



## Visual Network Exploration for Data Journalists

Tommaso Venturini, Mathieu Jacomy, Liliana Bounegru, Jonathan Gray

► **To cite this version:**

Tommaso Venturini, Mathieu Jacomy, Liliana Bounegru, Jonathan Gray. Visual Network Exploration for Data Journalists. Scott A. Eldridge II; Bob Franklin. The Routledge Handbook of Developments in Digital Journalism Studies, Routledge, 2018, 9781138283053. hal-01672282

**HAL Id: hal-01672282**

**<https://hal.archives-ouvertes.fr/hal-01672282>**

Submitted on 23 Dec 2017

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

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

# VISUAL NETWORK EXPLORATION FOR DATA JOURNALISTS

TOMMASO VENTURINI, MATHIEU JACOMY, LILIANA BOUNEGRU, JONATHAN GRAY

*Networks are classic but under-acknowledged figures of journalistic storytelling. Who is connected to whom and by which means? Which organizations receive support from which others? What resources or information circulate through which channels and which intermediaries enable and regulate their flows? These are all customary stories and lines of inquiry in journalism and they all have to do with networks. Additionally, the recent spread of digital media has increasingly confronted journalists with information coming not only in the traditional form of statistic tables, but also of relational databases. Yet, journalists have so far made little use of the analytical resources offered by networks. To address this problem in this chapter we examine how “visual network exploration” may be brought to bear in the context of data journalism in order to explore, narrate and make sense of large and complex relational datasets. We borrow the more familiar vocabulary of geographical maps to show how key graphical variables such as position, size and hue can be used to interpret and characterise graph structures and properties. We illustrate this technique by taking as a starting point a recent example from journalism, namely a catalogue of French information sources compiled by Le Monde’s The Decodex. We establish that good visual exploration of networks is an iterative process where practices to demarcate categories and territories are entangled and mutually constitutive. To enrich investigation we suggest ways in which the insights of the visual exploration of networks can be supplemented with simple calculations and statistics of distributions of nodes and links across the network. We conclude with reflection on the knowledge-making capacities of this technique and how these compare to the insights and instruments that journalists have used in the Decodex project – suggesting that visual network exploration is a fertile area for further exploration and collaborations between data journalists and digital researchers.*

## INTRODUCTION

Few people know as well as journalists that the world is made of relations. Following alliances, unveiling links, unravelling threads is, and has long been, a central part of their investigations. If social scientists can speculate about longstanding structures and global arrangements, journalists have no such leisure. Their work consists in tracing the specific associations that connect individuals and institutions to uncover how lumps of money, influence and knowledge are exchanged through them and where unethical behaviour, corruption, fraud or unfair political influence may occur. The advent of digital technologies has made such work both easier and more difficult. Easier, because it has increased the traceability of economic and political associations. More difficult, because it has submerged journalists with more information than their investigative toolkit is used to handling.

When, for example, the reporters of the International Consortium for Investigative Journalism (ICIJ) received the 2,6 terabytes and 11,5 million documents composing the so-called 'Panama Papers', they obviously could not process them manually (Baruch & Vaudano, 2016). Note that this is not just a 'big data' problem. The trouble with the leak was not only its size, but the fact that its interest came from the links it established between specific individuals and particular tax-havens. Extracting “key” figures through statistical aggregation or abstracted computational models would miss the point of many of the stories that journalists were most keen to explore. The inquiry could not simplify the dataset, but had to explore each and every one of the connections it exposed. This was done, among others ways, through a tool called *Linkurious* (<http://linkurio.us>), whose interest comes less from its computational power than from the way in which it allows its users to see and follow the connections of a network.

The Panama Paper case is interesting, but also interestingly isolated. Despite longstanding interest, the use of networks in journalism remains comparatively marginal (cf. Bounegru *et al.*, 2016 for an overview of the emerging uses of networks in journalism). The reasons are not difficult to imagine. Graph mathematics is more demanding and less widely known than traditional statistical approaches and does not come with the same readily accessible and publicly recognised vocabulary of visual motifs. With all its

computational power, graph mathematics does not fit journalistic needs because it tends to be obscure for both reporters and their readers.

In this chapter, we address this difficulty by suggesting a technique for the visual exploration of networks. As we will try to show, when performed correctly, the visual representation of network translates some of the most important graph structures into graphical variables (thereby supporting investigative work) and allowing the interpretation of networks with conventions similar to those developed for geographical maps (thereby remaining legible for a large audience). After having introduced the mathematical and historical bases of our approach, we will present our technique for the visual exploration of networks. Using as an example, a the network of the French information sphere, we will illustrate the recursive work of interpretation and categorisation that allow to read the network as an organised territory. Visual network exploration, which is growing in prominence amongst digital methods researchers for social and cultural research, may be useful not only for studying media landscapes, but also for digital journalism practitioners who are interested in exploring and telling stories with networks and relational data.

## UNDERSTANDING FORCE-DIRECTED LAYOUTS

Far from being merely aesthetic, the graphical representation of networks has an intrinsic hermeneutic value, which you will have experienced if you have ever used a public transportation map. Such maps are distinctively different from road maps or city maps. It is not only that transportation maps are simpler (the level of details depending only on the resolution of the map), it is that they represent a network and not a geographical territory. An illustration of this difference can be found in the famous map of the London tube as designed by Harry Beck in 1933. Before Beck's redesign, the diagram was a classic geographical map locating stations according to their coordinates. After the redesign, it became a network of correspondences in which stations are positioned according to their relative proximity and connectivity. The gain in legibility is evident as the function of the transportation map is not to situate stations in urban space, but relative to each other, so as to help users to move from one to another (a type of orientation that resembles strikingly to one used by of traditional sea navigators, see, for example, Turnbull, 2000, pp.133-165).

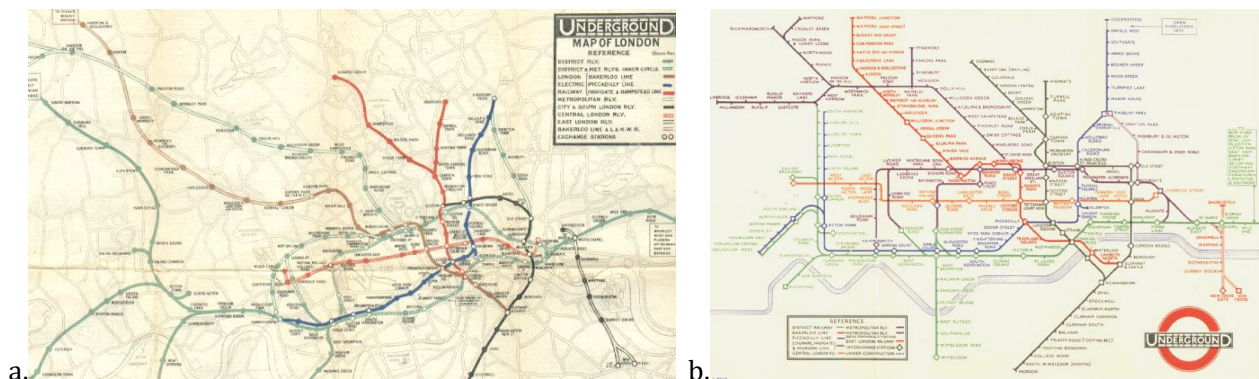


Figure 1. London tube map (a) in 1920 before Beck redesign and (b) in 1933 after Beck redesign.

