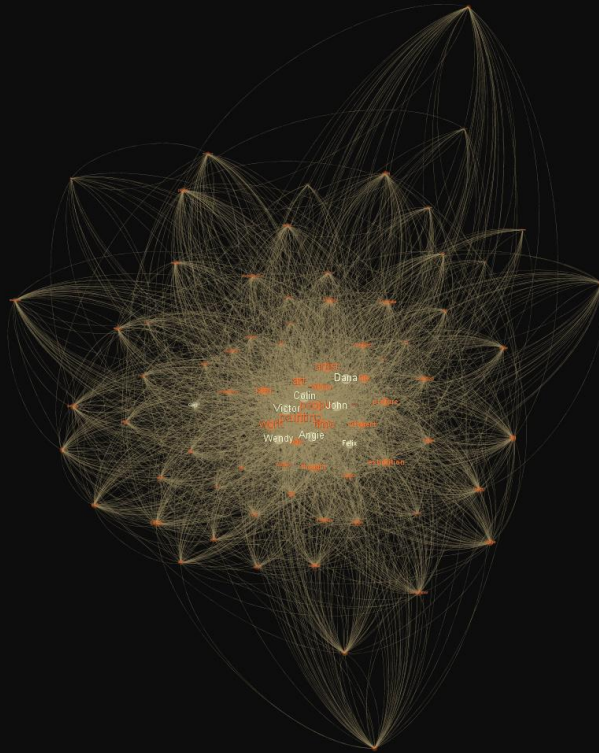


OPEN SOURCE GRAPH VISUALISATION AND ANALYTICS

Jacob L. Cybulski

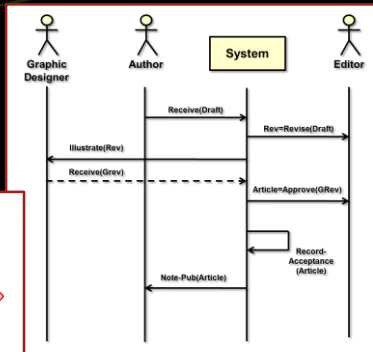
*SAS Visual Analytics Collaboratory
Dept of Info Sys and Bus Analytics
Faculty of Business and Law
Deakin University*



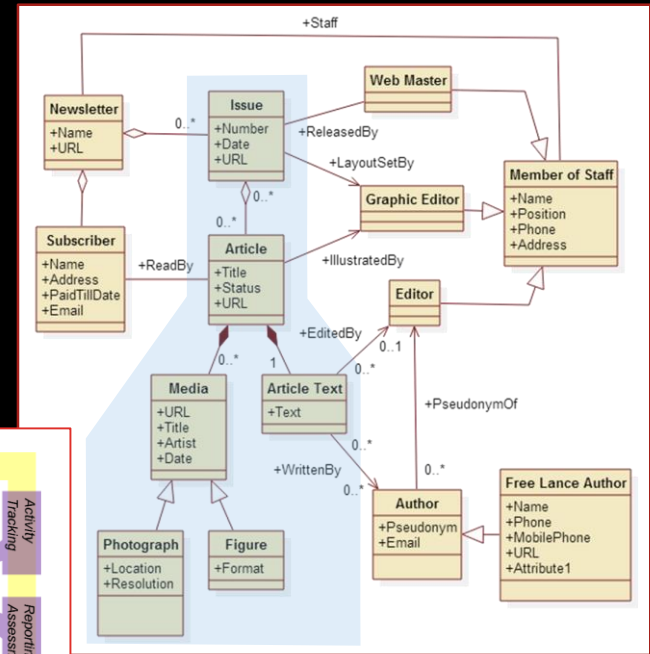
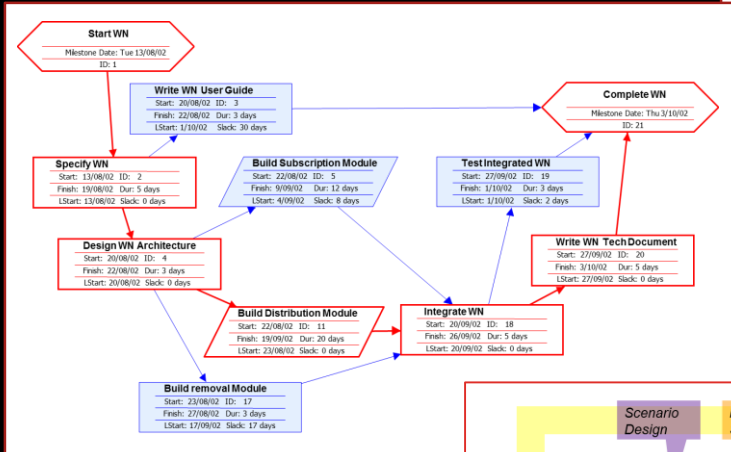
*Art and mass entertainment and
propaganda, they can all be
plotted on the same graph, but
there is a difference.*

*David Alan Mamet
American playwright*

GRAPH?

