Tracing Culinary Discourse on Facebook

A Digital Methods Approach

Munk, Anders Kristian

Published in:
Research Methods in Digital Food Studies

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

Tracing Culinary Discourse on Facebook: A Digital Methods Approach

Anders Kristian Munk

"Alain Ducasse says you can stop your Dry January and go back to enjoying wine!" (FineDiningLovers on Facebook, January 17, 2020)

"Anthony Bourdain's Final Book Will Be Released This Year." (FineDiningLovers on Facebook, January 17, 2020)

"The Paul Bocuse Restaurant Has Lost its Third Star...are Michelin correct?" (FineDiningLovers on Facebook, January 17, 2020)

If you follow a page like FineDiningLovers on Facebook, you will be familiar with the feeling that names of certain chefs and food personalities always appear in the same places. It is, for example, no surprise to find Alain Ducasse and Paul Bocuse mentioned by the same food blog. Similarly, when a restaurant writes a post that acknowledges the achievements of a chef like Magnus Nilsson, one could reasonably expect this restaurant to also write posts about Massimo Bottura, Joan Roca or some other superstar from the San Pellegrino best restaurants list. When a blogger mentions Rick Stein, other British TV-chefs like Keith Floyd or Marco Pierre White could likely be next. Spheres, foams and agar pearls rhyme with Ferran Adriá, Heston Blumenthal and molecular gastronomy. Distinct ways of talking about food come with distinct gastronomic reference points (chefs, ingredients, cooking techniques, terroirs, restaurants, you name it).

In this chapter, I will show you how to map such discursive patterns with digital methods. In a dataset of 102M posts collected from 242K food-related Facebook pages worldwide, I track mentions of almost 700 chefs. I construct a 'co-chef' network (a network of chefs connected to each other if they are mentioned by the same pages) which I subject to a visual network analysis. I provide the network file so that you can follow my steps but it is of course also possible to take a different direction on your own (for example based on a smaller dataset, differently curated, or from a different medium). I also provide a series of intermediary visualizations annotated with my field notes which will hopefully make my thought process clearer.

I am interested in chefs, and Nordic chefs in particular, an interest I have been cultivating as part of a wider effort to understand how the New Nordic Food movement has developed and diffused over time (Munk & Ellern 2015). I have previously mapped that development by following different ingredients, concepts and practices (for example sea buckthorn or fermentation) in a multi-sited ethnographic study across Scandinavia (in the style of Marcus 1995). Social media makes it possible to associate discussions about food with specific actors, in specific places, at specific points in time, which has opened up new opportunities for this kind of mapmaking (Munk & Jensen 2014). Digital traces make it possible to follow how ingredients and cooking techniques are being adopted into culinary practices as these are exhibited and performed in text and images on the Web.

Chefs and other food personalities are, in this respect, a relatively straightforward trace to follow. When the same food blogger cites René Redzepi or Claus Meyer, it is typically a sign of recognition and we can be fairly certain that they are also referencing the New Nordic movement. Other traces are comparatively more challenging. We can for example have a presumption that sea buckthorn is an indicator ingredient for the New Nordic Cuisine, but there could be other sources of inspiration that would cause a food blogger to post a recipe with this particular ingredient.

The simplest approach to understanding how Nordic chefs have become household names outside of Scandinavia would probably be to search for mentions of them by pages outside the region. However, as stated in the opening paragraph, I already have a clear impression from casually browsing these pages that certain chefs tend to be mentioned together in distinct discursive patterns. So, rather than presuming in the research design that geography is what matters here I am going to adopt what Richard Rogers calls a 'post-demographic' approach to social media (Rogers 2009a) and assume that food talk on Facebook is organized around communities of interest rather than geography or language. Instead of asking where in the world Nordic chefs are talked about, I ask to what extent Nordic chefs have broken out of their local Scandinavian food-talk bubbles and migrated into other and more international discourses.

The difference is central to understanding the contribution of network analysis to this kind of food study. If we presume that discourses about food are national or regional, in the sense that there is Danish conversation about food on Facebook which is qualitatively different from the French or the American conversation about food, then there is no need for networks. I know where the food-related pages in my dataset are located geographically, what remains is simply to count how much Nordic chefs are mentioned by each of them. If, on the other hand, we do not presume to know where to draw the boundary between different food discourses – if we turn this boundary drawing into a central empirical question for the project – then network analysis becomes useful as a way to show emergent patterns in the way food is being talked about. What is at stake for the analysis is the ability to claim, based on very large volumes of text, that some chefs are talked about in qualitatively different ways than others, not just that they are talked about in certain places.

Digital methods and digital food studies

The field of digital methods has evolved over the past 20 years as new kinds of natively digital empirical material have become available and new computational techniques for exploring large volumes of unstructured data like text and images have become more accessible (Manovich 2011, Rogers 2019). One of the characteristic features of this development is the idea that data intensive analysis is no longer the remit of quantitative methods alone. Indeed, the combined ability to find patterns in messy data without discarding the rich information that allows researchers to make deeper forays into the context and meaning of a statement, a link, a 'like', or other seemingly shallow digital traces has opened up new avenues for asking and answering qualitative questions with the aid of relevant computational techniques (Blok & Pedersen 2014, Munk 2019).

Besides a more general sociological/anthropological interest in bridging the quali-quantitative divide (see also Latour et al. 2011), digital methods have typically been advanced as a way to study how issues are debated online (Venturini 2012, Marres & Moats 2015, Burgess & Matamoros-Fernandez 2016). With a few exceptions (e.g. Munk & Ellern 2015, Munk et al. 2016), food-related issues have mostly been absent from this literature, and yet it is a literature that is highly relevant to digital food studies. Indeed, there is perhaps no other topic so photographed and discussed online as food and the fact that this material is natively digital (i.e. brought into being on digital platforms) as opposed to digitized (e.g. scanned documents or images) produces a particular set of methodological challenges that online issue mappers have already been facing and dealing with for a couple of decades (for an early example see Marres & Rogers 2000). Any digital food study that engage with debate or discussion from online sources can therefore productively look to issue mapping and digital methods for inspiration.

One of the key challenges, that is shared between digital food studies and online issue mapping, is our reliance, as researchers, on the way platforms organize and make data available. We are implicitly or explicitly soliciting methodological decisions to the architects of these platforms all the time. The case I introduce below relies on two types of natively digital material, namely posts from pages on Facebook and names from lists and categories on Wikipedia. Just to take the former: A Facebook page is meant to be the public profile for a business or an organization, as opposed to your personal profile or a group that you can be a member of. A page can have an address and a geolocation, it can be categorized by its owner according the kind of business it is doing, and it is allowed to 'like' other pages. A page can post content, typically images, links or videos accompanied by text, and it can allow its followers to comment or react to that content. If I want to find patterns in the way food actors on Facebook cite chefs, and especially if I want to do it at scale and therefore with some level of automation, I have to turn these features of the medium into affordances for the research I am doing. Rather than trying to come up with my own definition of what counts as a food actor and what does not (i.e. who should be included in the dataset), I might as well make creative use of the page categories that are already in the data and thus ensure that my research design can actually be operationalized within the affordance space of Facebook. Similarly: rather than having an ambition to understand how these food actors really feel about their gastronomic heroes, it makes more sense to ask a question that recognizes how Facebook pages are in fact used, for example as strategic marketing tools where we will be able to see who food actors find it opportune to associate with.

