Introduction

Social media refers to online tools that enable users to create and share content and support social interaction between users. Facebook is a social media tool that allows users to keep in touch with family and friends in a social network. It provides a public space for community groups to have members with common interests find each other and interact. It allow allows members, groups and businesses to create and promote public pages built around specific topics or products.

There are many ways people connect to each other in Facebook, such as "Friending", "Liking" pages and posts and "Posting" messages and comments (Hansen, 2011, p. 25). This paper focuses on the public Facebook Page **Empowering Motherhood Network**, a page familiar to this author.

The Empowering Motherhood Network Facebook page description is as follows:

> *Helping women thrive in motherhood and to give their children the absolute best start to life by discovering conscious parenting, creating a clean living lifestyle and awakening their women's wisdom.*

The page owner works as a doula and teacher of natural therapies to women pre and post birth. Page likes between group pages are collected and analysed using visual tools to uncover relationships, and interesting properties and features of the network.

Netvizz

Netvizz is a tool that extracts data from different sections of the Facebook platform (groups, pages) for research purposes. Collection outputs can be analysed using commonly accessible visualization tools. (Rieder, 2013)

Netvizz is written and maintained by **Bernhard Rieder**, Associate Professor in Media Studies at the **University of Amsterdam** and researcher with the **Digital Methods Initiative**.

It was developed as a practical attempt to study the Facebook API, and to investigate the potential of using digital methods to study social networking services (SNS)

Netvizz is run from within the Facebook environment, and requires Facebook users to log in with a Facebook account to be able to access data.

Privacy and Security

Netvizz has privacy and security limitations imposed by Facebook and by the tools author (Rieder, 2013).

The ability to collect data on personal networks (friendship groups) was disabled by Facebook on 30th April 2015, preventing the automatic collection of friend data by the API. Data from closed or secret Facebook groups cannot be collected. User privacy settings can also prevent access to data .The Netvizz author has also anonymized user names with the aim to protect users from inadvertent harm (Rieder, 2013).

Data

As Netvizz uses unique Facebook account, page and group identifiers, it is easy enough to collect data in separate files and combine them into larger networks.

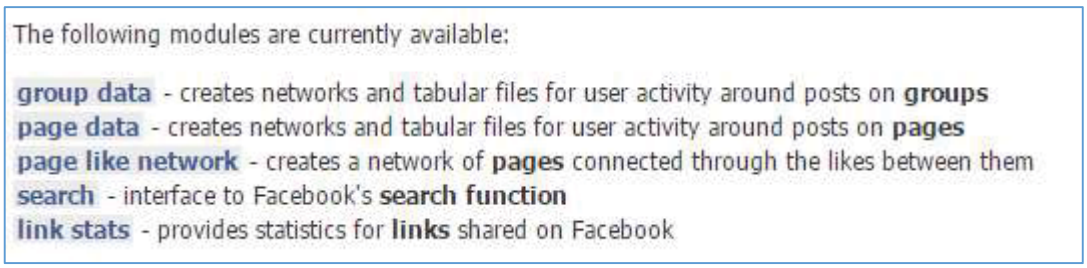

Figure 1. . Netvizz start page.

Page like data is collected from the Facebook page **Empowering Motherhood Network**. Netvizz uses the Facebook page ID, which is a separate entity from the group ID. The Empowering Motherhood Network page ID is 146913795360504, URL https://www.facebook.com/groups/empoweringmotherhood/

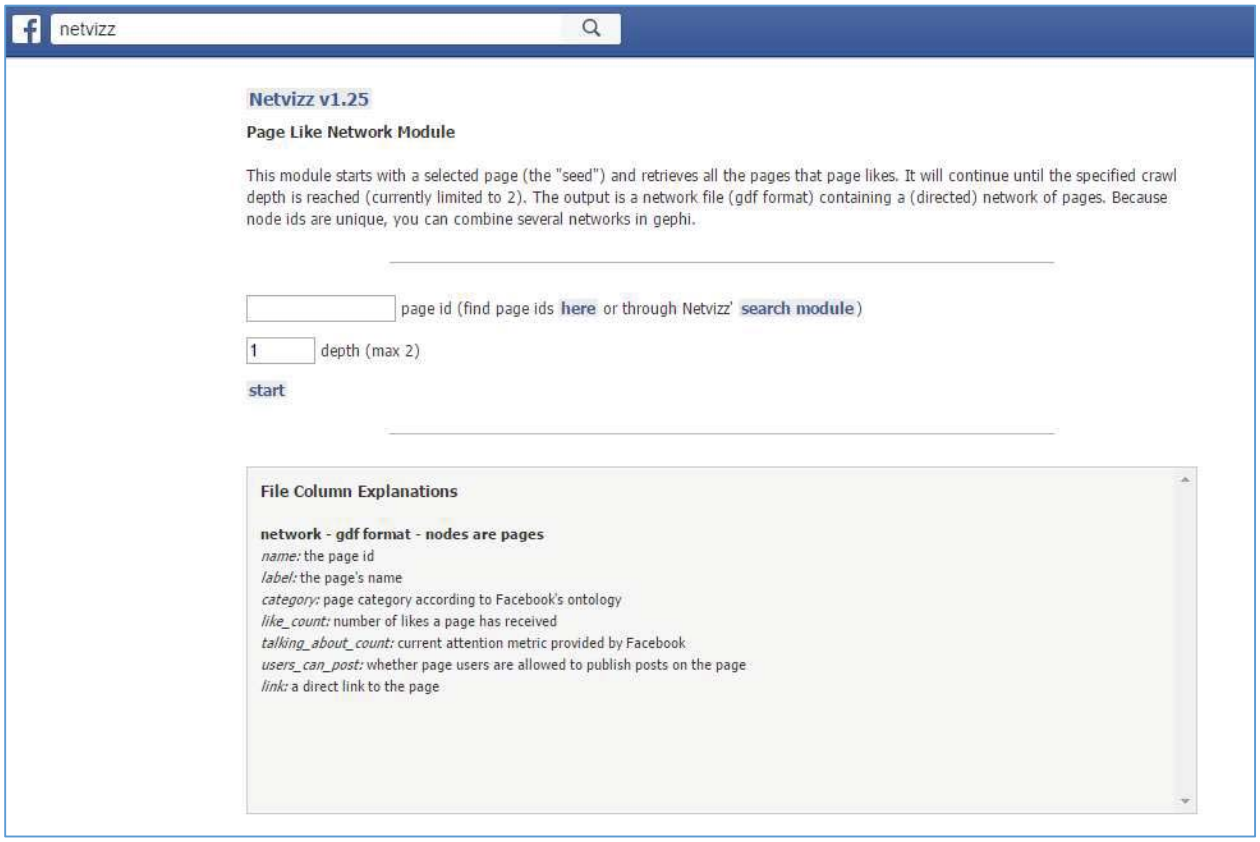

Figure 2. Netvizz Page Like collection page

Data such as node and edge attributes are compiled into a GDF (Graph Data File) file. GDF is a common network graph format, built like a database table or CSV file, which can be opened in many open source visualization tools. Each element is on a line and values are separated by columns.

Nodes in this data set represent unique Facebook pages. Edges represent the “Likes” between pages. A page can either like a page, or is liked by other pages. This means that the graph is directed, and the edges have a source and a target page.

Gephi

Gephi is a tool for data analysts used to explore and understand network graphs (Gephi.org). Users interact with the representation manipulate the structures, shapes and colors to reveal hidden patterns. The goal is to help analysts make hypothesis, intuitively discover patterns, and isolate structure singularities or faults during data sourcing.

Accepted Gephi file formats include:

- GEXF
- GraphML
- Pajek NET
- GDF
- GML
- Tulip TLP
- CSV
- Compressed ZIP

Data

As Netvizz uses unique Facebook account, page and group identifiers, it is easy enough to collect data in separate files and combine them into larger networks.

The **Empowering Motherhood Network** page like data (data) was collected to a network depth of 2. This means that data was collected for individual page likes, plus also likes for those subsequent pages. This can result in increasing complexity.

For example, the data network to a depth of 1 contains 18 nodes (Facebook pages) and 59 edges (likes between pages). As a comparison, the data network to a depth of 2 contains 808 nodes and 28396 edges!

Network graphs can often get too large to make out any meaningful patterns. Obstacles like node occlusions and edge crossings make creating well ordered and readable graphs a challenge (Hansen, 2011). To finish with a meaningful result, careful preparation, layout and filtering techniques need to be used.

The resulting graph data can be used to create reports about community health, comparisons of subgroups, and identification of important individuals and communities (Hansen, 2011). Graphing social networks gives the ability to ask and answer questions that are not available to other visualization methods.

Visualisation and Analysis

Graph Output

The initial loading of the Netvizz data set, via GDF format is shown below.