Another example of such mapping approach comes from early works in Social Network Analysis (Freeman, 2000). Jacob Moreno, founder of SNA, is explicit about the importance of visualization: 'A process of charting has been devised by the sociometrists, the sociogram, which is more than merely a method of presentation. It is first of all a method of exploration' (1953, pp. 95-96). In an interview released by Moreno to the New York Times in 1933, network analysis is presented as a 'new geography'. More important than the title, however, is the figure that accompanies that interview, depicting friendships among fourth grade pupils. The sociogram presented by these figures powerfully reveals how friendship is not equally distributed in the class. One only need to know that triangles represent boys and circles girls to see how inter-gender relationships are discouraged at that specific age (or at least the declaration of such friendships). The trick, of course, only works because the nodes are not positioned

randomly in the space, but in a way that minimizes line-crossing (in Moreno's own words 'the fewer the number of lines crossing, the better the sociogram', 1953, p.141). It is because triangles are pushed on one side and circles on another that it is easy to spot the existence of a single inter-gender connection.

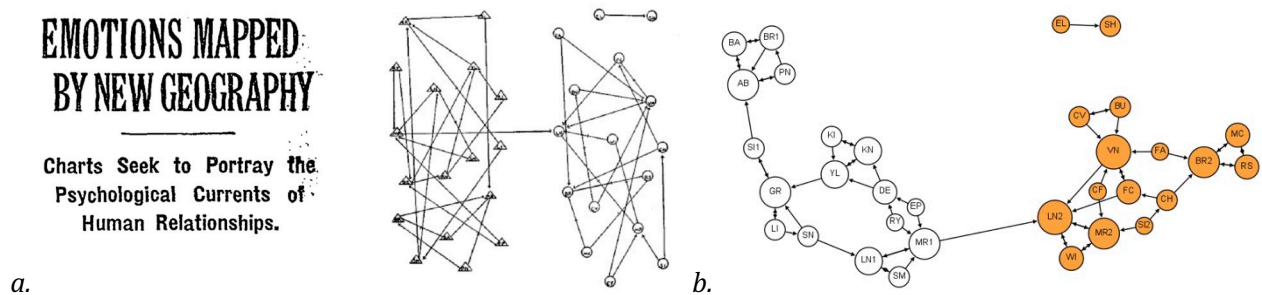


Figure 2. Sociogram representing friendship among school pupils (original title and image accompanying Moreno's 1933 New York Time interview) (a) in the original version and (b) in the modern force-directed spatialisation.

Moreno's rule of spatialisation is easy to follow on a graph of a few dozen nodes and edges but impracticable on larger networks. Graphs with thousands of nodes and edges are so intricate that the direct counting of line-crossings becomes prohibitively time-consuming. An indirect approach consists of drawing closer the connected nodes to minimize the length of the edges and therefore the possibility of crossings. But even in this case, since each node may be connected to several other nodes which are themselves connected many other nodes, minimizing the length of the edges is far from a trivial exercise.

Thus we might explore the network using a technique called 'force-directed spatialisation'. Such spatialisation follows a physical analogy: nodes are charged with a repulsive force that drives them apart, while edges act as springs binding the nodes that they connect. Once the algorithm is launched it changes the disposition of nodes until it reaches a balance such of forces (Jacomy *et al.*, 2014). Such equilibrium reduces line-crossings and improves the legibility of the graph. Fruchterman and Reingold (1991), who proposed the first efficient force-directed algorithm, cite line-crossing as the second of their aesthetic criteria.

Yet, scholars working with networks soon realised that avoiding line-crossing is not the most interesting effect of force-directed layouts. At equilibrium, the visual density of nodes and edges becomes an approximate but reliable proxy of the mathematical structure of the graph (for a detailed mathematical proof, see Venturini *et al.*, forthcoming). Groups of nodes gathering in the layout tend to correspond to the clusters identified by community-detection techniques (Noack, 2009); structural holes (Burt, 1995) tend to look like sparser zones; central nodes move towards middle positions; and bridges are positioned someway between different regions (Jensen *et al.*, 2015).

The trick of force-directed algorithms is all the more remarkable, given that the space of networks is relative rather than absolute (it can be rotated or mirrored without distortion of information) and that it is a consequence and not a condition of element positioning. In traditional geographical representation, the space is defined *a priori* by the way the horizontal and vertical axes are constructed. Points are projected on such pre-existing space according to a set of rules that assign a univocal position to a pair of coordinates. The same is true for any Cartesian diagram (scatterplots for instance), but not for networks, in which the space is defined by the position of the nodes and not the other way around.

Despite such differences (which should not be forgotten), force-directed algorithms allow reading networks as geographical maps, translating complicated mathematical concepts into more conventional vocabulary of regions and margins, path and landmarks, centres and peripheries (Lynch, 1960). This is a crucial advantage that explains why force-directed algorithms have become the *de-facto standard* of network visualisation: they facilitate the exploration of networks and relations by means of more familiar and intuitive spatial metaphors, as well as through less familiar computational and statistical metrics.

## THE DÉCODEX: A CONTROVERSIAL CASE STUDY

In the following pages, we will illustrate the technique of visual network exploration drawing on a concrete example. Our case study is a network of websites extracted from a listing compiled by the French journal *Le Monde*. Since 2009, a group of journalists gathered under the name of *Les Décodeurs* ([www.lemonde.fr/les-decodeurs/article/2014/02/12/l-equipe-des-decodeurs\\_4365082\\_4355770.html](http://www.lemonde.fr/les-decodeurs/article/2014/02/12/l-equipe-des-decodeurs_4365082_4355770.html)) has verified the accuracy of thousands of stories circulating in the French blogosphere and in social media. In January 2017 (at the beginning the French presidential campaign), *Les Décodeurs* have launched an online tool called the *Décodex* ([www.lemonde.fr/verification](http://www.lemonde.fr/verification)), allowing readers to search for the most important sources of online information relevant to French public debates (though not necessary in French). Each source is accompanied by a short description and, more crucially, by an evaluation of its trustworthiness according to the journalists of *Le Monde*.



