Network Analysis

Inspiration:

- https://introdh2022.commons.msu.edu/networks-9-28/
- https://ieeexplore.ieee.org/document/9373218
- https://introdh2023.commons.msu.edu/networks-9-22/
- https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book.pdf
- https://miriamposner.com/classes/dh101f16/tutorials-guides/data-visualization/network-analysis/
- https://evelinag.com/blog/2016/01-25-social-network-force-awakens/#.VwAKUxMrJE5
- https://www.nypl.org/blog/2014/07/31/networked-catalog
- https://kindred.stanford.edu/#
- Mazanexc article
- https://github.com/miriamposner/network_analysis_workshop/blob/master/Creating the dataset for this workshop.md
- https://www.martingrandjean.ch/digital-humanities-on-twitter/

Sturcture

- Portraying winston curchill

portraying a person

Painters, photographers, and illustrators know that portraying someone is not an easy task. It requires a mix of abilities such as taste, delicacy, and technical ability. Even when these qualities excel, it might happen that the subject does not agree with the representation.

as it was the case for Winston Churchill.

Last century **Graham Sutherland** portrayed Churchill at the age of eighty. The artwork to celebrate Churchill's 80th anniversary was commissioned by the House of Commons and House of Lord, which named the most famous English painter of that time. However, considering Churchill's temper, it was not a surprise that something went wrong. His disappointment was so bitter that the portrait disappeared in his country

house [10] and, years later, it seems that the painting was destroyed by his wife Lady Spencer-Churchill.

This story shows the complexity of representing personal identity, which is also true in contemporary media. With the recent, massive quantification operated by personal devices [11], large datasets bring into question the way in which individuals are represented, as well as the perception of their own self [12], Severe ethical issues relate to the problem of self-representation, as it intimately concerns all human beings. Nevertheless, although the scientific debate on personal data is going on [13], the discussion about the ethic of visualizing individual information is still in an embryonic state.

a connected world

[slide-1]

In the days following the 9/11 terror attack, an FBI agent visited the Whitney Museum of American Art to see **Mark Lombardi's 1996 drawing 'BCCI-ICIC & FAB, 1972–91 (4th version)'** (Hobbs, 2003:11–12, 95–8). The web-like image comprises a meticulously researched diagram of individuals and groups with ties to a money-laundering organisation that operated under the name of the **Bank of Commerce and Credit International (BCCI)**, which included Osama bin Laden and others associated with al-Qaeda. As Lombardi himself described, BCCI 'was used not only by drug dealers and con men but also by the governments of the US, UK, Saudi Arabia and the Gulf Arab states to funnel support to Afghan guerrillas fighting Soviet occupation, to pay off friends and adversaries alike and conduct secret arms sales to Iran' (Lombardi, 2001). In other words, in black and red ink, Lombardi traced a terrorist network that reached the centres of government.

[slide-2]

He had grasped the power of the **network perspective to reveal conspiracy**, adapting graphical traditions associated with the study of social networks developed in the first half of the twentieth century.

Tragically, **the significance of his research would be uncovered only after his death**: in early 2002, a year after the artist's suicide, the FBI's Operation Green Quest raided the offices prominent and powerful people, featured in Lombardi's work (Goldstone, 2015).

[slide-3]

Three years after Lombardi produced his artwork, Albert-László Barabási and Réka Albert published a scientific article entitled **'Emergence of Scaling in Random Networks' (Barabási & Albert, 1999)**. It argued that a wide variety of seemingly heterogeneous networks, such as *power grids*, *social networks*, and the *World Wide Web*, exhibit **nearly identical distributions of connectivity**, and it offered an elegant model that explained how these distributions might arise. This particular distribution of connectivity was different from those most scientists expected at the time.

The significance of Barabási and Albert's findings was that they provided a compelling case for analysing seemingly disparate systems and kinds of data using the same mathematical models and tools. For this reason, their article is regarded as one of the **founding publications** of the **interdisciplinary field of modern network science**.

