Visual Network Analysis

Tommaso Venturini, Mathieu Jacomy, Débora Pereira

Introduction

In the last few years, a spectre has been haunting our academic and popular culture — the spectre of networks. Throughout social as well as natural sciences, more and more phenomena have come to be conceived as networks. Telecommunication networks, neural networks, social networks, epigenetic networks, ecological and economic networks[1], the very fabric of our existence seems to be made of lines and points.

Our fascination for networks is not unjustified and it is not new. Since Euler's walk on Königsberg's bridges[2], networks have proved to be powerful mathematical objects, capable of harnessing the most diverse situations where the connection of discrete elements is at stake. Yet, the recent fortune of networks derives less from their computational power than from their visual affordances. In the last years, the increasing availability of software for network manipulation has turned graphs into something that can be seen and manipulated. Turning graphs into maps and interface, this software has made network analysis available to more and more scholars particularly (but not exclusively) in the social sciences.

Yet the visualization of networks has so far lacked of reflexivity and formalization. Though all network analysis packages propose rich libraries of visualization functions, most literature on networks analysis is still centered on mathematical metrics[3] and does not detail how to *read* visualized network[4]. We painfully lack the conceptual tools to think about the projection of graphs in the space. The very vocabulary we use has been borrowed from mathematics (e.g. cluster, structural equivalence…) and geography (e.g.

---

1 In the words of Fritjof Capra "The organic societies, like anthills and beehives, are metaphors that project the natural environment in the technological social space, as well as the structure of neurons and cells are models for understanding the networked world. In fact, networks naturally reflect the (dis)organization of the universe and nature." (1996)

2 *Solutio problematis ad geometriam situs pertinentis*, 1736.

3 The lack of interests of scholars working on graph mathematics for network visualization is not surprising. In solving to the problem of Königsberg's bridges, Euler performed the most classical of mathematics operations. He abstracted the formal structure of the problem from its empirical features: he took a city and turned it into a table of number. In doing so, Euler laid the foundation of discrete mathematics at the cost of separating the idea of network from its physical materializations.

4 A notable exception can be found at the very beginning of the tradition of *social* networks analysis. Jacob Moreno, founder of this approach, was very 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). Though crucial for the founders of social network analysis, the reflection on network design progressively lost interest for their followers. Understandably fascinated by the parallel developments of graph mathematics, later social networks' analysts focused on statistics and progressively neglected networks design. On the history of social network visualization see Freeman, 2010.

centrality, bridging…) and need to be adapted to the new visual paradigm. This paper means to contribute to such reflection and propose a tentative framework for the visual analysis of networks.

To do so we will draw on the visual semiotics of Jacques Bertin (1967) and in particular on three of its variables: positions, size and hue. The papers will therefore be divided in three main sections, each addressing one of the three variables. Each section will explain how to project one variable on networks (using Gephi software as an example) and provide guidance on how to make sense of the resulting image. As position is, by far, the most important variable (for reasons that will be extensively explained), its discussion will occupy a largest part of the paper and will be divided in three sub-sections. To exemplifying our method of visual analysis, we will discuss a specific case study: a network of some 600 websites and hyperlinks related to the 2012 United Nations Conference on Sustainable Development of Rio de Janeiro (aka Rio+20)[5]. For each step in the analysis of this network, we will a) introduce the conceptual principle employed to read the network; b) exemplify the application of the principle on our case study; c) provide a tentative interpretation of the patterns observed on the network.

Visualizing node positions

How to give a position to nodes

Like geographical maps, graphs are generally two-dimensional representations, but unlike maps they cannot rely on a predefined set of projection rules. In a 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 them a pair of coordinates and thereby a univocal position. The same is true for any Cartesian coordinate system, but not for network graphs. Nothing in network data predetermines where nodes have to be located in the graph. This has to do with the essentially discrete nature of graphs. Unlike geographical maps, graphs do not represent a continuous phenomenon (such as the distance between two landmarks), but a discrete one: two nodes are either connected or not. Therefore, as long as the edges are correctly drawn and link nodes that are connected in the dataset, nodes can assume whatever position without affecting the way the graph is read.

As a consequence, many different ways of positioning networks' node have been proposed through the years. In this article we will focus on a family of spatialization algorithms called "force-vector". Not only because these algorithms are, by far, the most commonly used in network spatialization, but they also because these algorithms have very interesting features. Force-vector algorithm work simulating a system of a physical forces: nodes are charged with a repulsive force that drives them apart, while edges work as

---

[5] 'Rio+20' is the short name for the United Nations Conference on Sustainable Development which took place in Rio de Janeiro, Brazil in June 2012 – twenty years after the landmark 1992 Earth Summit in Rio. At the Rio+20 Conference, world leaders, along with thousands of participants from the private sector, NGOs and other groups, came together to shape how we can reduce poverty, advance social equity and ensure environmental protection on an ever more crowded planet» (http://www.un.org/en/sustainablefuture/about.shtml).
The websites that compose the corpus that we will analyze have been selected according to two criteria:
1. if they are issued by organizations and groups active on environmental issues;
2. if they contain contents specifically related to Rio+20 or if they authors were present and active in the Conference.

springs bounding the nodes that they connect. Once the algorithm is launched it changes the disposition of nodes until reaching the equilibrium that guarantees the best balance of forces. Such equilibrium minimizes the number of lines crossings and thereby maximizes the legibility of the graph.

There is, however, a most interesting by-product of such visualization techniques: not only do force-vector algorithms minimize lines crossings, but they also give sense to the disposition of nodes in the space of the graph. In a spatialized network, spatial distance becomes meaningful: two nodes are close if they are directly connected or connected to the same set of nodes. Because of the very logic that drives them, force-vector algorithms assure that the distance among nodes is roughly proportional to their structural equivalence, that is to say the number of neighbors that they have in common (divided by the total number of their neighbors). Spatialization deliver an amazing result, it turns the discontinuous mathematics of graphs into a continuous space.

To spatialize our example network we used one of the many force-vector algorithms available in Gephi and called ForceAtlas2 (<add reference>) with the following parameters – LinLog mode, scaling 0.35, gravity 0.2, prevent overlap. Here is the result:

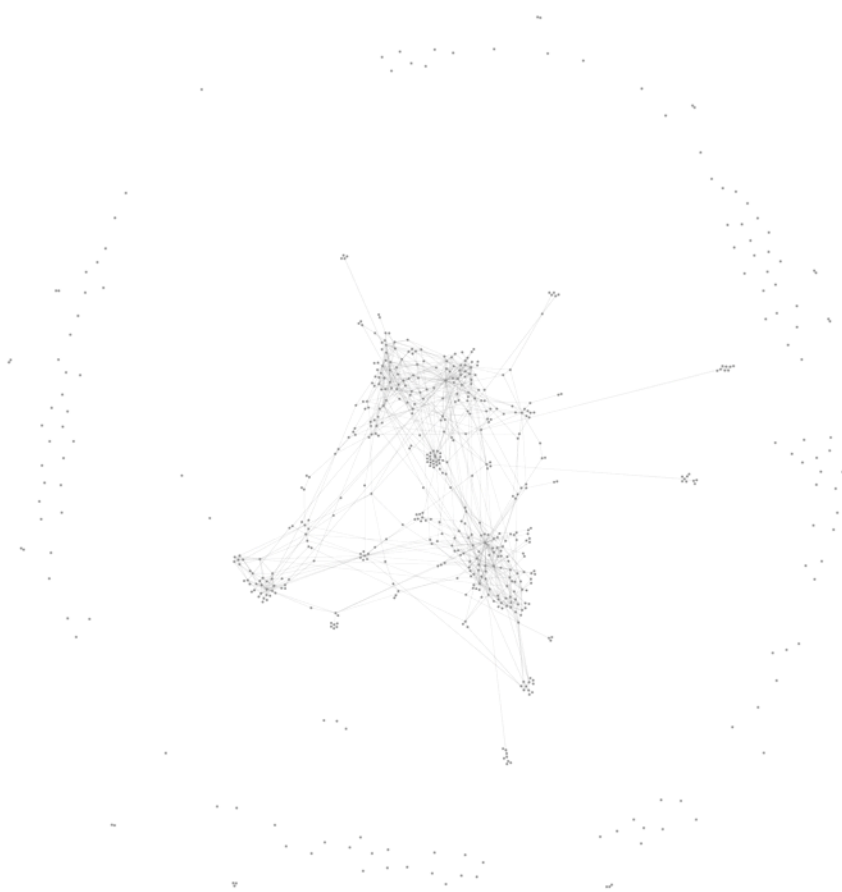

Figure. 1. The network after the spatialization. The main component is the most interesting part. The disconnected nodes form the ring (size and colors have not be modified).

How to interpret difference in density

Reading principle

In most networks, the spatialization reveals regions in which numerous nodes are assembled and regions that are empty or almost. These differences of densities (determined by the uneven distribution of the connectivity in the network) are revealed by the force-vector algorithm like different light exposures are revealed by chemical agents in photography. Spatialization generates visual patterns that translate the mathematical properties of the network. This translation is not free from distortions. Some properties are clearly visible, others are not. Some of the things that can be observed are meaningful, other are not. For example, the absolute position of nodes and cluster (at the top or bottom, left or right of the image) is completely arbitrary. What counts is the *relative* position of the nodes, their agglomeration and their separation. What matters is the clustering of the network.

To be sure, clusters could be detected in other ways. Andreas Noack, in particular, has shown that the mathematically mechanism of force-vectors corresponds to the computation of the clusters by modularity <citation Noack>, a technique often used to detect communities in networks <citation Newman>. Mathematical clustering however imposes a dissection of the network that is often too clear-cut. The advantage of visual techniques discussed in this article is that their fuzziness allows negotiating the frontiers of the clusters. These frontiers are naturally blurred, since clusters are not exclusive categories, but shades of density. Clusters may have clear boundaries, like cliffs separating a plateau from the valley, but most of the time their borders are gradual as the slopes of a mountain. The fuzziness of clusters' frontiers, by the way, is no obstacle to their recognition: a mountain is easy to see even is it impossible to say exactly where it starts and ends.

What is important is to be able to distinguish the clusters and to identify the empty zones between them. These zones are called "structural holes". The larger these holes are, the more they denote the absence of connection between the clusters. In dense graphs (such as those designed by hyperlinks' web or scientometrics networks), such absence is particularly significant and can be interpreted as a symptom of an opposition. Finally, we can remark that large clusters are often composed by smaller (and less distinct) sub-clusters. If large structural holes can be read as oppositions, smaller holes among sub-clusters may denote distinctions without opposition.

We can then summarize the reading principle of the first step by four questions:

- Which are the main clusters?
- Which are the main structural holes that separate them?
- Which are the sub-clusters within each cluster?
- Which are the smaller structural holes separating them?

Example

Which are the main clusters? In our example network it is easy to identify three main clusters at the top (A), at the bottom right (B) and at the bottom left (C). The clusters A and B are the largest and the easiest to identify. The cluster C is smaller and does not contain more nodes than the plurality of smaller sub-clusters scattered through the graph. The cluster C, however, is clearly distinguished from A and B and occupies *its own space*. This is why we count it as one of the main cluster of the network. The triangular shape of our network is thus the result of the three main clusters *pulling in three different directions*.

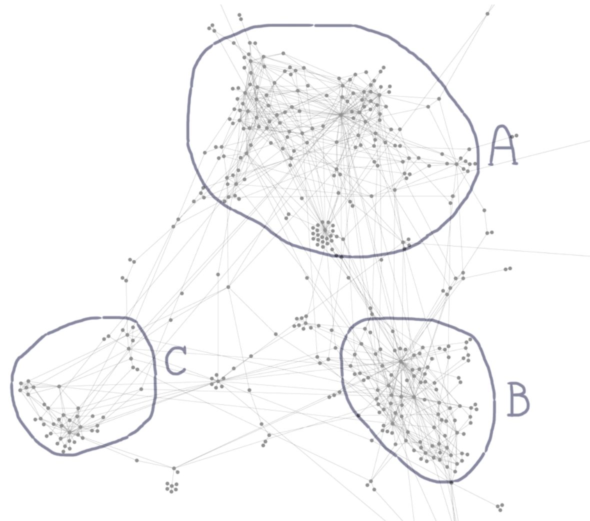

Figure. 2. The three main clusters (A-C)

Which are the sub-clusters? It is easier to distinguish the sub-clusters in the clusters A and B than in the cluster C that is significantly more compact. In A, we have identified two main sub-clusters a1 and a2 and three smaller groups of nodes. In C, we have identified three sub-clusters. In B, we decided not to separate any sub-clusters.

In the identification of the sub-clusters, there is always a part of subjectivity. Some sub-clusters are pretty evident (a1, a2 and c1), but most are not. Sub-clusters by definition smaller and less clear-cut than the main clusters and this can raise doubts on their existence: does c3 contains enough nodes to be interesting? Is a3 really separated from a1? We can leave these questions open for the moment. So far, the sub-clusters are only suggested to provide insights for the analysis.

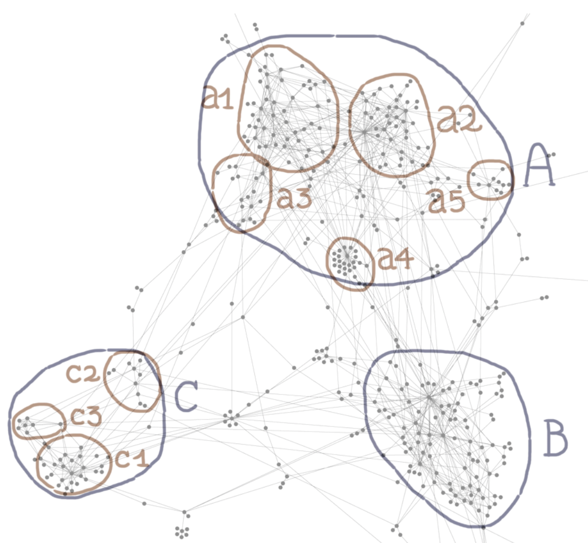

Figure. 3. The sub-clusters of A, B and C

Which are the main structural holes? In our network, there are four main structural holes: one at the center of the graph and three more separating the main cluster two by two. The cluster C is more isolated than the other two. The structural holes are very evident in this network: the absence of links between the main clusters is radical and demands to be explained.