Noortje Marres encourages us to acknowledge, and think critically with, this distributedness of methods (Marres 2012). Digital media are not just archived online interaction, but methods in their own right. They nudge our informants to perform in certain ways, suggesting to them paths to be taken (and paths not to be taken), offering them ways to express themselves and interact with each other (while discouraging other ways to do so). We cannot ignore or work around this already existing methods ecology. Richard Rogers therefore mobilizes the notion of 'repurposing' to remind us that we should never uncritically rely on whatever the medium wants us to do, but rather creatively reuse digital media for purposes that fit our research interests (Rogers 2018). There is of course no point in *re*purposing anything if the original purpose of that thing is not properly understood. It follows that doing research *with* data from the web invariably involves doing some research *on* and *about* parts of the web itself. As Rogers puts it, we have to be able to 'follow the medium' (Rogers 2009b). Marres makes a distinction between 'discursivist' and 'empiricist' approaches, the latter embodying the commitment to unpack how digital media platform make a difference to issues rather than simply following what is being said by different actors about these issues (Marres 2015).

1. Building a corpus of Facebook posts from food-related pages

The first task of any digital methods project is to be build a dataset or a 'corpus'. In this case I needed a corpus of posts from food-related Facebook pages, and to some extent I already had one. By casually browsing from one post in my feed to the next I had manually collected a list of almost 200 pages that I thought could be interesting for the project to look at (in the sense that they frequently mentioned chefs when they posted, including the occasional Nordic chef). But since this collection was based on what was presented to me in my feed and hence heavily influenced by my actions as a Facebook user, I needed a better way to find pages. Besides, I wanted a larger corpus to have a better chance of detecting patterns in the way chefs were cited together. This poses a problem since Facebook, like most other digital platforms, does not offer a collection of carefully curated material for research. It is not an archive where you can know the principles of collection and search it accordingly. Neither is it possible to know the full population of anything, and so even if it was possible to systematically search for food related pages, proper sampling would still be impossible since we would not be able to know what it was a sample of. Again, this is why Richard Rogers calls social media 'post-demographic machines' (Rogers 2009a) and why re-purposing is important. If sampling and representativeness, at least in any demographic sense of the word, is impossible, then what can we do within the possibilities offered to us by the platform?

As mentioned, owners of a Facebook page are allowed to 'like' other pages on behalf of their own page. This feature can be repurposed for a 'snowball' strategy where you start data collection with a small seed of material that is allowed to point you to more relevant material, in this case a seed of pages pointing to other potentially relevant pages through 'likes'. The seed can be very subjective and idiosyncratic – remember that my list of 200 pages was compiled relatively intuitively from browsing posts that I found interesting in my feed – but as the snowball picks up speed this idiosyncratic starting point is gradually counterbalanced as the question of what counts as interesting material is increasingly outsourced to the informants (in this case to more and more Facebook pages). It is possible to experiment with several snowballs from different starting points and compare when the corpora begin to converge.

In order to iteratively ask a platform like Facebook how an evolving collection of pages provide 'likes' to other pages it is necessary to interact with the API (Application Programming Interface). This is the method through which third party applications pull data from the platform and hence another way in which we repurpose the platform. APIs are everchanging machines designed for other purposes than academic research, which makes it hard to replicate datasets, but it is normally the best option available to researchers who want to work with social media data (Lomborg & Bechmann 2014). Most of the popular tools for data harvest are therefore also API-based. The corpus I have built for this project was constructed through the publicly available API endpoints in July 2018. To automate the process, I built a so-called script in the programming language Python. A script is a set of conditional instructions that tells the computer what to do in different situations. My script did the following:

1. Load the seed list of 200 pages
2. For each of the seed pages, call the API and ask what other pages it likes.
3. For each of the newly found liked pages, decide if it is food-related and should be included in the corpus. For this task I repurposed the page categories which allow page owners to self-classify under themes like 'Chinese Restaurant', 'Organic Grocery Store', or 'Dairy Farm'. I manually went through all the page categories that were available on Facebook at the time, identifying 317 of them as food-related. The script calls the API to ask for the page category, checks if it belongs to the 317 food-related categories, and, if it does, includes the page in the corpus.
4. Iterate the process on the next layer of the corpus: for each of the food-related pages liked by a page on the seed list, call the API, ask for page likes plus categories, and include in corpus if relevant.
5. Keep iterating and add new layers to the corpus.

I kept running the script until I had accumulated a corpus of 242.000 Facebook pages from around the world that I knew belonged to a food-related category. This was a practical cut-off point as the

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

project was running out of time, but in principle the snowball could have continued. I then called the API to ask for all the posts made by each of the pages in the corpus. The result is a collection of 102M text documents in different languages. In the table below, I provide a schematic overview of my protocol for all the steps in the project and suggest what you could do differently at each step.

Step	**What I did**	**What you could also do**
1. Build a text corpus	*Collected 102M Facebook posts from 242.000 food-related pages found by snowballing page likes from a seed list and evaluating if page categories pertained to food.*	*Collect Facebook posts from pages you already know to be interesting (no need for a snowball). You could use a tool like the FacePager which is free and easy to use.* *Collect other kinds of online food talk such as tweets (you could use the TCAT tool), instagrams (you could use InstaLoader), or Reddit threads (you could use 4CAT).* *Collect other kinds of documents, such as magazine articles or scientific papers. Any kind of textual corpus will work with the next steps in the protocol.*
2. Query the corpus	*Collected names from lists of chefs on Wikipedia and searched the corpus for mentions of each of them.*	*Curate your own list of names that you know are interesting.* *Collect names from different lists of chefs or other food personalities.* *Use natural language processing to recognize names from the text in the corpus.*
3. Find patterns	*Built a network of chef names connected to each other if mentioned by the same page. This was done in Python using NetworkX. Then used a force vector layout algorithm in Gephi to find clusters of chefs that tend to be mentioned together.*	*Build a network of pages connected to each other if the mention the same chefs. This would make it possible to group the pages rather than the chefs.* *Build a network of pages connected to the chefs they mention. This would make it possible to explore directly in the graph visualization which pages mention a particular chef or which chefs are mentioned by a particular page.* *You could use Table2Net as a user-friendly alternative to building networks in Python.*

4. Interpret the results	*Qualitatively analysed each of the clusters.*	*Although you can extract quantitative metrics for any given network, a qualitative analysis of clusters, centers and bridges will probably be more interesting for your research interests.*

2. Building a query design for Nordic and international chefs

Having built the corpus, I now have to find a way to record when a chef is mentioned. Given that the size of the corpus does not permit manual coding, I have to do this detection automatically. This task would be the same regardless of how the corpus had been constructed (from Instagram, Twitter, lifestyle magazines, news media, cookbooks, whatever) as long as the time required to read and code the material exceeds what is possible inside the resource constraints of the project. It would also be the same challenge if I had been looking for ingredients, place names, or cooking techniques instead, although, as we shall see, the fact that personal names are usually spelled the same way in different languages means that more solutions are available for chefs than for, say, ingredients.

Essentially, there are two options. The first is to use a type of *Natural Language Processing* (NLP) called *Named Entity Recognition* (NER) to automatically detect personal names in the text. This option comes with the benefit of being able to discover bottom-up who is cited by the pages, but with the disadvantage of not knowing which of these discovered people are chefs. The list of extracted names would therefore have to be manually curated (again an impossible task considering the size of the corpus). Had I been looking for ingredients or place names instead, there are good models available. NLP also works to different standards in different languages (a consequence of differences in the volume and quality of available training data) and since I am working with a multi-lingual corpus I should reasonably expect the solution to disadvantage food talk in smaller languages.