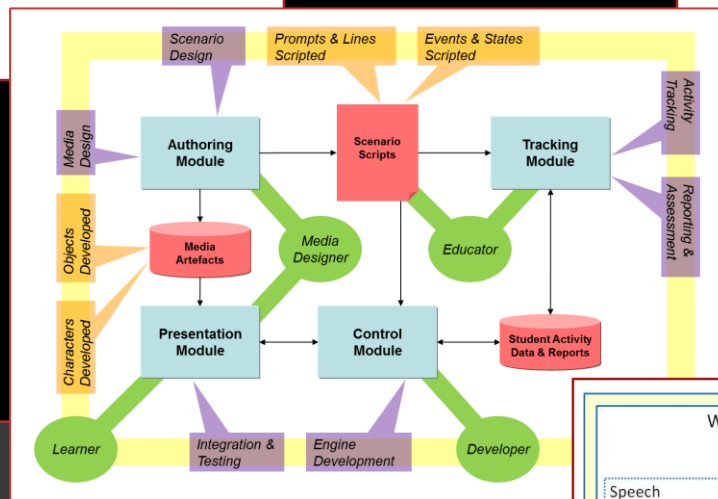


Representation Linked Matrices

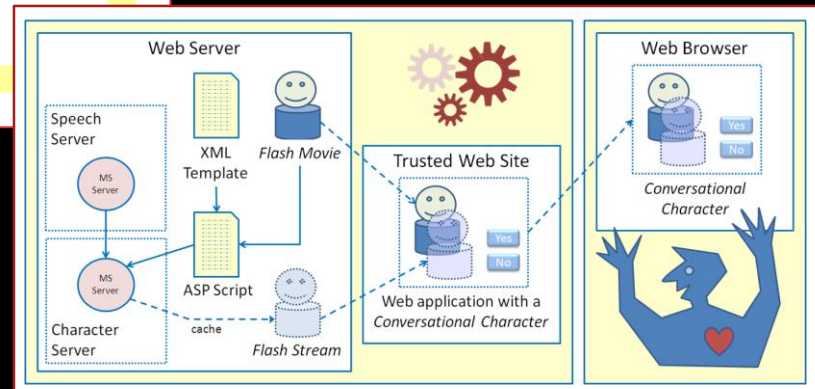


Graph Theory Vertices/Nodes, Edges/Links

Terminology Graph Network Flow Hierarchy Mesh



Visualisation Labels and Colours Layout Algorithms

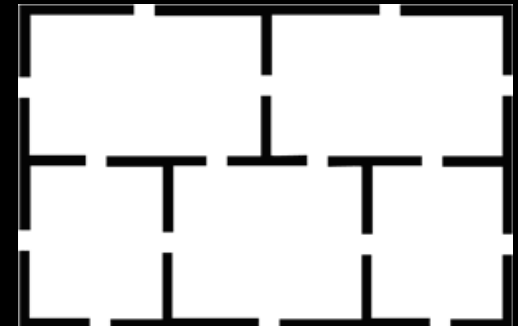


GRAPHS MORE FORMALLY

- Mathematically, a graph is:
 - a set of vertices $V = \{v_i\}$, and
 - a set of edges $E = \{(v_i, v_j) : v_i, v_j \in V\}$.
- Now we can define some useful properties of graphs, e.g.
 - Two vertices are **adjacent** if they belong to the same edge.
 - Vertices of an edge are **incident** to that edge.
 - A **loop** is an edge with the same vertex on both ends.
 - A **degree** of a vertex is the number of edges incident to it (loops being counted twice)
- Graphs have some interesting properties which have puzzled people for millennia. For example:
 - For a graph, the sum of degrees of all its nodes equals twice the number of edges.
 - In a party of people, some of whom shake hands, prove that an even number of people must have shaken an odd number of other people's hands.
 - Draw a continuous path through each of the doors of 5 rooms depicted in the diagram to the right, without going through any door twice, and without the path crossing over itself.

Most interestingly graphs can be drawn and large part of their mathematical properties are to do with their layout.

House of 5 Rooms



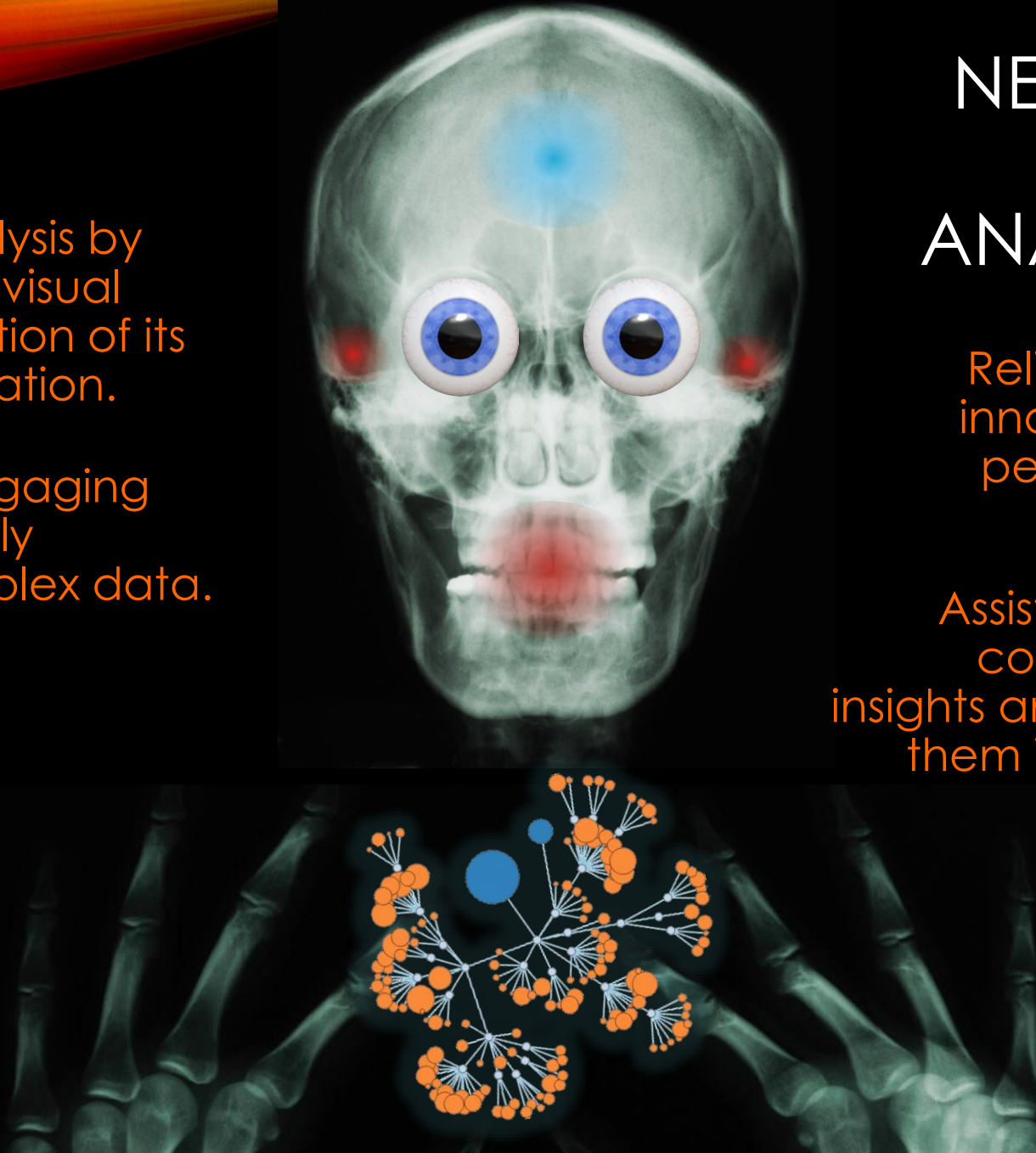
NEED FOR VISUAL ANALYTICS

Data analysis by means of visual manipulation of its representation.