Figure 3. User interface of the *Décodex* tool by *Le Monde*

Not surprisingly, the classification provided by *Les Décodeurs* has stirred much debate in the French media spheres. Several of the sources categorized as imprecise or unreliable, along with other newspapers and blogs, have contested the *Décodex*, with critique spanning from challenging the way in which websites are over-simplistically classified; to questioning the right of *Le Monde* (which is itself a rival source of information) to note the reliability of other websites; to disputing the legitimacy and interest of such classification in general (arguing that some of the websites in the list means to circulate opinions rather than information). *Les Décodeurs* themselves admitted the difficulty of their exercise, the many ambiguities that they were obliged to decide on and the errors and inaccuracies that may have derived from them. At the same time, they defended their work by pointing at the increasing quantity of false or partisan information circulating online and by affirming their openness to discussing their classification and revising it if necessary.

The controversy around the *Décodex* is a good example of difficulties connected to the detection of fake news online (Bounegru *et al.*, 2017), but also of the more general debates surrounding all kind of classifications. Categorizing things is never a self-evident or innocent practice (Bowker & Star, 1999) and should always be carried out with the greatest caution. This is true for the initial classification of the *Décodex*, but it is also true for the network extracted from it. As we will see in the following pages, the visual exploration of network involves a constant toing and froing of categorization and observation, typology and topology.

To build our example network, we have extracted, in collaboration with *Les Décodeurs*, all the websites



contained in the *Décodex* and investigated the way in which they cite each other. To do so, we employed Hyphe (<http://hyphe.medialab.sciences-po.fr>) a web crawler developed by the médialab of Sciences Po, which facilitates the exploration of websites and following the hyperlinks present in their pages. All the websites comprising the *Décodex* corpus have been crawled at a depth of one click starting from the homepage. We so obtained a network with 653 nodes and 5943 edges. Whilst *Les Décodeurs* focus on editorial judgements about how to *classify* websites in the French media landscape, our network exploration examines the relations between them and other websites by means of their *linking practices*. While some researchers focus on how networks are held together through financial ties, organisational affiliations, business relationships and family and social relations – we consider their relations according to the hyperlink, in accordance with a longer tradition of digital methods, digital sociology and new media studies research (see, e.g. Marres & Rogers, 2005; Rogers, 2013)

The treatment of social platforms (such as Facebook, Twitter, YouTube...) in our crawl requires some additional explanation. These platforms are both sources of information as a whole and containers of multiple individual sources in the form of pages or accounts. Since extracting all the hyperlinks from a site as large as Facebook would have been impossible, we only crawled the accounts that were specifically mentioned in the *Décodex*. We have, however, kept a record of all the links pointing toward the main social media platform to investigate how they are cited by the other websites of our corpus.

## A VISUAL EXPLORATION OF THE DÉCODEX NETWORK

The visual exploration of networks exploits three visual variables to graphically represent their features: position, size and hue (for a definition of these variables and their semiotic affordances, see Bertin, 1967). For the reasons discussed above, position is crucial in translating the mathematical characteristics of the graphs. Force-directed layouts create regions where numerous nodes are densely assembled and regions that are less crowded. These differences of density, determined by the uneven distribution of links, reveal the uneven association between the entities of the network. Everything may be connected in this world, but not everything is *equally* connected.

Discerning the spatial structure of networks, however, is not always straightforward. In the easiest cases, the difference in the density of association is such that clusters appear as well defined knots of nodes and edges separated by empty (or almost empty) zones. These zones are called 'structural holes' (Burt, 1995) and, when they exist, they provide a crucial guidance for the interpretation of the network. Thanks to the ruptures created by structural holes, the boundaries of clusters can be easily detected, like cliffs separating a plateau from a valley. Most of natural and social networks, however, do not exhibit such a clear separation and the borders of their cluster tend to be gradual as the hillside slopes. The fuzziness of clusters' frontiers is not necessarily an obstacle to their recognition (one can point at a hill even when it is impossible to say exactly where it starts and ends), but it certainly make their identification more difficult. This is why visual network analysis is often more like an exploratory expedition - where meanings and findings are progressively and hermeneutically generated - than to the statistical confirmation of a set of pre-existing hypotheses (on the difference between exploratory and confirmatory analysis see Tuckey, 1997 and Behrens and Chong-Ho, 2003).

This is certainly the case for our *Décodex* network, which, at a first look, does not present any manifest structural hole or any clear spatial structure. To visualise our network we used two main tools: Gephi (<https://gephi.org>) for filtering and spatializing the network (using in particular the force-driven algorithm ForceAtlas2) and Graph Recipes (<http://tools.medialab.sciences-po.fr/graph-recipes>) to tweak the visual rendering of the network. Though no structural holes are evident in the *Décodex* network, looking closely at the layout makes it is possible to notice that the network does not spatialize as a perfect circle, but rather in an avocado-like shape with a smaller top and and a larger bottom. These irregularities (as weak and subtle as they can be) often suggest the presence of polarising effects which can be interesting to investigate further.

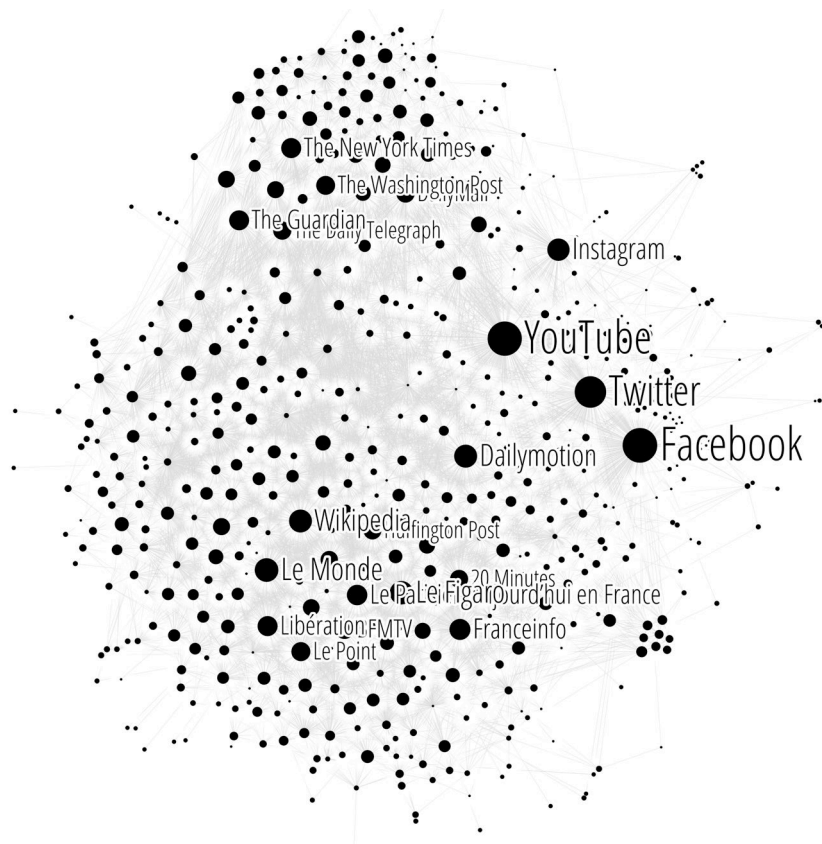


Figure 4. The Décodex network spatialized by ForceAtlas2. The size of nodes is proportional to in-degree.

