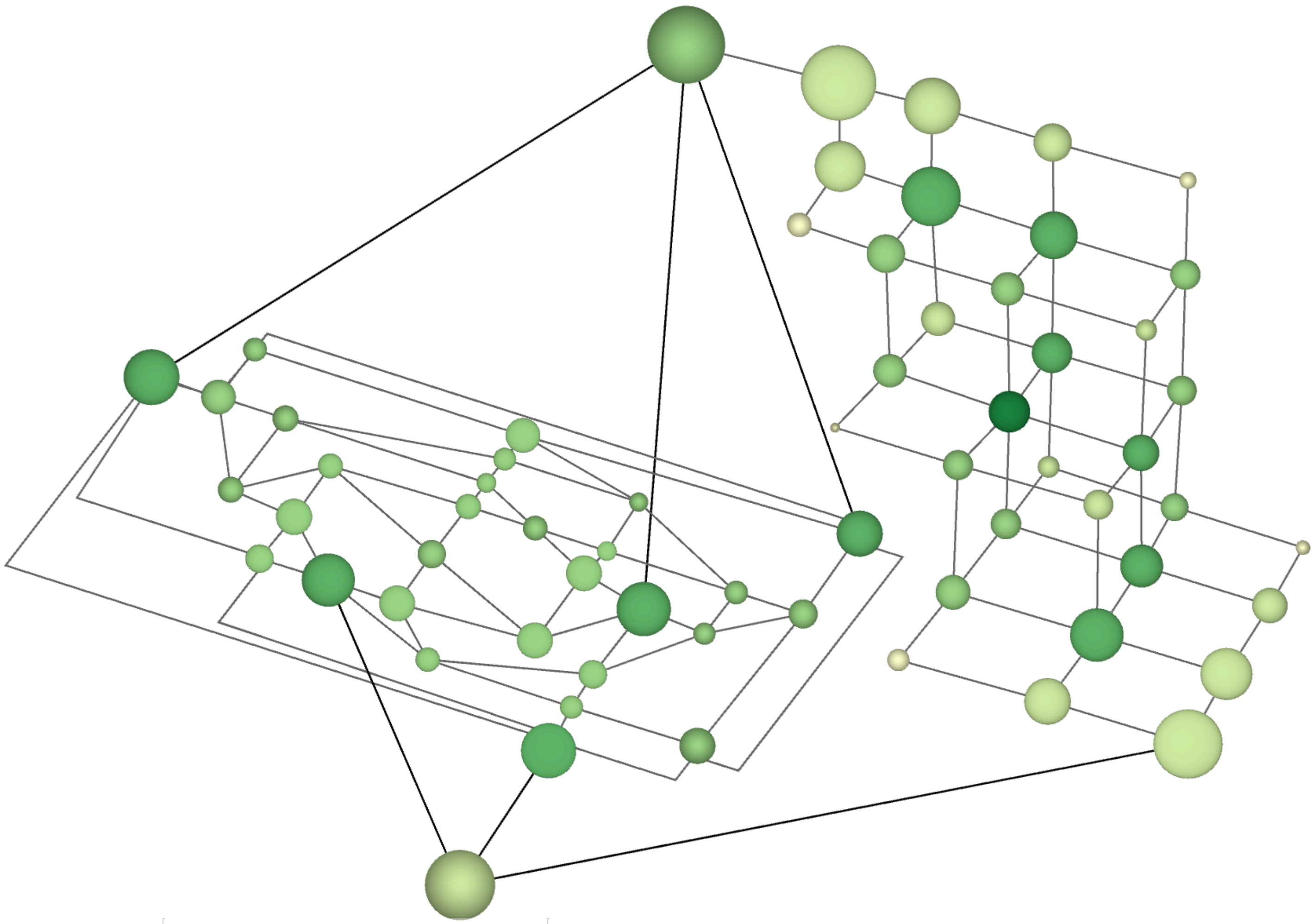
COMP5048: Visual Analytics Assignment 1

Ereina Camille Gomez
SID: 4100 2527



Visualisation of B3 Artificial Graph

Description

Tools and Layouts

To improve upon the winning graph, an overall 2.5 dimensional perspective was the objective to ease the viewer's navigation of the graph overall.
Firstly, I separated the two highly connected subgraphs using yEd (subgraphs identified in figure 1) .

