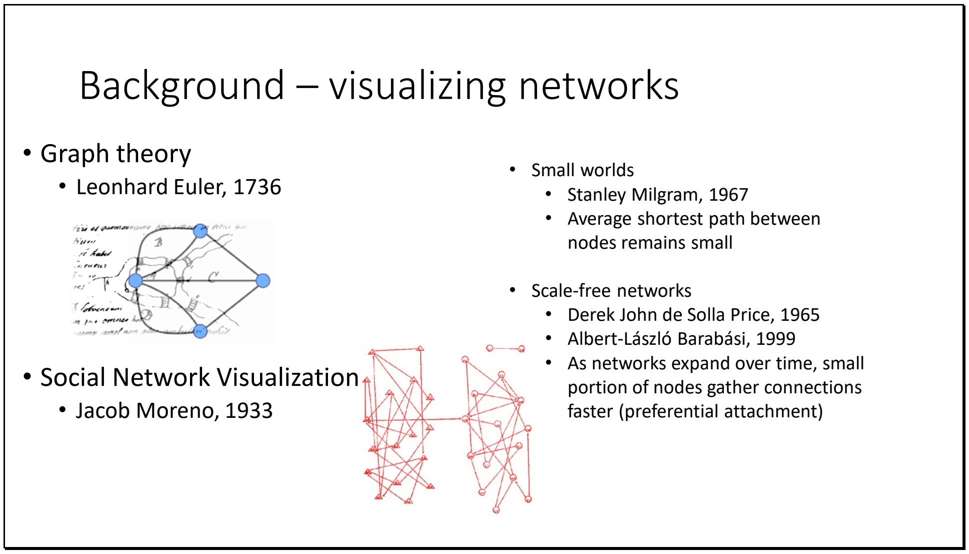
Tomasz Neugebauer
Speaking Notes

Neugebauer, Tomasz, Tayler, Felicity and MacDonald, Corina (2015) *Metadata as a Complex Network: A Case Study of Data Visualization for Art Historical Research.* In: Constellations, Clusters, Networks. 2015 AHGSA (Art History Graduate Student Association) Annual Conference, March 6-7, 2015, Concordia University.
https://spectrum.library.concordia.ca/980078/



Background – visualizing networks

- Graph theory
  - Leonhard Euler, 1736
- Social Network Visualization
  - Jacob Moreno, 1933
- Small worlds
  - Stanley Milgram, 1967
  - Average shortest path between nodes remains small
- Scale-free networks
  - Derek John de Solla Price, 1965
  - Albert-László Barabási, 1999
  - As networks expand over time, small portion of nodes gather connections faster (preferential attachment)

The background literature about networks and their properties is vast and multidisciplinary, including:

- its roots in **mathematics**/graph theory, starting with the work of Euler in the 18th century
- social network statistics and visualization in **sociology** and **sociometrics,** starting with the work of Moreno and Milgram,
- scientometrics/information science an the study of the Web, with the recent work of Barabasi.

In doing this research, we wanted to check if/how some of the previous notable findings apply to our metadata network. So we wanted to check what would be the average shortest path length in the network, following the work of Milgram. Milgram is known for his “small world” experiment in 1967 tracking chains of acquaintances in the US. He sent packages to random people living in Omaha, asking them to forward the package to a friend or acquaintance who they thought would bring the package closer to a set individual. He was able to determine the average shortest path to be 6, in that context.

We also wanted to see if we would find a similar power-law node degree distribution in our network, with some nodes forming “hubs” of high degree and preferential attachment, as was found by Barabasi and others in the study of the Web.

The background of this work includes mathematical roots in graph theory, beginning with Leonhard Euler.

Euler:

- Konigsberg - find a route where one could cross all seven bridges, without crossing the same one twice.
- 1736: Solution of a problem relating to the geometry of position : proves that such a path does not exist.
    - Instead of listing all possibilities (too many), he created an abstraction of it, where bridges are edges (links), and lands are nodes.
    - This simplified scheme is a graph
    - Proof was based on the node degree – for such a path to exist, there had to be enough nodes of a certain degree – since there weren't, it was a proof (by contradiction).

Euler inspired much research into graphs and the properties of the nodes and paths in graphs. We are also representing a bibliographic metadata set as a graph.