The second option is to compile a list of chef names that can be used as search terms to query the corpus. This comes with the advantage of knowing in advance that the names on the list are indeed chefs, and potentially, depending on the information available to the list builder, different kinds of metadata such as which country the chefs are from, what style of food they cook, etc. Since this option relies on a simple search, rather than language technology, it is also less sensitive to language differences. Once the name of a chef is known, it does not have to be translated (although non-Latin alphabets still pose a challenge). The drawback, however, is that we have to decide *a priori* who are relevant chefs to search for rather than soliciting this task to the food actors we are studying on Facebook.

Given the practical challenges associated with the NLP approach I am, in this case, going with the simpler query option (but with a smaller English language corpus the benefits of the NLP approach could likely have outweighed its disadvantages). The next step is therefore to acquire a list of chefs to work from, and this is where it becomes necessary to repurpose Wikipedia. As it turns out, getting a good list of chefs is more difficult than you might think. There are national and international chef associations, but they typically do not list their members online, and, if they do it, it is debatable whether these members are in fact the top tier chefs that would be cited in online food talk. There are also privately curated lists of chefs. San Pellegrino, for instance, publishes a ranking of the best chefs in the world, but I suspect that it is itself part and parcel of a very distinct culinary discourse already. Wikipedia, on the other hand, offers what appears to be a quite encompassing list of chefs from different centuries (https://en.wikipedia.org/wiki/List_of_chefs). I am here scraping it for names under the 20th and 21st century headings.

Believing that Wikipedia is not a perspective, however, would be a mistake. The point about repurposing is still valid here. On the 'talk' page behind the list (https://en.wikipedia.org/w/index.php?title=Talk:List_of_chefs) the editors are discussing the criteria for inclusion. Who should be counted as a chef? Who is important enough to be on the list? It turns out that the original version of the list was copy pasted from French Wikipedia and since elaborated. It is this online community of Wikipedians and their ideas about good list building that I am now repurposing as part of my method. The same is true for the categories of Norwegian (https://no.wikipedia.org/wiki/Kategori:Norske_kokker), Swedish (https://sv.wikipedia.org/wiki/Kategori:Svenska_kockar), and Danish chefs (https://da.wikipedia.org/wiki/Kategori:Kokke_fra_Danmark) from which I compile the Nordic part of the query design.

3. Building a co-chef network

The process of turning the lists of names from Wikipedia into a network of chefs cited by the same pages is essentially a very large search operation. For each of the 683 chefs I am searching the text of each of the 102M posts in the corpus to see if the chef is mentioned in the actual post text on Facebook. I am using another Python script for this purpose [link to script repository]. Whenever two chefs are mentioned in posts by the same page, the script registers a link between them. The script also associates each link with a weight that becomes heavier as more pages make the same connection. In network parlance I am representing the chefs as 'nodes' and the links as 'edges' (see Figure 1), which simply means that I have converted the results from the query to a relational data structure where chefs are always related with different weights to one another. Using the Python library NetworkX I am able to export the network as a .gexf file that can be opened in Gephi and explored visually. Gephi is an open source tool for explorative network visualization (Bastian et al. 2009) and can be downloaded freely at https://gephi.org/. It allows easy layout and manipulation of graphs, such as sizing, coloring, and filtering nodes and edges, as well as placing them visually in space. As a user you are encouraged to experiment with these settings and explore their implications.

In order to keep visual track of my Nordic chefs I am giving my nodes a triangular shape (useful when color is not an option). I then apply a layout algorithm called ForceAtlas2 (Jacomy et al. 2014) that helps me spot clusters of chefs that are frequently cited together. A 'force vector' or 'spring based' layout works as a repulsive force that pushes all nodes apart from each other unless they are held together by an edge (i.e. edges act as springs). In Figure 1, I am spatializing a smaller sample of the network with ForceAtlas2 to demonstrate how nodes become visually grouped into clusters. Figure 2 also shows how the network is constructed: in a situation where FB Page 1 mentions Chef X and Chef Z, these two chefs will be represented as nodes connected by an edge. When FB Page 2 then also mentions both Chef X and Chef Z, the weight of the edge between the nodes increases.

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

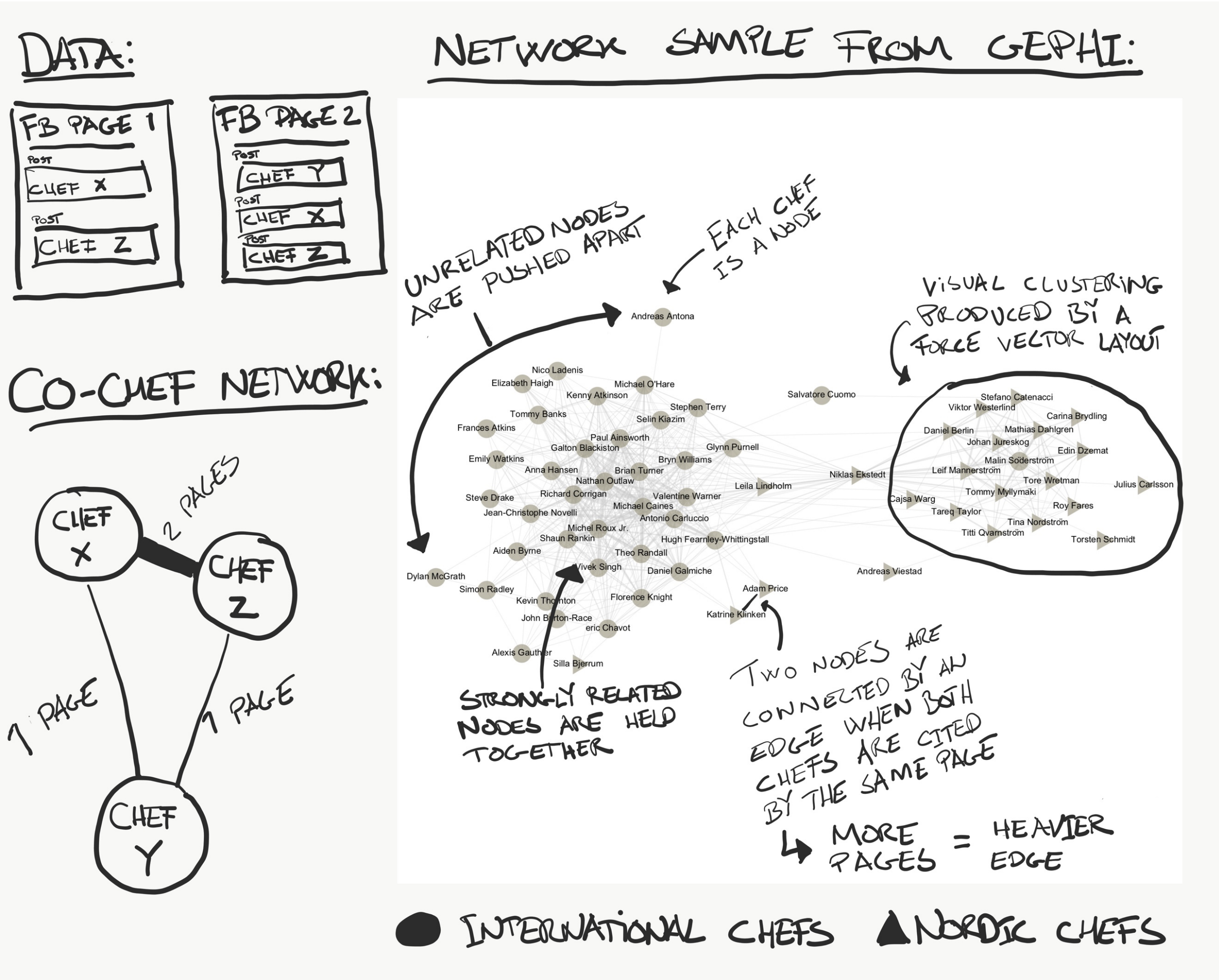

Figure 1: First concept of how to build the co-chef network and a sample testing the idea in Gephi.