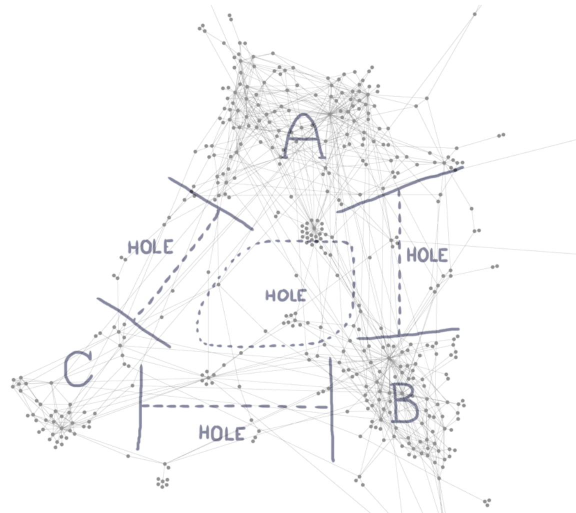

Figure. 4. The structural holes separating A, B and C.

Besides A, B and C, ten smaller clusters occupy different positions. These clusters can be divided in two groups: (1) The intermediary clusters M, E, F and L located among the main clusters. (2) The peripheral clusters K, G, I, D, J and H pushed towards the margins of the graph by the scarcity of edges that connects them to the three main clusters (some of them are so detached from the graph that we are led to consider the possibility of excluding them from the corpus).

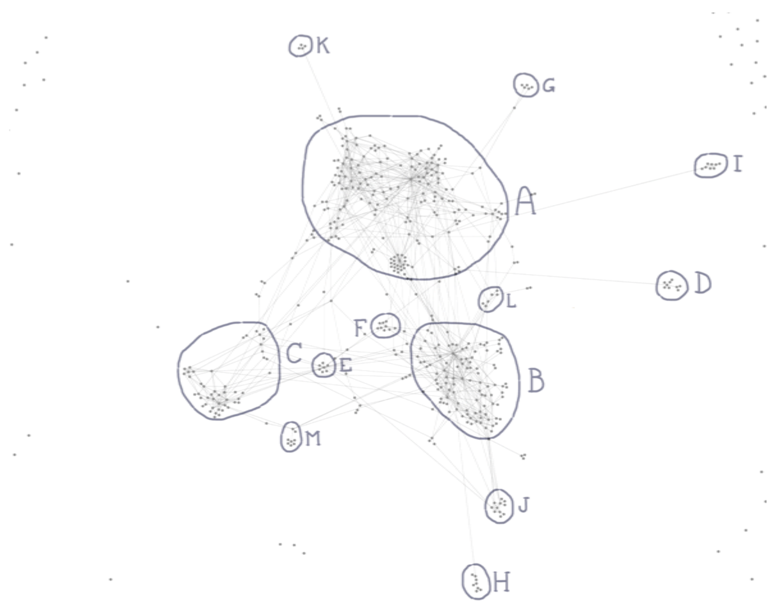

Figure. 5. Three main clusters (A-C) and ten smaller clusters (D to M)

Interpretation

After having spotted the clusters and sub-clusters of our network, we can try to make sense of them. The crucial aim of this phase is to find a suitable 'collective name' for each group of nodes. This is done by examining the nodes each cluster and trying to find what they have in common (at least what most have in common). Clearly some knowledge of the websites and their contents is necessary to discover similarities and that why visual network analysis should always be accompanied by some qualitative enquiry. In our case, all the websites had been visited and analyzed at the moment of the constitution of the corpus. Clustures constitute the main landmarks of the reading process and we will intensively refer to them in the next sections. The table below presents the main clusters and sub-clusters of our example networks:

Cluster	Actors	Contents
A "NGOs and social movements"	Social movements, environmental and human rights NGOs (mostly in Brazil).	Manifestations and social conflicts, indigenous issues, oppositions to dams, cultural events, courses agro-ecology, environmental education, forest management, 'Peoples Summit' event
A1 "marxist eco-socialism"	Main actors: Via Campesina, Movimento dos Sem Terra, Movimento dos Atingidos por Barragens	Ecological discourses inspired by the theories of the Marxist eco-socialism
A2 "environmental politics"	Actors active in southern and central highlands of Brazil	More heterogeneous than a1, its members do not exhibit the same militant politics, but a softer version of environmental politics, which does not engage in the struggle for human rights
A3 "Xingu river"	Dominated by Xingu Vivo movement, shows a connection between local entities in the north and northeast of Brazil and transnational NGOs, such as International Rivers and Conservation Strategy	Struggles against the construction for dams in the river Xingu
A4 "People's Summit"	Dedicate to the People's Summit, an event in Rio de Janeiro organized by social movements during the official United Nations Rio +20 summit	Protest against the official negotiations.
A5 "Brazilian government"	Centered around the Brazilian government, referred to by INRA in France, government of Bogotá and Forest Stewardship Council	
B "international institutions"	UN-related agencies and NGOs working on green economy and sustainable development. Main sites: ONU, official Rio+20, Unep	Reports on UN conferences and debates proposed. Recurring themes are alternative energy, clean water, carbon market, biomaterials and green ICT
C "environmental and climate NGOs"	NGOs for the preservation of forests, indigenous movements and scientific groups who advocate global warming as caused by humans.	Scientific articles, longer texts, campaigns and appeals for donation. Images of animals, forests, landscapes or nature, but without the presence of the man (often described as harmful to nature).
C1 "scientific websites"	Main actors: Real Climate blog, EcoEquity, Skepticalscience, Climateaudit, and Simondonner Indigenize	Debates on climate change and its causes
C2 "Mongabay"	Centered around Mongabay NGO	Information and pictures about nature and protest against its destruction
C3 "ecological Internet"	four websites (Forests, Rain Forests Portal, ClimateArk and Water Conservation) drawing contents from the Ecological Internet	Ecological Internet, self-define as a "non-profit organization that specializes in the use of the Internet to achieve conservation outcomes".

In general, the main clusters A, B and C form three coherent ensembles of websites. Despite internal differences exist, the websites in each of the main clusters are connected by hyperlinks because they share analogous interests and worries and a similar language. These specificities are also the reason of the separation between the three clusters. The NGOs in A and the institutions in B differ on every aspect: movements (A) VS. establishment (B); protest (A) VS. policy-making (B); mobilization (A) VS. planning (B). The institutions in B are also strongly opposed to the NGOs in C because of a deep difference in their values that opposes a pragmatic conception of modern societies and economies (B) to a radical questioning of the place of the humankind in the world (C). Finally, the websites in C and A are separated by the object they defend: social groups and communities for A and environment and ecosystems for C. A and C also employs different forms of engagement: the rationality of scientific knowledge and distant donation of money (C) against the emotions of social movements and first person participation (A). These differences explain why there are few bridges between our three main clusters. The actors tend not to link each, because of their ideological and practical opposition. They are more than different thematic clusters, they are opposed communities of interest.

How to interpret the size and density of clusters

Reading principle

Network clusters have two main properties that we can observe directly: their size and their density. Making sense of these properties is crucial to understand the balance of forces in the network.

The size of clusters is defined as the number of nodes they contain. The biggest clusters are the most visible on the Web and it is interesting to investigate the offline counterparts of such online significance. The Web is always a deforming prism. Minorities are sometime over-represented in online debate and some groups exist almost exclusively in the cyberspace <add citation to “Visibles mais peu nombreux”>. It is often relevant to compare the size of clusters on the Web to the presence of these actors in other public spaces (newspapers, institutions…).