The argument for the application of analytical tools across domains was extended in Barabási's best-selling book *Linked*, in which he argued that many challenges in our world, such as managing the spread of epidemics, fighting terrorism, and handling economic crises, can be cracked by understanding these systems as networks (Barabási, 2002). As such, networks appear in Barabási's study as *a kind of Rosetta Stone*. This message reached 70,000 readers and thus played a small part in the rise of the 'network' perspective in the modern consciousness following the new millennium.

the network turn

[slide-4]

Lombardi and Barabási's work is part of what we call the **'network turn'**. This turn cannot be attributed to either the artist or the scientist; they are but two examples of a whole host of converging thoughts and practices around the *turn of the new millennium — the zeitgeist of the networked age*.

The World Wide Web had become available to the public only in 1991, but by 2004, the web-based view of relations manifested itself in an entirely new kind of communication platform when 'The Facebook' was launched. The subsequent **proliferation of social networking platforms** has profoundly shaped the way we understand connectivity in the world today.

[slide-5]

- I would like to argue that **arts and humanities scholars can use the same kind of visual and quantitative analysis of networks to shed light on the study of culture**;

while at the same time:

the **critical skills native to humanistic inquiry are vital to the theorisation and critique of our networked world**.

[slide-6]

Network analysis, as we define it in this book, is a set of practices and discourses that sit at the interface of the natural sciences, humanities, social sciences, computer science, and design. [show-venn-diagram]

We contend that networks are a category of study that **cuts across traditional academic boundaries** and that has the potential to unite diverse disciplines through a shared understanding of complexity in our world – whether that complexity pertains to the nature of the interactions of proteins in gene-regulatory networks or to the network of textual variants that can reveal the lineage of a poem.

[slide-7]

Moreover, this shared framework provides a compelling case for **collaboration across those boundaries**, for bringing together computational tools for quantitative network analysis, together with theories, discourses, and applied techniques from the social sciences, the humanities, visual design, and art practice.

[slide-8-differences]

The cases of Lombardi and Barabási provide an instructive way of grasping that **shared framework** because, superficially, their work has very little in common.

Barabási and Albert explicitly cite the *computerisation* of data acquisition as essential to their research. Their research process was *digital* and *algorithmic*.

- By contrast, **Lombardi**'s research process was *analogue*. He gathered his data on three-by-five *notecards*. There is no evidence that Lombardi read Barabási and Albert's groundbreaking work in

statistics and physics; rather, his inspiration was *panorama and history painting*. He used the term *'narrative structures'* to describe his handdrawn webs of connection. Produced through an iterative process of refinement, the work is human in scale, legible visually in its entirety. Perhaps more importantly, it is his interpretation of a *carefully researched* but inevitably *incomplete* record. It *does not pretend to objectivity*.

- In stark contrast, **Barabási and Albert**'s method is scientific: it proposes a *model* to predict the behaviour of systems and to understand complex topologies 'independent of the system and the identity of its constituents' (Barabási & Albert, 1999). They used the term 'scale-free networks' to describe the distribution of connectivity they observed. They used *mathematical models* to predict the behaviour of systems and to understand complex topologies. They produced results that are *invisible* in their entirety, only legible through the lens of statistical analysis. It pretends to *objectivity*.

Thus, where Lombardi is analysing past events, Barabási and Albert offer a *predictive* model; where Lombardi is visual, Barabási and Albert use *algorithms* designed to detect patterns in data sets too large or complex for the human eye to detect. These approaches seem to occupy two very separate worlds.

[slide-9-similaries]

Nevertheless, Lombardi's art and the scientific approaches of Barabási and Albert have **much in common**.

- Lombardi distils the composition of relationships in history painting and **the comprehensive 'at one view' of the panorama** into a formal abstraction rooted in the conceptual art movement of the mid-twentieth century, and reflects the overlapping concerns, discourses, and methods of art and science.

The artist and scientists use connectivity to make sense out of data: a representation of knowledge that relies on abstraction.

Both produce **results that are seductive in their elegance and simplicity**.