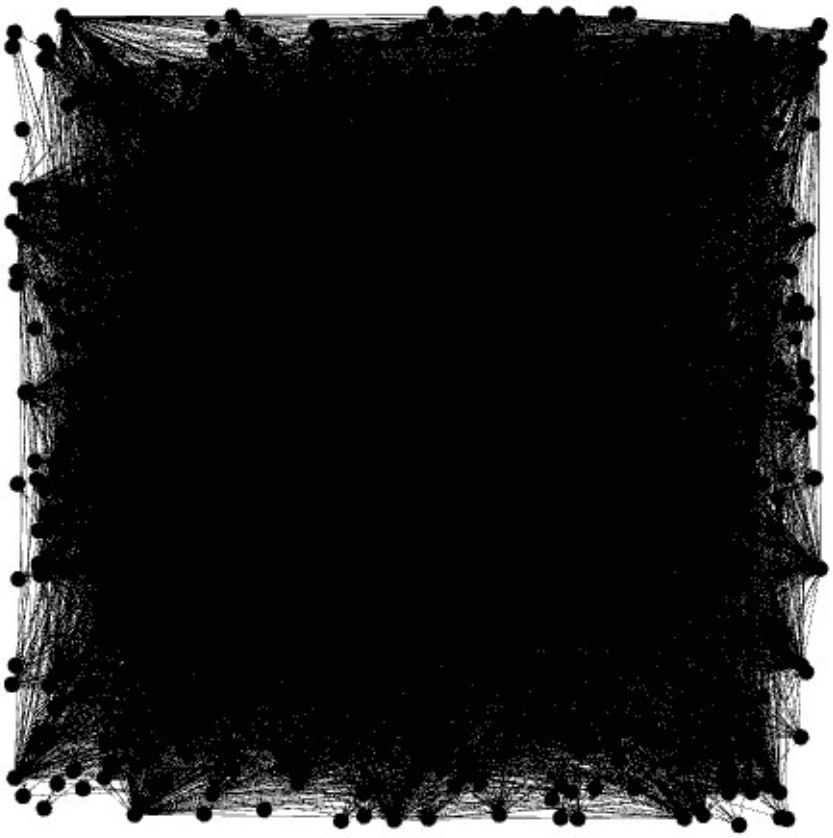

Figure 3. Initial Gephi network visualisation

The data set to a depth of 1 is shown below. This author decided to follow through with the more complex data set, as the smaller network did not have much ability to find communities or relationships outside the single network. Only 2 distinct communities can be seen.

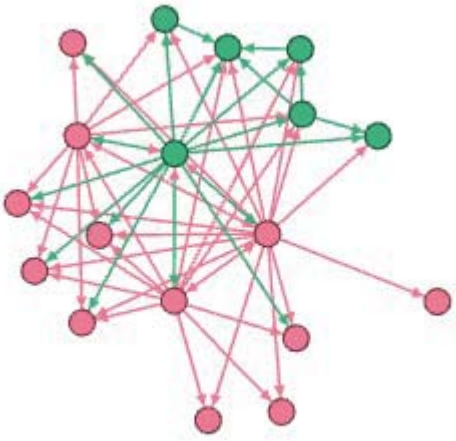

Figure 4. Initial Gephi network visualisation, depth 1.

Network graphs predominantly use force-based layout algorithms. Linked nodes attract each other and non-linked nodes are pushed.

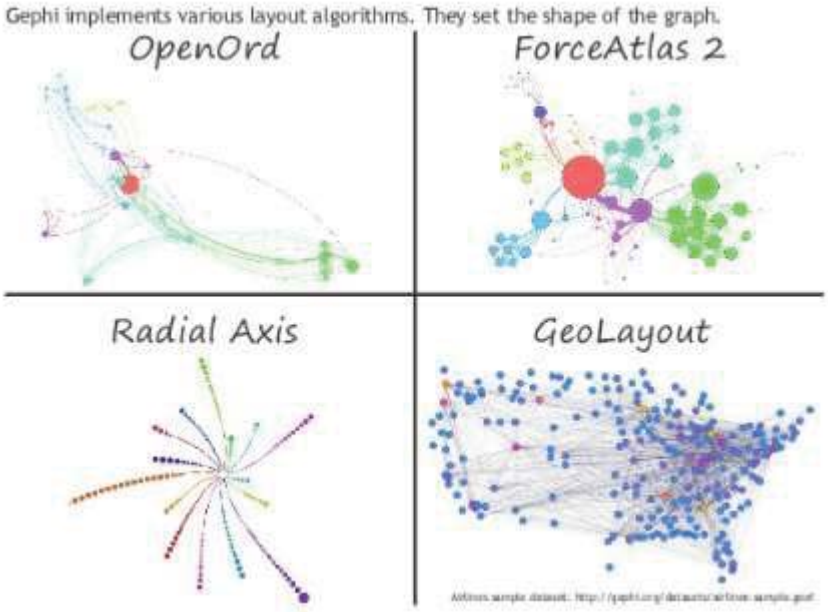

Figure 5. Gephi tutorial layouts (Gephi.org, Gephi Tutorial Layouts, 2016)

In order to emphasis divisions, choose the **OpenOrd** layout, to emphasis ranking, use the **Circular**, or **Radial Axis** layout. To focus on geographic partitioning, the **GeoLayout** can be used. In this case, the emphasis is placed on complementary relationships, and **Fruchterman-Reingold, Yifan Hu**, and **ForceAtlas** layouts can be used.

The **Fruchterman-Reingold** layout simulates the graph as a system of mass particles, which treats nodes as particles, and edges like springs that move nodes closer or further from each other in an attempt to find an equilibrium that minimizes the energy of the system.

The **ForceAtlas** layout is the go-to layout of Gephi, made to spatialize Small-World / Scale-free networks. (Gephi.org) It is focused on quality (meaning "being useful to explore real data") to allow a rigorous interpretation of the graph and a good readability.

The **ForceAtlas** 2 layout is an improved version of **ForceAtlas** to handle large networks while keeping a good quality (Gephi.org). Nodes repulsion is approximated with a Barnes-Hut calculation, which reduces the algorithm complexity. Replace the "attraction" and "repulsion" forces by a "scaling" parameter.

The treatments of the data (node and edge table data) are as follows.

1. Give the nodes a size proportional to their degree (number of connections) by applying degree of connectedness to the node size.

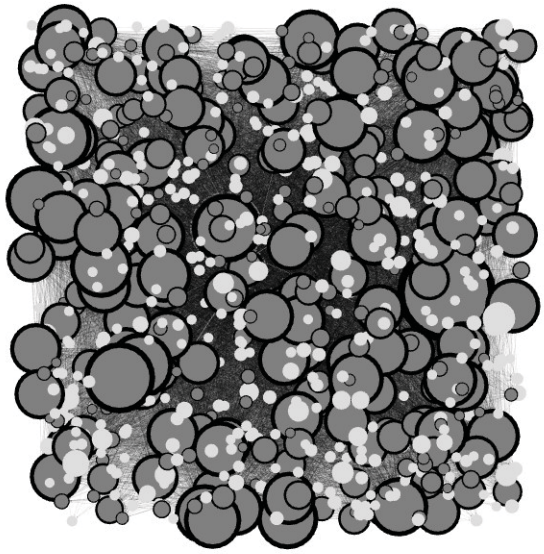

Figure 6. Node degree size

2. Apply a spacialisation that gives more space to the graph, but maintains it in a decided area.

Fruchterman Reingold,

The **Fruchterman Reingold** visualization separates nodes in a gravitational way. Communities are able to start being distinguished, viewing the denser parts of the network. The layout is run until the graph is stabilized. (Fruchterman, 1991).

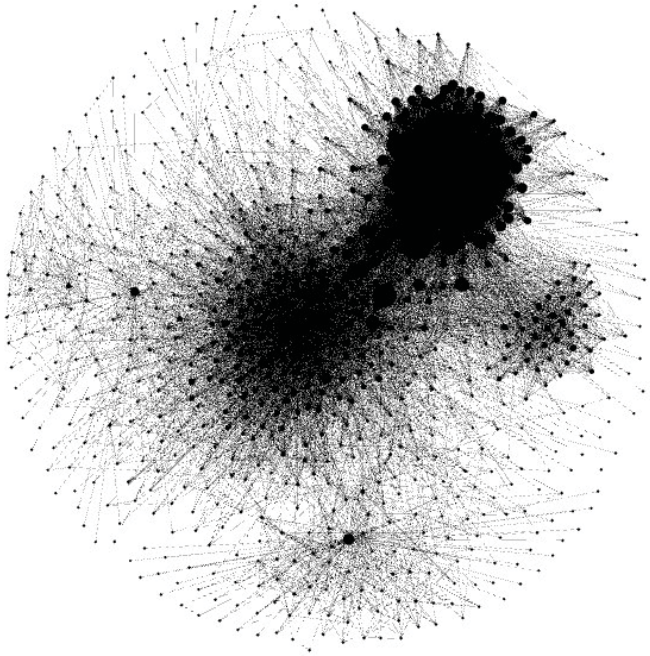

Figure 7. Fruchterman Reingold layout

You can see that initial node clusters are starting to form, but there is still a great deal of overlap.

3. Use the Force Atlas 2 layout to disperse groups and allow space around larger nodes.

The following options are available:

- LinLog mode = checked/unchecked

  Linear attraction and logarithmic repulsion (default lin-lin), to make clusters tighter.

- Scaling – Increase to make the graph sparser.
- Edge weight influence. From 0 (no influence) to 1 (normal). Set 0 to calculate forces without edge weight.

The "Prevent Overlap" option is used, and "Scaling" is set to 50. This is how much repulsion is required for the graph. The layout function is run until the graph is stabilized.

Force Atlas 2 can be run directly without using the **Fruchterman Reingold** first, but a better result is produced when the network is untangled before applying a strong force-algorithm.

4. We can focus in on our original node "Empowering Motherhood" by selecting in the Data Laboratory, and switching back to the Graph Overview.

Edge colours are then selected to show in/out relationships. A page can either like a page, or is liked by other pages. This means that the graph is directed, the edges have a source and a target, the direction is shown by arrows on the edges in the graph overview.

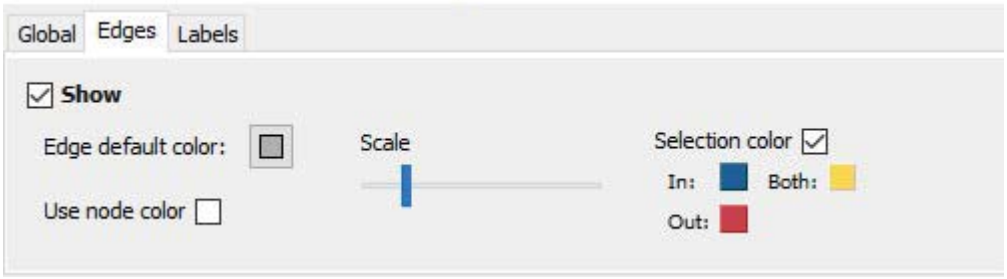

Figure 8. Gephi edge visualisation settings

Figure 9.

