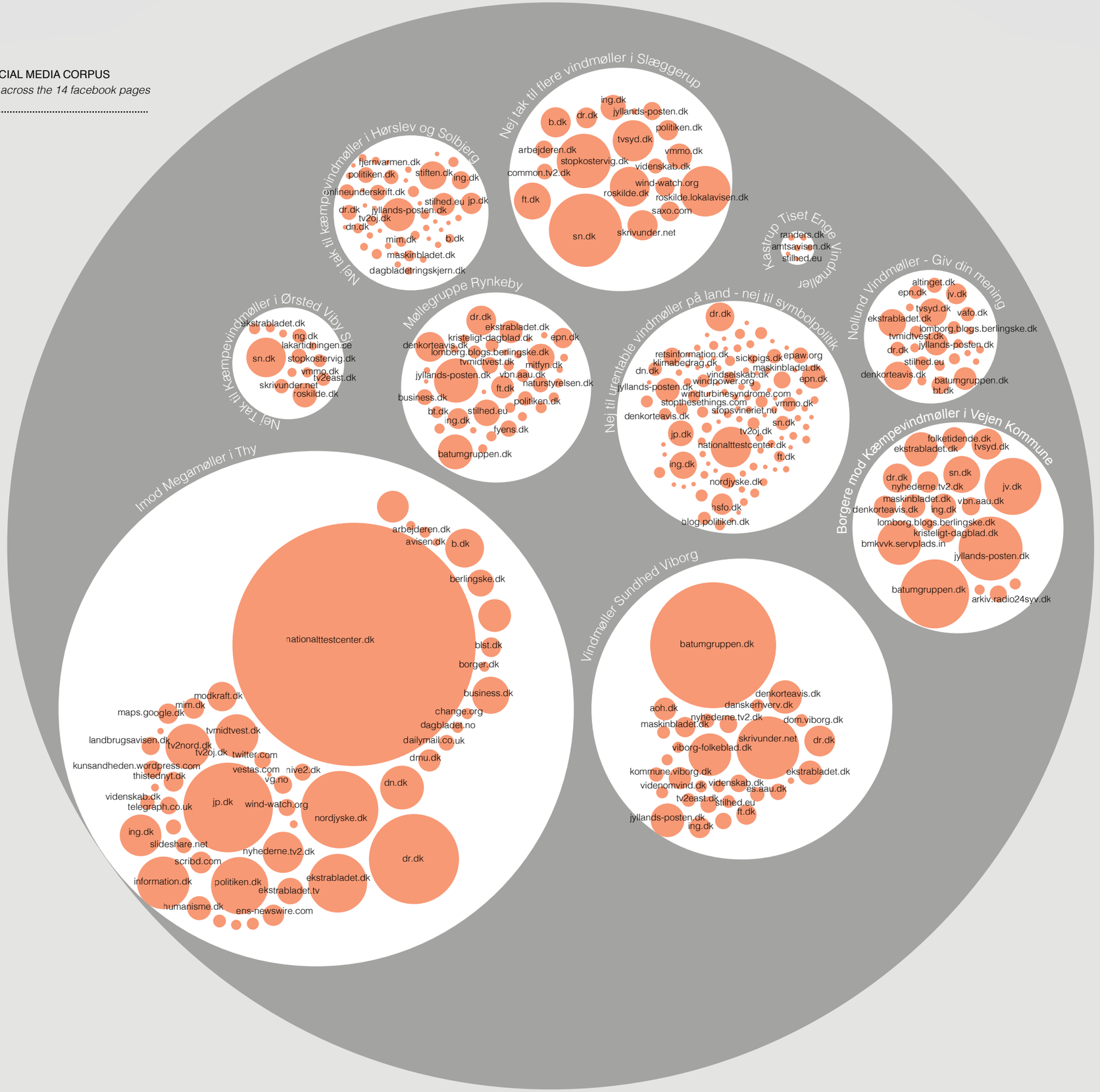
Finally, it is also possible to make overall comparisons of the contents of the interactions on the pages. Although we have not done so yet it would for example be possible to query each page with the issue dictionary and see how the resonance with different issue terms have changed over time.