The density of a cluster is a measure of its cohesion. Clusters are tight when they contain many edges and loose when few edges connect their nodes. In the case of the Web, a high number of in-cluster links may denote a the activity of a community: the actors know and acknowledge each other through their citations. A low density is also interesting. It may denote that the nodes do not know their neighbors or actively disregard them (because of competition or controversy). In low-density clusters, it is not the connections among their members that keeps them together, but the stronger separation with the rest of the network that *digs a ditch* of structural holes around them.

Example

The three main clusters count several dozens of nodes. Among them, A is the biggest and C the smallest. All the other clusters are definitely smaller with less than a dozen nodes.

As for the density, B is the only among the main clusters that is relatively dense. A and C are more spread out and clearly separated in smaller and denser sub-clusters (which appears to be more interesting than they larger parents). Cluster A is largely defined by a1 and a2. These two sub-clusters are large and dense and contains most of the nodes of A. a3 and a5, on the contrary, are scarcely dense and are distinguished from a1 and a2 only by their separated position. a4 is special case because of its star structure (see fig. 8).

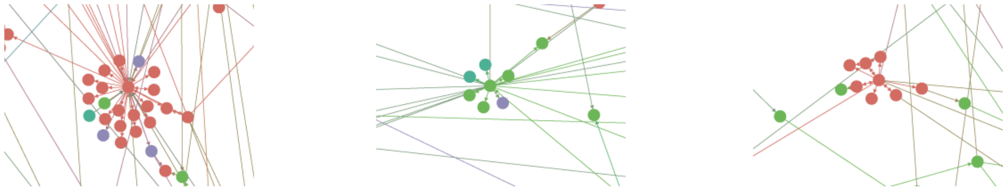

Figure. 6. Star structures: the sub-cluster a4 and the smaller clusters E and F

Similar little stars appears everywhere in the network (see fig. 8). They all are characterized by a central point surrounded by nodes that are only connected to the center but not among them. This is one case where the visual analysis of the network can be misleading: though they look compact, the stars are not particularly dense from a mathematical point of view. In the case of the Web, we will interpret these patterns as the symptom of the activity of an authority or a hub (more on this later on) and not as the sign of a communitarian activity. It is therefore important to distinguish stars from clusters: though they may look equally dense they are produced by very different mechanisms.

Cluster C contains a larger sub-cluster c1, a very sparse cluster c2 and another special case c3 that appears to be a clique. Cliques are group of nodes that are all connected to each other. It is rare to observe large cliques in natural networks, but it is possible to find small ones (with less than ten nodes). Quasi-cliques are similar structures where *almost* all the nodes are connected (see fig. 9). Cliques and quasi-cliques have no center, even if some nodes may appear more central than the others on the image (if all nodes are connected, no nodes is more connected than the others).

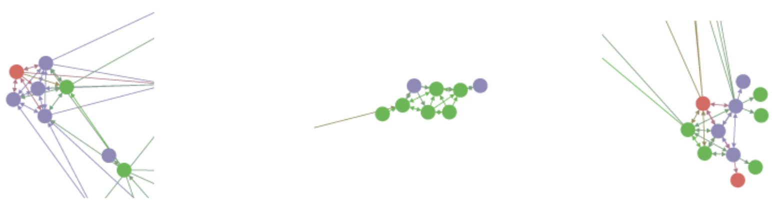

Figure. 7. Clique of the sub-cluster c3 and quasi-cliques of the clusters I and J

Interpretation

The **cluster A** "NGOs and social movements" is the largest of our network and this is probably due to the fact that it corresponds to an active community. The 'occupation of the Web' is an important issue for activists, both to assure their internal communication and to win the support of public opinion. The phenomenon of minorities' over-representation may also be at play and this specific community may be particularly visible on the web. **Cluster C** "Environmental and climate NGOs" is also composed predominantly by NGOs and associations, but its lesser size may indicate a smaller or less active community. Finally, the **cluster B** "International institutions" is composed mainly by the numerous institutions gathered around the site of the United Nations (which explains the absence of sub-clusters). The links between institutions are often the simple mark of formal partnership. The sizes and shapes of the larger clusters seem therefore to be consistent with the different types of social organizations present in our network. The same can be said for the sub-clusters, which correspond to the divisions among the different types of associative groups present in A and C.

In the **sub-cluster a1** “Marxist eco-socialism”, a lot of information is circulated and discussed. This activity creates many links among the actors explaining the relatively high density. Blogs, in particular, play a central role in this group of sites (more on this later). The **sub-cluster a2** “Environmental politics” display a similarly intense communitarian activity (which explain the high number of links), but has a different thematic focus and is composed by a different type of actors, mostly NGOs. The separation between a1 and a2 may be in part explained by the fact that blogs tend to cite other blogs while NGOs prefer citing other NGOs. Though permeable, these two spheres remain relatively separated.

The **sub-cluster a4** “The People’s Summit” has the form of a star which can be explained by the fact that the social ecology scheme of the management of nature is well articulated, though it does not have the strengths nor the excessively referenced informational authorities. At the center of cluster C, the **sub-cluster c1** “Scientific debate on climate change” gathers an active and connected community. **c2** "Mongabay" and **c3** "Ecological Internet" represent two smaller groups of different types of actors (most NGOs) that though referring to the scientific blogs, are not confused with them. The clique structure of the **c3 sub-cluster** (experts in forest preservation, ecosystems and indigenous peoples) is explained by the fact that, with the exception of the sites Peoples Issues and New Earth Rising, all the other sites mirror the contents of the website Ecological Internet. Knowing the clique structure reinforces each actor notably in the search engine visibility, it would be interesting to investigate whether this strategy is deliberate.

How to detect centers and bridges

Reading principle

Now that we have identified the clusters, we can use them as landmarks to analyze two remarkable positions in spatialized networks: centers and bridges.

Centrality can be global (referred to the whole network) or local (referred to a single cluster). These two types of centrality are different. While the elements that are globally central are pulled in this position by the fact of being evenly linked to all the regions of the network, the element that are locally central tend to be linked predominantly within one cluster. Central positions (local or global) can be occupied by single nodes or by (sub-)clusters. In many cases, the center of the network (or the center of a large cluster composed by several sub-cluster) is just empty.

Bridges, on the other hand, are nodes or clusters that have connections with several clusters (two or more, but not all the clusters of the network). Bridges can be located outside the clusters they connect, if their connections are evenly distributed among them, or they can be located within one of the clusters, if they are more connected to it than to the others.

The following questions may help to identify systematically central and bridging elements:

- Which nodes or clusters (if any) are located in the center of the network?
- Which nodes or sub-clusters (if any) are located in the center of each cluster?
- Which nodes (if any) are located in the center of each sub-cluster?
- Which nodes or clusters (if any) are located among the main clusters?

- Which nodes (if any) are located among the main sub-clusters?

Example

Six nodes have been identified as centers of clusters and sub-clusters. Nature.com is the only node occupying the center of our network and the only node to connect the three main clusters. UN.org the website of the United Nations, is at the center of the cluster B. CupulaDosPovos.org.br is at the center of the sub-cluster a4 "People's Summit" and RealClimate.org is at the center of c1 "scientific debate on climate change". Finally, two nodes are central in smaller clusters IUCNWorldConservationCongress.org for E and Demilitarize.org for F. The presence of many edges around each of these nodes has helped us to detect their centrality.