This zoom shows the page "Empowering Motherhood" and the in/out relationship between neighbour nodes.

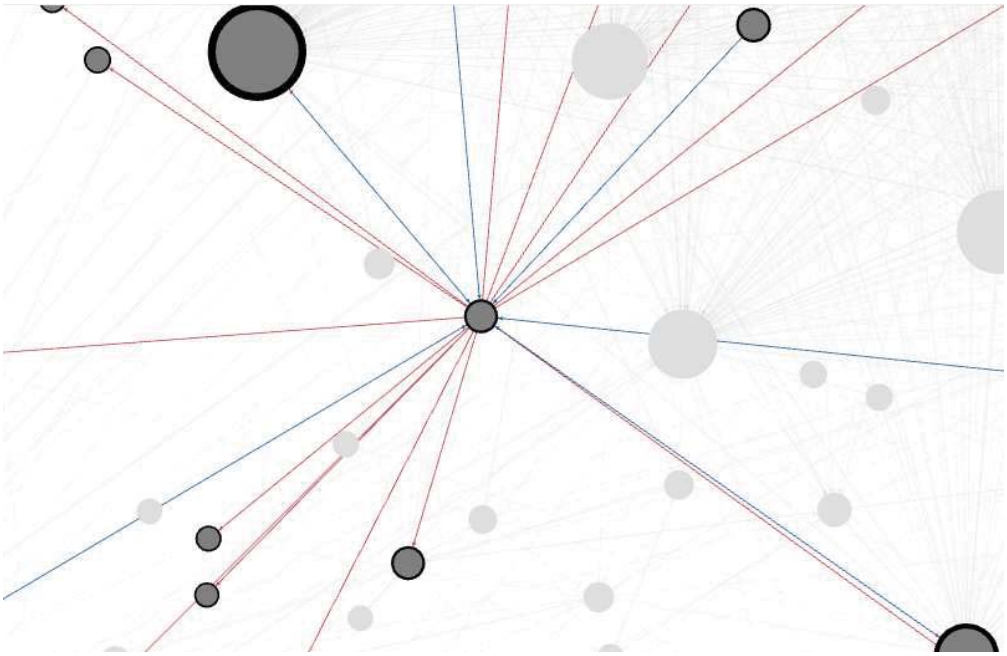

Figure 10. Edge relationships

As an example of how the complexity of networks with a search depth of 2 can also become unreadable, see the following image with the "Empowering Motherhood" node briefly coloured red and expanded in size.

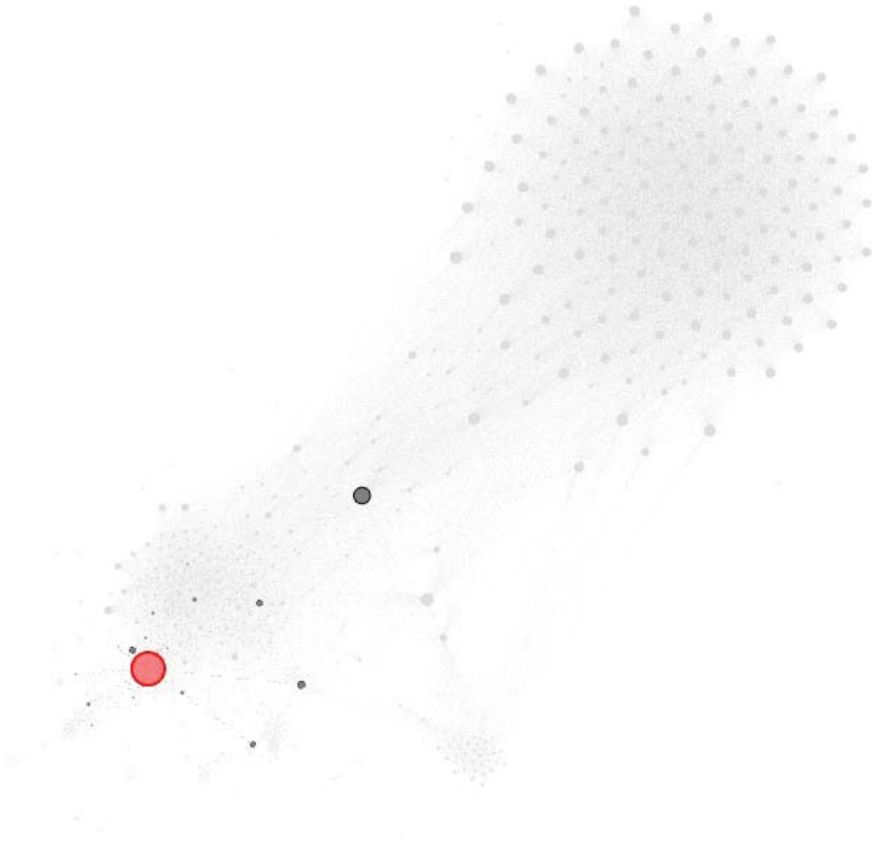

Figure 11. Data complexity

The graph is still a blobby "hairball" now, so more treatment is required to calculate attributes, influencing colour and size.

The edges are all of value 1, showing a single weight connection between pages, either the pages is liked, or not liked, so focusing on giving the nodes new attributes will enhance the graph.

5. Calculate the "Average Weighted Degree" statistic to distinguish the in and out connections.

Then colour the nodes with the "Weighted In-Degree" attribute.

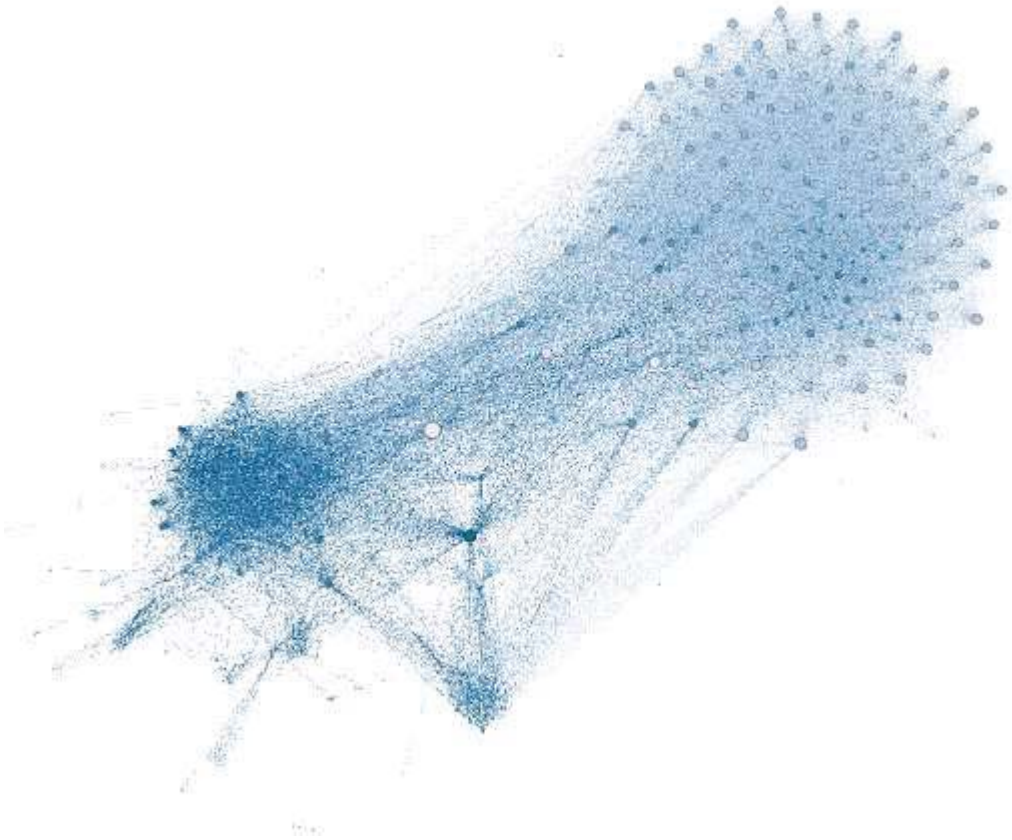

Figure 12. Coloured with "In-degree"

The biggest nodes (larger degree) are not always those with the biggest weighted in degree.

A network contains internal sub-divisions called communities. A network is said to have community structure if the nodes of the network can be grouped into densely connected sets. Non-overlapping community networks divide naturally into groups of nodes with these connection properties.

6. To highlight communities, use the "Modularity" statistic. Higher resolution settings produce bigger communities. Colour the nodes to show community relationships.