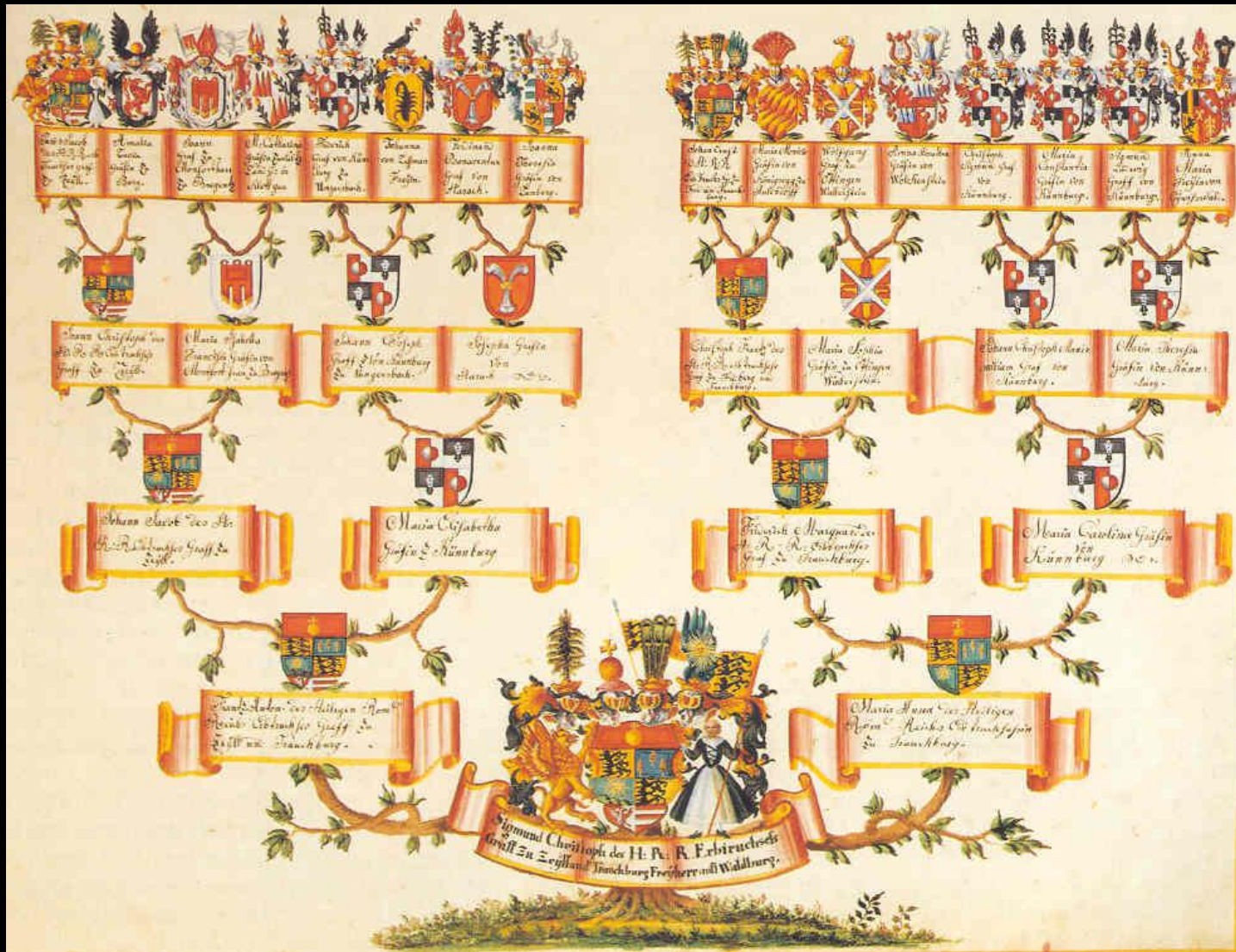
Allows engaging instinctively with complex data.

Relies on human innate abilities of perception and cognition.

Assist gaining and communicating insights and translating them into decisions and actions.