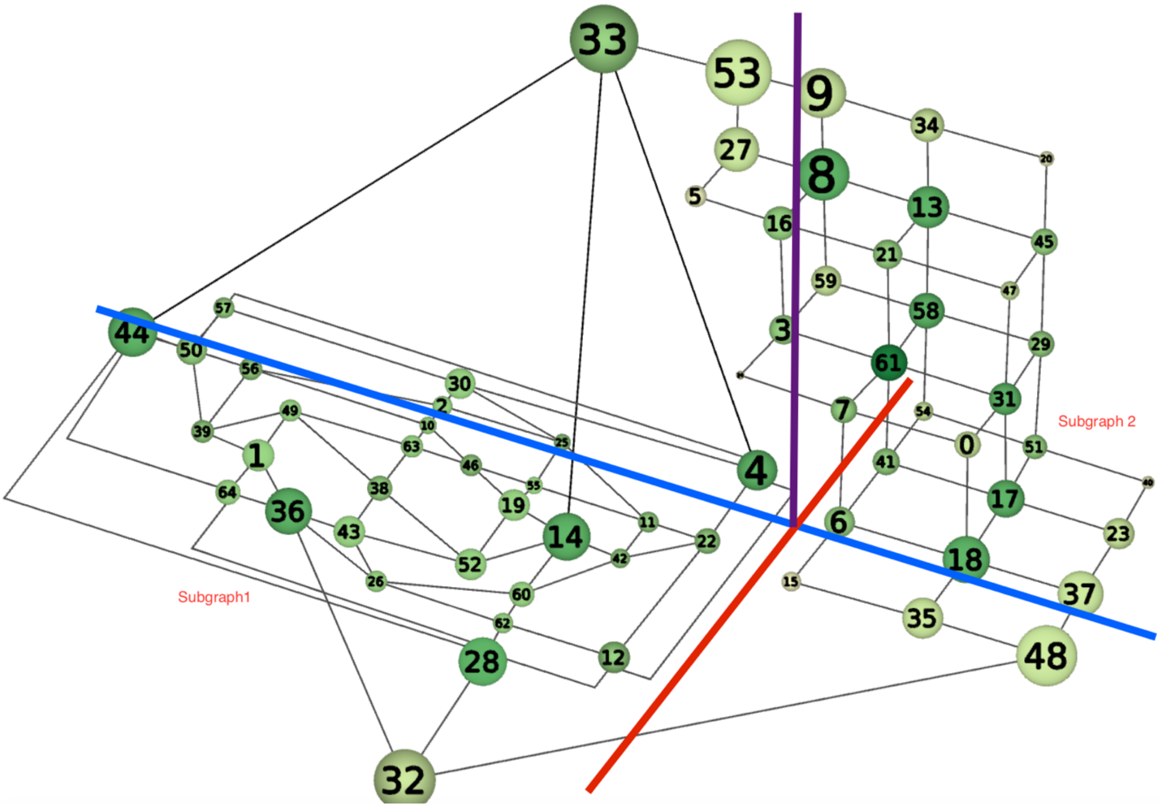


Figure 1: The defined x,y and z planes and subgraphs

The classic Orthogonal layout in yEd was used on subgraph 1. This resulted in a planar graph with several edge bends as seen in figure 2. Manual editing was done to minimise the edge bends yet still retain the bended edges that minimised possible edge crossings or helped with the overall rectangular area of the subgraph.