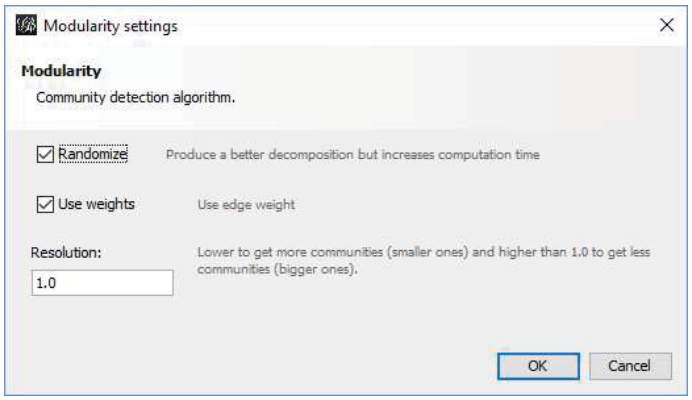

Figure 13. Modularity settings

From the resulting Modularity Reports, the results are shown in the following table

Resolution	Modularity	Number of Communities
0.5	0.075	34
1.0	0.410	5
2.0	1.285	4

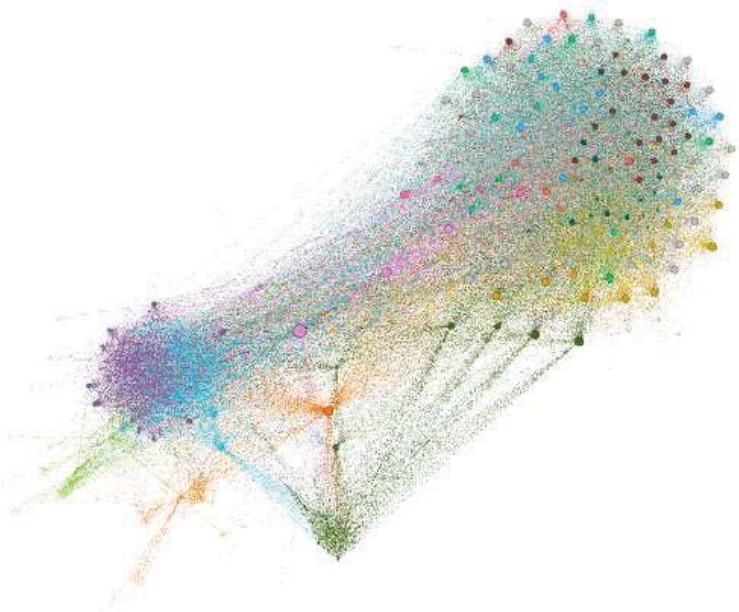

Figure 14. Modularity Class, resolution 0.5.

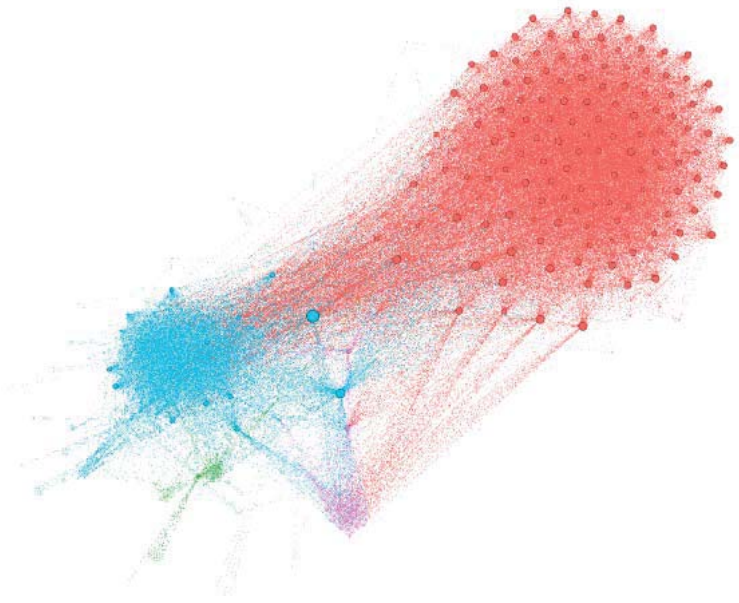

Figure 15. Modularity Class, resolution 2.0

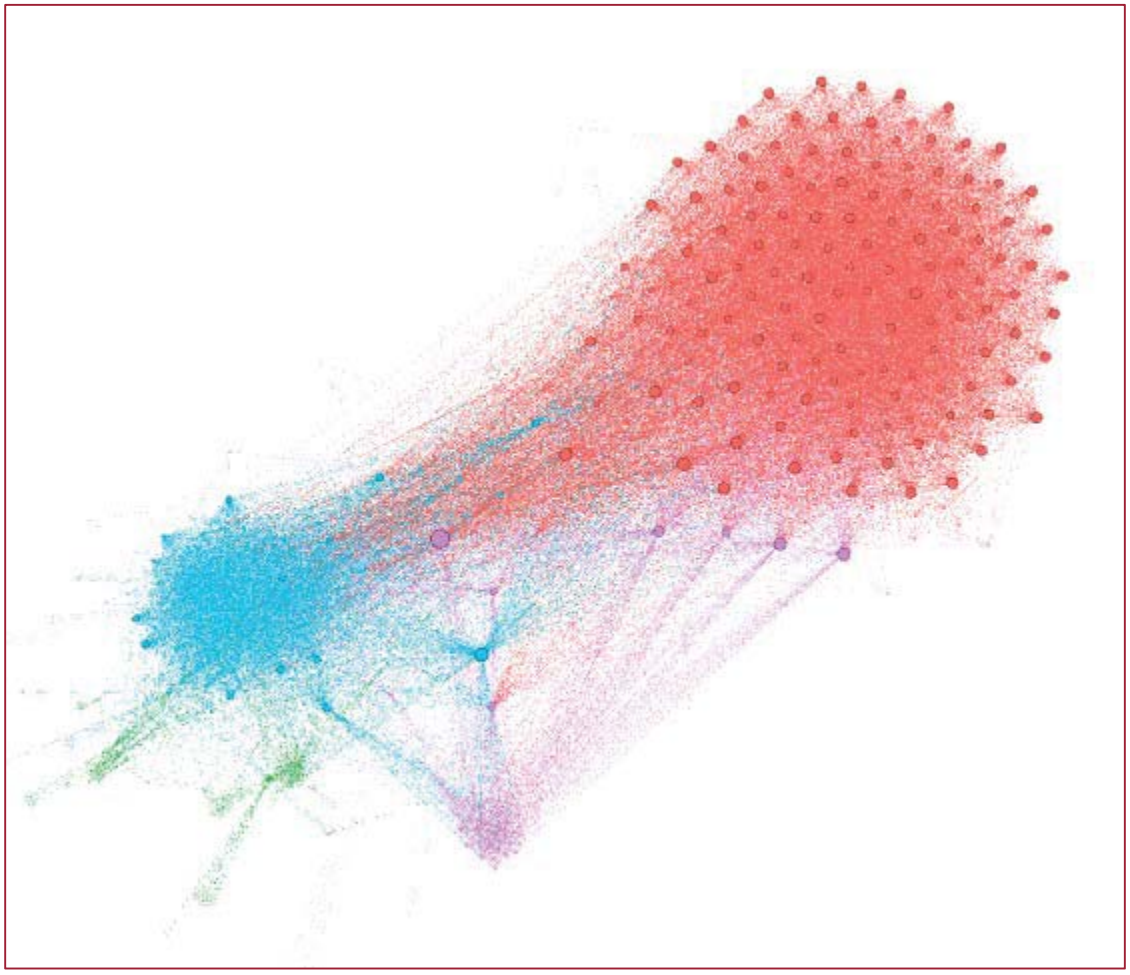

Figure 16. Modularity Class Graph, resolution 1.0.

The author has chosen the modularity resolution of 1.0, with the number of communities as 5. Now we can start seeing the relationships between liked pages.

Another way to visualize the data set is colouring the network by category.

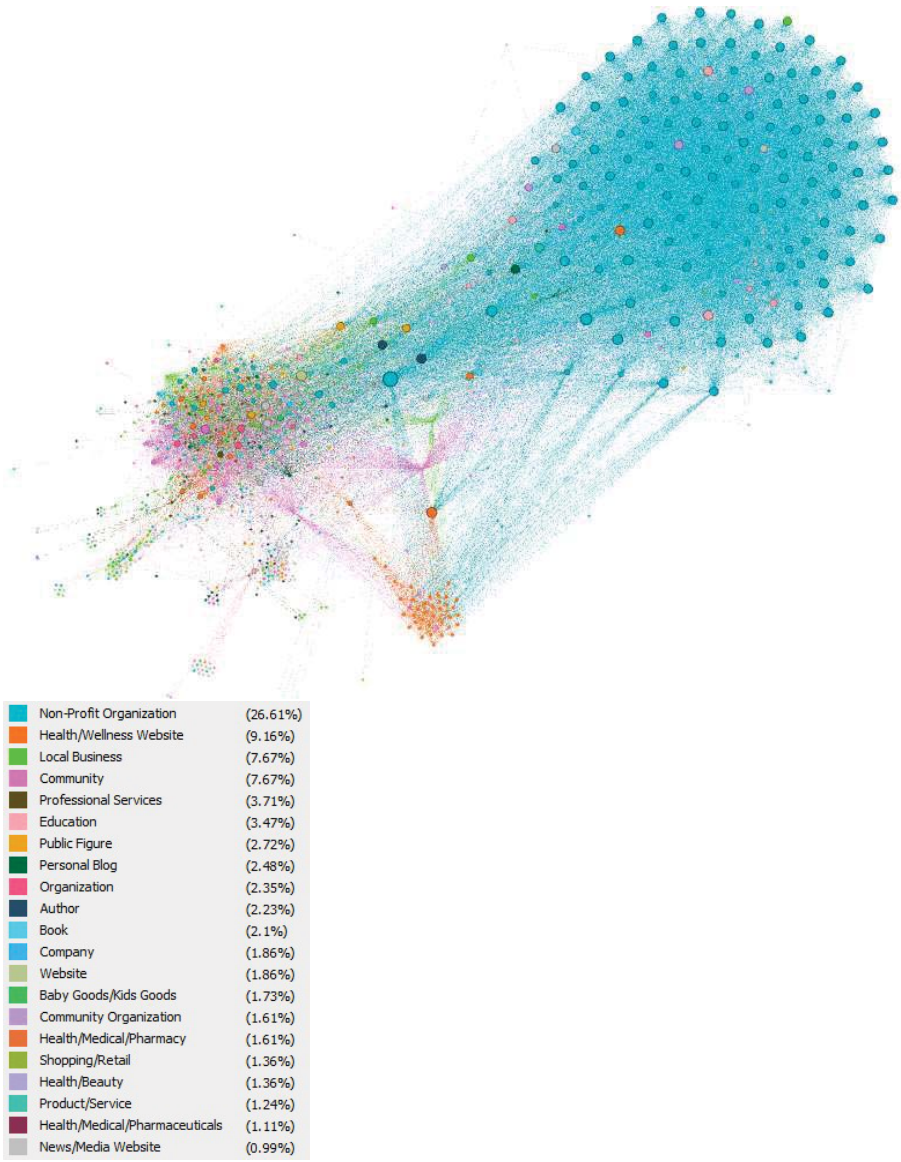

Figure 17. Network by category

Now we can pick out more interesting sub-divisions of the network.