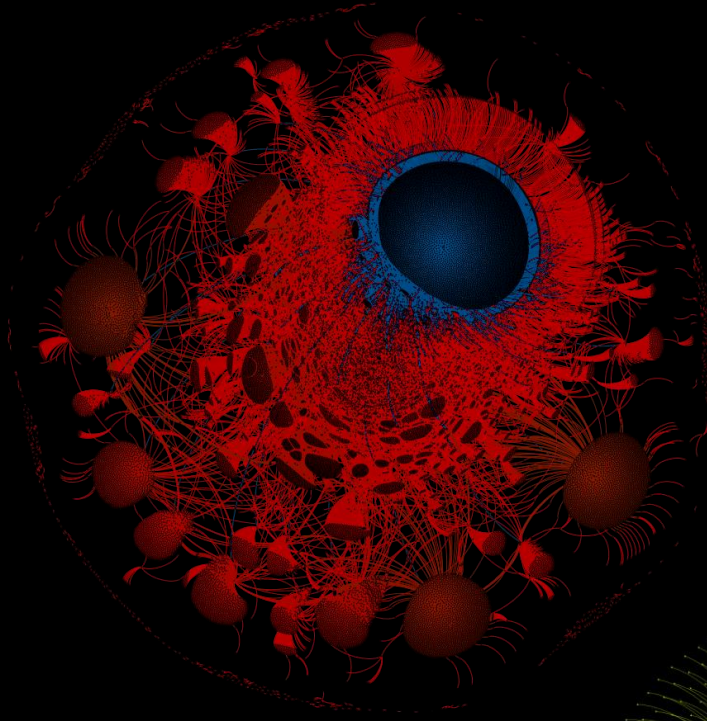


THE BEAUTY OF GRAPHS

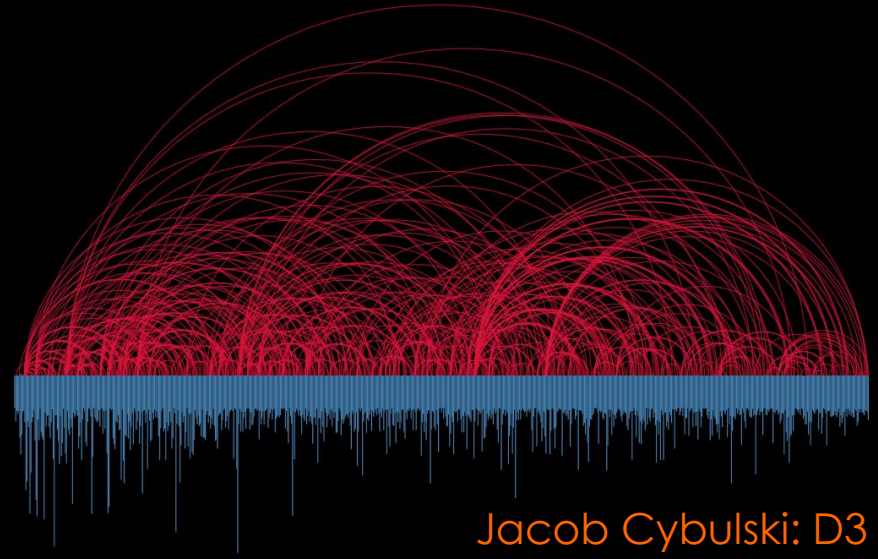


The family tree of Sigmund Christoph von Waldburg-Zeil-Trauchburg
 Wikimedia

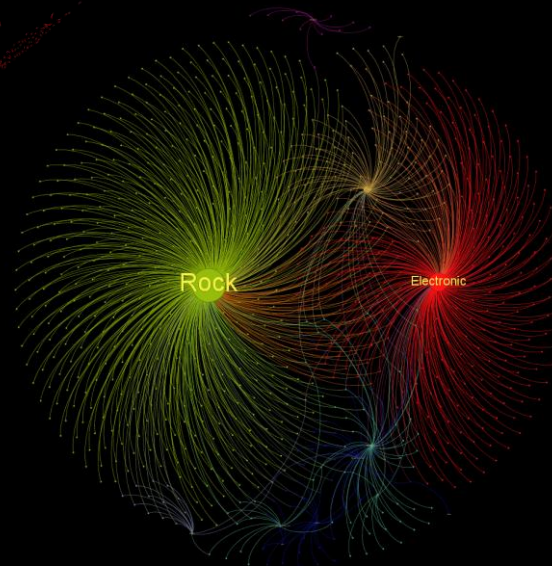
THE BEAUTY OF GRAPHS



Stuart Palmer: Gephi
Astronomy in Tweets
Comet Ison (73k+ Tweets)



Jacob Cybulski: D3
Persistence of Music
Discogs (1000 Best Songs)



Jacob Cybulski: Gephi
The Web of Musical Genres
Discogs (1000 Best Songs)

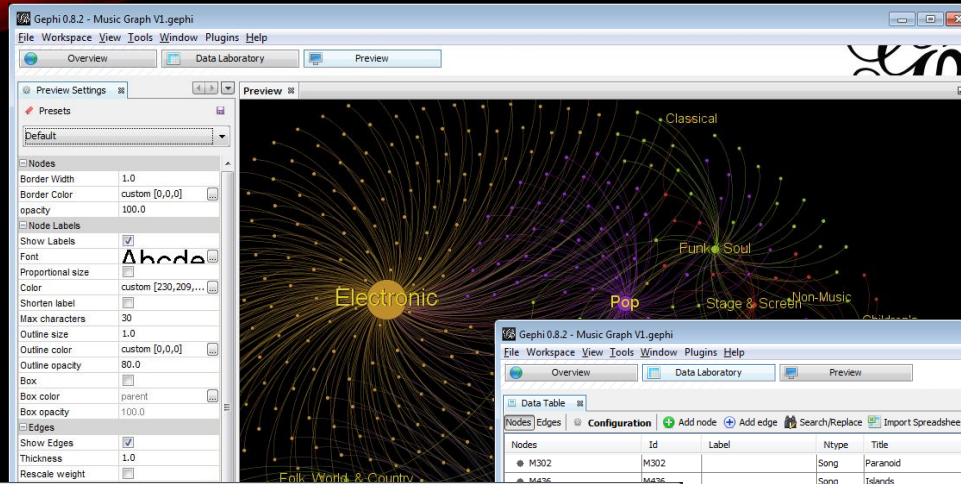
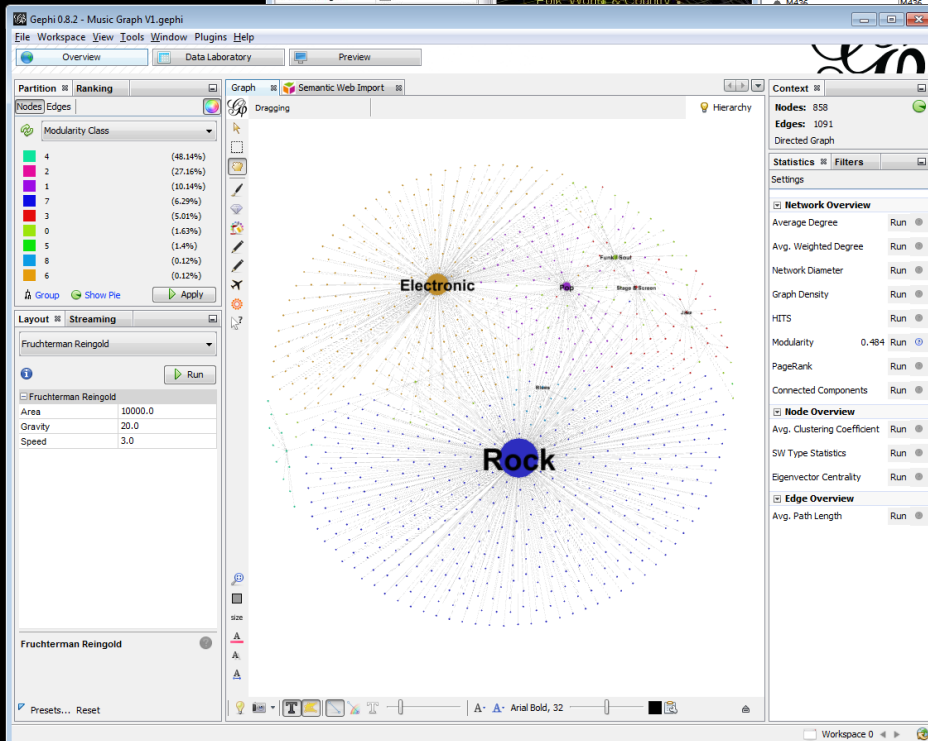
Gephi?

Data
Presentation

WHAT IS GEPHI?

Java
Netbeans
IDE for...