Moreno (Jacob) – Romanian doctor

- 1933 New York Times article: "Emotions Mapped by New Geography: Charts Seek to Portray the Psychological Currents of Human Relationships"
- Includes a visualization of relationships in a class of fourth graders
- Inventor of sociograms - graphic representation of social links
- Early visualizations of social networks.

We are also attempting to provide novel compelling graphic representations of links between actors (authors, artists, publishers, critics, etc.)

Stanley Milgram's

- "small world experiment" in 1967 that tracked chains of acquaintances in the United States. In the experiment, Milgram sent several packages to 160 random people living in Omaha, Nebraska, asking them to forward the package to a friend or acquaintance who they thought would bring the package closer to a set final individual.
- "six degrees of separation"

We wanted to confirm if we will find the same 'small world' property in the metadata catalogue represented as a network.

John de Solla Price

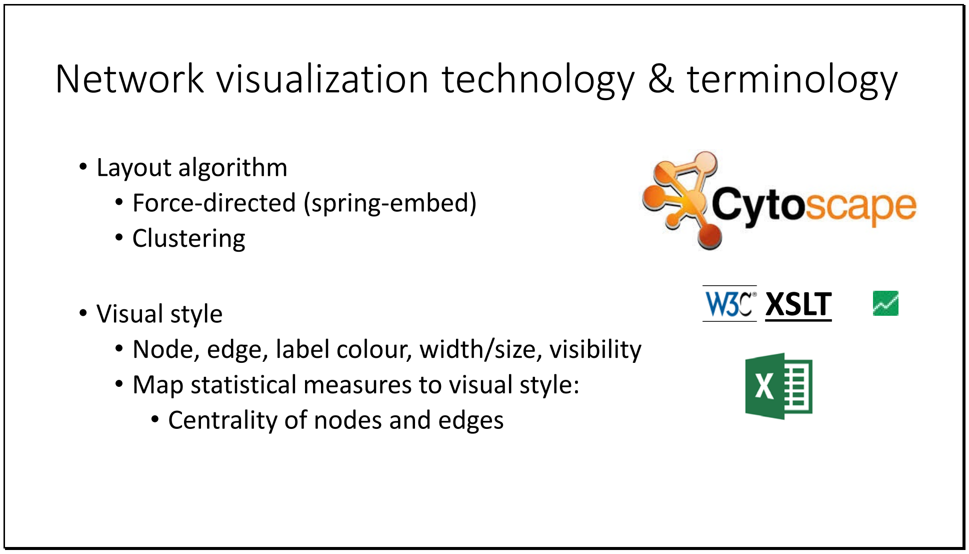
- Inventor of scientometrics

- In studies of the networks of citations between scientific papers, showed in 1965 that the number of links to papers—i.e., the number of citations they receive—had a heavy-tailed distribution following a Pareto distribution or power law.

Albert-László Barabási

- mapped the topology of a portion of the World Wide Web, finding that some nodes, which he called "hubs", had many more connections than others and that the network as a whole had a power-law distribution of the number of links connecting to a node.
- After finding that a few other networks, including some social and biological networks, also had heavy-tailed degree distributions, Barabási and collaborators coined the term "scale-free network" to describe the class of networks that exhibit a power-law degree distribution.

We also wanted to confirm if we find the same heavy-tailed (power law) distribution of node degree in our network.



Network visualization technology & terminology

- Layout algorithm
  - Force-directed (spring-embed)
  - Clustering
- Visual style
  - Node, edge, label colour, width/size, visibility
  - Map statistical measures to visual style:
    - Centrality of nodes and edges

Cytoscape
W3C XSLT
X

A few words about the technology that I used for this work:
Data transformation/shaping was done using XSLT programming.
I used the open source Cytoscape Software for generating the statistics and the images.
While Cytoscape is most commonly used for biological research applications, it can be used to visualize and analyze network graphs of any kind involving nodes and edges (e.g., social networks).

I also used a plugin for tapping in to the GPU of my computer to accelerate the computation time required to process the data.   The networks contained hundreds of thousands of nodes and edges.

We also developed a web based viewer for publishing our networks to the web and navigating the results, with direct links back to e-artexte catalogue. We extended the Cytoscape Web application to achieve this.

When rendering networks, there are two main sets of parameters/options/configurations:

1. the layout algorithm for the positioning of nodes and edges
2. Visual style of nodes and edges.