Viewing the original Facebook page node, and changing this colour and all its neighbours to a more distinguishable black, we can see how the page is linked to the Facebook community.

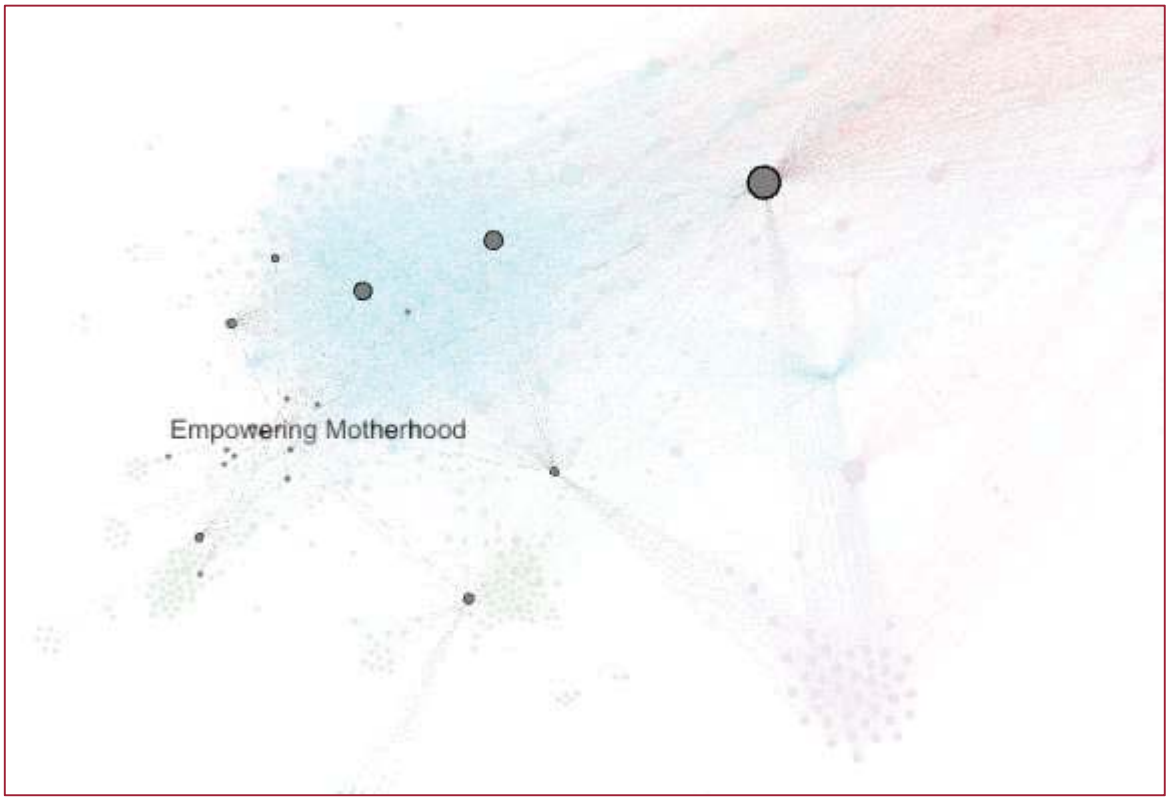

Figure 18. Facebook overview

Empowering Motherhood node selected, highlighting immediate (depth 1) neighbours.

Compared to the greater Facebook network, the community is quite small. Below is an extracted network graph of the initial community.

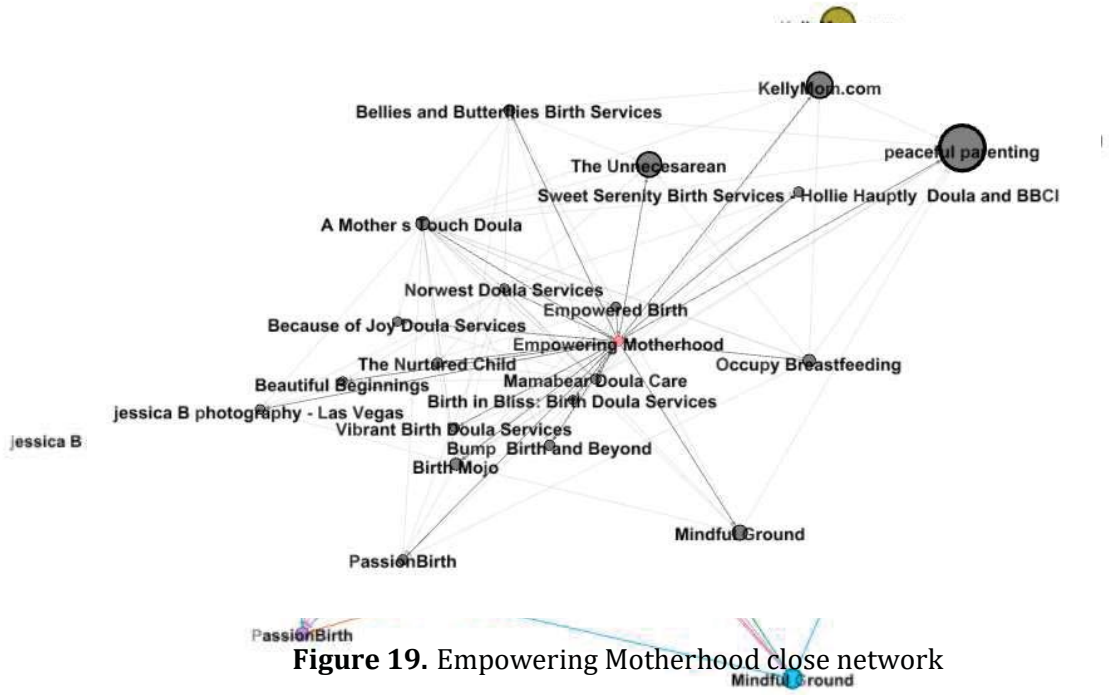

Figure 19. Empowering Motherhood close network

Several highlights pop out when categorizing nodes by community type.

Graph Analysis

Occupy breastfeeding, which is a Facebook community, is linked to several breastfeeding support groups, such as:

- Human Milk for Human Babies Global Network and related branches
- Peaceful Parenting a non-profit organisation
- KellyMom.com, also non-profit

