Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics
Martin Grandjean, Mathieu Jacomy

To cite this version:
Martin Grandjean, Mathieu Jacomy. Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics. Digital Humanities, 2019, Utrecht, Netherlands. halshs-02179024

HAL Id: halshs-02179024
https://halshs.archives-ouvertes.fr/halshs-02179024
Submitted on 10 Jul 2019

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.
Translating Networks
Assessing correspondence between network visualisation and analytics

Digital Humanities 2019, Utrecht, Netherlands

Martin Grandjean (1) and Mathieu Jacomy (2)
1. University of Lausanne, Switzerland
2. Aalborg University, Denmark

Network interpretation is a widespread practice in the digital humanities, and its exercise is surprisingly flexible. While there is now a wide variety of uses in different fields from social network analysis (Ables et al., 2017) to the study of document circulation metadata (Grandjean, 2016) or literature and linguistic data (Maryl and Elder, 2017), many projects highlight the difficulty of bringing graph theory and their discipline into dialogue. Fortunately, the development of accessible software (Bastian et al., 2009), followed by new interfaces (Rosa Pérez et al., 2018; Wieneke et al., 2016), sometimes with an educational dimension (Beaulieu, 2017; Xanthos et al., 2016), has been accompanied in recent years by a critical reflection on our practices (Weingart, 2011; Kaufman et al., 2017), particularly with regard to visualisation. Yet, it often focuses on technical aspects.

In this paper, we propose to shift this emphasis and address the question of the researcher’s interpretative journey from visualisation to metrics resulting from the network structure. Often addressed in relation to graphical representation, when it is not used only as an illustration, the subjectivity of translation is all the more important when it comes to interpreting structural metrics. But these two things are closely related. To separate metrics from visualisation would be to forget that two historical examples of network representation, Euler (1736) and Moreno (1934), are not limited to a graphic reading (the term “network” itself would only appear in 1954 in Barnes’ work). In the first case, the demonstration was based on a degree centrality measurement whereas in the second case the author made the difference between “stars” and “unchosen” individuals while qualifying the edges as inbound and outbound relationships.

This is why this paper propose to examine the practice of visual reading and metrics-based analysis in a correspondence table that clarifies the subjectivity of the translation while presenting possible and generic interpretation scenarios.

Visual approach: making the global structure readable

The way we read networks has changed over time. Historically the question of network readability was asked in terms of aesthetic criteria. In the word of Jacob Moreno “the fewer the number of lines crossing, the better the sociogram”. Even in the nineties, when giving birth to the modern layout algorithm, Fruchterman and Reingold (1991) aimed at “minimizing edge crossings” and “reflecting inherent symmetry”. However, these criteria do not seem so crucial to practices observed nowadays in digital humanities (and beyond).
Looking at recent papers in digital humanities, networks appear to have a wide range of usages. Their visualisations are either self-sufficient [fig. 1.a.] (Algee-Hewitt, 2018; Pino-Diaz and Fiormonte, 2018; Verhoeven et al., 2018; Marraccini, 2017), an optional help to understanding [fig 1.b.] (Colavizza et al., 2016) or strongly connected to the text. Some authors use them to highlight the position of a specific node [fig. 1.c.] (Moretti et al., 2016), to compare layouts [fig. 1.d.] (Sozinova, 2016) or the layout of the same graph in time [fig. 1.e.] (Wright, 2016). They may aim at visualising communities [fig. 1.f.] (Rybicki et al., 2018; Torres-Yepez and Zreik, 2018), mapping a general structure [fig. 1.g.] (Gao et al., 2017), tracking density patterns [fig. 1.h.] (Gao et al., 2018) or monitoring algorithms like modularity clustering [fig. 1.i.] (Choinski and Rybicki, 2017). These usages reveal a different perspective in network visualisation where we expect the visual to translate underlying relational structures. It helps to give different names to these two different approaches. We call diagrammatic the perspective where the network is a diagram that we read by following paths. We do not want the edges to cross and we use aesthetic criteria to bring clarity. It was Moreno’s perspective, and is still relevant to small networks and local exploration. Then we call topological the perspective where the network is a structure that we read by detecting patterns. We expect the visualisation to help us retrieve structural features like clustering or centralities. It is a common practice in digital humanities, more holistic and relevant to larger networks. Aside or in