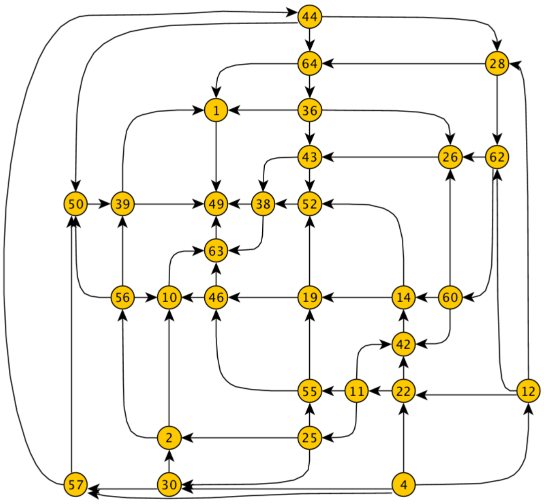


Figure 2: Initial orthogonal layout of subgraph1

For Subgraph 2, I did not change the overall positioning as it came how it was during the import where nodes operated as corners of squares. Instead, I combined both subgraphs and changed the orientation of the nodes to reflect the spatial planes as defined in Figure 1.

The whole graph was then imported into Tulip where Betweenness Centrality measures were calculated and node size was manipulated (from range 1 – 7) with larger nodes showing higher Betweenness centrality scores.

InOut degree was calculated in Tulip and colour mapping on a green spectrum was done to reflect nodes with a higher degree as having a darker shade of green, with nodes of lower degree having lighter shades.

Finally, the node shape was set to sphere to improve visibility of the lighter shaded nodes in the graph by way of shadowing.