The network sample in Figure 1 can be seen as a proof of concept that force vector spatialization, i.e. the process of visually pushing nodes apart in the layout unless they are connected by edges, allows us to see relevant clusters of co-cited chefs. Two such clusters can be clearly delineated. Most of the Nordic chefs (triangular nodes) tend to be cited together by some pages, while all the international chefs (circular nodes) tend to be cited together by other pages. All the chefs in the Nordic cluster to the right are Swedish which would suggest that they are part of a national culinary discourse. This is further reinforced by the fact that most of the Nordic chefs found outside this cluster (Katrine Klinken, Silla Bjerrum and Adam Price) are Danish. Leila Lindholm, who is Swedish, clusters with the international chefs, which would suggest that she has become part of a different culinary discourse than her countrymen in the sample, a reasonable interpretation given that she has had her cookbooks translated into multiple languages and has also appeared in cooking shows on the BBC.

Although the force vector layout places most of the Swedish chefs in close proximity to one another their nationality never has an influence on the layout. The only information the algorithm relies on to produce the layout is the relations between the nodes. When most Swedish chefs in the sample end up in the same cluster it is simply an expression of the fact that they tend to be cited by the same pages. Rather than assuming that nationality determines the way chefs belong to culinary discourses, which would prevent us from seeing how a chef like Leila Lindholm belongs differently, the network approach allows us to map discourses relationally. In fact, since nationality is not a factor in the clustering, we can also use the visualization to identify chefs that have been miscategorized by Wikipedia. Malin Söderström, for instance, is Swedish, but the Wikipedians who curated the lists from which I got the names in the first place did not recognize her as such. Yet, based on the way she is cited alongside other Swedish chefs, the force vector layout is able to place her in the cluster emerging as a distinctly Swedish culinary discourse to the right in Figure 1.

4. Analyzing the co-chef network

The spatialized sample thus promises a method for delineating culinary discourse relationally and from the bottom up. However, as we move from the sample to the full network it becomes clear that additional manipulation is necessary for the force vector layout to keep showing clusters. In Figure 2, which is an unfiltered version of the full network, we still see some smaller clusters on the right but the vast majority of the nodes are grouped in a densely connected 'hairball' where it is possible to discern some regions (nodes on the left side of the hairball seem more related to each other than to nodes at the bottom or at the top of the hairball) but not at all clear where to delineate distinct discourses. In network terms this hairball can be explained by the fact that many of the nodes are connected to almost all other nodes in the graph. When we run a force vector layout, the edges of these highly connected nodes act as springs that prevent the rest of the nodes from being pushed

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

apart. In Figure 2 I have sized the nodes by their 'degree' (a measure of how many connections they have) to make the effect clear.

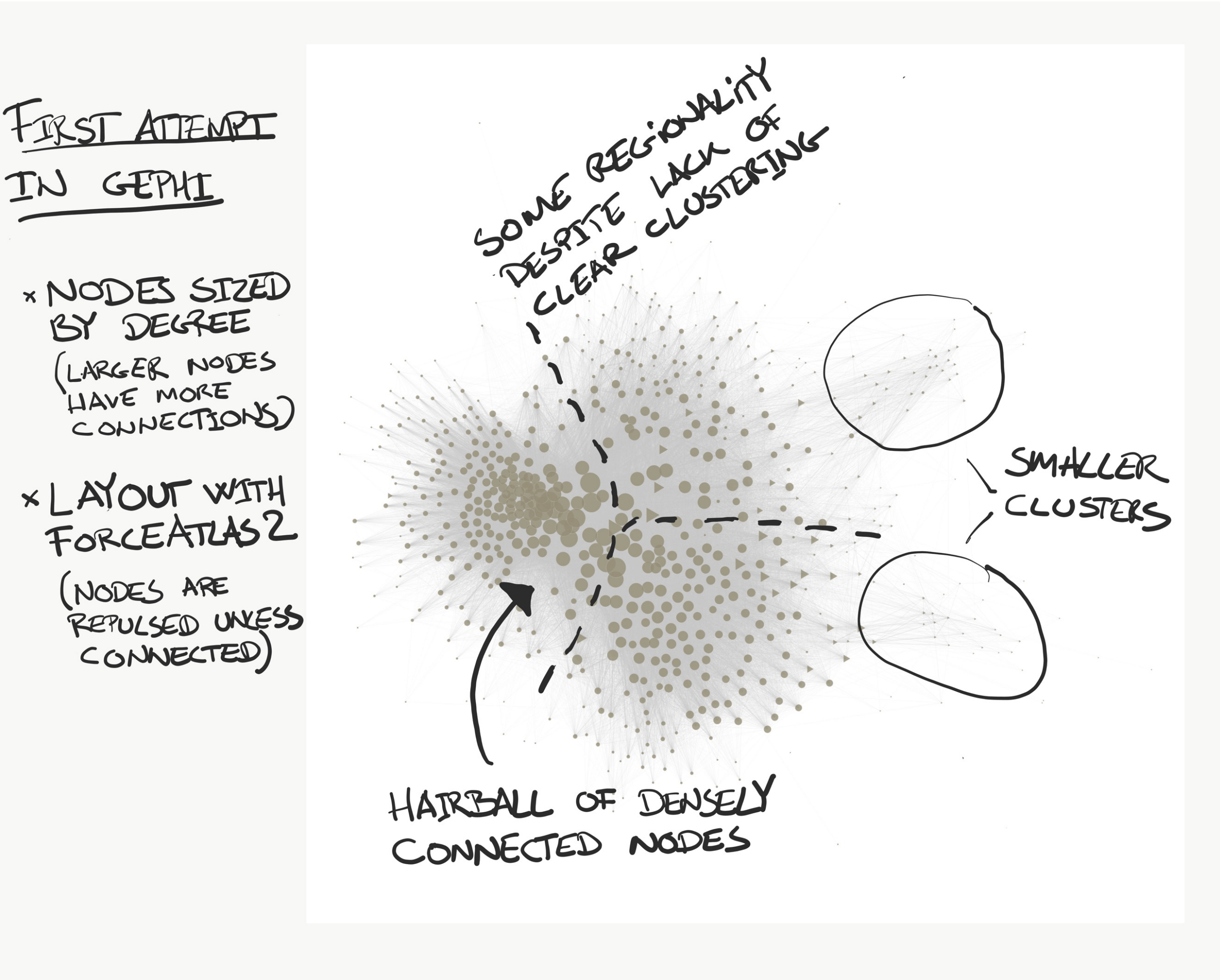

Figure 2: The full network of 669 chefs connected by co-occurrence on 242.000 food pages analysed in Gephi.

Translated to our dataset we can surmise that some chefs are cited so widely that they have also been cited, at some point, by the same page as almost any other chef on our list. There is no particular pattern to the way these omnipresent chefs are being talked about. Or more precisely, there is no particular pattern to the company of chefs they are being talked about in. It will become clear later that such patterns among chefs do indeed exist if we disregard the most omnipresent among them – there are very clear spheres of food talk across our 242,000 Facebook pages, each of them characterized by a distinct universe of cited chefs – and so the fact that some chefs are so widely talked about that they transcend this pattern constitutes a finding in itself.