Data
Exploration



Nodes	Id	Label	Ntype	Title	Year	Country	Re...	M...
M302	M302	Paranoid	Song	Paranoid	1970	UK	152	4
M305	M305	Islands	Song	Lark's Tongues In Aspic	1973	UK	50	4
		Red	Song	Red	1974	UK	47	4
		Lizard	Song	Lizard	1970	UK	52	4
		Starless And Bible Black	Song	Starless And Bible Black	1974	UK	47	4
		Face Value	Song	Face Value	1981	UK	61	1
		Joyride	Song	Joyride	1991	Scandinavia	51	1
		Dare	Song	Dare	1981	UK	55	2
		Shades Of Deep Purple	Song	Shades Of Deep Purple	1968	UK	57	4
		Black Sabbath	Song	Black Sabbath	1970	UK	120	4
		Love	Song	Love	1985	Canada	60	4
		The Book Of Talesyn	Song	The Book Of Talesyn	1969	UK	46	4
		Deep Purple	Song	Deep Purple	1969	UK	48	4
		Sabbath Bloody Sabbath	Song	Sabbath Bloody Sabbath	1973	UK	103	4
		Undertow	Song	Undertow	1993	US	47	4
		In Rock	Song	In Rock	1970	UK	105	4
		The Rise And Fall Of Ziggy Stardust And The Spiders From Mars	Song	The Rise And Fall Of Ziggy Stardust And The Spiders From Mars	1972	UK	108	4
		Fireball	Song	Fireball	1971	UK	75	4
		Sonic Temple	Song	Sonic Temple	1989	Australia	56	4
		Hunky Dory	Song	Hunky Dory	1971	UK	79	4
		Machine Head	Song	Machine Head	1972	UK	112	4
		The Man Who Sold The World	Song	The Man Who Sold The World	1970	US	65	4
		Made In Japan	Song	Made In Japan	1972	UK	91	4
		Aladdin Sane	Song	Aladdin Sane	1973	UK	77	4
		Café Del Mar	Song	Café Del Mar	1993	Germany	123	2
		Who Do We Think We Are	Song	Who Do We Think We Are	1973	UK	49	4

Data
Management

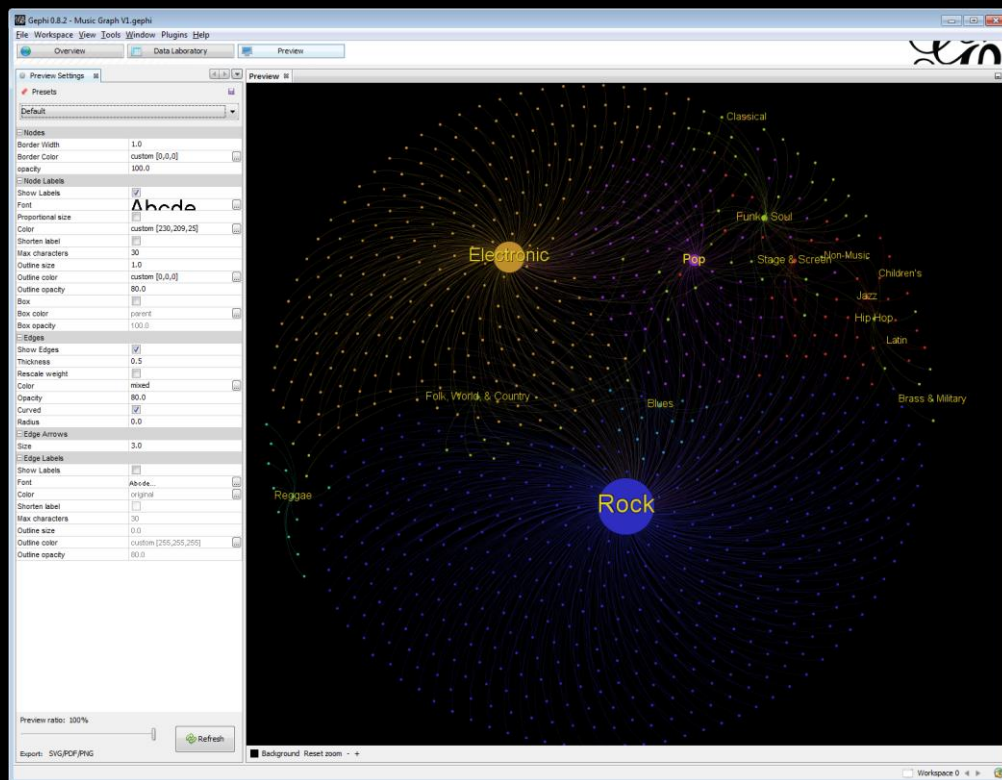
DISCOGS MUSIC RELEASES

Exploring data and its structure

	A	B	C	D
1	Source	Target	Type	
2	M302	G3	Directed	
3	M436	G3	Directed	
4	M444	G3	Directed	
5	M463	G3	Directed	
6	M498	G3	Directed	
7	M520	G3	Directed	
8	M553	G7	Directed	
9	M553	G3	Directed	
10	M563	G7	Directed	
11	M563	G3	Directed	
12	M669	G1	Directed	
13	M687	G3	Directed	

	A	B	C
1	Id	Label	Ntype
2	G1	Electronic	Genre
3	G2	Jazz	Genre
4	G3	Rock	Genre
5	G4	Hip Hop	Genre
6	G5	Stage & Sc	Genre
7	G6	Folk, Worl	Genre
8	G7	Pop	Genre
9	G8	Classical	Genre

	A	B	C	D	E	F
1	Id	Ntype	Title	Year	Country	RelTotal
2	M302	Song	Paranoid	1970	UK	152
3	M436	Song	Islands	1971	UK	53
4	M444	Song	Larks' Ton	1973	UK	50
5	M463	Song	Red	1974	UK	47
6	M498	Song	Lizard	1970	UK	52
7	M520	Song	Starless A	1974	UK	47
8	M553	Song	Face Valu	1981	UK	61
9	M563	Song	Joyride	1991	Scandinav	51
10	M669	Song	Dare	1981	UK	55



Let's Gephi?

MUSIC DATA - DISCOGS

Run 1 - Force Atlas:

- Inertia = 0.3
- Repulsion = 800.0
- Attraction = 9.0
- Adjust by Sizes = Yes

Nodes:

- Size = 5..100
- Colour = Shifted
- Label Size = 1..2.5
- Modularity = Yes (Partition)

Edges:

- Thickness = 0.5
- Opacity = 80%

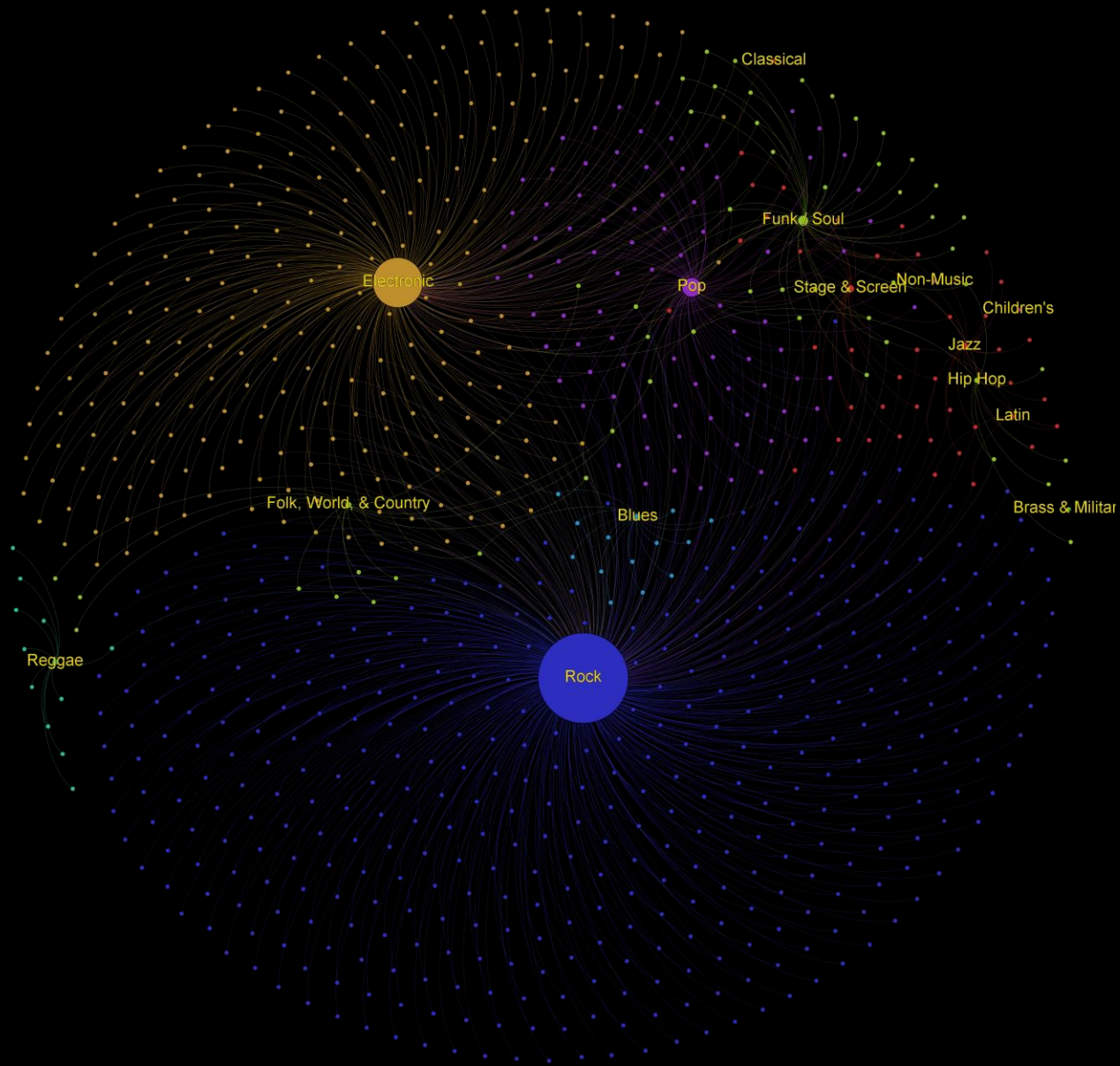
Preview:

- Show Labels
- Font = 24 pts
- Outline = 0.5 Light Grey
- Edges = 0.5, Rescale, Mixed

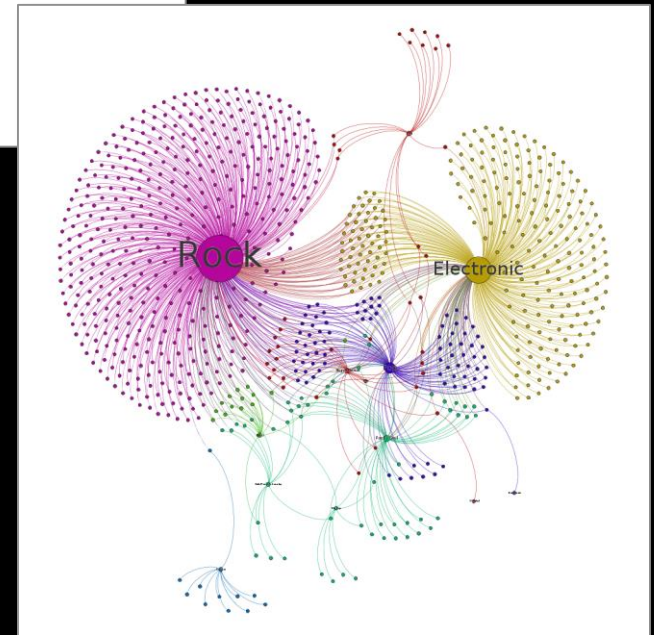
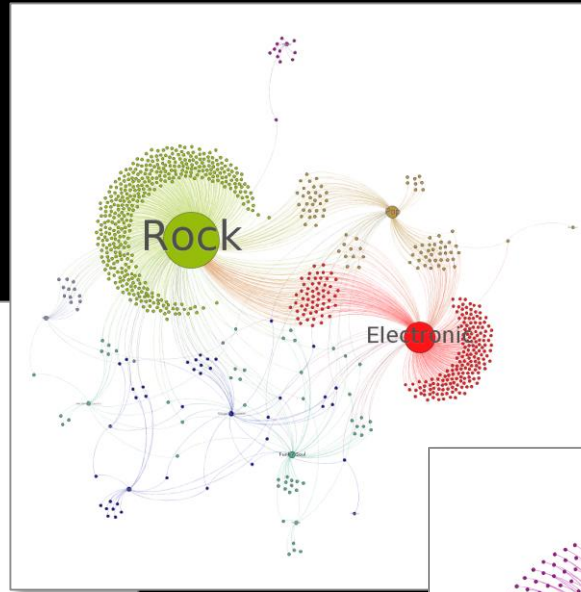
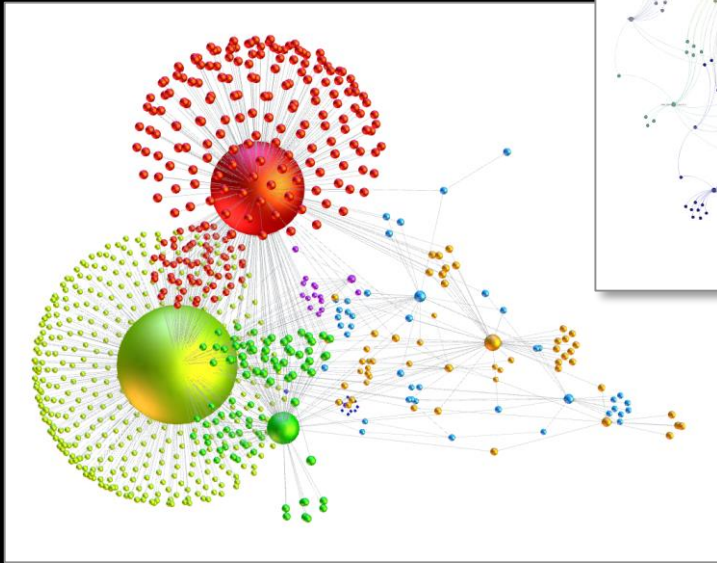
Steps:

1. Import nodes songs + genres
2. Import edges
3. Change node colours
4. Modularise and apply
5. Select Force Atlas layout
6. Run and change layout options
7. Define node and edge options
8. Switch to preview
9. Adjust preview options
10. Try Curved and Straight edges
11. Try node opacity and label boxes
12. Go to Data Lab see new column
13. Save project
14. Save PNG from Preview
15. Scramble
16. Try other layouts (Force Atlas 2)
17. Try helper layouts to rotate
18. Try Yifan Hu
19. Try Circular
20. Then Fruchterman Reingold
21. Save and export

RESULTS



RESULTS



(Job 14766) Interview with Angie, Part 1
 Note that noise and illegible parts have been removed by Jacob
 (I1: Interviewer David I2: Interviewer Jacob P: Participant Angie)

INTERVIEW TRANSCRIPT
 (Job 14828) Interview with Colin – Part 2
 (I: Interviewers/CS: and DW: Participants)

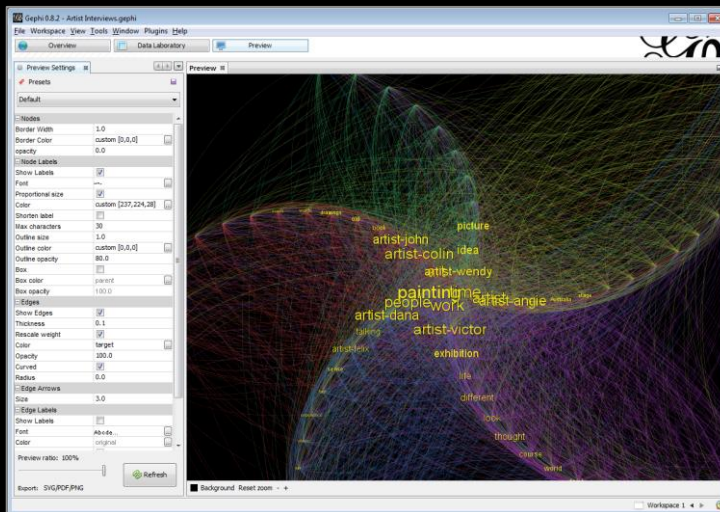
- I: So here's an interview with Colin Singleton and Deidre Walsh, and it's the 7th of February and we'd like to hear some of Colin and Deidre's views of the world in the art world, the impact of the society on the arts, the impact of technology on the arts but I think we'll start with maybe Colin's journey, personal journey into the art world.
- CS: You're going to start with me. I've got drawings I did in 1949, so you know I was only six or seven so drawings of whales actually. I've copied them into a diary which has got the date on it, 1947, so I know when I did it. So I've always wanted to draw, and it's always something that attracted my parents' approval. Most of my life was attracted to their disapproval and that was about the only area where I actually had continuous affirmation of who I was.
- I: Did the family have an artistic background?
- CS: Yeah yeah, on both sides. They were professional artists on one side and on the other side they were amateurs, quite capable amateurs. My grandfather who died at 32 was quite good at it and his daughter, my father's sister, could also paint. But my grandfather could and his brothers could, there's some of their paintings out there. And Charlie Burge who was the head artist for the Herald Sun, and he painted prolifically until a trip across the Nullarbor, the car turned upside down and he was under it for a couple of days and lost his right arm and he had to re-learn to paint with his left. Which he did do, but not anywhere near the same sort of capacity that he did without it. I've got some of his work up, a lot of his work upstairs.

INTERVIEWS

conceptcc-inverse-prominence-2-commas.csv - Excel

TF-IDF Matrix

	A	B	C	D	E	F	G	H	I	J
1		artist-colin	painting	artist-victor	artist-angie	artist-dana	people	time	art	artist-wendy
2	artist-colin	0.17728332	1.3233824	0	0	0	0.5682563	1.231134	0.7233974	0
3	painting	1.3233824	0.0770797	0.3283024	1.4627423	1.2726715	0.7095285	0.65692925	1.6096056	1.4940615
4	artist-victor	0	0.3283024	0.0836242	0	0	0.68229747	0.5501592	0.5208174	0
5	artist-angie	0	1.4627423	0	0.24011053	0	1.2492707	1.1847559	2.5070364	0
6	artist-dana	0	1.2726715	0	0	0.10805701	0.90215033	0.8104276	0.65017354	0
7	people	0.5682563	0.7095285	0.68229747	1.2492707	0.90215033	0.052210584	0.8512596	0.54513997	2.6400397
8	time	1.231134	0.65692925	0.5501592	1.1847559	0.8104276	0.8512596	0.05453752	0.74464697	1.0935713
9	art	0.7233974	1.6096056	0.5208174	2.5070364	0.65017354	0.54513997	0.74464697	0.05162885	1.5801146
10	artist-wendy	0	1.4940615	0	0	0	2.6400397	1.0935713	1.5801146	0.16913903
11	artist-john	0	0.7608512	0	0	0	1.2953869	1.3175117	1.2472445	0
12	work	1.264107	0.9363756	0.6857184	2.6252084	0.78761554	0.74456835	0.8944154	0.65131783	0.566098
13	artist	0.9159637	0.86889845	0.64808756	1.4886853	0.45420575	0.517929	0.67626524	0.45728415	2.9961772
14	look	0.9273281	0.714739	0.7417103	0.8029926	1.574545	0.44379	1.3907068	1.0532286	0.9080095
15	artist-felix	0	0.83427435	0	0	0	0.8005624	0.62718153	0.6551523	0
16	talking	0.6281995	0.8974279	0.6490831	1.7790009	0.97851634	0.56735504	0.555601	0.56103355	1.0603716
17	life	0.8697962	1.2101512	1.8755713	1.6390154	0.3769989	0.63054323	0.9513768	0.54038197	1.6596768
18	idea	0.5071127	0.4514668	0.20571554	2.6848722	4.4303374	0.5838093	0.47915107	0.5773045	4.1608205
19	picture	2.600155	0.67830133	0.6132441	1.7608105	0.30347925	0.53013825	0.65445024	0.9735727	5.5815883
20	different	1.1943296	0.4484637	0.41168836	1.5367073	0.69156486	0.47735396	0.5817336	0.47203523	1.0824898
21	thought	0.7851118	0.6827059	0.51847005	0.82704735	0	0.9248733	0.52020407	1.0669963	1.23372
22	exhibition	3.1910994	0.83246076	0.2580404	1.8522811	0.4668063	0.626527	0.45308095	0.79655945	3.653403
23	course	1.7551047	1.9077226	0.63683045	1.828534	0.39620903	0.64610606	0.49083772	0.5679174	1.8605294