The first and most crucial way to explore our network is to look at the identity of the nodes that occupy its different regions. This may seem trivial, but it is not. It is a distinct advantage of visual exploration compared to other form of statistical analysis, that it does not aggregate the individual entities that compose its corpus: each and every node is visible in the layout and can be interrogated by the researcher. Even on a small network as the one in our example, however, the quantity of nodes can make it difficult (and time consuming) to look at all of them.

This is where the second variable of our visual exploration, *size*, comes in handy. Since, in networks, nodes are defined first and foremost by their connections, we have ranked the nodes according to the number of edges pointing to them. In the jargon of network analysis this number is called 'in-degree' and nodes with an elevated in-degree are called 'authorities', because they are recognised and referred to by many others. In the previous figure and in all following, we have sized the nodes according to their in-degree so that a greater authority literally translates into increased visual prominence.

Reading the names of websites that occupy the two poles of our avocado, it seems natural to suppose that their separation derives from a linguistic fracture. The websites in the lower part are predominantly French, while those in the upper part are more international. A way to highlight this is to show the uneven distribution of TLD (Top Level Domain) in the network.





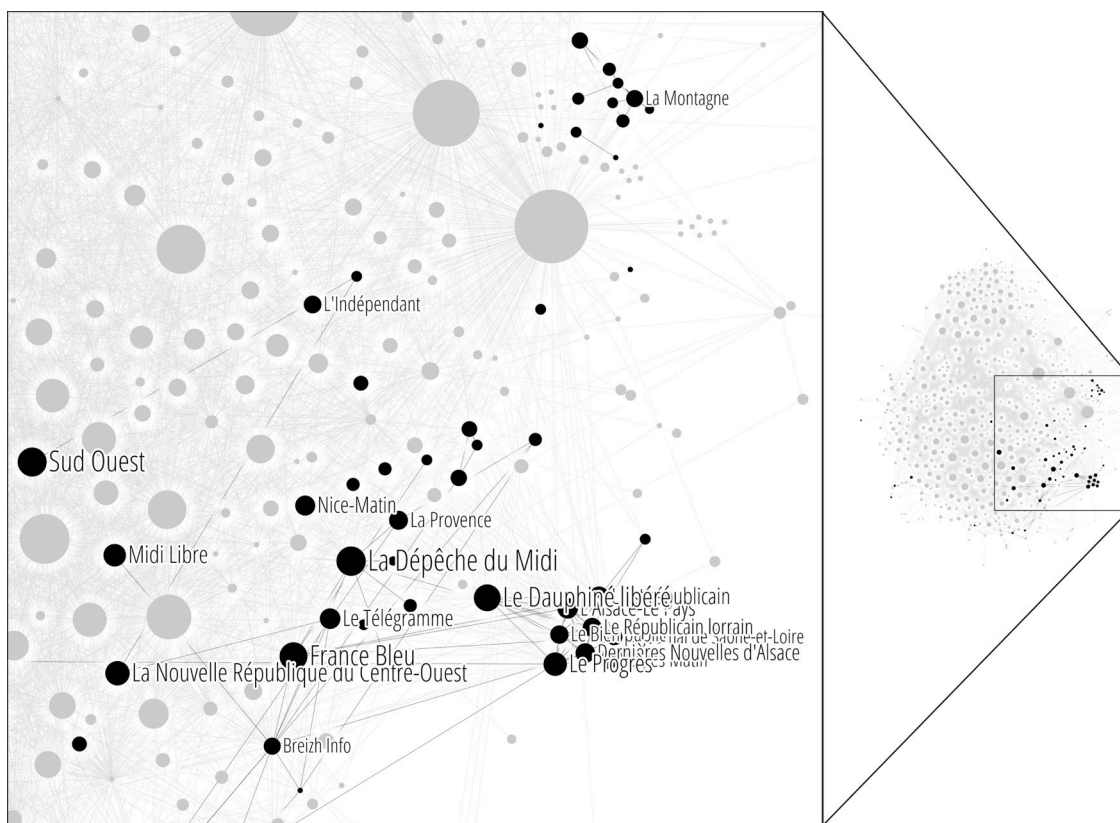
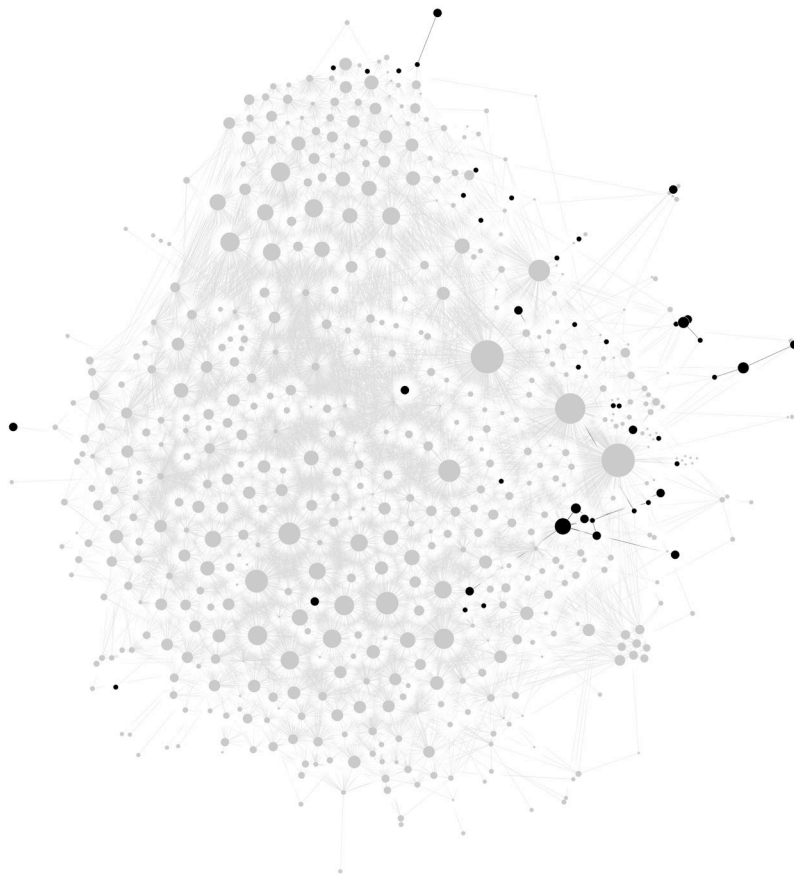