<table>
<thead>
<tr>
<th><strong>a. Image speaks for itself</strong></th>
<th><strong>b. Helps to visualise the argument</strong></th>
<th><strong>c. Highlights a specific node</strong></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>d. Comparing layouts</strong></th>
<th><strong>e. Same network over time</strong></th>
<th><strong>f. Visualising communities</strong></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>g. Image presented as a map</strong></th>
<th><strong>h. Comparing densities</strong></th>
<th><strong>i. Visualising modularity clustering</strong></th>
</tr>
</thead>
</table>

**Fig. 1** Different contexts for network visualisation in DH2016, DH2017 and DH2018 abstracts.
complement, classic data visualisation is also employed to visualise non-relational structures (node attributes, etc.).

In the topological perspective, a standard procedure is to assign nodes a position using a force-driven algorithm. This family of algorithms is known for displaying clusters that match a widely used measure of community detection, modularity clustering (Noack, 2009). Its translation remains however difficult to interpret locally, as we can never give a simple explanation for a node’s position. Classic data visualisation also translates non-relational structures, by itself or combined with a relational perspective. Different structural features may require different visualisations: the examples of fig. 2 shows curated visualisations using categories [fig. 2.a boys and girls, in the famous example of (Moreno, 1934)], temporality [fig. 2.b] (Jänicke and Focht, 2017) or hierarchy [fig. 2.c] (Grandjean, 2017). Though very different from force-driven placement, they display better certain structural features.

Objectifying the structure with metrics

Often opposed to visual interpretation, of which they would be a more objective and reliable representation, centrality measures have a history that goes back to more than half a century and shows that they are not immutable and require constant adaptation to usage. Moreover, Freeman (1979) insists on the fact that the notion of “centrality” is the result of several intuitive conceptions. To remind that these metrics are based on “intuition” means to recognize that they have no meaning in themselves and that their interpretation must be rediscussed - and therefore translated - according to the context. This paper thus proposes to list and evaluate to which extent these metrics are applicable to humanities and social science data and can, if necessary, be “translated” into this language to complement visual analyses.

Global properties

Statistical analysis allows for comparing networks across multiple dimensions at once (Tank and Chen, 2017). For instance, comparing the number of nodes and edges of different graphs of the same type (Trilcke et al., 2016) can be a ranking tool that is directly translatable into natural language. In addition to that, studies suggest that density (the number of edges in relation to the number of nodes) is relevant to analyse character networks, especially when compared within a homogeneous collection (Evalyn and Gauch, 2018; Grandjean, 2015). This is also the case when measuring average path length (Trilcke et al., 2016).

Fig. 2 Various layouts do not follow a force-driven algorithm to make non-relational dimensions of the data explicit.
Local properties
With regard to local measures, the **degree** (number of neighbouring nodes) is the simplest centrality, and the only one systematically used between the late 1950s and early 1970s, before the development of more diversified metrics (Freeman, 1979). Its simplicity allows for a transparent translation: in a literary network, for example, it counts the number of times one character speaks to another (Jannidis et al., 2016).

The notion of **betweenness centrality** disrupts the conception of what the “centre” of a network may consist of. Its ability to reveal structural elements bridging large, immediately visible clusters makes it popular in the social sciences since the emergence of Granovetter’s concept of “weak ties” (Granovetter, 1973). Betweenness is very closely linked to the notion of circulation: it counts the shortest paths to detect intermediate “bridges” or “key passages” capable of opening or locking certain parts of the network (Tayler and Neugebauer, 2018). Depending on applications, these are therefore both positions of power and vulnerable places.

The **closeness centrality** allows to highlight the “geographical” middle of the graph. In networks of a certain density and when they are not divided into several distinct communities, the closeness is generally fairly evenly distributed and allows a good translation of the notions of “center” and “periphery”.