Networks are by definition an abstraction into a system of nodes and edges.

```
- Nodes are entities;
- edges are the relationships between them.
```

Two examples can be seen in Figure 2. **Such an abstract system is inherently intuitive**. These two elements, nodes and edges, are the simple building blocks of **an obviously abbreviated rendering, a malleable geometry that can range in complexity from a direct and declarative schematic to a dense, indecipherable web of connections.** [image of example of complicated network vs. simple network]

[slide-10-networks-in-arts-and-humanities]

Around 2010 there has been a slow but steady increase in scholars in the **arts and humanities employing network visualisation**, **social network analysis theory**, and **quantitative measures from network science** to address their research questions, e.g. to the study of cultural artefact.

The case for scholars from the arts and humanities engaging with networks is compelling on a number of levels.

```
- The use of computational network analysis can lead to

- It makes it possible, with relative ease and speed, to

- It has proven to be an effective tool for understandir

- The ability to analyse data across scales has been rer

- Networks further offer the ability to contextualise th
```

[slide-11]

Moreover, and perhaps most importantly, scholars from the arts and humanities **already have the conceptual framework** to make this leap: they have been writing about networks for centuries, albeit from **the metaphorical perspective**:

- examining communities of practitioners (artists, writers, and thinkers)
- the dissemination of ideas
- the relationships between certain texts, images, or artefacts

Although researchers with standard humanities training will likely need to acquire some new skills to engage with the computational challenges of network visualisation and quantitative analysis, I am convinced that **you already have a set of skills that are key** to the development of the interdisciplinary practice of network analysis.

I don't want to call for arts and humanities scholars to accept unquestioningly frameworks and methods developed in the field of network science. Rather, it argues that the discourse and analysis of networks can move forward only through collaboration and exchange at the interface of computational method, humanistic inquiry, and design practice.

Networks are always metaphorical

[slide-13]

Again, the differences...

In **network science**, researchers utilise networks as a formalised abstraction that permits computational analysis.

In the **humanities**, by comparison, scholars largely employ networks as a metaphor.

Despite these methodological differences, there are important **continuities** between **the act of abstraction/modeling (computer science)** and **use of metaphor**.

George Lakoff and Mark Johnson argue that **metaphors are not just linguistic embellishment**, but rather provide a **conceptual framework that structures our most basic understandings of the world** (Lakoff & Johnson, 1980).

To see the network as a metaphor, by contrast, we have to be fully aware of the process of abstraction: what information it prioritises and what the abstraction elides.

- In what follows, I want to make a case for **anyone who employs networks as a metaphor** (traditionally those in the arts and humanities) they **ought to be familiar with the mathematical formalisations**.

- Conversely, scholars wedded to the computational power of **quantitative network analysis** should be **aware that its power derives from its reliance on the metaphorical dimension** and an **act of interpretation**.

[slide-14]

In 1929, the popular Hungarian author **Frigyes Karinthy** wrote 'Láncszemek' ('Chain-Links'), a short story musing on the shrinking social world during a period of rich international trade in Hungary. In the novel, Karinthy's characters create a game:

```
 *One of us suggested performing the following experiment
```

[slide-15]

Karinthy's game became reality in **Stanley Milgram**'s research into the **'small world' phenomenon** just over three decades later. Building on his earlier work with the mathematician Manfred Kochen and the political scientist Ithiel de Sola Pool, Milgram undertook a **series of experiments** that sought to determine the **degrees of separation between people in realworld networks**, which he reported in a 1967 issue of the popular magazine *Psychology Today*.

Milgram invited members of the public to forward a parcel to close acquaintances in their immediate social (but not necessarily geographical) neighbourhood with the goal of eventually reaching a particular individual on the other side of the country.

Although the methods and findings have since been **disputed**, Milgram claimed his study showed that *'only five intermediaries will, on average, suffice to link any two randomly chosen individuals, no matter where they happen to live in the United States'* (Milgram, 1967: 66).

Milgram's article generated enormous publicity, thereby connecting in a public consciousness the concept of the network and that of the small world.

[slide-16]

More recently the findings have been popularised as the theory of **'six degrees of separation'**.

Conceptually this phrase makes sense **only if you view the world in terms of the network**. The network, then, is a pre-existing framework upon which the concept of six degrees of separation is drawn.