Fig. 6. Zoom on the French regional press

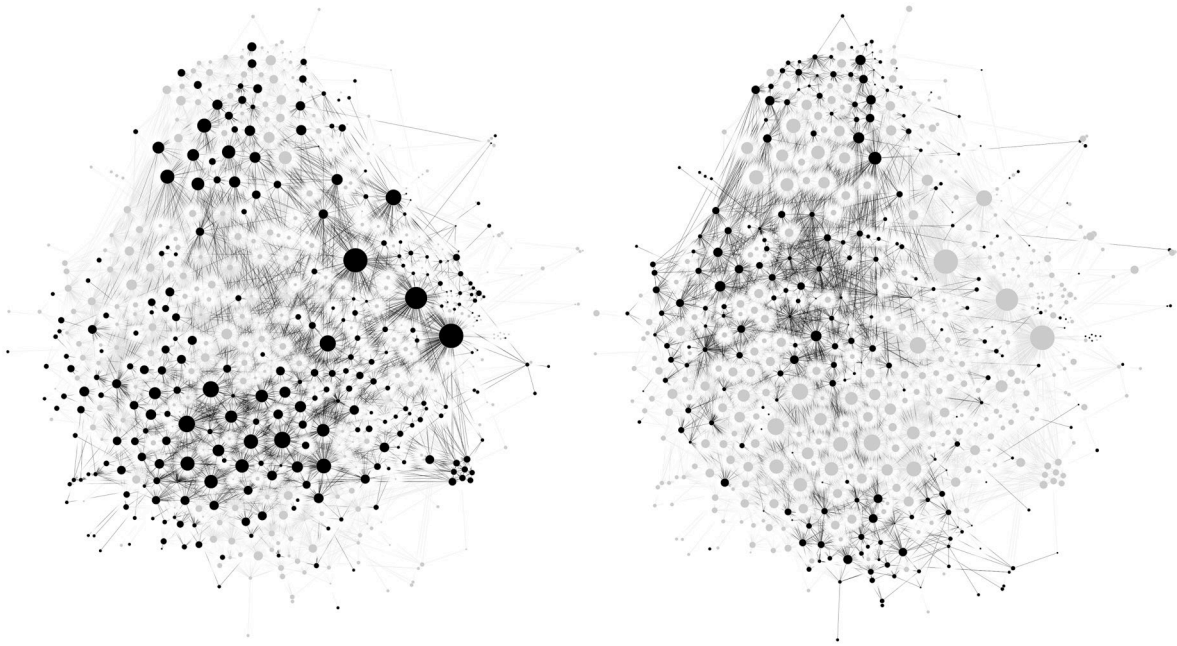
The distinctive position of the platforms and the national/regional press are both interesting and nontrivial findings, but we can push our analysis further. The way to do so is by playing with the third visual variable exploited by visual exploration of network: the hue of the node. This is a laborious but revealing part of our visual exploration. It consists in categorizing the nodes of the network according to multiple classifications and visualizing these classes on the network as different colors or (as in this paper) as different shades of grey. It is important to notice that the operation of classifying the nodes and of reading the disposition of classes are not separated, but performed at the same time. As it will become clear in the next pages, our technique does not consist simply in the projection of a set of pre-existing categories on a connectivity-based layout, but on recursively using the categories to make sense of the layout and the layout to define the categories. It is important to remember that the color is a 'non-mixable' visual variable. A node can be red or blue, for example, but not the two at the same time. When categorizing nodes, it is therefore necessary to employ exclusive categories. A website, for example, can be classed in the category 'news' or 'satire', but not in both. In the (not uncommon) case of nodes resisting a unique classification, researcher can introduce a residual category such as 'multiple' or 'misc'.

As a first step in our combined exploration of topology and typology, we will color the nodes of the network according to the original categories of the *Décodex*. These categories refer to the trustworthiness of the sources as manually assessed by the journalists of *Le Monde* in the four categories are 'reliable', 'imprecise', 'unreliable' and 'satirical'. Precisely because these categories have been defined before and independently from the extraction of the network, their disposition does not follow the spatial articulation of the network. Rather, it is possible to find nodes of every categories in almost of regions of the network. A remarkable exception are the satirical websites that are to be found on the right side of the layout both in its upper and lower part. Arguably, this position is not due to the hyperlinks between the satirical websites (which do not cite each other very much), but by their strong connection with social media platforms to which all these sites extensively link.



*Fig. 7. The 'satirical' websites according to the original Decodex classification (node have been emphasized by the black color and by doubling their radius despite their low degree)*

The other classes are distributed more evenly but not randomly. The 'reliable' websites tend to occupy the center of both in the international and French pole, while the 'imprecise' and 'unreliable' take a more marginal position. More interestingly, looking at the lower part of the network, we observe two groups of 'imprecise' and 'unreliable' sources - while a majority of these nodes are positioned above the core of national and reliable websites (and hence in-between the French and the international website), a significant minority is located below them.



*Fig. 8. Highlight of the 'reliable' websites (left) and 'unreliable' and 'imprecise' websites (right)*

To account for this separation, we introduce an additional categorisation based on the political leaning of the websites. In particular, we distinguish the websites that disseminate unreliable or imprecise information because they pursue a right-wing or extreme-right agenda (which occupy the center of the network) and the websites exhibiting a more general conspiritorial attitude (which occupy the bottom of the network).



*Fig. 9. Highlight of the 'conspiritorial' websites (left) and 'right' and 'extreme right' websites (right)*

Through our iterative exploration of typology and topology we have eventually revealed a partitioning of

the network that, while invisible at first glance, allows to interpret some of the main contours of the French media landscape. Though these territories are not separated by clear structural holes, the nodes that they contain are fairly consistent. Interestingly, our final classification produces a homogeneous partition of the layout *not in spite, but because of* its heterogeneity, which mixes linguistic categories, trustworthiness classes and political leanings. The fact that a non-homogenous categorization turns up to offer the best characterization of the structure of our network should not come as a surprise. Networks are complex objects which articulate diverse elements through disparate logics. In this, they remind us of a passage by Jorge Luis Borges cited by Foucault as a perfect example of a heterogenous classification that, while defying our traditional categories, is nonetheless highly efficient to describe the culture in which it has been elaborated:

“[Borges] quotes a ‘certain Chinese encyclopaedia’ in which it is written that ‘animals are divided into: (a) belonging to the Emperor, (b) embalmed, (c) tame, (d) suckling pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) *et cetera*, (m) having just broken the water pitcher, (n) that from a long way off look like flies’. In the wonderment of this taxonomy, the thing we apprehend in one great leap, the thing that, by means of the fable, is demonstrated as the exotic charm of another system of thought, is the limitation of our own”. (Foucault, 1970 p. XV).