For its part, the **eigenvector centrality** is quite complicated to translate since it works iteratively and is very much dependent on the structural context at short and medium range around a node. “Prestige” or “influence” centrality, named “power” centrality by its author (Bonacich, 1972),

---

**Fig. 3** Three levels of interpretation that can be articulated: visual analysis (examples top left), use of global metrics (examples bottom right) and use of local metrics (highlighted nodes).
it qualifies a node’s environment while operating in cascade: a well-connected node gives its neighbours a part of its authority capital, and so on. It is therefore particularly useful when trying to analyse the hierarchy of the nodes in a graph (Piper et al., 2017). The most well-known use of this measure is the backbone of the Google search engine: the PageRank algorithm (Brin and Page, 1998).

Towards mixed approaches

This contribution proposes a table of correspondence between the concepts of graph theory and the practice of visual network analysis in the social science and humanities. This effort must not be understood as a demarcationist attempt at telling the right method from the wrong. The “dictionary” is not exhaustive and only aims at helping to bridge two worlds that have more in common that what meets the eye. By focusing on translating methods, we want to stress that crossing points are real even though they do not come without issues, and thus require our methodological attention.

We also note that the analysis should not be limited to a catalogue of well-known methods (basic centralities, etc.) but that approaches combining several of those should be encouraged to obtain an optimal and innovative “translation”. In this way, we could compare metrics (Escobar and Schauf, 2018) or combine them to establish rankings (Fischer et al., 2018; Grandjean, 2018: 328). Furthermore, the enrichment of the networks by means of categories that are not dependent on the structure, like the gender of individuals in a social network (Dunst and Hartel, 2017) or the discipline of projects in a scientometric analysis (Grandjean et al., 2017), allows to test translation and interpretation hypotheses by avoiding the blind approach of testing all possible graph metrics.

References


Euler L. (1736). Solutio Problematis ad Geometriam Situs, Opera Omnia, 7, 128-140.


# Network visual and topological patterns

This table of correspondence between network analysis concepts and interpretations or "translations" is a work in progress. The authors propose this document to open a discussion on the most relevant translation scenarios/examples/references from the different disciplines applying these methods.

<table>
<thead>
<tr>
<th>NOTION</th>
<th>VISUAL ANALYSIS</th>
<th>COMPUTATIONAL ANALYSIS</th>
<th>INTERPRETATIVE POTENTIAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Graph size (nodes)</strong></td>
<td><strong>SIMPLE DEFINITION</strong></td>
<td>How many nodes are there in the graph?</td>
<td><strong>HERMENEUTIC CRITERIA</strong></td>
</tr>
<tr>
<td><strong>Graph size (edges)</strong></td>
<td><strong>SIMPLE DEFINITION</strong></td>
<td>How many edges are there in the graph?</td>
<td><strong>HERMENEUTIC CRITERIA</strong></td>
</tr>
<tr>
<td><strong>Density</strong></td>
<td><strong>SIMPLE DEFINITION</strong></td>
<td>How connected are the nodes overall?</td>
<td><strong>HERMENEUTIC CRITERIA</strong></td>
</tr>
<tr>
<td><strong>Diameter</strong></td>
<td><strong>SIMPLE DEFINITION</strong></td>
<td>How far are the most distant nodes?</td>
<td><strong>HERMENEUTIC CRITERIA</strong></td>
</tr>
<tr>
<td><strong>Average path length</strong></td>
<td><strong>SIMPLE DEFINITION</strong></td>
<td>On average, how close are nodes to each other?</td>
<td><strong>HERMENEUTIC CRITERIA</strong></td>
</tr>
</tbody>
</table>
### NOTION | VISUAL ANALYSIS | COMPUTATIONAL ANALYSIS | INTERPRETATIVE POTENTIAL
--- | --- | --- | ---
**Connectedness**
Is the graph a connected system where there is a path between every node?
- **GLOBAL**
  - Shortest paths: Removing that node disconnects the graph.
- **SIMPLE DEFINITION**
  - Being a bridge, connecting otherwise disconnected groups of nodes.