The accessibility of the concept of six degrees of separation is shown by the way it has been seized on in popular culture. The parlour game **'Six degrees of Kevin Bacon'** challenges players to find the shortest path between a given actor or actress and prolific actor Kevin Bacon (which is in turn referenced by the digital project Six Degrees of Francis Bacon):

```
Six Degrees of Kevin Bacon or Bacon's Law is a parlor g

https://oracleofbacon.org/movielinks.php
```

[slide-17]

That sense of wonder, however, is arguably a residue of the analogue era. By contrast, from the late 1990s onwards the rapid growth of both the Internet and computational processing power has made it possible to gather and analyse network data on an unprecedented scale.

Now in a few lines of code we can construct and measure networks of various kinds and extract information about their global structure: **how big** they were, **how densely clustered**, and in the case of the small world phenomenon, **how many degrees** on average any randomly selected node was from any other node in a network.

[slide-18]

In the foundational 1998 publication *'Collective Dynamics of Small World Networks'*, Duncan J. Watts and Steven Strogatz showed how small world properties are not limited to social networks...

- the neural network of the worm Caenorhabditis elegans,
- the power grid of the western United States
- and the collaboration graph of film actors

... are all small world networks (Watts & Strogatz, 1998).

but why?

[slide-19]

But why would scholars from the arts and humanities bother themselves with the task of theorising and operating within this network turn?

Just because we can challenge the divisions between intellectual and conceptual domains with network thinking, it does not follow that we should.

- Perhaps our motivation to think through networks should be that it allows us to see features in history, language, literature, or art that would otherwise remain inscrutable.

- Perhaps we yearn for the ability to recognise patterns across disparate domains as simply kaleidoscoped manifestations of the startlingly predictable human condition. P

- Perhaps the seamless connection between part and whole, the ability to connect the microhistory with the longue durée, explains the allure of this siren's song.

DH as a network

IV. During the Digital Humanities Conference in 2014, when speakers were invited to recognize themselves in a network visualization representing themselves, reactions were different. Someone was offended for not being part of the visualization and someone was pleased to be there, sharing portraits on social media [14]. The ENAC Research Day caused the same reactions in a similar setting: when faculty members were invited to recognize themselves in a data visualization, responses ranged from engagement to indifference.

Although most of the complaints were related to technical problems easy to solve, something else was not completely focused. After a period of reflection, it was evident that the reactions originated from self-recognition were influenced by network centrality. Different interviews clearly indicated that the satisfaction was proportional to the closeness to the center: the more a person was placed in the middle of the network, the more the self-perception was flattering.

The origin of centrality stays in the concept of being the "star" of a social group, which is being popular and worthy of attention. But in network analysis, centrality also determines the node's relevancy [15]. Such a measure can be mathematically calculated but also diagrammatically interpreted, as in networks it is rather easy to see whether a node is central or not.

Grandjean, *A social network analysis of Twitter: Mapping the Digital Humanities community* (2016)

excercise

Kindred Britain, chinese poems, etc. Pick one that interests you. play with .... what How does it play with network as a metaphor? How does it play with network as a computational tool? Determine what information you can reasonably extract from this graph. Now toggle between modes. Does this change your understanding?

- **Kindred Britain**: https://kindred.stanford.edu/ This project shows the relationships between people in Britain from 1200 to 1900. It is a great example of a network visualization that is both beautiful and informative. The project is a collaboration between the Stanford Literary Lab and the King's College London Department of Digital Humanities.

- Linked Jazz: https://linkedjazz.org/network/ This project uses network analysis to explore relationships between jazz musicians.

- Another interesting example looks at the genre of "exchange poems" that were part of medieval Chinese culture. These had traditionally been characterized by schools and styles. But new research positioned them in social networks. To paraphrase the work of the project director, Tom Mazanec, it turns out that the Buddhist monks in the 7th to 10th centuries of the Tang dynasty were central "nodes" in the network of literary production (Mazanec 2017). Graphing these has changed the way this form of Chinese poetry is understood and its place in cultural and social life. Relations between literary forms and social activity that were not noted before were revealed through the analysis.