In search of the missing link(s)
 we will rely on graph adjacency matrix representation and TF-IDF measures, i.e. term frequency – inverse document frequency

INTERVIEWS - ARTISTS

Run 2 – Radial Axis:

- Scaling Width = 200.0
- Node Size = 5
- Group Nodes by: Modularity

Nodes:

- Size = 10..400
- Colour = Blues
- Label Size = 1..5
- Modularity = Yes (Partition)
- Opacity = 0%

Edges:

- Thickness = 0.1
- Opacity = 100%

Preview:

- Black
- Show Labels
- Font = 3 pts
- Outline = 0.5 Light Grey
- Edges = 0.1, Rescale, Target Colour

Preparation:

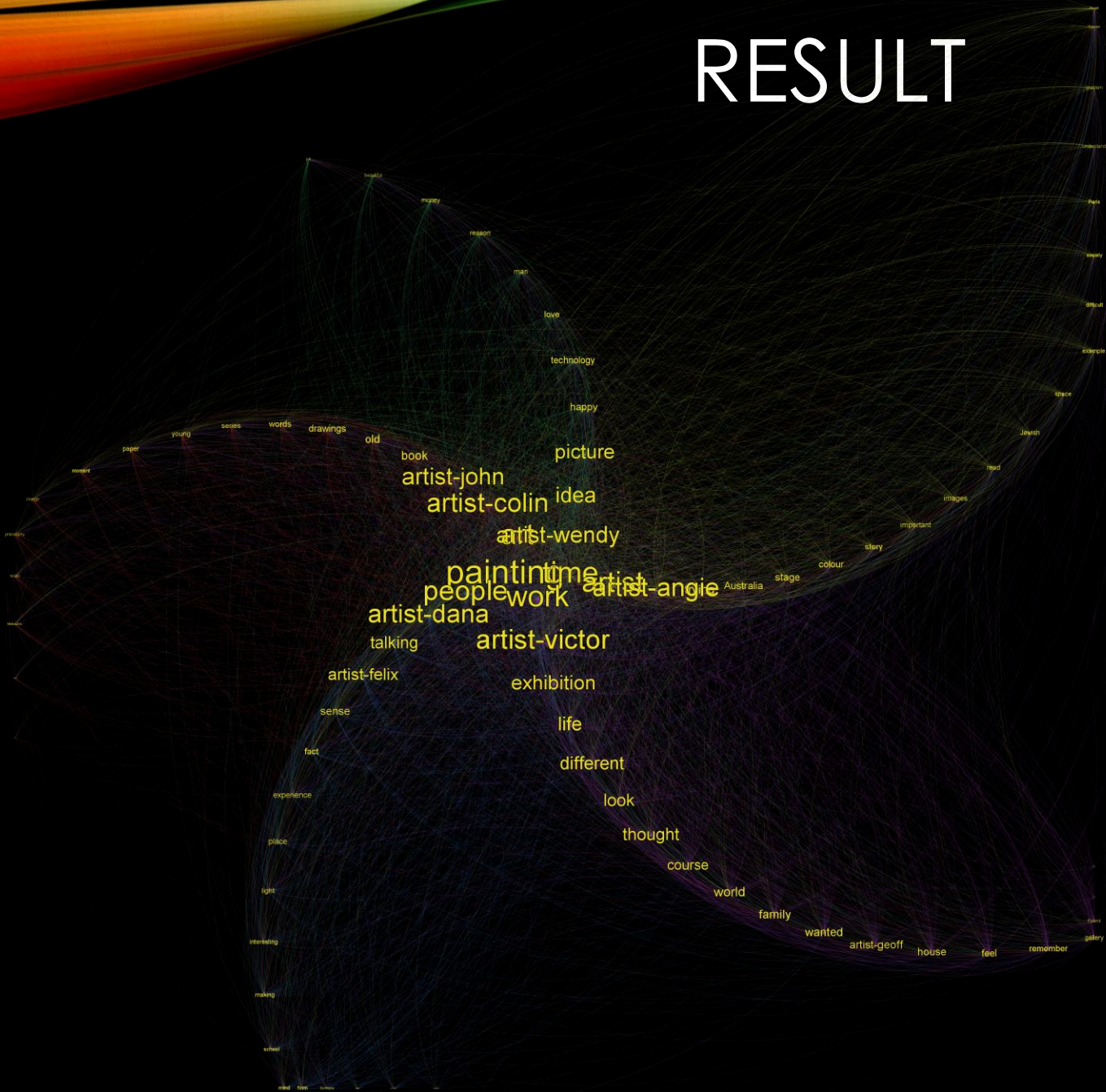
- Obtain interview documents
- Split them into statements attributed to interviewees (collection of interviewee statements are considered documents)
- Split all documents into paragraphs and then into words
- Convert all words into their root terms - remove tenses, plural forms, etc.
- Remove all stop words – a, the, etc.
- Identify N most frequent and useful terms
- Identify pairs of co-occurring frequent terms
- Calculate TF-IDF weights based on their overall frequency and their frequency in each document
- Create a TF-IDF matrix of term pairs

Steps:

- Create ; separated Excel CSV
- Load the S-CSV file
- Note edge weights
- Select Radial Axis Layout
- Run Modularity
- Save PNG
- Experiment with other layouts and their parameters

Or use
a tool to
do this for
you

RESULT



SUMMARY & REFLECTION

What is a graph?

Mathematical structure
Data structure
Consists of vertices and edges
Features layouts
Requires algorithms
Great many applications

What software assists graph visualization?

There are many
Some are open source

- Gephi
- Pajek
- Cytoscape
- KH Coder
- Blender

Current projects:

Teaching 1st year Business Analytics
Passer-By Analytics
Impact of sitting

Data sets:

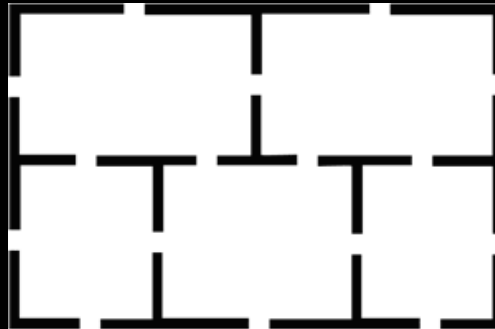
Sales of movie tickets
Music distribution
Real-estate sales
Employee office mobility

Future work:

Experiments
Collaborative analytics
Shared and private analytic spaces
Support for variety of interfaces
Back-end analytics
Dynamic data

HINT FOR THE SOLUTION TO THE HOUSE OF 5 ROOMS PUZZLE

House of 5 Rooms



If all fails, try
drawing the
house on a 3D
surface.