**Clusters/Communities**

- **GLOBAL**
  - Number of potential clusters.
  - Number of closed triplets.
  - Modularity.

- **SIMPLE DEFINITION**
  - Clusters (unclear distribution of nodes).

**Global or average clustering coefficient**

- **GLOBAL**
  - Average clustering coefficient: Average of the local clustering coefficient.

- **SIMPLE DEFINITION**
  - Triangles are easy to count visually in a small network, but the ratio between this result and the total number of potential triangles is almost impossible to calculate directly.

**Connectivity (degree)**

- **GLOBAL**
  - Average degree: Average of out-degree and in-degree.

- **SIMPLE DEFINITION**
  - Counting the edges converging to that node.

**Betweenness**

- **GLOBAL**
  - The notion of betweenness is more complicated for a directed graph than for an undirected graph.

- **SIMPLE DEFINITION**
  - Being a bridge, connecting otherwise separated groups of nodes.

**Visual pattern**

- **GLOBAL**
  - shortest paths: Removing that node
  - clusters

- **SIMPLE DEFINITION**
  - looking at empty areas (structural holes).

**Multilevel criterion**

- **GLOBAL**
  - looking for groups of nodes, as visually dense and separated as possible.

**Multilevel criterion**

- **GLOBAL**
  - Looking for clusters.

**Hermeneutic criteria**

- **GLOBAL**
  - The network is a continent, or, on the contrary, an archipelago.

**Visual pattern**

- **GLOBAL**
  - many links it has / how many neighbors

**Multilevel criterion**

- **GLOBAL**
  - connectedness is more remarkable if they contain many nodes.

**Hermeneutic criteria**

- **GLOBAL**
  - Risk is marginal.

**computational pattern**

- **GLOBAL**
  - Useful for exploration. It is tempting to take the result of a cluster calculation as a given. In some cases, it is interesting to compare these clusters with previously known groups (categories that do not depend on the structure obtained).

**Hermeneutic criteria**

- **GLOBAL**
  - The term cluster has become part of the common language, but we also like to talk about groups, communities or hubs. This notion of community is very directly related to the way in which the social sciences and humanities use the metaphor of the "network".

**Computational pattern**

- **GLOBAL**
  - Modularity is a measure of a graph partitioning, so it is necessary to partition the graph first.

**Hermeneutic criteria**

- **GLOBAL**
  - The network is a continent, or, on the contrary, an archipelago.

**Computational pattern**

- **GLOBAL**
  - In a directed network, we also distinguish indegree (inbound links) and outdegree (outbound links).

**Hermeneutic criteria**

- **GLOBAL**
  - The network is a continent, or, on the contrary, an archipelago.

**Computational pattern**

- **GLOBAL**
  - Modern algorithms such as Leiden algorithm may often be used even for directed graphs.

**Hermeneutic criteria**

- **GLOBAL**
  - The network is a continent, or, on the contrary, an archipelago.

**Computational pattern**

- **GLOBAL**
  - In a directed network, we also distinguish indegree (inbound links) and outdegree (outbound links).

**Hermeneutic criteria**

- **GLOBAL**
  - The network is a continent, or, on the contrary, an archipelago.

**Computational pattern**

- **GLOBAL**
  - The notion of bridge is often a good proxy for authority/influence.

**Hermeneutic criteria**

- **GLOBAL**
  - The network is a continent, or, on the contrary, an archipelago.

**Computational pattern**

- **GLOBAL**
  - The notion of bridge is often a good proxy for authority/influence.