Layout algorithm:

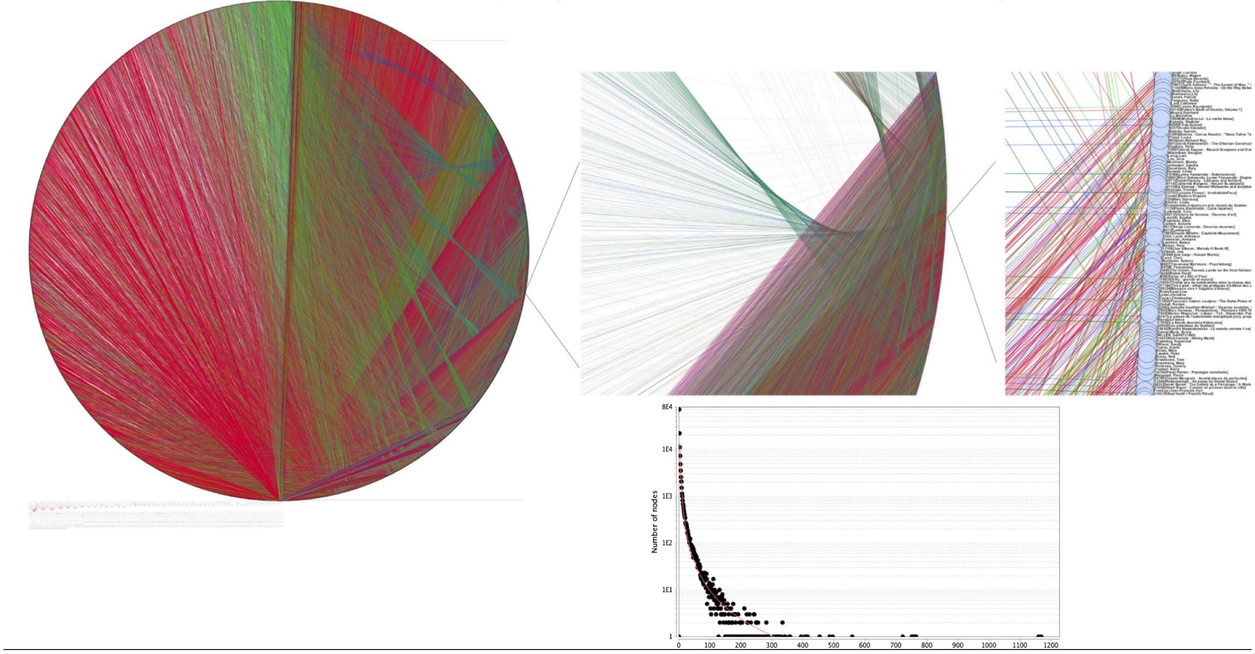
I used variations of Force-Directed, circular and grid based layouts/algorithms.
Force-directed algorithms simulate the network as a physical system.

Visual Style:

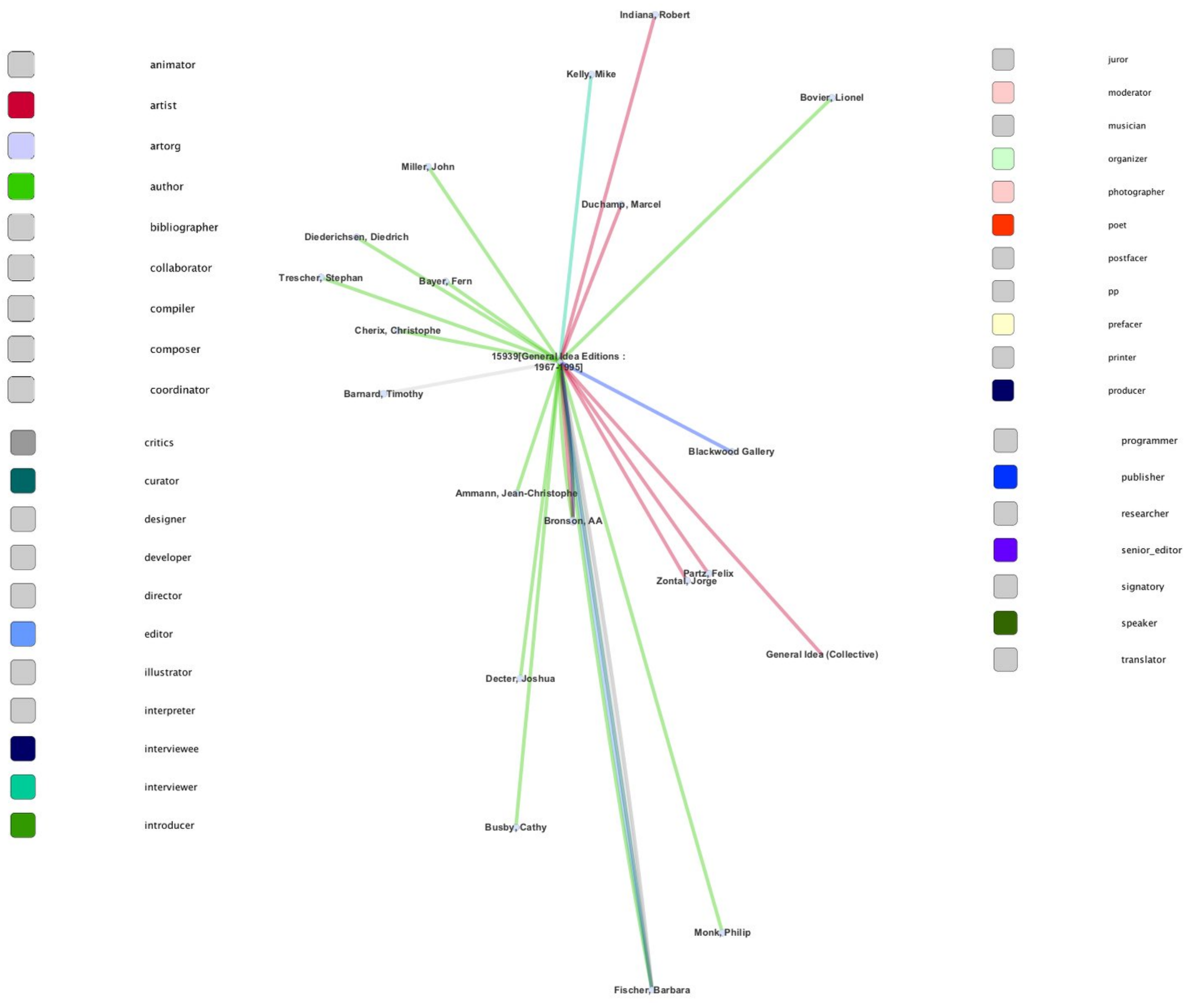
We used colour of edges to communicate the role (artist, author, editor, publisher, etc.)

We also mapped network statistics, such as node centrality to the visual property of node size and edge width.   Centrality is an attempt to measure the importance of a node in the network.

Circular (degree sorted) layout of e-artexte catalogue



Part 1 – Single node:

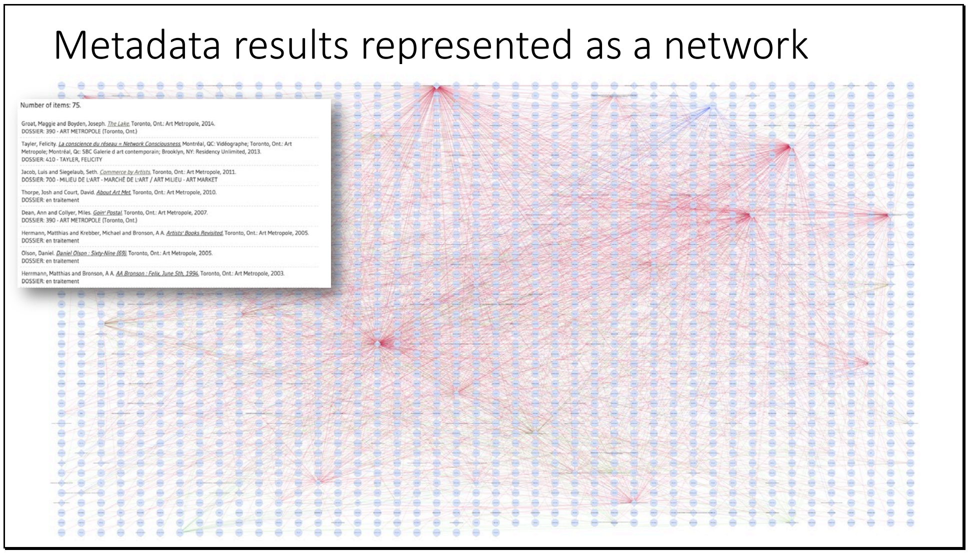


We chose to treat the graphs as bi-partite, where there are nodes for documents or actors (authors, artists, editors, publishers, etc.) and map the role types to edge colours.   Corina already showed you what a single node looks like.