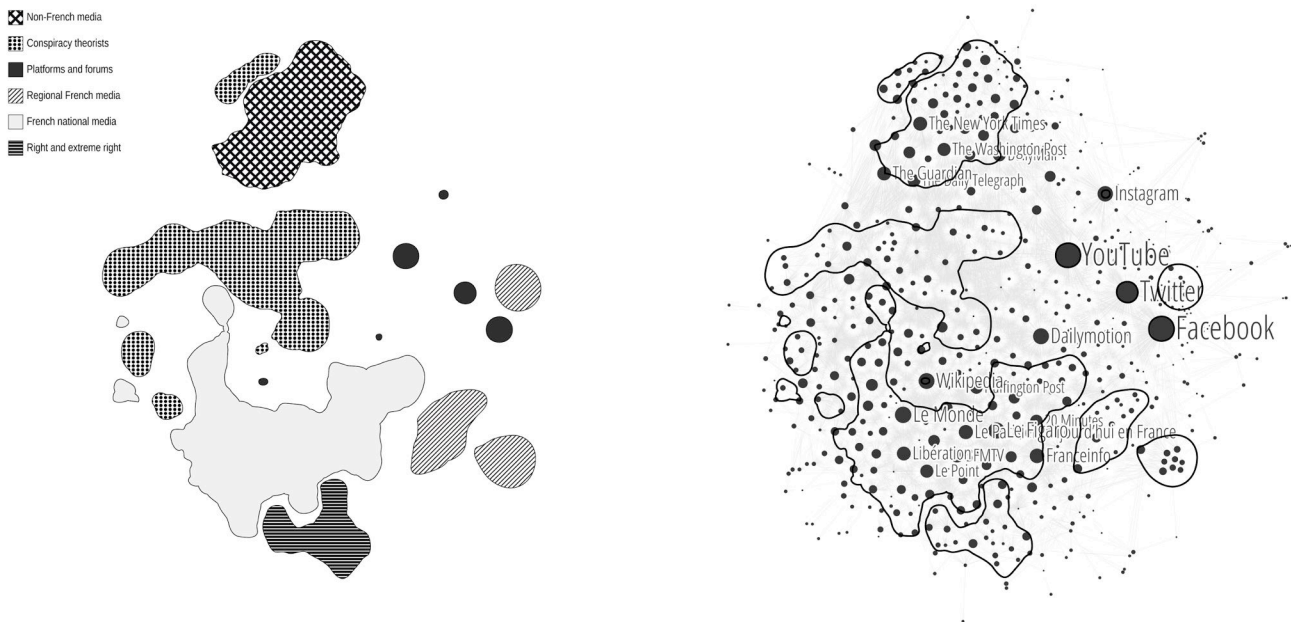
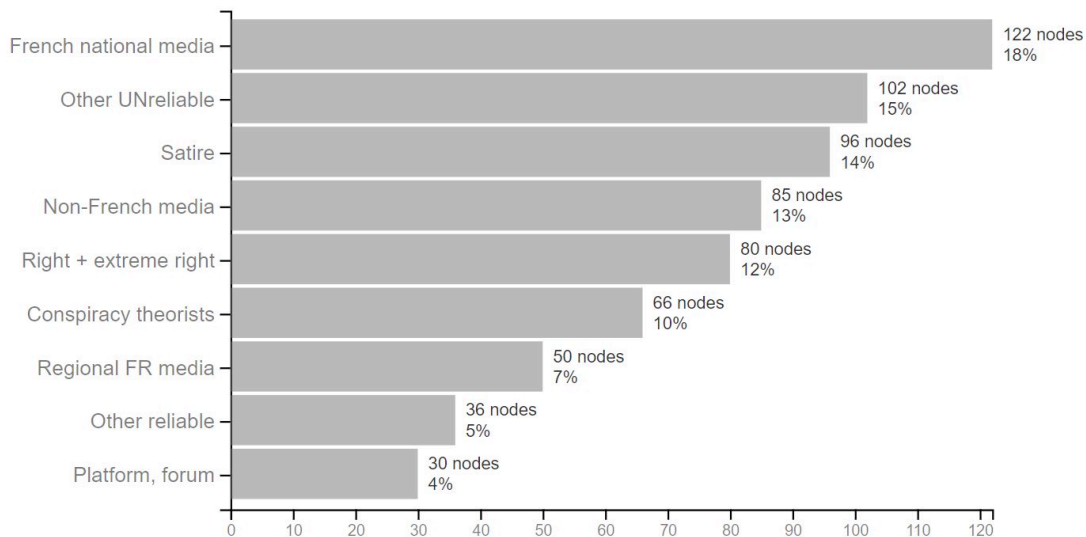


Fig. 10. The heterogeneous territories of the Décodex network.

## LINKING PATTERNS IN THE DÉCODEX NETWORK

Now that, by means of visual exploration, we have defined a heterogeneous but hermeneutically robust partitioning of our network, we can use it as a basis for a statistical analysis. While praising the advantages of the visual interpretation, we are also aware that not all structural properties can be rendered visually. The direction of edges or the connection between different classes, in particular, are not easily read in network images. These questions, however, can be investigated by other means once the partitioning of the network has been defined.



*Fig. 11. Distribution of the number of nodes per category*

Fig. 11 shows the distribution of nodes in the regions identified in our final classification (see figure 10), to which we have added the ‘satirical’ websites (which we discussed above but not included in figure 10 for the sake of legibility) as well as “other reliable” and ‘other *unreliable*’. These two residual categories comprise together about one fifth of the nodes of the network. This relatively high figure is not uncommon. Given the heterogeneity of the networks they work with, social scientists and journalists should aim at classifications that are robust and insightful (capable of delineating homogenous zones in the graph) rather than comprehensive.





*Fig. 12. Connectivity between the categories of our final classification.*

*Rows convey how many time the nodes of a given category cite the nodes of other categories.*

*Columns convey how many time the node of a given category are cited by the nodes of other categories.*

Our empirical categories are powerful tools to unveil different linking strategies in the network. Figure 12 above presents the links in the corpus aggregated by categories. As we can see, not all categories cite or are cited the same way. ‘French national media’ and ‘platforms’ are much cited and by various actors (their columns contain larger circles), while ‘satirical’ websites are scarcely cited (their column is almost empty). Platforms do not cite much, but this is merely a consequence of our method since (as explained above) most of them had not been not crawled. ‘Right-wing’, conspiracy theorist and other ‘unreliable’ websites are on the contrary the origins of the highest number of citations and, very interestingly, they seem to favour “reliable” sources over “unreliable” ones. As expected, the reliable websites do not link back to them, and this asymmetry reveals an important hierarchy. To investigate this linking pattern, we will compare the incoming and outgoing links of some of the most interesting categories.