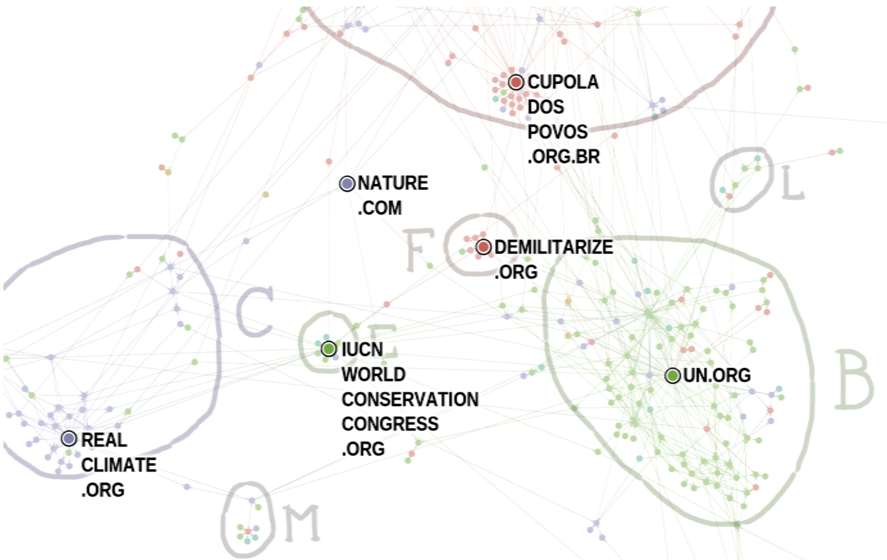

Figure. 8. Nodes in central position in the network or in their cluster.

Three clusters and seven nodes and are in bridge position. The clusters are the easiest to identify: E and F are located between B and C and L is located between A and B. Seven nodes are in a position to bridge between two clusters. Three nodes are simply 'stretched' between two clusters: GlobalVoicesOnline.com (A and C), NoGreenEconomy.org (A and B) and EffetsDeTerre.fr (B and C). Care2.com and IndianCountryTodayMediaNetwork.com *together* form a (3 links-long) bridge between A and C. Finally, Bndes.gov.br is an "internal" bridge located inside cluster C and connecting it to cluster B.

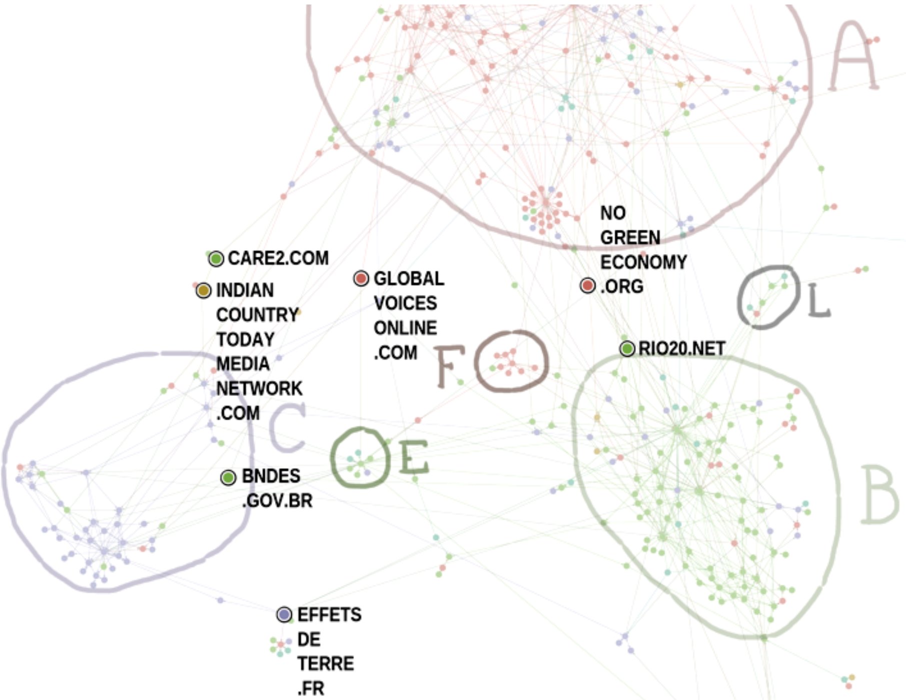

Figure. 9. Nodes and cluster in a bridging position.

Interpretation

Centers

Nature.com is the site of the famous scientific journal. It is an academic authority and is cited by all the many clusters, but not much (6 links in total). Of course, other websites exist that are cited by the nodes of the three main clusters (probably the major news websites), but they have not been included in the corpus because their content was not sufficiently focused on Rio +20 and its issues.

UN.org UN.org is a large portal linked to most of the numerous specialized institutions that constitute cluster B (and this explains its central position). The site contains static information on the United Nations: its mission, structure, Charter of Principles, the list of the member states and more. There is also a section of updates, daily news and highlighted dossiers.

RealClimate.org is a commentary website on climate science run by a group of renowned climate scientists and addressed to journalists and to the general public.

CupolaDosPovos.org.br is the center of the sub-cluster a4 “People’s Summit’ and its connections keep this star of nodes together. **IUCNWorldConservationCongress.org** and **Demilitarize.org** are in a similar position in clusters E and F.

Bridges

Interestingly, in our network, the role of bridge is played not only by nodes but also by small clusters. **Clusters E** “IUCN congress” is a bridge because of its focus on global warming mitigation. Mitigation is a key theme in conferences and events organized both by the institutions of the clusters B and the NGOs of the cluster C. As such it connects two distant regions of the network and facilitate their interaction.

Another site that promotes connections between clusters B "international institutions" and C "environmental and climate NGOs" is the blog **Effets de Terre** (an independent version of the blog that the journalist Denis Delbecq maintained from 2005 to 2007 in the French newspaper Libération on climate change and environmentalism).

Rio20.net present the program of the Cupola dos Povos event and is cited by several actors in clusters A and B. **NoGreenEconomy.org** is not an important website (the site seems 'under construction' and it contains only 5 posts). Its position between A and B is explained by the fact that the website is maintained by a group of activists, whose position are close to those of the social movements in A while criticizing punctually the approach of the institutions in B (and thereby citing them).

For all other cases, we have found no convincing interpretation. When the bridging position of a node cannot be confirmed by the qualitative analysis of its content, the best option is simply to ignore it. Unlike the larger patterns visible in the networks, single edges are not always significant. The aim of the visual analysis of networks is not to explain the position of each and every node, but to detect large trends and notable nodes.

Visualizing node sizes

How to give a size to nodes

We have now completed the part of the analysis based on the spatial position of nodes and we will start mobilizing the two other visual variables employed in the visual analysis of networks, starting from the size. In particular, we will now visualize the number of the edges arriving to or leaving from a node by changing the diameter of point that represent it. We will first change the size of the nodes according to number of incoming links (the in-degree) and then according to the number of their outgoing links (the out-degree). To do so, we have used the ranking palette of Gephi and set the diameter 1 for the smallest degree and 20 for the largest.

It is also possible (and indeed useful) to just look at the list of nodes sorted by their in-degree ou out-degree. Projecting the ranking on the spatialized networks, however, is also interesting as it allows identifying where the hubs and authorities are: are they central in a cluster or do they bridge different regions? Are they uniformly distributed or do they concentrate in some part of the graph? We could even go as far as to detect the local hubs and the authorities for each cluster (but we will not be so detailed in this article).