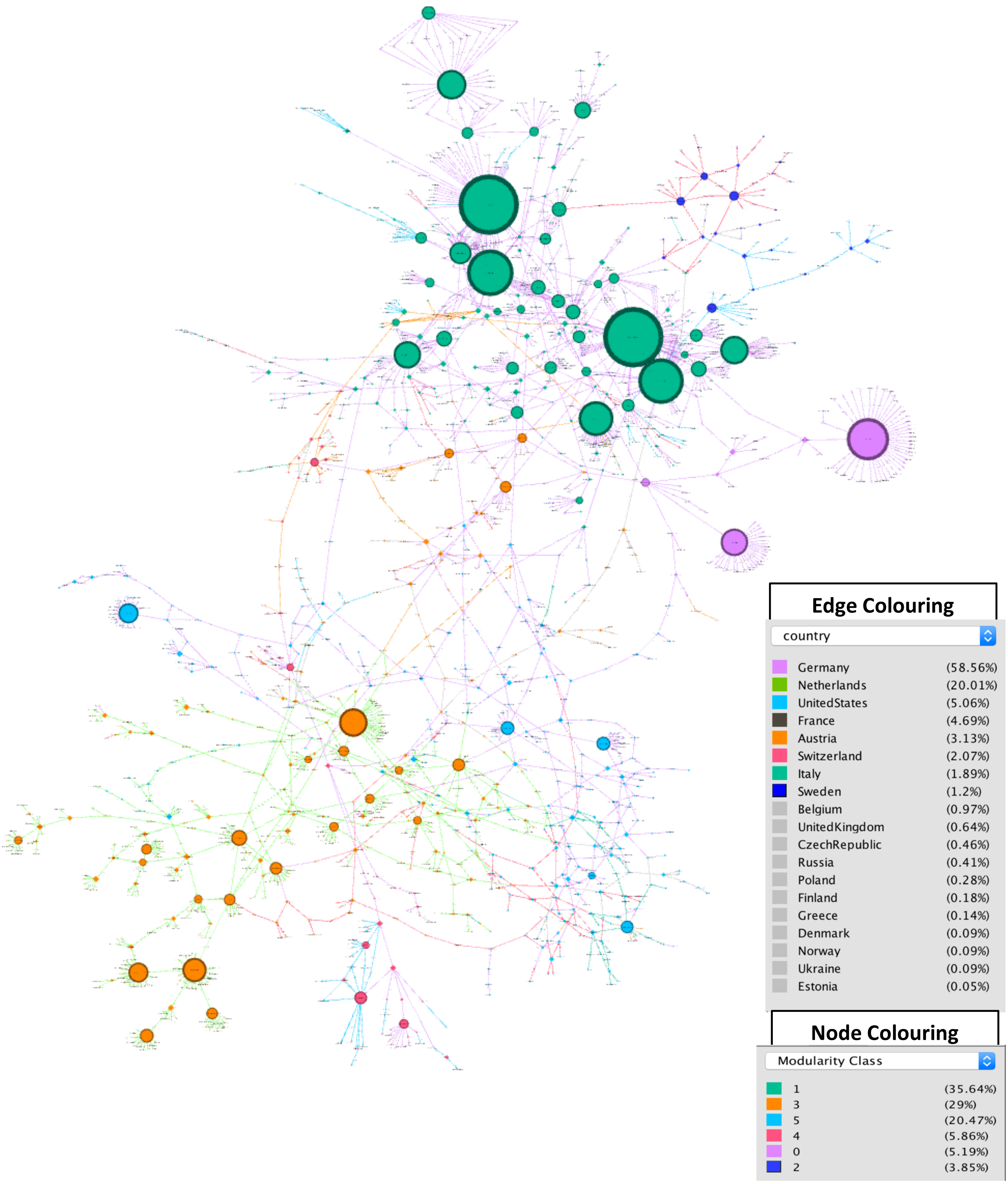
Self-evaluation

Strengths
As the whole graph is non-planar, the use of an overall 2.5D projection helps the viewer distinguish the connections of the graph. For example node 33 appears as if it is situated above subgraph 1 with its edges pointing downwards to the "surface". While in subgraph 2, retaining the original spacing gives the illusion of cubes with depth which helps the viewer navigate the edge connections.
The use of visual variables of node size (for Betweenness Centrality) allow us to see which nodes are passed through the most within the full network. Additionally, node colour highlights the nodes with the most degree. These two variables combined, we can find interesting observations such as weak spots within the full network (e.g. nodes with both a high betweenness centrality and low degree such as node 48 and 53 suggest that their few ties are crucial for the network to flow).

Weaknesses
My solution introduces edge bends which is not ideal. This is a trade-off for minimising the number of edge crossings. Moreover, the edges with bends mostly have a long length compared to the straight edges. This was a result of retaining a rectangular area of subgraph 1 to mimic the surface of subgraph 2 to ease visualisation in this projection.
Finally, there is variability in the spacing of nodes which is a trade-off to minimise overall edge lengths. Unfortunately overall, my visualisation introduced one more edge crossing compared to the winning graph. Future work would involve improving the node spacing of the entire graph and exploration of geometric shapes to minimise edge bends.

Visualisation of Math Genealogy Graph

Interactive version : comp5048_assignment1/network/index.html



Description

For the math genealogy graph I have chosen to explore the visualisation task:
Is there a relation between country of graduation for advisors and their advisees?

Tools and Layouts

Initially, MultiGravity ForceAtlas 2 layout using Gephi was used to expand the nodes of the network. The resulting image consisted of a giant connected component in the centre, with smaller components surrounding it.