In Figure 3 I have produced a so-called 'ego network' for the three chefs that co-appear with most other chefs in the network. An ego-network is a subset of the graph consisting only of the nodes connected directly to the node in question (in our case, the chefs that are mentioned on the same pages as the chef in question). The first thing to notice here is that there is almost no noticeable difference between these subgraphs and the full graph shown in Figure 3. It makes very little difference to the graph if we introduce a filter that excludes all chefs that have not been cited on the same page as Jamie Oliver, Gordon Ramsay or Paul Bocuse.

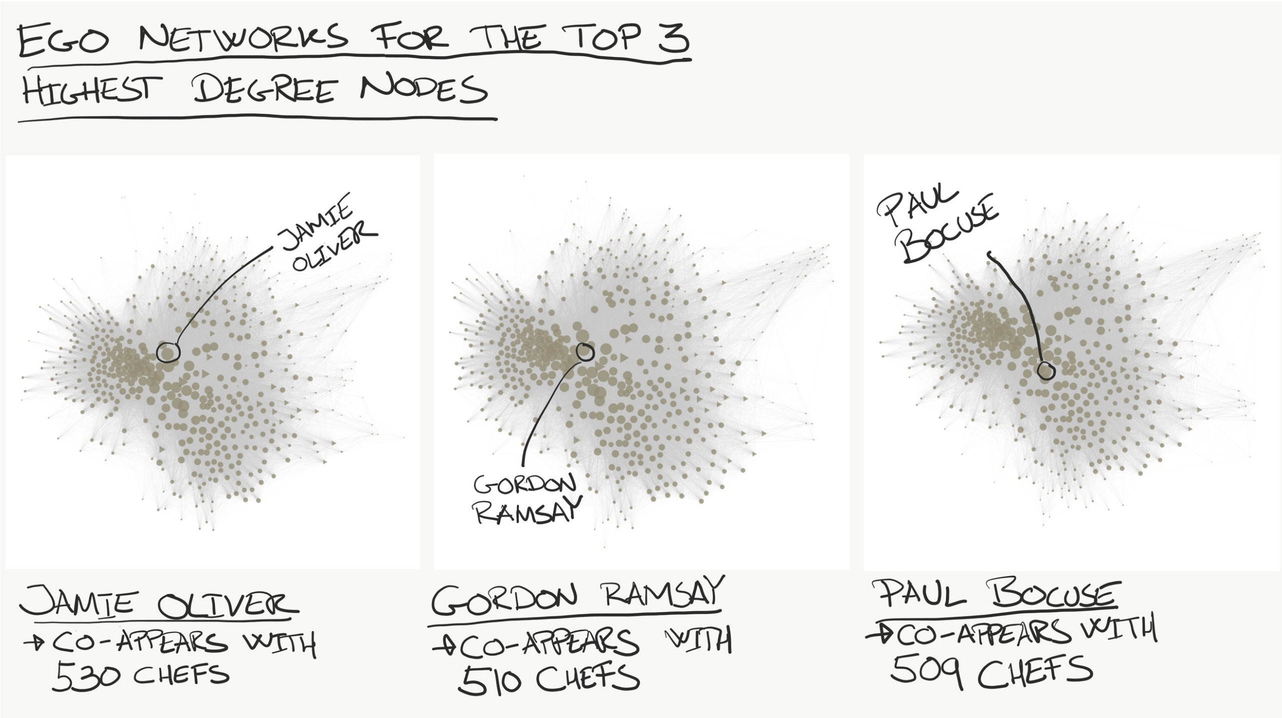

Figure 3: Exploring the neighbourhoods of the most omnipresent chefs and the effects of these nodes on network topology.

This is a good time to remember why we are doing visual network analysis in the first place. Typically, and especially when we are working with force vector layouts, the idea is to discover a topology where clusters of connected nodes can be distinguished from one another. In our case the idea is to spot clusters of chefs that tend to be cited by the same pages. Such clustering tendencies in complex relational data can be hard to make visible, but force vector layouts are designed to help us do it. They are not, however, particularly helpful if we want to spot the top 100 most connected nodes. These nodes are much better shown on a simple list where the items can be ranked. In fact, trying to retain this information as part of a network visualization can actively prevent us from appreciating the underlying topology (as we have just seen above). Instead we could just produce a list of the high degree chefs, like so:

1. Jamie Oliver (co-cited on pages with 530 other chefs)
2. Gordon Ramsay (co-cited on pages with 510 other chefs)
3. Paul Bocuse (co-cited on pages with 509 other chefs)
4. James Beard (co-cited on pages with 490 other chefs)
5. Thomas Keller (co-cited on pages with 486 other chefs)
6. Alain Ducasse (co-cited on pages with 481 other chefs)
7. Julia Child (co-cited on pages with 475 other chefs)
8. Ferran Adriá (co-cited on pages with 471 other chefs)
9. Massimo Bottura (co-cited on pages with 457 other chefs)
10. Daniel Boulud (co-cited on pages with 456 other chefs)
11. Heston Blumenthal (co-cited on pages with 451 other chefs)
12. René Redzepi (co-cited on pages with 446 other chefs)

13. Anthony Bourdain (co-cited on pages with 436 other chefs)
14. Eric Ripert (co-cited on pages with 425 other chefs)
15. …

Now that we know that these chefs are very widely talked about across almost any context, we can freely filter them out of the visualization. There is, in fact, nothing more an analysis of the network topology can teach us about them. In Figure 4 I therefore experiment with different thresholds for such a filter, subsequently redoing the force vector layout for the remaining graph in each instance (this is easily done in Gephi). A topology with distinct clusters gradually emerges as more high degree nodes are filtered out. With 75% of the original graph visible, 6 clusters can be clearly delineated.