Below on page 43-44 we have explored the sources quoted by users in their interactions on 9 of the pages (there was not enough linked sources in the interactions of the last 5 to make for an interesting comparison). Users normally quote a source when they link to a news article, a blog post, or a video. The dataset makes it possible to follow the deep links to news stories or videos posted in this manner by users. Here, however, we are exploring links shared by users on the level of their source domains. It enables you to see which websites, blogs or media outlets typically feed the discussions on a page and compare the differences between pages.

Each page is represented by a meta bubble and each source quoted on that page is represented as a bubble within it. The source bubbles have been sized according to the frequency with which they are quoted on the page. Some source names have been edited out to improve the legibility of the chart. They exist in the dataset.

TOP SOURCES IN THE SOCIAL MEDIA CORPUS

Most engaged link domains across the 14 facebook pages



References

Aitken, M. (2010). Why we still don't understand the social aspects of wind power: A critique of key assumptions within the literature. Energy Policy, 38(4), 1834-1841.

Beck, G., & Kropp, C. (2011). Infrastructures of risk: a mapping approach towards controversies on risks. Journal of risk research, 14(1), 1-16.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008.

Callon, M., Lascoumes, P., & Barthe, Y. (2009). Acting in an Uncertain World: An Essay on Technical Democracy.

Devine-Wright, P. (2005). Beyond NIMBYism: towards an integrated framework for understanding public perceptions of wind energy. Wind energy, 8(2), 125-139.

Haggett, C. (2011). Understanding public responses to offshore wind power.Energy Policy, 39(2), 503-510.
Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PloS one, 9(6), e98679.

Marres, N. (2004). Tracing the trajectories of issues, and their democratic deficits, on the Web: The case of the Development Gateway and its doubles.Information Technology & People, 17(2), 124-149.

Marres, N., & Rogers, R. (2008). Subsuming the ground: how local realities of the Fergana Valley, the Narmada Dams and the BTC pipeline are put to use on the Web. Economy and Society, 37(2), 251-281.

Nowotny, H. (2003). Democratising expertise and socially robust knowledge.Science and public policy, 30(3), 151-156.

Neuhaus, F., & Webmoor, T. (2012). Agile ethics for massified research and visualization. Information, Communication & Society, 15(1), 43-65.

Møller, L., Korsgaard, L.W. & Jørgensen L.Ø. (2014). Methodology Development for Creating and Issue Dictionary - a Contribution to the Mapping of the Wind Turbine Controversy", semester project, M.Sc. Techno-Anthropology, University of Aalborg

Rieder, B. (2013). Studying Facebook via data extraction: the Netvizz application. In Proceedings of the 5th Annual ACM Web Science Conference(pp. 346-355). ACM.

Rogers, R. (2013). Digital methods. MIT Press.

Venturini, T. (2010). Diving in magma: How to explore controversies with actor-network theory. Public understanding of science, 19(3), 258-273.

Venturini, T. (2012). Building on faults: how to represent controversies with digital methods. Public Understanding of Science, 21(7), 796-812.

Whatmore, S. J. (2009). Mapping knowledge controversies: science, democracy and the redistribution of expertise. Progress in Human Geography.

Wolsink, M. (2006). Invalid theory impedes our understanding: a critique on the persistence of the language of NIMBY. Transactions of the Institute of British Geographers, 31(1), 85-91.

Wolsink, M. (2007). Wind power implementation: the nature of public attitudes: equity and fairness instead of 'backyard motives'. Renewable and sustainable energy reviews, 11(6), 1188-1207.

Yaneva, A. (2012). Mapping controversies in architecture. Ashgate Publishing, Ltd..