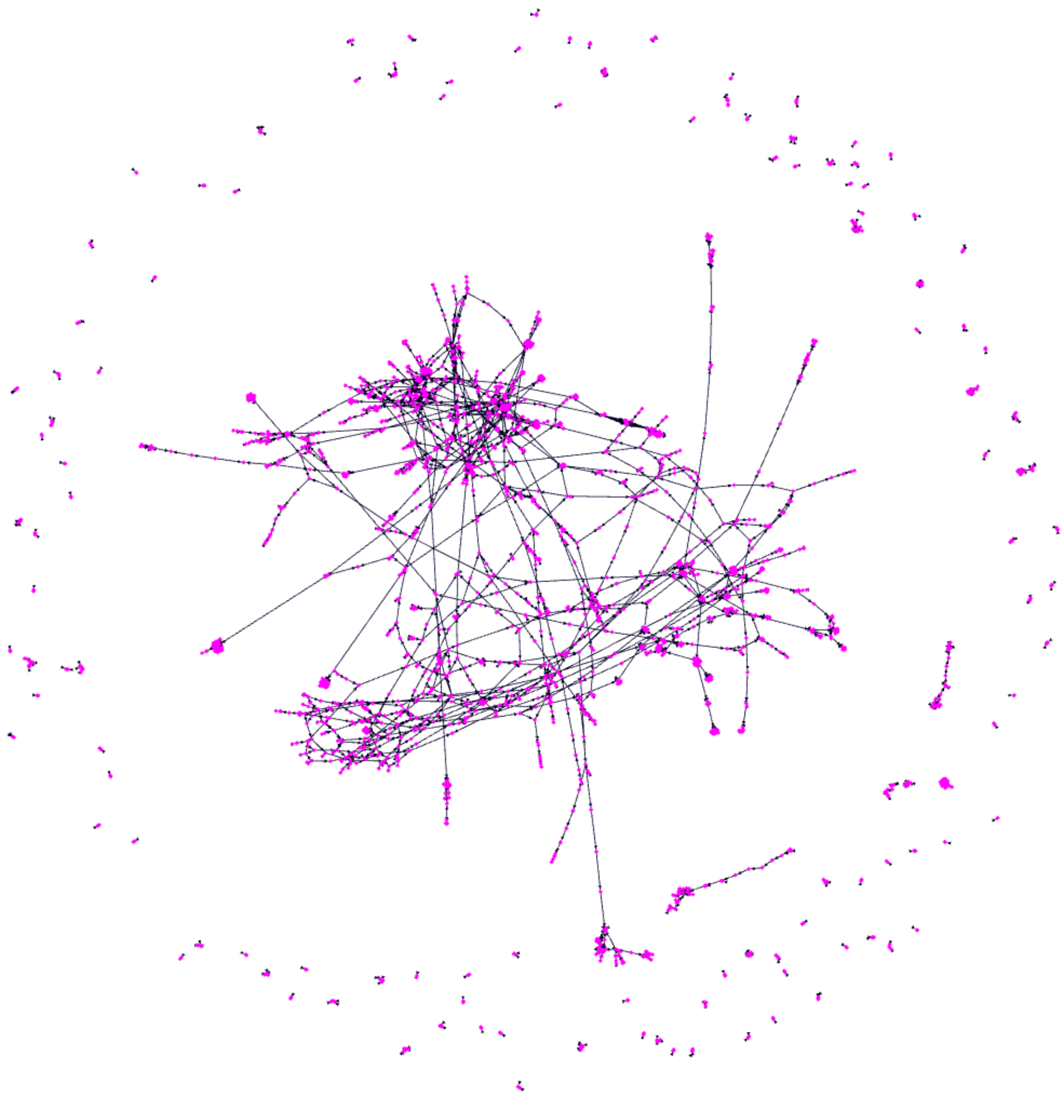


Figure 3: The full network after MultiGravity ForceAtlas 2 layout

The Giant Component filter was used to discard the smaller components. I chose to focus on the giant component where most advisor-advisee relations were concentrated. Modularity Class statistic was then computed to find clustered communities inside of the graph which uses the Louvain method for community detection. The resolution parameter was set to 10.0 (larger communities >= 1.0, smaller communities <= -1) which resulted in detecting 6 unique communities at a modularity score of 0.702. The visual variable of node colour was set to reflect these communities, with one distinct colour belonging to one community. To minimise edge crossings, the Yifan Hu Proportional layout was applied which expanded the nodes outwards and then the Noverlap layout was applied which helped minimise some edge crossings. Manual editing and spacing was done to further minimise edge crossings.