**Hermeneutic criteria**

- **GLOBAL**
  - The network is a continent, or, on the contrary, an archipelago.
<table>
<thead>
<tr>
<th>NOTION</th>
<th>VISUAL ANALYSIS</th>
<th>COMPUTATIONAL ANALYSIS</th>
<th>INTERPRETATIVE POTENTIAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Closeness (eigenvector)</strong></td>
<td>Proximity to well connected nodes (often inside a cluster).</td>
<td>Clustering coefficient is generally hard to see and visual interpretation is considered unreliable. Exceptions are small networks, nodes that have only a few neighbors that we see well, and nodes that are only connected to a very dense cluster.</td>
<td>This form of centrality is not only adapted to directed networks, and can be related to the functioning of a search engine (the Page Rank principle was used in the first Google search engine) or a system where information flows.</td>
</tr>
<tr>
<td><strong>Local clustering coefficient</strong></td>
<td>Are the neighbors of a node also connected together?</td>
<td>Clustering coefficient refers to social groups of individuals who know each other. They exist.</td>
<td>Excellent to describe the center or the middle of a network, especially when the latter is described in topographical terms. Low values of this metric are very appropriate for the use of concepts which are the opposite of the center: the periphery, the margins, etc.</td>
</tr>
<tr>
<td><strong>Shortest path</strong></td>
<td>Two nodes are connected by a path</td>
<td>Requires that we can follow the links in practice, which is possible only for small (undirected) networks and depends on the graphic settings. Finding a path can be difficult, and ensuring that the path is the shortest can be too difficult. However the visual distance is a loose approximation of the shortest path length.</td>
<td>The iterative nature of this notion makes it difficult to translate (and difficult to use in some contexts). It is confused with the notions of prestige, authority, influence and, sometimes, power and elites. This measure distinguishes nodes that are “well” connected (and not just &quot;a lot&quot;). It opposes the notion of assortativity.</td>
</tr>
<tr>
<td><strong>Cliquences</strong></td>
<td>Groups of nodes where all possible edges exist between them.</td>
<td>A group of nodes has a density of 1. Clique detection algorithms exist.</td>
<td>The number of cliques, their size and distribution are metrics that are complementary to the clustering coefficients (local and global). They can be used as a more strict community detection algorithms.</td>
</tr>
</tbody>
</table>

**SIMPLE DEFINITION**

- In the middle of the network.
- Being in the middle of the network.
- Being well connected nodes without necessarily having a large number of neighbors itself.
- Are the neighbors of a node also connected together?
- Two nodes are connected by a path.
- Groups of nodes where all possible edges exist between them.

**VISUAL PATTERN**

- The geographical center of the graph.
- The visual estimation of centrality is considered acceptable, but it remains an evaluation. It is harder to find in very sparse graphs.
- The score of closeness centrality is the average length of the geodesic distances to all the other nodes.
- Nodes are inside a cluster.
- Following the series of edges from one node to another to find the shortest.
- Very dense clusters.

**COMPUTATIONAL CRITERIA**

- Finding the center (the center of the "hard masses") of the graph.
- The average path length is not always the shortest distance. The steps between them and that the perception distances is not always the shortest path.
- This notion measures the location of bridges.
- Clique (group of nodes with density of 1).

**TOPOLOGICAL PATTERN**

- Being connected to well connected nodes being well connected nodes.
- Nodes with density of 1.
- A group of nodes has a density of 1. Clique detection algorithms exist.
- Clique (group of nodes with density of 1).

**COMPUTATIONAL ANALYSIS**

- Clique detection (following the notion of a bridge).
- Eigenvector centrality or Page Rank.
- Shortest path(s).
- Geodesic distance, algorithms for shortest path detection.

**LANGUAGE**

- Meets a notion of redundancy in the local connections, comparable to centrality but at a very local scale. Tells if a node is in a clustered environment. Complex networks are often characterized by a high average clustering coefficient.
- The iterative nature of this notion makes it difficult to translate (and difficult to use in some contexts). It is confused with the notions of prestige, authority, influence and, sometimes, power and elites. This measure distinguishes nodes that are “well” connected (and not just "a lot"). It opposes the notion of assortativity.

- Excellent to describe the center or the middle of a network, especially when the latter is described in topographical terms. Low values of this metric are very appropriate for the use of concepts which are the opposite of the center: the periphery, the margins, etc.

- The number of cliques, their size and distribution are metrics that are complementary to the clustering coefficients (local and global). They can be used as a more strict community detection algorithms.
- The term clique itself refers to social groups of individuals who know each other. They can be translated as communities, neighborhoods, closed societies.