How to read the hierarchy of connectivity

Reading principle

We will now consider the hierarchy of the most connected nodes. Following the tradition of network analysis, we will call 'authorities' the nodes that are the destination of many edges (inbound links) and 'hub' the nodes that are the origins of many edges (outbound links). Authorities are websites with a high visibility and toward which much of the traffic is addressed. Hubs are portals or websites that reference many other sites in the network. Both authorities and hubs tend to be influential nodes in the corpus.

Please remark that, in counting in-bound and out-bound links, we only take into consideration the connections within the corpus (and not all the hyperlinks that one website receives or sends). The website Nature.com, for instance, is certainly an authority in the World Wide Web, but it is not in our smaller network (despite its very central position).

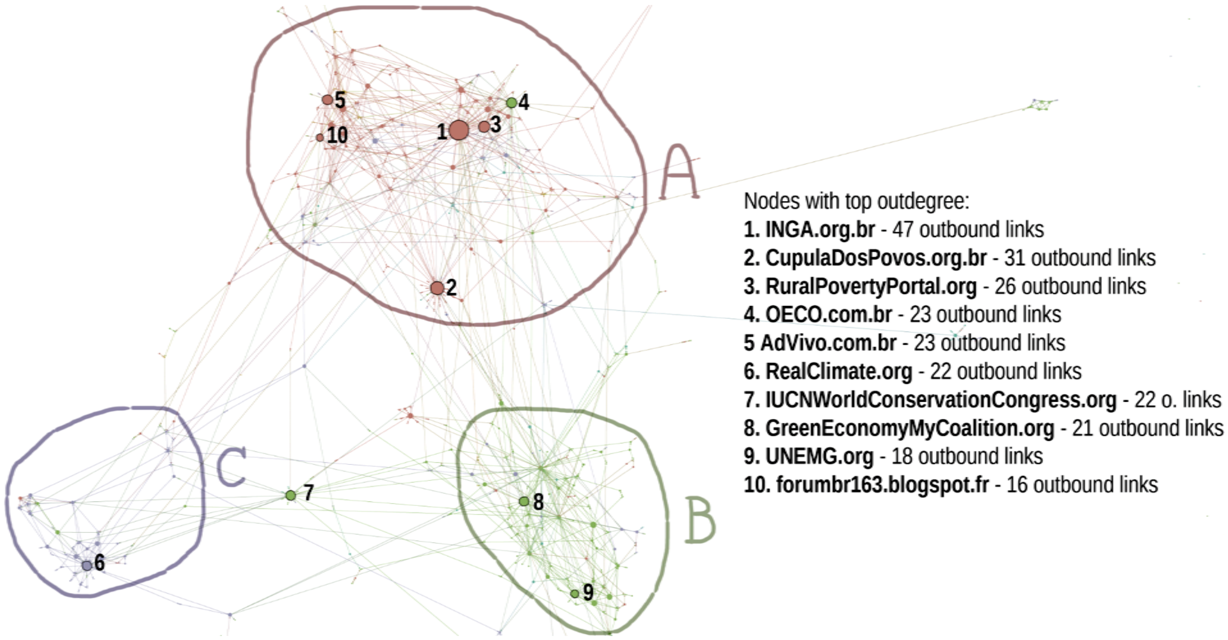

Figure 10. Top10 authorities: the ten most cited sites in the corpus [top in-degree].

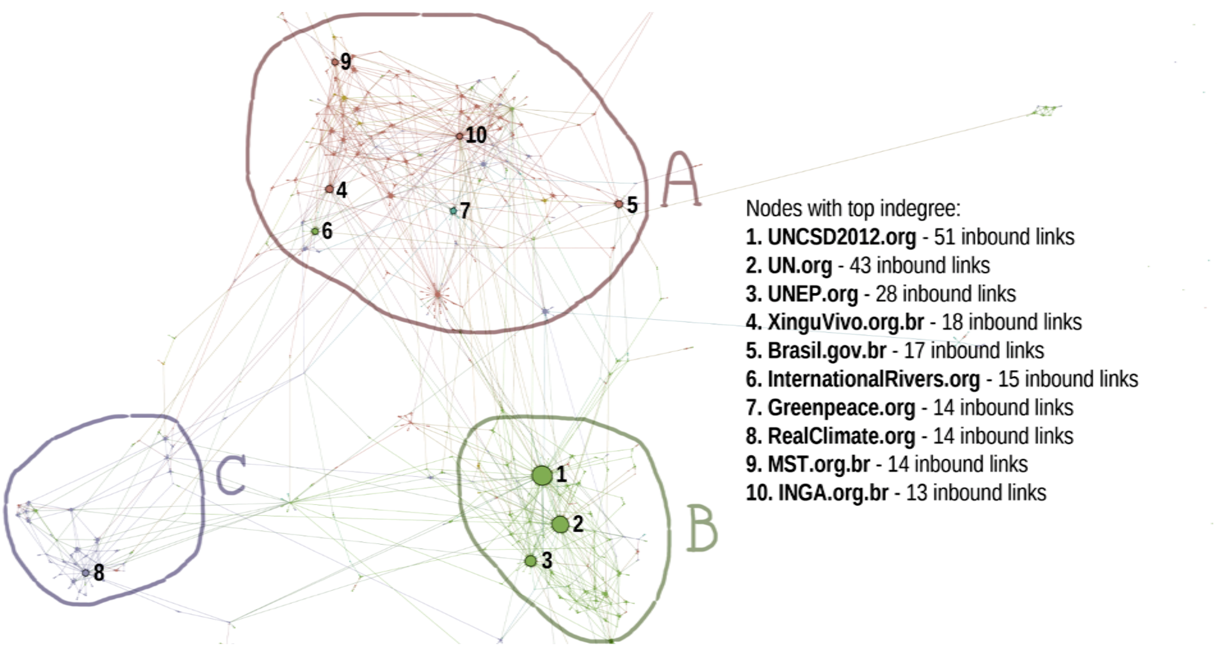

Figure 11. Top10 hubs: the ten sites citing the most other sites in the corpus [top out-degree]

Example of analysis

According to a power law often found in the Web (and in general in natural networks according to Barabasi, 2003), the distribution of the in-degree is very skewed in this graph, with the three most cited websites having 28 to 51 incoming links, while the rest of the top10 varies from 18 to 13. It is remarkable that all three main authorities of the graph are located in cluster B "international institutions". The rest of the top10 (except for one authority in cluster C) is located in cluster A "NGOs and social movement", whose high density of connection naturally produce local authorities.

As for the hubs, more than a half of the top10 (and the 5 biggest hubs) is located in the cluster A and again the density of such cluster can provide an explanation. It interesting to remark the presence of an important hub 'IUCN Congress' in a bridging position between B and C.

Interpretation

The main authorities of the network are all international institutions and this status seems to drive a large amount of hyperlinks to them. The three main authorities of the graph (uncsd2012.org, un.org and unep.org) are located in cluster B. The high density of this cluster and the lack of sub-clusters are largely due to the centripetal force of these three websites. These three websites are however local authorities: even if they receive links from other clusters, the largest part of their neighbors remains within cluster B.

Looking at outgoing links, given the high digital mobilization we observed in "NGOs and social movements", it is not surprising that most hubs are in cluster A and that these sites correspond to very active communities: INGA.org.br, RuralPovertyPortal.org and OECO.com.br are strongly engaged on rural ecology; AdVivo.com.br and ForumBr163.blogspot.fr on Marxist questions.