The average degree statistics was computed for all nodes. The visualisation task was to find the advisors with the most offspring. Due to the directed nature of the network, I used each node's out-degree to assign node size for each node (from range node size = 10, to node size = 300) where larger sizes indicated more offspring.

To reflect the country variable for advisee graduations, I manipulated the edge colours to each represent a particular country. In the dataset there were 19 distinct countries, with those of proportion more than 1% each being assigned a unique colour and the rest being assigned grey. This was done as the minor countries provided little data for exploring advisor-advisee cluster relations.
The graph was then exported with the SigmaExporter plugin to create the interactive version.

Self-Evaluation

Strengths
The nodes with the most offspring are clearly visible due to their size and area coverage in the graph, making it easy to identify advisors with the most advisees.
The edge colour partitioning visually highlights the significant proportions of graduations by country – even at the global level, we can clearly see most advisees graduated in Germany.
The use of edge colour and node modularity colour partitioning combined leads to interesting observations. While edge colourings show instances of network homophily (particularly for Germany and the Netherlands, contrasted to the US which little clusters are spread around), modularity class unfolds communities within the network which allow us to see inter-country communities (involving more than one country).
The zooming function allows the viewer to manipulate their perspective, gives the option of both the global view and localised views of the graph.
Click function highlighting a chosen node's neighbours aims to disambiguate the possible confusion brought on by edge crossings.

Weaknesses
It is not easily clear at first glance who is the advisor and the advisee as my graph has no hierarchical structure and there is a high variance in edge directions (each edge can be directed at any angle). Although my visualisation task was pre-defined, I did not make use of the year attribute of edges which would provide more information and may yield interesting observations. I attempted an additive-like "small multiples" to display temporal information, however, since the graph has lots of connections and many smaller sized nodes it is hard to see. Moreover, small multiples is not easily scalable considering this is only a subgraph of the real Math Genealogy graph. Future work would be to build upon the interactive graph by exploring ways to reflect temporal information. For example a possible slider function which allows the viewer to filter by year ranges and still zoom to get localised views. Finally

there are many edge crossings. Future work would involve experimenting with different spacing of the clusters and individual nodes to best minimise edge crossings.

Appendix

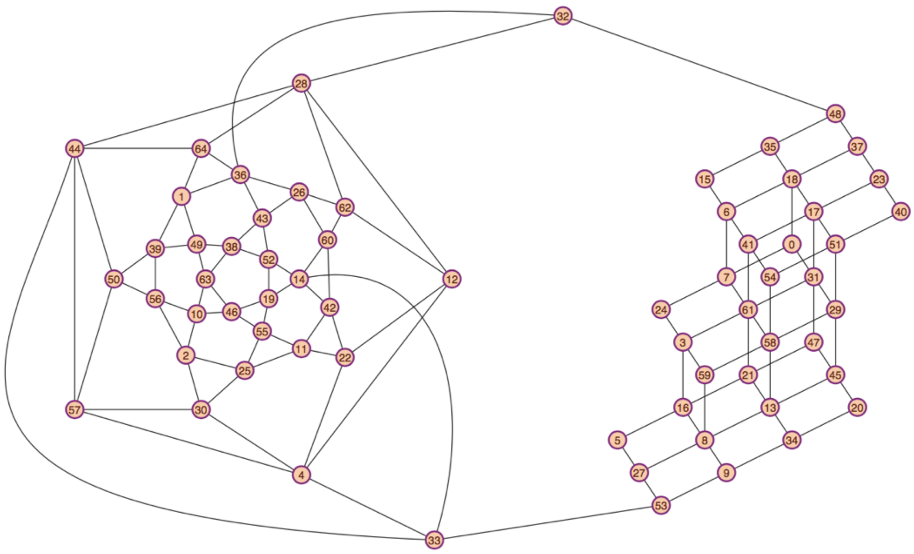


Figure 4: Winner, Graph C (original in color).

Graph 1: The winning B3 graph entry