Art historians Pamela Fletcher and Anne Helmreich used network analysis to study the London art market, and found surprising insights from sales records and auction catalogs (Fletcher and Helmreich 2012). Artists and styles that have not necessarily been seen as important by later art historians turned out to play a significant role in the markets of the time, even if they have largely vanished from the canon.

glossary

Adapted from

Brian V. Carolan, "Key Terms," Social Network Analysis and Education: Theory, Methods & Applications (SAGE, 2014, http://www.sagepub.com/carolan/study/materials/KeyTerms.pdf);

Datavu, "Introduction to Network Analysis terminology" (http://datavu.blogspot.com/2013/10/sna-social-network-analysis-basic.html);

Katharina Zweig, "An Introductory Course on Network Analysis" (https://sites.google.com/site/networkanalysisacourse/schedule/an-introduction-to-centrality-measures).

Social network

A finite set (or sets) of actors and the relations defined on them. It consists of three elements: (1) a set of actors; (2) each actor has a set of individual attributes; and (3) a set of ties that defines at least one relation among actors.

Graph

critical approach to network diagrams is useful. Technically, networks are graphs, not visualizations. The distinction is important because graphs can include the feature of directed or undirected connections.

A common way to visually represent social networks, consisting of two dimensions: actors and relations (also called nodes and edges).

The computational process by which graphs are produced requires that data be structured in a specific way: source > relationship > target. The vocabulary of nodes and edges is used to differentiate entities (source and target) from their relationships (edges).

Node

Nodes are the entities in graph (also called vectors). For example, if we consider Facebook friends as a graph, then every friend is a node.

Edge

These are the relationships between nodes. For example, if we consider Facebook friends as a graph then every friendship is an edge.

Types of Graphs

Undirected graph

When the relationship is always valid in both directions, then it is called undirected graph. If Dave is friends with Raj on Facebook, then Raj is also friends with Dave.

Directed graph

When the relationship may not be valid in both directions (connecting nodes), then it is called a directed graph. If Bill is following Steve on Twitter and Steve is not following Bill, the relationship is directed.

Image via http://www.codediesel.com/wp-content/uploads/2012/02/d-graph1.gif

Weighted network

A network in which the edges between nodes have weights (corresponding to, e.g., the strength of a relationship) assigned to them.

Unweighted network

A network in which the edges between nodes do not have weights assigned to them.

Image via http://blogs.sas.com/content/sascom/files/2011/10/weighted-network2.jpg

Single-mode graph

A type of graph in which all nodes belong to the same category. For example, in a graph of Facebook friends, each node is a person.

Multimode graph

A type of graph in which all nodes are not of same type. For example, a graph that includes both "buyers" and "sellers" is a multimode (or two-mode, or bimodal) graph.

Ego network

When you perform ego network analysis, you select a focal node (an "ego") and determine its connections to other nodes (which are called "alters"). Each ego is treated as its own case.

Image via

Measures of Graphs or Nodes

Size

A measure of the number of actors (nodes) in a complete or egocentric network.

Density

The number of ties in the network reported as a fraction of the total possible number of ties.

Reciprocity

The proportion of mutual ties in a network.

Distance

The number of "steps" between any two actors in a network.

Clustering coefficient

A measure of a network's actors' tendency to "group together" into pockets of dense connectivity.

In the first image, the blue node has a clustering coefficient of one, because all possible connections among its neighboring nodes have been realized. In the second image, only one of the possible connections has been realized — so the blue node has a clustering coefficient of 1/3. In the third image, none of the neighboring nodes are connected, so the blue node has a clustering coefficient of 0. Image via Wikipedia.

Centrality

There are multiple ways to determine a node's importance, or centrality. The measure you use depends on how you define centrality. Several of these measures are:

Degree centrality

An important node is involved in large number of interactions. The number of edges connected with a particular node.

Eigenvector centrality

An important node is connected to important neighbors. This is a measure of influence of a given node in the whole network. The notion is how well-connected a given node is with other well connected nodes in the network. This is how, for example, Google determines page rank.

Betweenness centrality

An important node lies on a high proportion of paths between other nodes in the network. Model based on communication flow. A person who lies on communication paths can control communication flow, and is thus important.