Visualizing node colors

How to apply a color to nodes

The last transformation we would like to operate on our network is to color its nodes according to the categories to which they belongs. This stage, of course, is only possible if the nodes have been categorized beforehand. To be sure, the same nodes can and (when possible) should be classified according to different systems of classification. Each classification system is project on the network as different layers of colors projected on the same background map. In our case, the nodes of the network had been categorized at the moment of the harvesting according to two different systems of classification: the approach to ecology that inspires them and the language in which they are written. Drawing on these classifications, we can use the partition panel in Gephi to attribute a different hue to each type of nodes.

It is important to remind that the color is a non-mixable visual variable. A node can be red or blue but not the two at the same time. When categorizing nodes, it is therefore necessary to employ exclusive categories. A website, for example, should be categorized as French or English, but not as both. If both languages are present on the same websites, researchers can add an additional category 'multi-lingual' (which is also exclusive).

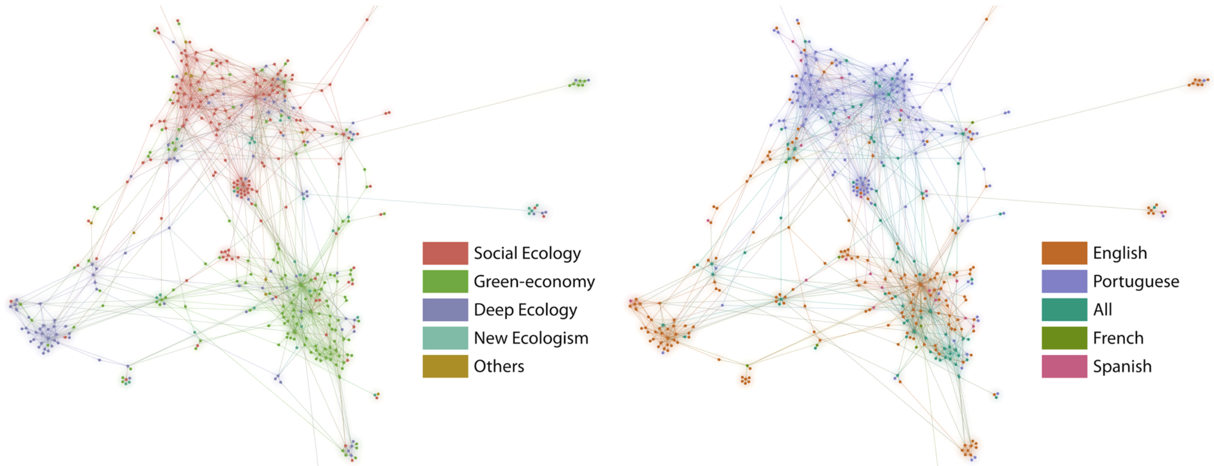

Figure. 12. The nodes of the network colored by approach and language

How to read the distribution of colored partitions

Reading principle

Having colored the nodes of our graph, we can now examine how the colors are distributed in the different regions of the network. In particular, it is interesting to observe if the nodes of the same color tend to be closer than nodes of different colors – creating a correspondence between the typology and the topology of the network. When such correspondence is observed it can be used as a basis to explain the patterns observed in the network.

Of course, the correspondence between categories and clusters is not always bijective: one category does not always correspond to one cluster. One category may colonize more than one cluster and two or more categories can associate to form a single cluster. Still, if the nodes of the same color tend to be closer than others, there is ground for interpretation. A interesting example of this situation is provided by the so-called 'hairball networks'. These are graphs that do no show any visible clusterization and are therefore difficult to analyze visually. However, when their nodes are colored, effects of polarization may appear. Even though the density of connection is homogenous all over the graph, nodes of different categories may still be visually separated.

Finally, when different layers of classification are present in the network, it is interesting to compare them and observe wether the different classifications produce the same borders in the network. Often this is not the case and sometime this explains why the correspondence between categories and clusters is not bijective.

Example

The most interesting differences between the websites of our network concern their different approaches to ecology. In particular, it is possible to find in our corpus websites corresponding to the three main 'schools' describe in the literature on ecology (Diegues, 2000; Koppes, 1989; Simonnet, 1979; Lipietz, 2012; Latour, 2004; Herber, 1964; Bramwell, 1989; Lynton, 1989; Lash Et Al, 1996; Zencey, 1989; Daly; Cobb, 1989):

- **Social Ecology**: explains environmental degradation as a result of capitalism and hierarchical division in society (between rich and poor, old and young, white, black and yellow). It advocates a return to primitive communitarian systems.
- **Deep Ecology**: deep ecology argues that nature was not given to humans, who have no right to use or exploit it. The objective of this type of ecology "is to preserve the nature of a hostile, essentially aggressive humanity" (Lipietz, 2012, p. 45).
- **New ecology**: emerged in the 60s, new ecology is directly opposed to consumerism.

An additional category, **Green Economy**, has been added to these to account for a large number of websites (32% of the corpus) that do not to fit in any of the previous categories and seem to be unified by the fact of proposing a synergy between ecology and market economy. Finally, as always, there are cases than cannot be pigeonholed in any category and are therefore classified as "**Others**".

Fig. 15 shows a clear correlation between our categorization and the topology of the network, as each of the three main clusters has a different dominant color. The cluster A "NGOs and social movements" is dominated by the social ecology approach; the cluster B "international institutions" is dominated by the green ecology; and the cluster C "environmental NGOs" by deep ecology. It is worth to remind that the spatialization algorithm we used do not take into consideration the categories of the nodes. The correspondence between hue and position is therefore a sign of strong correlation, so strong that we can make the hypothesis that the ideological agreement is a major driver of the connectivity in our network.

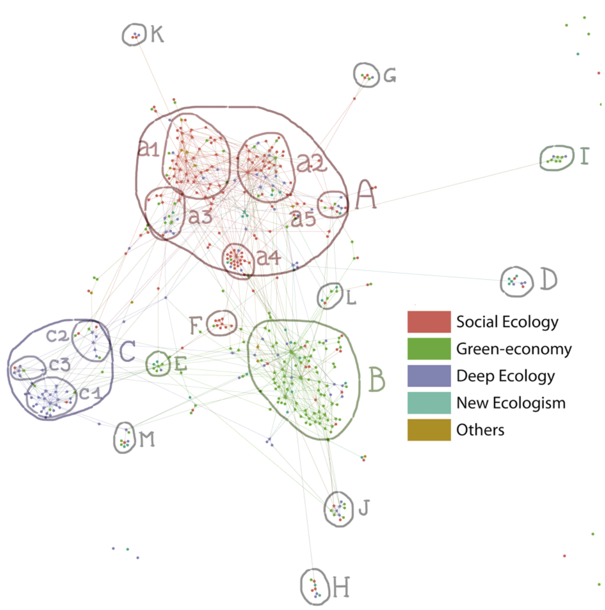

Figure. 13. All main clusters have a distinctive color.

Coloring the websites according to their language (fig. 16), we observe again a strong correspondence between typology and topology. Cluster A is largely composed by Portuguese websites, cluster B is divided between English and multilingual websites and cluster C is mostly in English. Since the proximity in the image indicates the (direct or indirect) connection, we observe a (not surprising) tendency to link websites of the same language.

It is interesting to observe the linguistic polarization of cluster B “international institutions", with English sites at the top and of multilingual sites at the bottom. Though clear-cut the linguistic separation in the cluster is not strong enough to produce a structural hole separating two different sub-clusters.