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

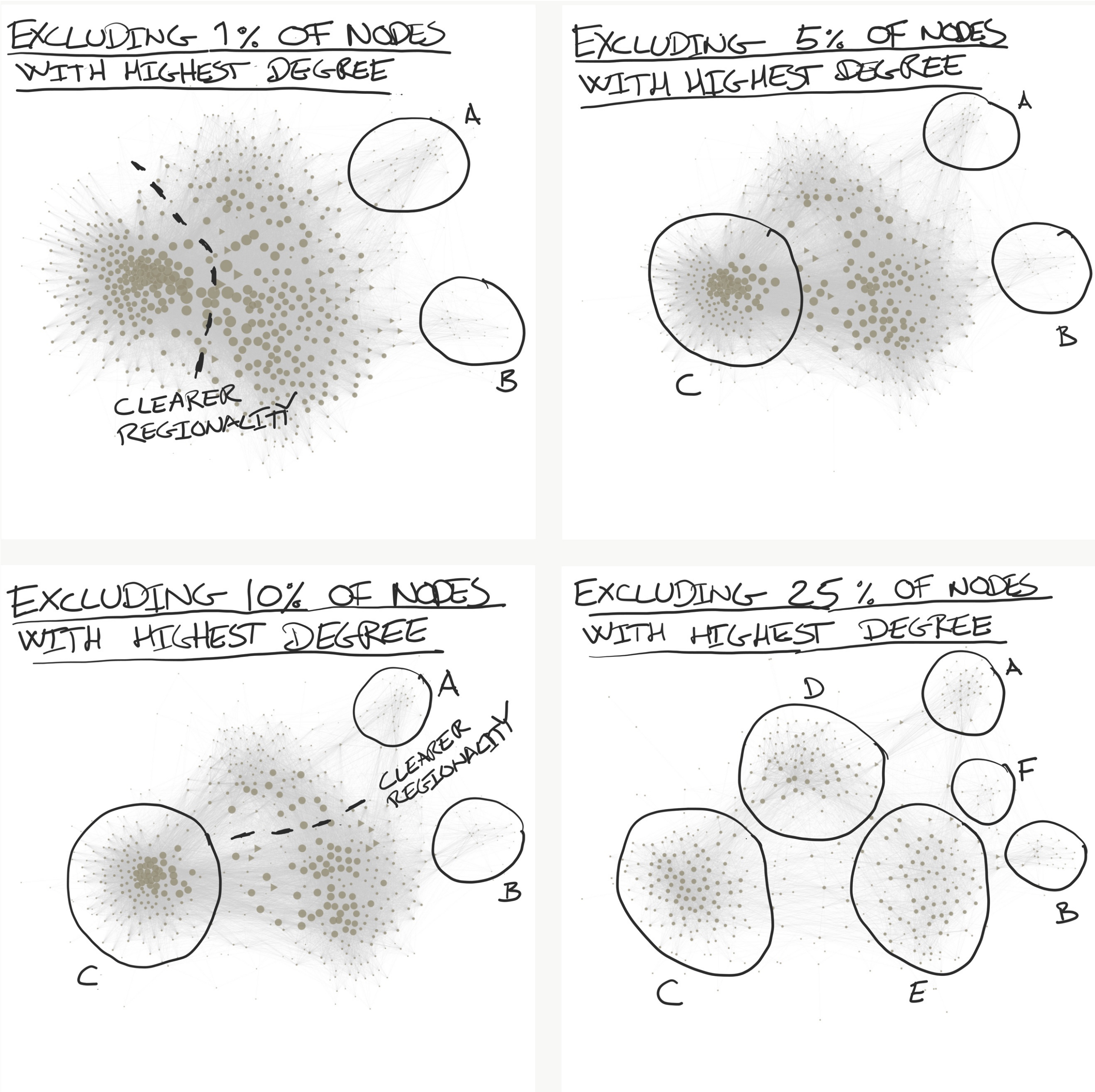

Figure 4: Experimenting with different thresholds for a filter based on node degree.

Eventually, through trial and error, I settle for a filter that removes the 20% highest degree nodes from the graph (see Figures 5 and 6). This setting gives the clearest clustering (again, the point of removing the highest degree nodes is to see clustering) with the least loss of information (the point of not filtering too heavily). It remains, however, a qualitative cut that could justifiably have been made differently. As I lower the threshold for the degree filter the chefs that are left in the graph become less and less omnipresent. The graph produced by removing the top 10% highest degree nodes, for example, has three clear clusters and a large hairballed center from which the remaining three clusters will emerge when we lower the threshold further. Some of the chefs that produce this hairball and get filtered out with a lower threshold (Atul Kochhar, Tom Kitchin or Angela Hartnett, for instance) only have strong connections across two or three of the eventual six clusters. They are not omnipresent in the same way as Paul Bocuse, Jamie Oliver or Gordon Ramsay (the highest degree chefs). Choosing a different filtering threshold would have allowed me to explore these differences in omnipresence further. Indeed, using an entirely different filter, such as the betweenness metric which measures the degree to which a node is on the shortest path between all other nodes in the network, I could have reoriented the analysis towards chefs that bridge two or three otherwise unrelated clusters without necessarily being omnipresent in all food talk on Facebook.

No matter how I set the filter, though, I want to get to a point where I can meaningfully interpret the resulting topology. Having set the threshold at 80% I am left with 6 clearly delineated clusters. The question is why? What could be the reason why the chefs on my list tend to be mentioned together in this particular pattern across the 242,000 food pages in my dataset?

At this point I can either dive straight into the chefs found in the different clusters, search for the textual contexts in which they are mentioned in my dataset, read about their background online, and thus try to stitch together a narrative that makes sense. Qualitative work, in other words. Or, if it is available to me, I can try to project some additional information onto the network, for instance by sizing or coloring the nodes according to some variable. In this case I know the country where each of the 242.000 food-related pages comprising the dataset are located geographically (my script got this information from the Facebook API when it was snowballing). I was therefore also able, when I was searching through my list of chefs, to count how much each chef is mentioned by pages from each country. In Figure 5 I have projected some of this information onto the network by sizing the nodes according to how much they are cited by pages from six different countries.

The result provides a good clue to an interpretation. Two of the larger clusters are dominated by chefs that tend to be cited by American and British pages respectively. This does not prevent the chefs in these clusters from being cited by pages from other countries as well, but the fact that the clustering corresponds to a geographical delimitation tells us that American and British pages produce two different patterns of co-cited chefs, i.e. two distinct culinary discourses. The same is true for Danish, Swedish and Norwegian pages, which produce the three smaller clusters on the top right. If we take a qualitative look at the chefs comprised in these five culinary discourses it turns out that the vast majority of them are also American, British, Swedish, Norwegian and Danish nationals respectively. In other words, when we filter out the chefs that everyone talks about in any context, it seems that Danish pages talk about Danish chefs only, Swedish pages about Swedish chefs only, and so forth. These five culinary discourses are essentially reflective of five national food scenes.

The sixth cluster at the bottom right, however, is more difficult to categorize. Parts of it corresponds well to food talk on French pages, but not all of it. Indeed, pages from countries like Germany, Italy, Spain or Switzerland are equally contributing to this cluster (although not shown in Figure 5). At closer inspection the chefs in this cluster also have different and not just European nationalities. There is, in other words, no national food scene here. What we are dealing with seems to be a sort of international gastrosphere where the entry ticket is a Michelin star or similar.