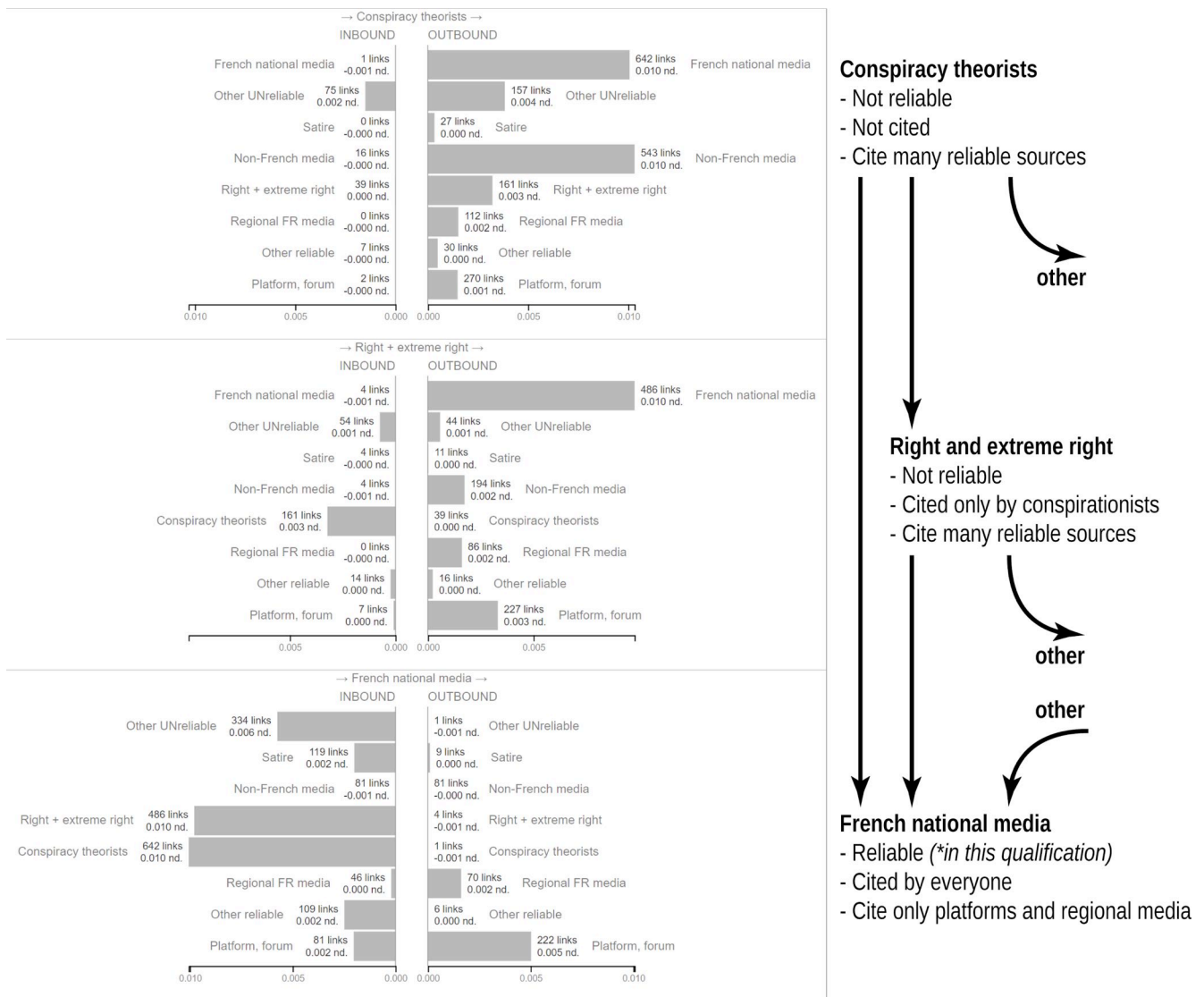


fig. 13. Hierarchical structure in the corpus, based on our final categories. Black arrows on the right side summarize the links structure between these categories.

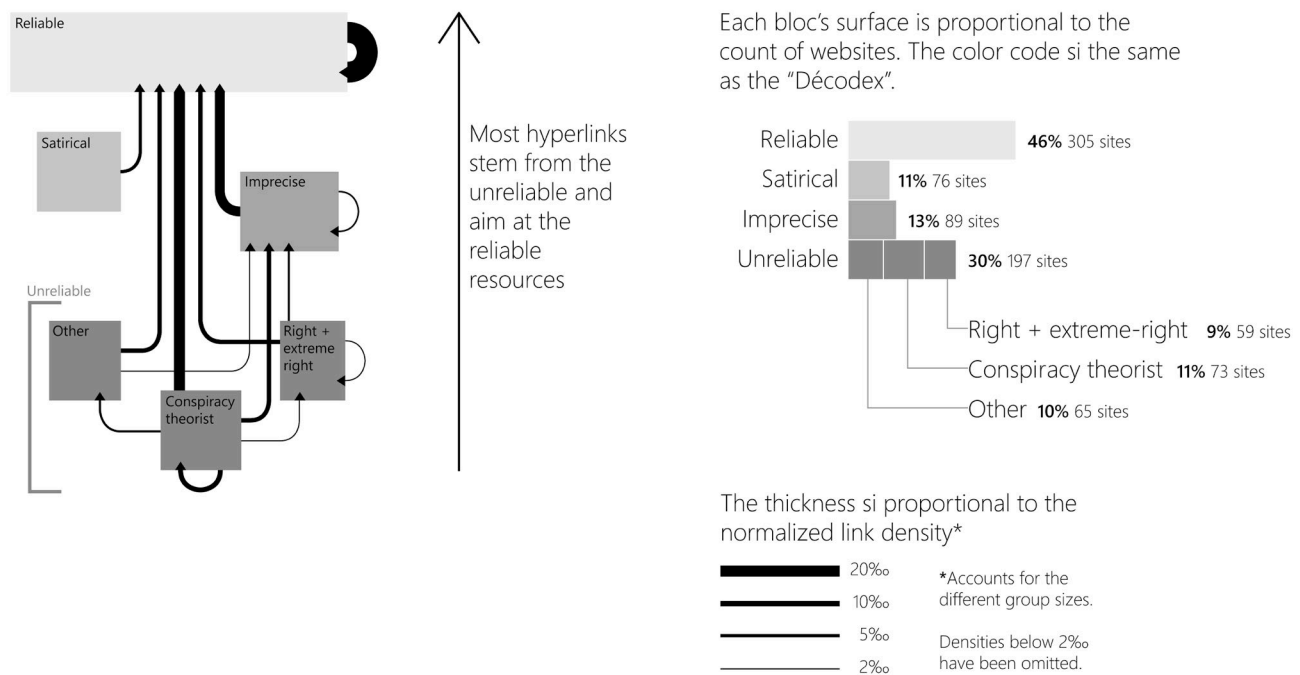


fig. 14. Simplified version of the statistical analysis presented in figure 13.