It also interesting to remark that cluster C “environmental NGOs” is also dominated by English websites and yet it does not merge to cluster B. Evidently, the organizational (NGOs VS international institutions) and ideological differences (deep ecology VS green economy) are, in this case, stronger than the linguistic bound.

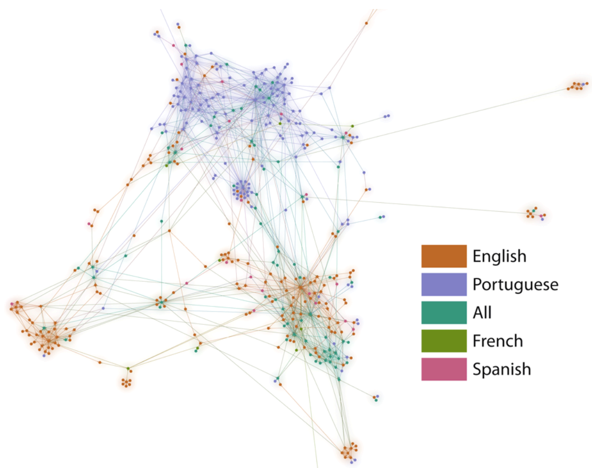

Figure. 14. Nodes colored by language

Interpretation

Drawing on the thematic and linguistic categorization, the separation between cluster A and B appears even deeper that we initially suspected. The two clusters are opposed by the type of organization that compose them (association VS institutions), by their approach to ecology (social ecology VS green economy) and by their language (Portuguese VS English and multilingual). This linguist difference seems also to imply a different geographical focus (local VS global). In this sense it is interesting to remark that the English pole of cluster B is closer to A (more connected) than the multilingual pole.

On the other hand, the structural hole between B and C cannot be explained by language and derives probably from the different positions in the debate. We can also observe that while the Portuguese websites cluster together the English websites do not. While Portuguese is the language shared by a community of local activists, English seems to be a neutral language used to address an international audience.

Conclusions

In this paper we have presented basic of the visual investigation of networks. This technique, we hope, will extend the ‘market’ of network analysis, by making the power of networks available to scholars with limited mathematical knowledge. By translating the key notions of graph mathematics (clustering, authority, bridging…) into the three visual variables of position, size and hue, we have tried to provide scholars with methods to analyze large and complex networks while sparing them most mathematical complications.

But there is more. Part of the interest of visual network analysis comes from the particular relationship between data and expertise that it proposes. Visual analysis entails a continuous iteration between the observation of the data and the interpretation of the findings. The continuous nature of two of the analytic variables (position and size) and the fact the third (color) depend on a manual categorization demand the constant engagement of the researcher's choice. Where are the limits of each cluster? Which nodes are central or more visible? Which are the bridges? Spatialized and ranked networks may suggest insights, but they never impose answers to these questions.

This has advantages and drawbacks. The main disadvantage of visual analysis is that it is impossible without some previous knowledge of the data and the phenomenon that they refer to. Without the help of Débora, who has constructed the hyperlink network and who has extensively studied Rio+20, there is no way we could have carried out such an insightful analysis. As every innovative research technique, visual research analysis is trapped in the "experimental regress" (Collins, 1975). Since both the method and its objects (the networks of hyperlinks, citations, words co-occurrence…) are still largely unexplored, it is hard to find a stable ground to establish their validity. How can we know that the patterns that we glimpse on the networks are not mere artifacts of the spatialization algorithm or projection of our previous knowledge? The only way out of these doubts is through the consistency between what we observe in the network and what we already know about the phenomenon it refers to. In our case, for example, we were comforted by finding a vast structural hole between social and deep ecology perfectly consistent with the long discussed difference between these two approaches. Also it was reassuring to find the websites of the organizers of the event around which the corpus was built (uncsd2012.org, un.org and unep.org) as the three largest authorities of the networks.

Our visual analysis, however, did not just confirm what we already knew about Rio+20 (little interest would have otherwise). It also offered a few notable surprises. The importance and separation of the green economy cluster was one of them, as well as the centrality of CupulaDosPovos.org.br and the bridging position of the site of its alternative summit, Rio20.net. We have already discussed these and other findings in the article and we will not come back to them in the conclusion. These examples serve only to illustrate the main advantage of visual network analysis. Precisely because it provides insights and not clear-cut answers, the visual investigation of network is primarily a method for exploratory analysis (Tukey, 1977). By encouraging scholars to engage with their networks (sometime to struggle with them), visual network analysis force researchers to assume an active attitude, to challenge and search ground for the previous knowledge and to open up to findings that they may not have thought to. Sometime visual network is frowningly compared to tasseography (the art of interpreting patterns in tea leaves, coffee grounds, or wine sediments). To a certain extent this comparison is not amiss: not unlike the best forms of divination, visual analysis is indeed meant to confront enquirer to their data, to explore their networks, to question their ideas. In this paper, we hope we have provided some guideline for it.

References

BARABÁSI, A. L. Linked. How Everything is Connected to Everything else and what it means for Business, Science and Everyday Life. Cambridge: Plume, 2003.

Bertin, J. (1967). *Sémiologie Graphique*. Paris - La Haye: Mouton.

BRAMWELL, Anna. **Ecology in the 20th Century**: a history. New York: Yale University Press, 1989.

Capra, F. (1996). *The Web of Life: A New Scientific Understanding of Living Systems*. *Colonial Waterbirds* (Vol. 20, p. 347). doi:10.2307/1521798

Collins, H. M. (1975). The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics. *Sociology*, 9, 205–224. doi:10.1177/003803857500900202

DIEGUES, Antonio Carlos Santana. **O mito moderno da natureza intocada.** São Paulo: Editora Hucitec, 2000.

DALY, Herman E.; COBB JR., John, B. **For the Common Good**: redirecting the Economy toward community, the environment and a sustainable future. Boston: Beacon, 1989.

Freeman, L. C. (2000). Visualizing Social Networks. *Journal of Social Structure*, 1(1).

HERBER, Lewis. **Ecology and Revolutionary Thought**. 1964. Disponível em: <http://ebookbrowse.com/gdoc.php?id=377184481&url=342a142f538180ceae5280ae1c3a84a3>. Acesso em: 3 jul. 2011.

KOPPES, C. 1989. "Efficiency, Equity, Esthetics; Shifting Themes in American Conservation". In: WORSTER, D. (ed.). **The Ends of the Earth: Perspectives on Modern Environmental History**. Cambridge: Cambridge University Press.

LATOUR, Bruno. **Politiques de la nature**: comment faire entre les sciences en démocratie. Paris: Éditions La Decouverte, 2004.

LASH, S.; SZERSZYNSKI, B.; WYNNE, B. **Risk, Environment & Modernity**: Towards a New Ecology. London: Sage, 1996.

LIPIETZ, Alain. **Qu'est-ce que l'ecolgie politique?** La grande transformation du XXI e siècle. Paris: Les Petits Matins, 2012.

LYNTON, Keith Caldwell. **International Environmental Policy**. Durham, N.C.: Duke University Press, 1990.

Moreno, J. (1953). *Who Shall Survive?* New York: Beacon House Inc.

SIMONNET, D. 1979. L'ecologisme. Paris: PUF (Que sais-je?)

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading: Addison-Wesley.

ZENCEY, Eric. Apocalipse now? Ecology and the perfidy of Doomsday visions. **Utne Reader**, n. 31, p. 90-93,1989.