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

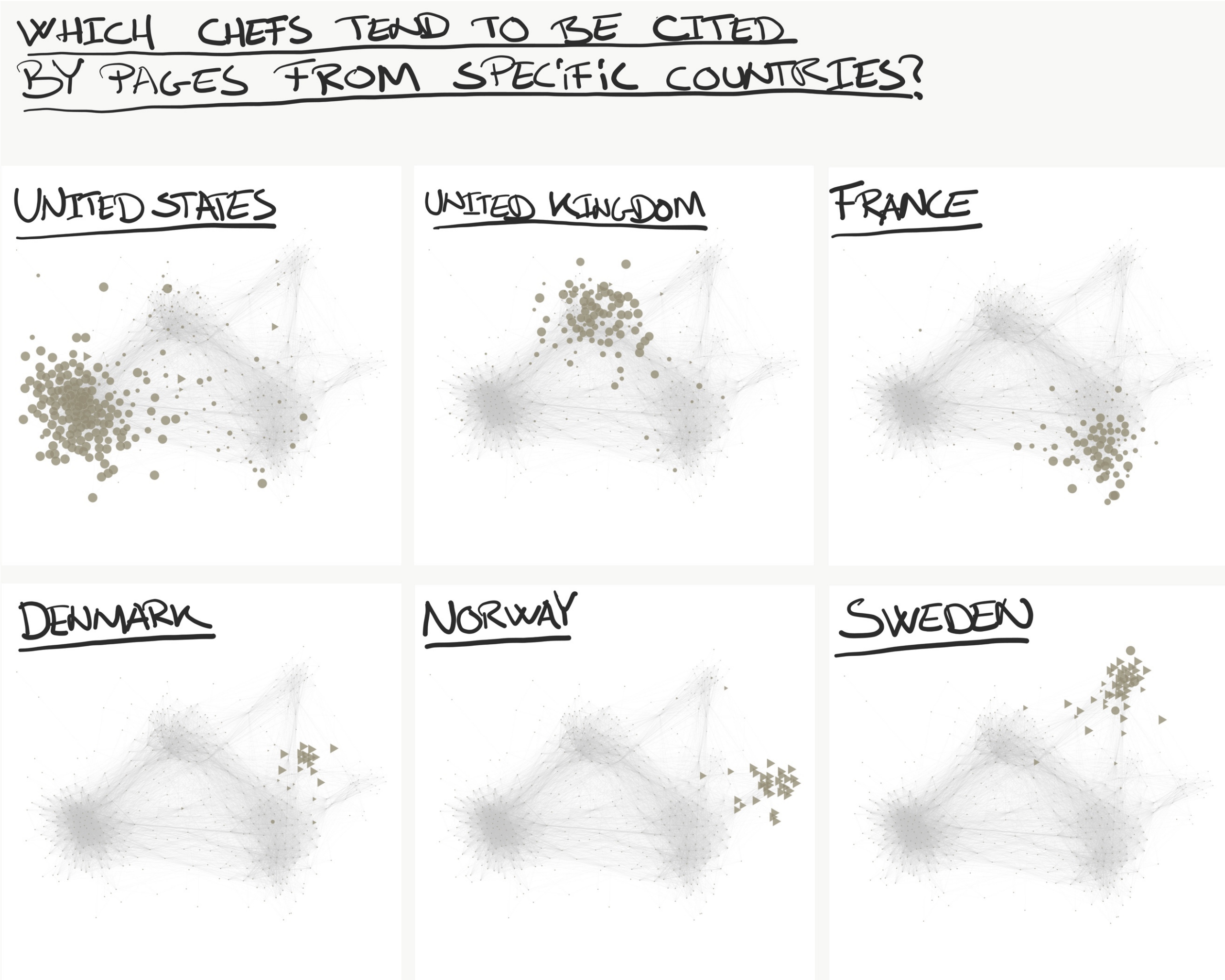

Figure 5: Interpreting the network topology: which chefs are mentioned a lot by pages from different countries.

With an interpretation of the topology in place (Figure 6) I can now explore the question about Nordic chefs that I was interested in to begin with. I already know from my degree filter (Figure 4) that some of the Nordic chefs have achieved so wide international recognition that they are mentioned across contexts and in the company of so many other chefs that they transcend any particular culinary discourse. These omnipresent Nordic chefs are Claus Meyer, René Redzepi and Rasmus Koefoed (Denmark), Magnus Nilsson, Marcus Samuelsson and Emma Bengtsson (Sweden) and Esben Holmbo Bang (Norway, although a Danish national).

Conversely, I can also assert that a good portion of the remaining Nordic chefs on my search list are not recognized outside the national culinary discourses in their respective countries. Some of them, like Frantz Knipschildt, who as a young chocolatier began his own production in New York, became widely renowned in the U.S. and never returned to a career in Denmark, are virtually unknown in their country of origin. Magnus Lindgren also made his career with Heston Blumenthal in the U.K. and was never on the gastronomic radar in Sweden. Others, like Niels Norén, Björn Frantzén, or Mads Refslund, have moved abroad to spearhead the New Nordic Cuisine movement with restaurants in places like New York or Hong Kong after claiming fame in Stockholm and Copenhagen first (Refslund was the co-founder of Noma). Something similar should be the case for

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

Kamilla Seidler, who was named the best female chef in Latin America after she opened her New Nordic restaurant in La Paz, but that seems to be a story that has mostly gained attention on the fringes between the national Danish discourse and the French/Michelin discourse.

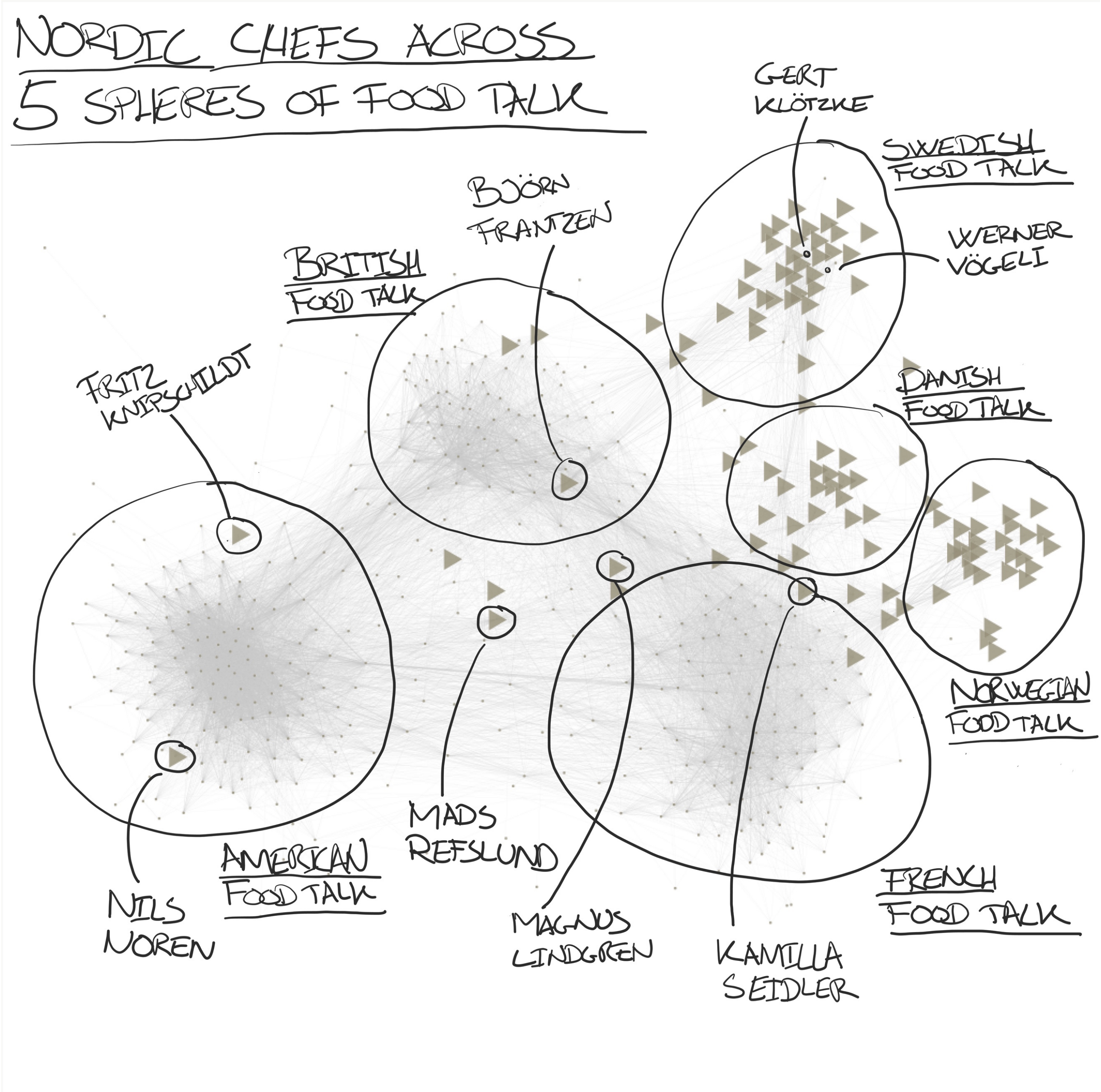

Figure 6: Interpreting the network topology: Annotated clusters and nodes, Nordic chefs enlarged as triangular nodes.