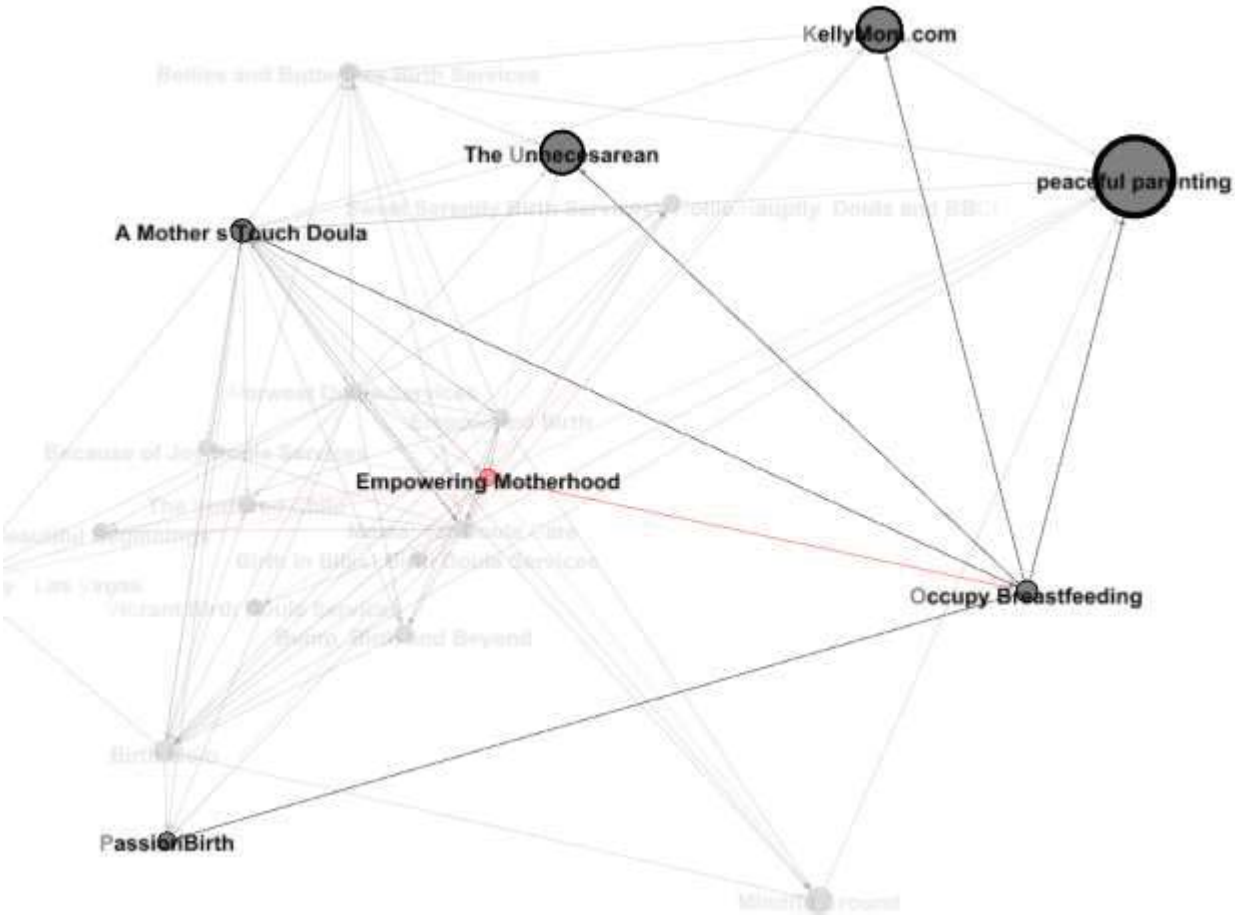

Figure 20. Occupy Breastfeeding

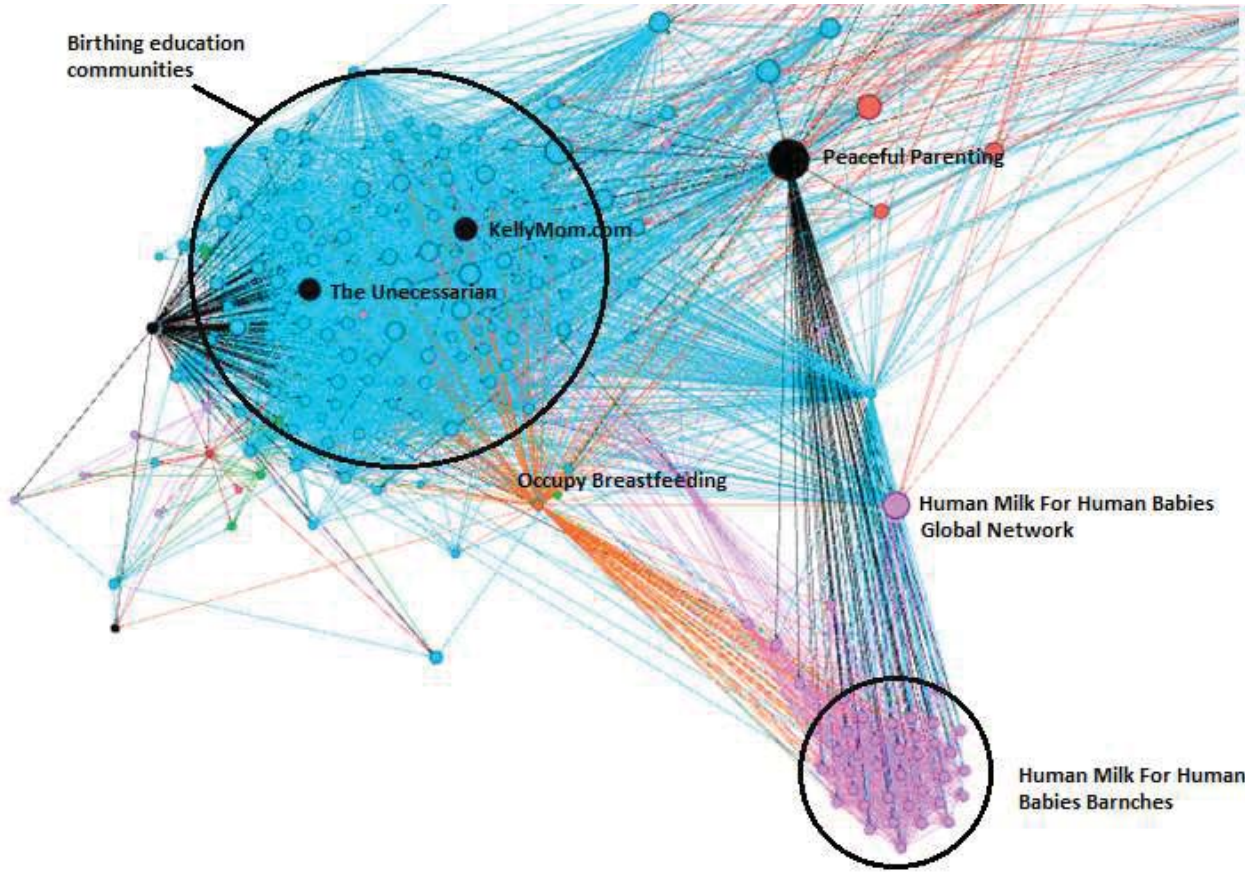

Figure 21. Occupy Breastfeeding communities

Mindful Ground, a meditation and mindfulness organization, links with Empowering Motherhood as follows.

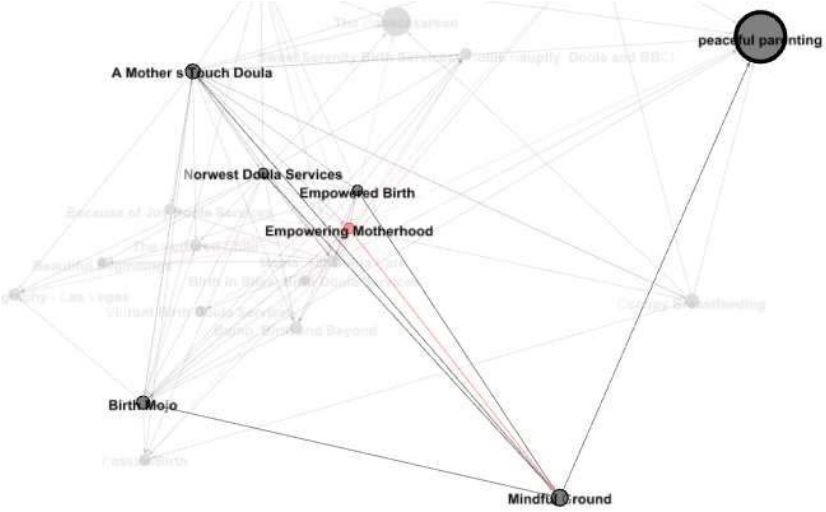

Figure 22. Mindful Ground

A number of smaller related communities can be seen branching off Mindful Ground, related to natural therapy providers and mindfulness motherhood groups.

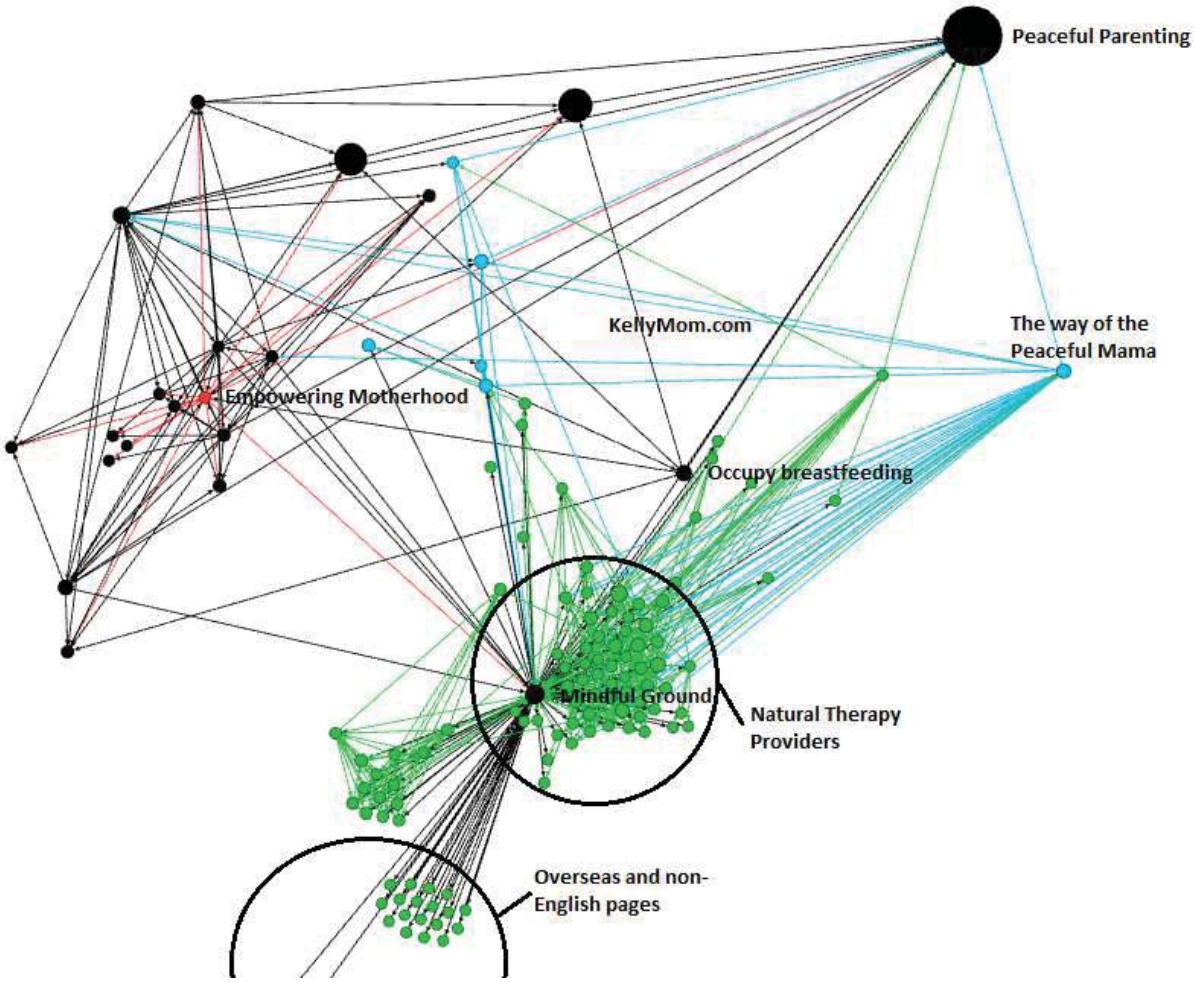

Figure 23. Mindful Ground communities

KellyMom.com is an interesting link with Empowering Motherhood, in that it provides accurate and detailed advice on the breastfeeding relationship. KellyMom.com also relates to several other nodes in the Empowering Motherhood network, Occupy Breastfeeding again, and Peaceful Parenting.