This kind of hierarchical structure is common on the web and has been explained as a consequence of preferential attachment (Barabási & Albert, 1999): actors tend to link to other websites that they perceive as higher in the hierarchy and avoid linking to those that they perceive as lower. This style of preferential attachment whereby smaller actors link to establishment actors without reciprocation of the linking act has elsewhere been called "aspirational linking" (Rogers, 2013). Links in a network do not always produce a hierarchy of categories but this behaviour does. This linking pattern and the way it fits our empirical categories, may suggest an alternative way to characterise the trustworthiness being investigated by *Le Décodeurs*: reliable sources are cited by all types of websites, while unreliable sources are only cited by few other types (if any).

This observation is in many ways at odds with what is often affirmed about "post-truth era" in which we have supposedly landed. While fake news is said to leverage the horizontality of digital media to blur the boundaries between true and false, the linking patterns of the (French) information spheres suggest a different picture. Despite their different ideological leanings, all websites agree on the overall hierarchy of reliability by citing in one sense and not in the other. The 'right-wing' websites, for example, try to blur the lines by citing both their peers and more reliable sources, but they also try to draw a line between them and the even less reliable 'conspiracy theorist' websites. Whatever its position in the pyramid of hyperlinking, every actor tries to improve its situation by linking upwards to authorities above, and not linking to less reputable websites below, thus reinforcing the hierarchy.

## CONCLUSION

This chapter discussed the visual exploration of networks with the aim of improving the understanding of one of the dominant visual-analytical forms of our digital age – the network diagram – and its potential role in relation to the study and practice of digital journalism. Drawing on graph semiotics and traditional

cartography, this chapter proposed a model whereby the interpretation of network topology with its regions, paths, cores and peripheries, is guided by three visual variables: position, size and hue. The process that we described is one that emphasizes the exploratory and iterative character of the investigation. While counter-intuitive at first, we emphasised that in order to surface the multiple logics that play out in the structure of a network graph, analysis should not limit itself to one classificatory principle. Multiple heterogeneous criteria of classification are often necessary to characterize the topology of a network map. Finally, we advocated for mixing methods, complementing visual network exploration with statistical analyses in order to further characterise network properties.

Through the case study of French media hyperlink map, we tried to show how the visual exploration of networks reveals new angles which other analyses may leave unexplored. In this case the chapter illustrated an alternative way to assess websites' reliability that complements the traditional fact-checking approach of qualifying content with an examination of the linking patterns between different regions of the network as reputational markers (Rogers, 2013). In this analysis thus we have combined the manual classification of reliability undertaken by *Le Monde's* journalists with the standing of a source according to the hyperlinks that it receives and gives. This approach enabled us to bring fresh findings to current debates around fake news. In spite of the proliferation of fabricated content of various shades, reputation hierarchies on the web seem to be maintained (at least to some extent), as fake and hyper-partisan sites deploy aspirational hyperlinking styles which favour, perhaps surprisingly, authoritative sources.

## REFERENCES

- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509. Retrieved from <http://www.sciencemag.org/cgi/content/abstract/sci;286/5439/509>
- Baruch, J., & Vaudano, M. (2016, April 8). « Panama papers » : un défi technique pour le journalisme de données. *Le Monde*. Paris. Retrieved from <http://data.blog.lemonde.fr/2016/04/08/panama-papers-un-defi-technique-pour-le-journalisme-de-donnees>
- Behrens, J. T., & Chong-Ho, Y. (2003). Exploratory Data Analysis. In I. B. Weiner (Ed.), *Handbook of Psychology* (pp. 33–64). London: Wiley. <http://doi.org/10.1002/0471264385.wei0202>
- Bounegru, L., Gray, J., Venturini, T., & Mauri, M. (2017). *A Field Guide to Fake News*. Retrieved from [fakenews.publicdatalab.org](http://fakenews.publicdatalab.org)
- Bounegru, L., Venturini, T., Gray, J., & Jacomy, M. (2016). Narrating Networks: Exploring the Affordances of Networks as Storytelling Devices in Journalism. *Digital Journalism*,
- Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences* (Inside Technology S.). Cambridge MA: MIT Press.
- Burt, R. S. (1995). *Structural Holes: The Social Structure of Competition*. Cambridge MA: Harvard University Press. Retrieved from <http://books.google.com/books?id=E6v0cVy8hVIC&pgis=1>
- Foucault, M. (1970). *The Order of Things*. New York: Pantheon Books.
- Freeman, L. C. (2000). Visualizing Social Networks. *Journal of Social Structure*, 1(1).
- Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(NOVEMBER), 1129–1164. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/spe.4380211102/abstract>
- 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. <http://doi.org/10.1371/journal.pone.0098679>

- Jensen, P., Morini, M., Karsai, M., Venturini, T., Vespignani, A., Jacomy, M., ... Fleury, E. (2015). Detecting global bridges in networks. *Journal of Complex Networks*, cnv022. <http://doi.org/10.1093/comnet/cnv022>
- Lynch, K. (1960). *The image of the city*. Cambridge MA: MIT Press. Retrieved from [http://books.google.com/books?hl=it&lr=&id=\\_phRPWsSpAgC&pgis=1](http://books.google.com/books?hl=it&lr=&id=_phRPWsSpAgC&pgis=1)
- Marres, N., & Rogers, R. (2005). Recipe for Tracing the Fate of Issues and Their Publics on the Web. In B. Latour & P. Weibel (Eds.), *Making Things Public: Atmospheres of Democracy* (pp. 922–935). Cambridge, MA: MIT Press.
- Moreno, J. (1953). *Who Shall Survive? (Second Edition)*. New York: Beacon House Inc.
- Noack, A. (2009). Modularity clustering is force-directed layout. *Physical Review E*, 79(2). <http://doi.org/10.1103/PhysRevE.79.026102>
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA: MIT Press
- The New York Times. (1933). Emotions Mapped by New Geography. *The New York Times*, 3 April.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Turnbull, D. (2000). *Masons, Tricksters and Cartographers*. London: Routledge.
- Venturini, T., Jacomy, M., & Jensen, P. (n.d.). What do we See, When we Look At Networks. Towards a Positive Measure of Spatialisation Quality for Force-Driven Network Layouts. Forthcoming.