Conclusion

The mapping thus allows me to show how chefs are talked about in separate spheres online, spheres that *turn out to be* predominantly but not entirely national in character. Rather than claiming *tout court* that there is a Danish, Swedish and Norwegian culinary discourse in which online tribute is

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

paid to Danish, Swedish and Norwegian chefs, Figure 6 both urges and supports a considerably more flexible analysis. Although an overall pattern of separate national spheres is clearly visible, we are forced to reckon with each individual chef and their different ways of belonging in this relational space. The nationality of a Nordic chef does not determine this belonging, that much is clear from Figure 6, and the same is of course true for non-Nordic chefs. Werner Vögeli, who is Swiss but has had a career in Sweden, is thus clearly part of the Swedish national discourse.

The strength of visual network analysis is that it does not presume any prior categories, except for the kind of relationality we choose to include as edges between the nodes. We can read the map in Figure 6 as a way to categorize the nodes based on the way they relate to one another rather than the predicates we have given them in advance. In our case, this means that we can read the map as a way to categorize chefs based on how they are part of the same discourses (mentioned by the same food-related pages) rather than their nationality, their style of cooking, their age or any other *a priori* category. In fact, once we have discovered the emergent categories represented by the clusters in Figure 6, and once we have concluded on the basis of our research design that these categories are in fact different discourses, we can use the network to discuss the role of nationality in this discursive landscape. We can use it to become curious about individual chefs who break a pattern or about clusters that are hard to explain. This is a fundamentally explorative process that is useful for setting a direction for further qualitative work. It is not a way to confirm or refute a hypothesis with some degree of probability. We are using quantitative techniques to do qualitative ground work on volumes of messy data that would otherwise be impossible to handle.

Now you try
Download the network of co-cited chefs from
https://github.com/akmunk/ediblenorth/blob/master/ChefNetwork.gexf
and install the latest version of Gephi from (www.gephi.org). Good tutorials are available on the website. Open the .gexf file and run the ForceVector2 layout algorithm (adjust the scaling if you need to separate the nodes further). From the filters tab select a degree filter and try to exclude the highest degree nodes. Run the layout algorithm again and notice how clustering becomes clearer.

Questions for reflection

1. How could your research problem be explored by visually clustering some aspect of your dataset with a force vector layout?
2. What could be the nodes? And what kind of relationships could define the edges between them?
3. How could you justify filtering the graph to make clustering clearer?
4. How could you repurpose digital media to build a dataset or make interesting query designs?

Further readings

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

1. *Doing Digital Methods* by Richard Rogers (2019)
2. The tool repository of the Digital Methods Initiative in Amsterdam: https://wiki.digitalmethods.net/Dmi/ToolDatabase
3. The blog of Martin Grandjean (http://www.martingrandjean.ch/) for good tutorials
4. The blog of Tommaso Venturini (http://www.tommasoventurini.it/wp/) for use cases and reflections
5. The Gephi community on Facebook (https://www.facebook.com/groups/gephi/) for peer group and support

Technical vocabulary

API (Application Programming Interface): A technology that makes it possible to for a computer to ask a website for certain types of information. Different social media platforms have different APIs with their own rules and syntax.

Force vector layout: A way to find clusters in network by visually placing the connected nodes close to each other. Works by introducing a repulsive force that push all nodes apart from each other, unless the nodes are connected by edges which act as springs holding the nodes together.

Gephi: A piece of open source software for doing visual network analysis.

Network: A relational data structure consisting of nodes and edges. Nodes can be any type of entity and edges any type of relation between these entities. For instance: authors (nodes) related by the way they cite (edges) each other in scientific journals; Or: hashtags (nodes) related by the way they co-occur (edges) in tweets.

Script: A piece of code, typically written in a programming language like Python, R, or JavaScript, that makes it possible to automate instructions to a computer and make these instructions conditional on various criteria being satisfied.

Ethical considerations
Internet research is riddled with ethical concerns, not least in relation to personal and sensitive information, and as a consequence of the fast pace of the political debate on data ethics the goal posts of are constantly moving. This is felt when platforms like Facebook decide to shut down large parts of their data access in response to GDPR, or when machine learning models become capable of predicting sensitive information, such as political or sexual orientation, from a few, innocent likes on Instagram or the writing style of a tweet. Some guidelines are available through the Association of Internet Researchers (https://aoir.org/reports/ethics2.pdf).

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

References:

Blok, A., & Pedersen, M. A. (2014). Complementary social science? Quali-quantitative experiments in a Big Data world. *Big Data & Society*, *1*(2), 2053951714543908.

Burgess, J., & Matamoros-Fernández, A. (2016). Mapping sociocultural controversies across digital media platforms: One week of# gamergate on Twitter, YouTube, and Tumblr. *Communication Research and Practice*, *2*(1), 79-96.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*.

Høyrup, J. F., & Munk, A. K. (2007). Translating terroir: Sociomaterial potentials in ethnography and wine-growing. *Ethnologia scandinavica*, *37,* 5-20

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.

Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). 'The whole is always smaller than its parts'–a digital test of G abriel T ardes' monads. *The British journal of sociology*, *63*(4), 590-615.

Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, *30*(4), 256-265.

Marres, N. (2015). Why map issues? On controversy analysis as a digital method. *Science, Technology, & Human Values*, *40*(5), 655-686.

Marcus, G. E. (1995). Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual review of anthropology*, *24*(1), 95-117.

Marres, N. (2012). The redistribution of methods: on intervention in digital social research, broadly conceived. *The sociological review*, *60*, 139-165.

Marres, N., & Moats, D. (2015). Mapping controversies with social media: The case for symmetry. *Social Media+ Society*, *1*(2), 2056305115604176.

Marres, N., & Rogers, R. (2000). Depluralising the Web, repluralising public debate-the case of the GM food debate on the Web. *Preferred placement: Knowledge politics on the Web*, 113-136.

Manovich, L. (2011). Trending: The promises and the challenges of big social data. *Debates in the digital humanities*, *2*, 460-475.

Munk, A. K. (2019). Four styles of quali-quantitative analysis: Making sense of the new Nordic food movement on the web. *Nordicom Review*, *40*(s1), 159-176.

Munk, A. K., & Ellern, A. B. (2015). Mapping the New Nordic Issuescape: How to Navigate a Diffuse Controversy with Digital Methods. *Tourism encounters and controversies: Ontological politics of tourism development*.

Munk, A. K., & Jensen, T. E. (2015). Revisiting the histories of mapping. *Ethnologia Europaea*, *44*(2), 31.

THIS IS A DRAFT
Final version published as Munk, A. K. (2021). Tracing Culinary Discourse on Facebook: A Digital Methods Approach. In S.G.S. Krogager & J. Leer eds. *Research Methods in Digital Food Studies*, London: Routledge.

Munk, A. K., Abildgaard, M. S., Birkbak, A., & Petersen, M. K. (2016). (Re-) Appropriating Instagram for Social Research: Three methods for studying obesogenic environments. In *Proceedings of the 7th 2016 International Conference on Social Media & Society* (pp. 1-10).

Rogers, R. (2009a). Post-democraphic machines. In A. Dekker, & A. Wolfsberger (Eds.), Walled garden (pp. 29- 39). Virtueel Platform.

Rogers, R. (2009b). *The end of the virtual: Digital methods* (Vol. 339). Amsterdam University Press.

Rogers, R. (2018). Otherwise engaged: Social media from vanity metrics to critical analytics. *International Journal of Communication*, *12*(732942), 450-72.

Rogers, R. (2019). *Doing Digital Methods*. SAGE Publications Limited.

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

Venturini, T., Baya Laffite, N., Cointet, J. P., Gray, I., Zabban, V., & De Pryck, K. (2014). Three maps and three misunderstandings: A digital mapping of climate diplomacy. *Big Data & Society*, *1*(2), 2053951714543804.