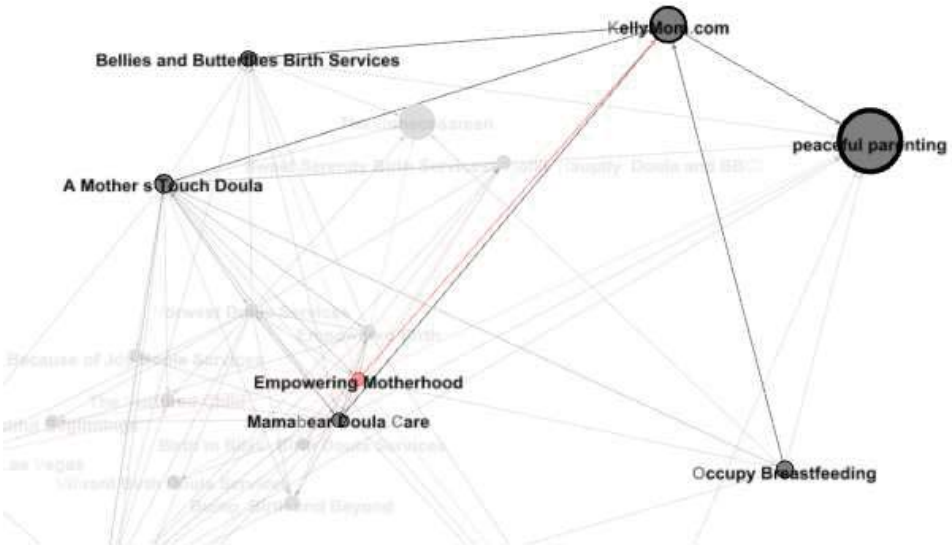

Figure 24. KellyMom.org

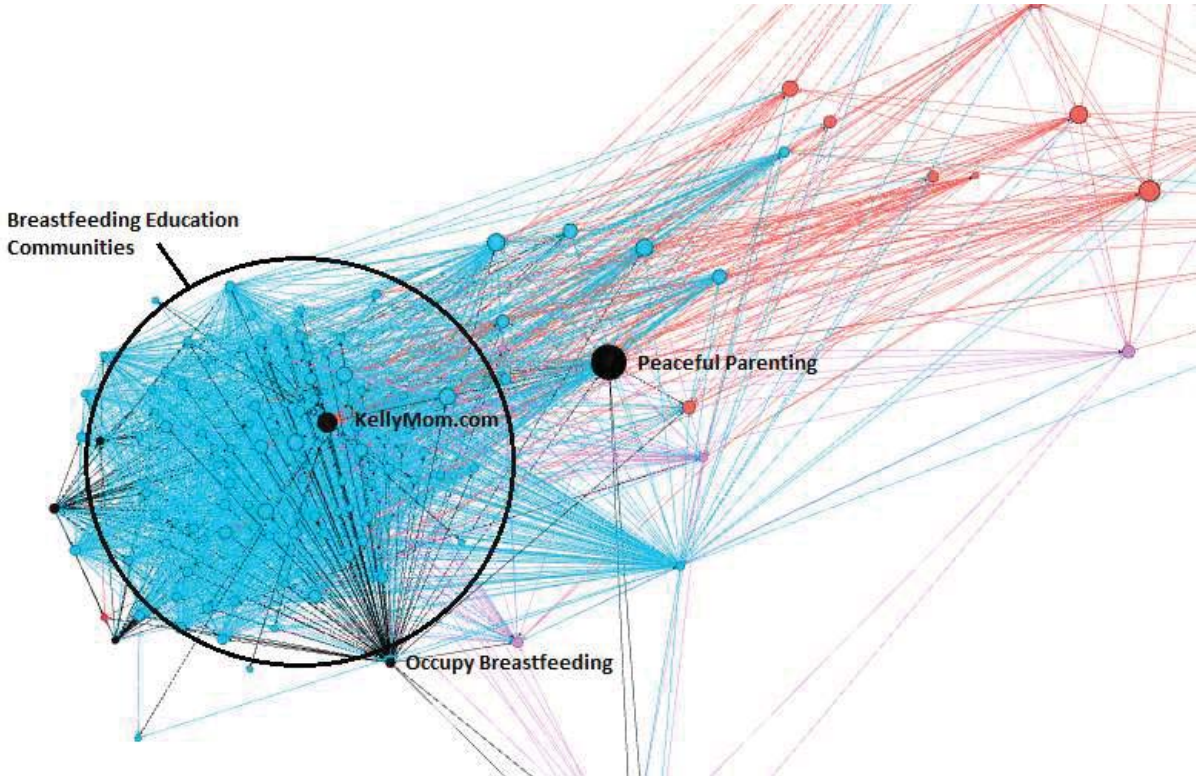

Figure 25. KellyMom.org communities

The Peaceful Parenting page node shows that it reinforces relationships with Empowering Motherhood, KellyMom.com, Occupy Breastfeeding and Mindful Ground.

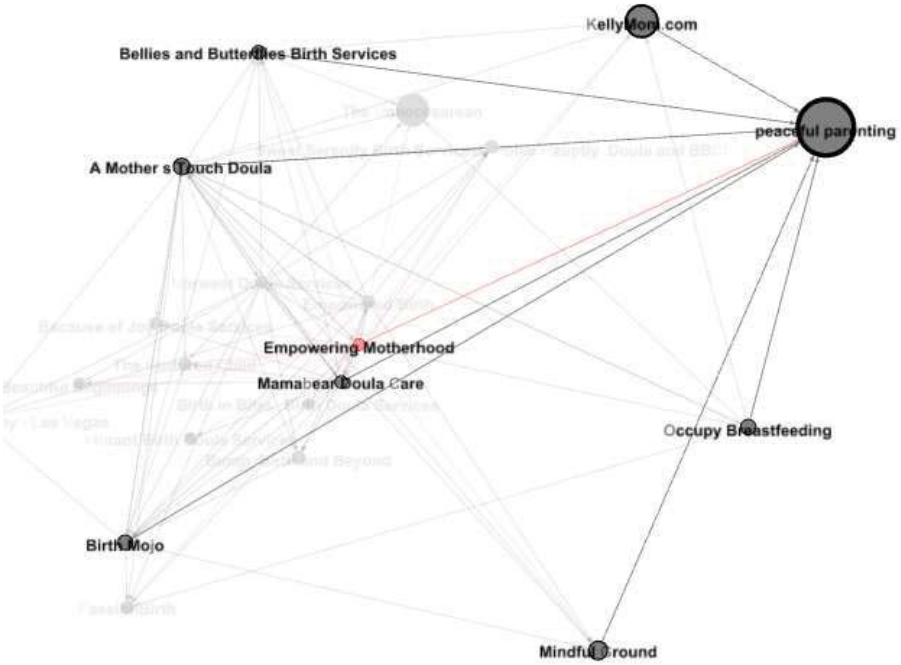

Figure 26. Peaceful Parenting

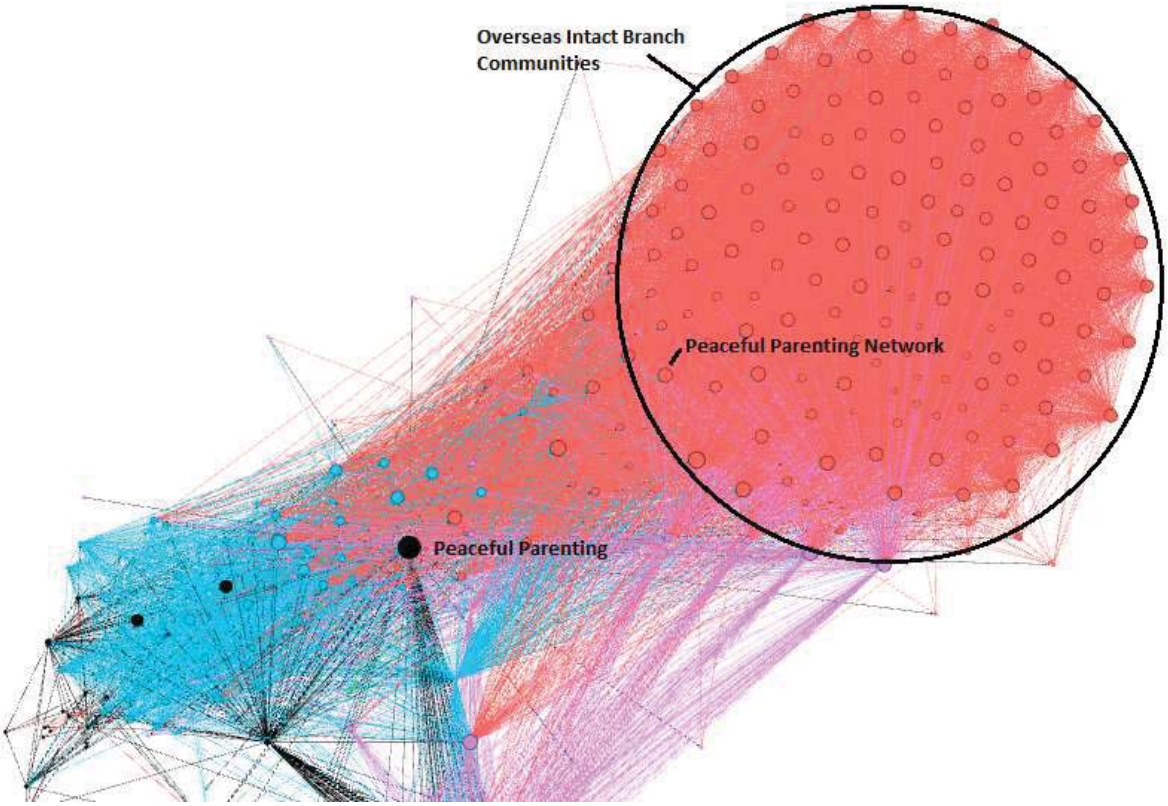

Figure 27. Peaceful Parenting network

The Peaceful Parenting Facebook page is an educational page, sharing "research based information" on all aspects of babies and children's health, development and wellbeing.