Closeness centrality

An important node is typically "close" to, and can communicate quickly with, the other nodes in the network. Length of the average shortest path between a given node and all other nodes in a graph.

tutorial harry potter and gephi, larger hairball fanfic

The following section heavily borrows from the standard Gephi tutorial that analyses the interaction between the characters of "Les Misérables". I am using version 0.9.2., so the screenshots and instructions will only pertain to this specific version. Please note that some of the procedures below use random components, meaning that your own visualizationd and results will probably deviate slightly from the ones I show here (e.g. the graph may be rotated). That is fine as long as the overall structure of the graph is similar.

In this section, we will work our way to a fairly standard visualization procedure that typically yields a stable result for different datasets, regardless of their size. Throughout the procedure discussed below, it is important to get an intuitive sense of the effect of certain "knobs" and parameters on the visualization. As with everything, practicing this sort of skill is important. First of all, open Gephi through double-clicking the Gephi icon (which can be found in whatever application folder you installed it.)

The program will launch (which might take a while, especially the first time).

Open our first network file (`uk1.gefx`): `File > Open....` Locate the file using the navigator that pops us. This will upon a small summary screen, where some general properties of the graph data and any errors will be reported. You can click `OK`. By default, we will land in the `Overview` pane (as opposed to the `Data Laboratory` and `Preview` pane which we will cover shortly). This where we can directly manipulate and design our graph and run calculations on it.

In the central `Graph` window you will see a representation of our graph data, that has been plotted in a random fashion. The first thing we will want to do is create a **layout** for your graph. Locate the `Layout` panel and select the `Force Atlas 2` layout algorithm. Click `Run`.

You will see that Gephi now uses an algorithm from Physics that tries to determine a sensible visualization for our graph as if it were a planetary system that is structured through the laws of gravity: the edge weights are taken to represent the powers of attraction between two nodes or planets. Gradually, you will see that the forces of attraction converge and stabilize: hit `Stop`. You can now use your mouse to navigate the graph structure, via zooming in and out on different location. (This is hard to explain in words, but if you play around a little with this navigation you will see that you get a hang of it.)

If you are completely lost, you can always click the magnification glass in the lower left to re-centre the graph. If you hover with your mouse over the individual nodes you will see that the network browser will highlight the corresponding node and some of its main neighbors which helps you inspect the data.

The first thing we will do is resize the nodes depending on their degree, so that the size of nodes reflects the number of edges of a node: more important nodes will be relatively larger. Visit the `Appearance` panel in the top left; then, select `Nodes, Ranking` and then the icon size (three circles). Select degree and click `Apply`. Now will now see that the node visualization shows a node's degree. In the `Appearance` pane you can control the size of the smallest and largest nodes (and even the spline between them).

If we now select the color palette tab (which targets the color, instead of the size of the nodes), we can also vary the color intensity of the nodes, depending on their degree. Don't forget to click `Apply` again.

As you can see, our network is taking shape.

However, as you can see, the network is very dense and nodes overlap. To prevent this, we can return to the Layout pane, and run `Force Atlas 2` again, however, this time with the option "Prevent Overlap" selected. Don't forget to hit `Stop` in due time. This will give more space around each node to unclutter the graph.

Now that our network is taking shape, it becomes interesting to find out to which characters some of these nodes actually correspond. For this we will add `text labels` to each node. In the lower bar of the graph canvas, click the dark T to add the ID of each node as a text label and under the black A, you should make the label size proportional to the node size. The font will probably be too large initially, but you can adjust this using the slider to the right. (The steps are visualized here below.)

The result should look a little like this:

Let us now calculate some more advanced statistics. Locate the `Statistics` panel and click the `Run` button next to the `Average Path Length`, click `OK` to use the default settings: this step will calculate, amongst others, the average path lengths for us. When the procedure is done it will produce a small report.

After you have closed the report below, we can visit the `Appearance pane` again. If we try to change the size again, you will notice that the dropdown now has a new criterion available for us: i.e. the betweenness centrality that we discussed previously. Select this option to make the `node` size reflect their betweenness centrality (instead of their simple degree). Don't forget to click `Apply`.