Going Further

Further analysis could be undertaken into how natural birthing and natural therapy groups relate to medical organisations, to view how the community interacts through breastfeeding and healthcare relationships and the relationship to current medical practices.

For example, the US/International breastfeeding support website KellyMom.com links back to Australian breastfeeding support networks through a link with the Australian Breastfeeding Association (ABA).

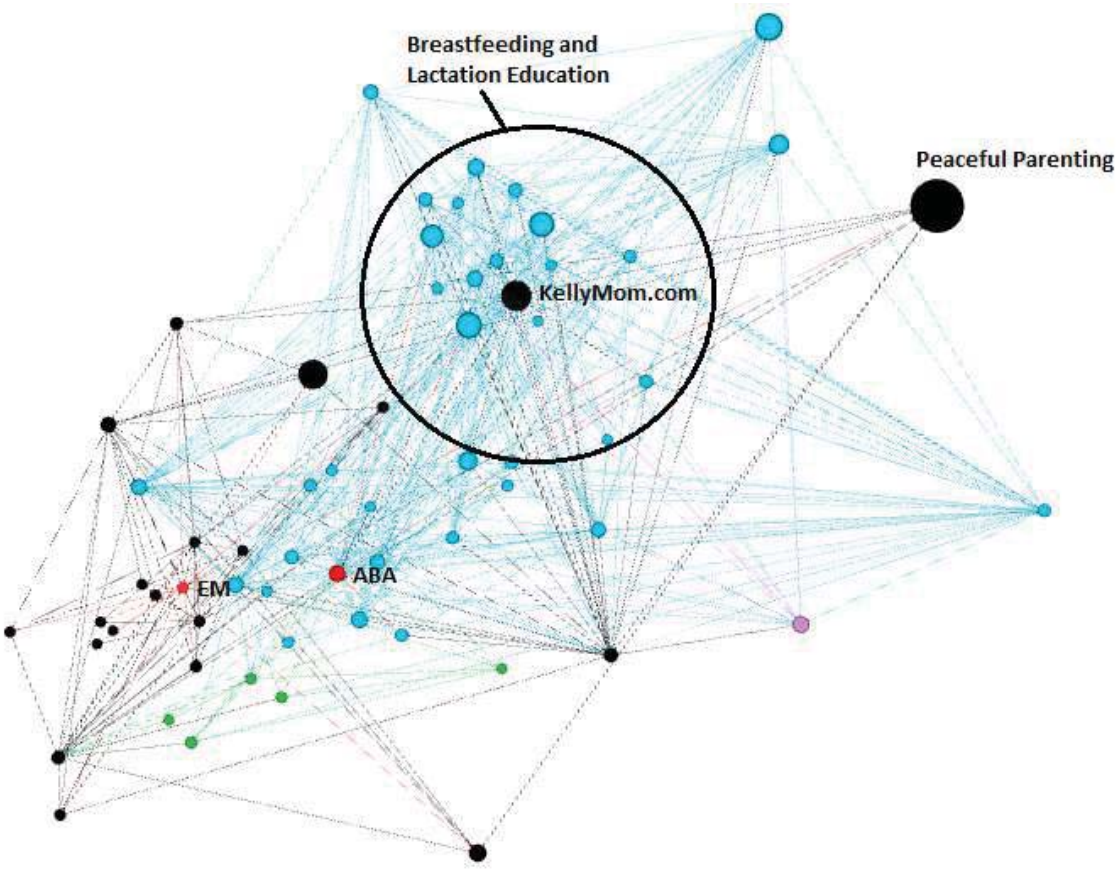

Figure 28. Breastfeeding support communities

The ABA links to other medical and health related communities through pages such as the Australian Medical Association, La Leche League, and BeyondBlue. Analysis of these networks would be time consuming and difficult due to the far ranging links to the wider Facebook community these organisations have.

Discussions

Visualisation Tools

The network analysis tool Gephi is just one of many visualisation tools used to examine networks. Through this authors interactions with the software, I found that it was difficult to zoom into nodes and analyze further detail, it was difficult to determine which nodes were useful and which should be removed, and the scale of the network was difficult to visualise on the screen, some hyperbolic or force-collapsible features would have been useful.

The following table describes just a sample of the tools available on the internet, and their advantages and disadvantages.

Visualisation tool	Description	Advantages	Disadvantages
Gephi (Gephi.org, 2016)	Used to explore and understand network graphs	Open source. Easy to install and use. Accepts majority of network graph file formats.	Difficulties with large (>1000 node) networks. Poor zooming.
NodeXL (About NodeXL, 2016)	A software template for Microsoft Excel	Built in data importers. Familiar environment for novices. Able to manipulate data using Excel table tools and functions	Pay for full license. Slow and difficult interface to handle
GUESS (graphexploration.cond .org, 2016)	A visualization tool for graphs and networks utilizing a domain-specific embedded language called Gython.	Open source. Simple to use	"A work in progress"
Cytoscape (Cytoscape.org, 2016)	Visualises molecular interaction networks and biological pathways and integrates these networks with annotations, gene expression profiles and other state data.	Good for large networks. Open source. Support for many network file standards	More complicated for analysis and larger learning curve.

The main barriers to utilization of each of these visualisation tools are the size of the network to be analysed, and the amount of knowledge and experience required to use the tools to their full extent.

Netvizz is also not the only way to extract data from Facebook. Data can be extracted via import calls from NodeXL, through the Facebook API directly using custom programming in languages such as Python, using NameGenWeb (now disabled due to Facebook policies on Friends data), through packages available in the R statistical programming language platform, and others.

Visualisation Methods

The **Empowering Motherhood Network** social network is visualised in this paper using force-directed graph drawing algorithms, **Fruchterman-Reingold** and **ForceAtlas 2**. Force-directed methods are a common way to visualise social networks as they position nodes of a network in a two or three dimensional space so that all the edges are mostly equal, there are as few crossing edges as possible and then using forces between the edges and nodes either to simulate the motion of the edges and nodes or to minimize their energy (Kobourov, 2012). The methods are well suited for graphs with different variations, they produce good results and are easy to implement. However, it is difficult to reproduce network layouts as the layouts are not predictable, they have a high running time and are not well suited to visualizing graphs with more than 1000 or so nodes.

Hierarchical methods are another way to represent relational data visually. However, this method is unsuitable for social networks, as most visual edges flow in the same directions and they do not scale well for large graphs. Hierarchical methods are more useful for the arrangement of flow diagrams, process charts and workflows.

Social Network Research Ethics

A large part data collection by third parties on Facebook is performed via software that uses similar technological strategies as the Netvizz data collection tool (Rieder, 2013, p. 8).

While the Facebook permission model aims to limit third party access by asking users explicitly for permission, there is often no possibility for users to actually modulate which rights are granted (Rieder, 2013, p. 9). A user granting rights to an application generally means that access is given not only to her data, but also to other user's data, and the same hold for groups.

Ethics in social networking research must balance the growth in social network analysis and its usefulness in myriads of applications, with the need to protect users from accidental harm caused by the publication of this research.

References

About NodeXL. (2016, 04 24). Retrieved from NodeXL Graph Gallery: http://nodexlgraphgallery.org/Pages/AboutNodeXL.aspx

Cytoscape.org. (2016, 04 24). *What is Cytoscape*. Retrieved from Cytoscape.org: http://www.cytoscape.org/what_is_cytoscape.html

Fruchterman, T. M. (1991). Graph Drawing by Force-Directed Placement. . *Software: Practice and Experience, 21(11)*, 21(11).

Gephi.org. (2016, 04 20). *Gephi Tutorial Layouts*. Retrieved from Gephi.org: https://gephi.org/users/

Gephi.org. (2016, 04 22). *Gephi.org/features*. Retrieved from Gephi.org: https://gephi.org/features/

graphexploration.cond.org. (2016, 04 24). *About GUESS*. Retrieved from GUESS The Graph Exploration System: http://graphexploration.cond.org/index.html

Hansen, S. a. (2011). *Analyzing Social Media Networks with NodeXL.* Burlington MA, USA: Morgan Kaufmann.

Kobourov, S. G. (2012). pring Embedders and Force-Directed Graph Drawing Algorithms,. *arXiv*, 1201.3011.

Qingyu Zhang, R. S. (2011). *Visual Analytics and Interactive Technologies.* Hershey PA: Information Science Reference.

Rieder, B. (2013). Studying Facebook via Data Extraction: The Netvizz. *University of Amsterdam*, 10.

Spence, R. (2001). *Information Visualization - An Introduction, 3rd Edition.* London, UK: Springer.

Zimmer, M. (2010). "But the Data Is Already Public": on the Ethics of Research in Facebook. *Ethics and Information Technology 12 4(2010)*, 313-325.