Finally, we can run a last, very useful procedure that is called **Community detection**. If you click the `Run` button next to the `Modularity` tab in the `Statistics` pane (click `OK` to use the default settings, which are typically fine). This procedure will use the so-called "Louvain method" to detect meaningful groups of nodes in the networks that can be considered members of the same **subcommunities** in the network in terms of the kinds of close neighbors they have. This feature is especially useful when you are dealing with large datasets and you would like Gephi to identify interesting subcomponents of the network. Again, you will get a report that also plots the size of the communities detected:

To actually show the networks in your graph, we return to the `Appearance` pane: in the dropdown for the styling of the node color (i.e. the color

palette icon) we now select `Partition`. In the dropdown, you will now find the `Modularity class` (that handily comes with a default color scheme):

After clicking `Apply`, you will now see the various communities which the method detected show up in different colors:

Interestingly, at the community level, a number of observations can already be made on the basis of this small toy network:

- The protagonists tend to form a tight community in the middle of the graph: these are mainly children, with the interesting exception of Hagrid - who seems to fulfill a bridge function between the children and adults, but also between the world of Hogwarts and that of the Muggles;
- The Weasley family (except Ron) are included in the same cluster.;
- The Dudleys (and other Muggles, such as the newsreader and wheatherman) form a cluster of their own. Note that the latter two occupy a highly **peripheral position in the network**, indicating that they do not take up a central role;
- The teachers at Hogwarts form a clusters of their own, together with e.g. the hat.

> ## Exercise
>
> 1. Run the `PageRank` algorithm in the `Statistics` pane and change the node (and text) size to reflect the PageRank score of each node.
> 2. The network contains a small, yet distinguished cluster of 'friar', 'seamus' and 'nick'. Does this cluster make sense? Why?

There are two additional visualization **filters** - small "hacks" really - that you might find useful to inspect your networks. They won't change the structure of your network as such, but they can be useful to focus on specific target areas of the network. The first filter we will discuss allows you to hide less well-connected (and thus potentially less important) nodes in the networks based on a simple frequency-based filtering criterion. Navigate to the `Filters` panel next to the `Statistics` panel. There, locate the `Degree Range` button in the `Topology` folder. Drag the `Degree Range` to the `Queries` panel underneath:

Through adapting the lower threshold (e.g. to 5, as above), we can now filter out nodes that have a degree of less than 5. They will unclutter the

visualization and remove these nodes from the visualization (but *not* from the underlying graph data, of course):

Another useful filter are so-called **ego networks**: this will limit the visualization to a single node and its direct neighbors (this is in fact similar to what happens when you hover over a particular node in the graph canvas). This is useful when you work with very large networks and are interested in the behaviour of a single node, but you have difficulties locating it in the giant hairball. To do this, go to the `Filters` panel again, and from the `Topology` folder, drag the `Ego network` to the Query panel. There, fill in the name (actually: the ID) of the node that you are interested in (e.g. "hermione", no caps) inse the "Node ID" box. Click `OK` (and `Stop` to remove the filter):

Data Laboratory

Visualizing graph data is great, but it is also useful to export some of the numbers calculated for further postprocessing. For this, it is useful to visit the `Data Laboratory` panel, where the node data can be inspected in a tabular, spreadsheet-like fashion. This will include the newly calculated statistics for each node, such as the various centrality measures or the modularity classes. Through clicking the PageRank column, we can rank the nodes according to that attribute: here, we see that the people who frequently interact with Harry, also get a higher PageRank score - a "Harry bonus", so to speak.

Through clicking the `Export Table` option, you can save this data as a CSV file that you can later load into other applications:

Final Layout

Once you are happy with the layout of your graph, we can start worrying about how to export it to an actual image. For this, we need to go to the `Preview` pane, where we can take care of the final steps in producing the ultimate layout. There:

- Check settings such as: `show labels`, `show edges`, `curved`, ...
- Adjust the minimum and maximum width of the edges, font sizes, etc.
- Change the background color (button lower left corner).
- ...

Don't forget to click `Refresh` whenever you change a setting to see its effect:

You can then export the final graph to the format of your liking (pdf, svg, jpg, ...) through clicking the corresponding export button in the lower left.