This was intuitive, fast and efficient in terms of data pre-processing – requiring little reshaping of the data – and allowing us to generate visualizations for the entire catalogue of 20,000 + items

A visualization of the entire e-artexte metadata set.
I included 3 zoom levels: complete zoom out, zoom about half way, and then zoom in so that you can get an idea for the scale.
More than 120,000 nodes, 300,000 edges.

This is a Degree sorted circular layout; the definition of node degree is the number of edges that a node has. This layout arranges the high degree nodes near the bottom of the circle, and as it circles around clockwise, the degree descends. As you go around the circle, you see the high-degree nodes on the bottom, then the individual artists only on the left side towards the bottom.



Metadata results represented as a network

Number of items: 75.

Groat, Maggie and Boyden, Joseph. *The Lake.* Toronto, Ont.: Art Metropole, 2014.
DOSSIER: 390 - ART METROPOLE (Toronto, Ont.)

Tayler, Felicity. *La conscience du réseau = Network Consciousness.* Montréal, QC: Vidéographe; Toronto, Ont.: Art Metropole; Montréal, Qc: SBC Galerie d art contemporain; Brooklyn, NY: Residency Unlimited, 2013.
DOSSIER: 410 - TAYLER, FELICITY

Jacob, Luis and Siegelaub, Seth. *Commerce by Artists.* Toronto, Ont.: Art Metropole, 2011.
DOSSIER: 700 - MILIEU DE L'ART - MARCHÉ DE L'ART / ART MILIEU - ART MARKET

Thorpe, Josh and Court, David. *About Art Met.* Toronto, Ont.: Art Metropole, 2010.
DOSSIER: en traitement

Dean, Ann and Collyer, Miles. *Goin' Postal.* Toronto, Ont.: Art Metropole, 2007.
DOSSIER: 390 - ART METROPOLE (Toronto, Ont.)

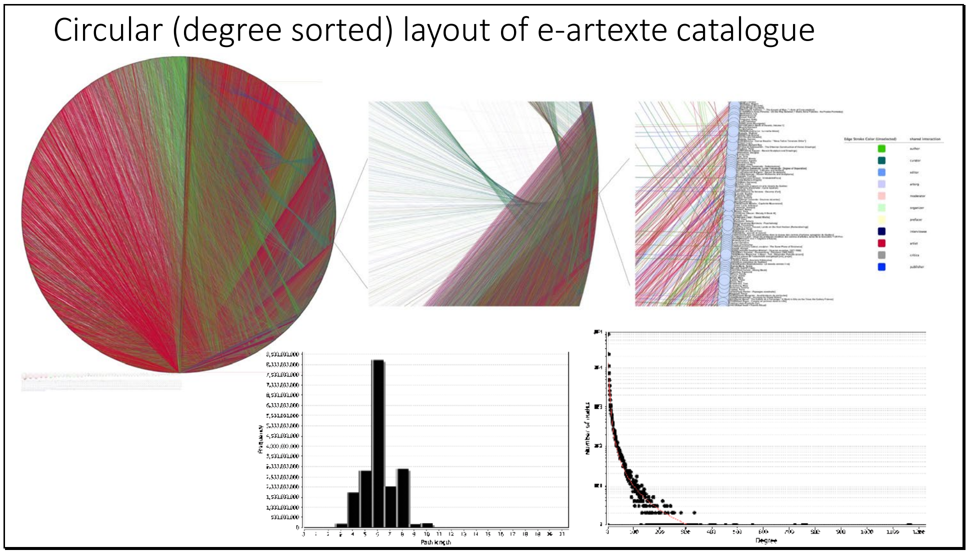
Hermann, Matthias and Krebber, Michael and Bronson, A A. *Artists' Books Revisited.* Toronto, Ont.: Art Metropole, 2005.
DOSSIER: en traitement

Olson, Daniel. *Daniel Olson : Sixty-Nine (69).* Toronto, Ont.: Art Metropole, 2005.
DOSSIER: en traitement

Herrmann, Matthias and Bronson, A A. *AA Bronson : Felix, June 5th, 1994.* Toronto, Ont.: Art Metropole, 2003.
DOSSIER: en traitement

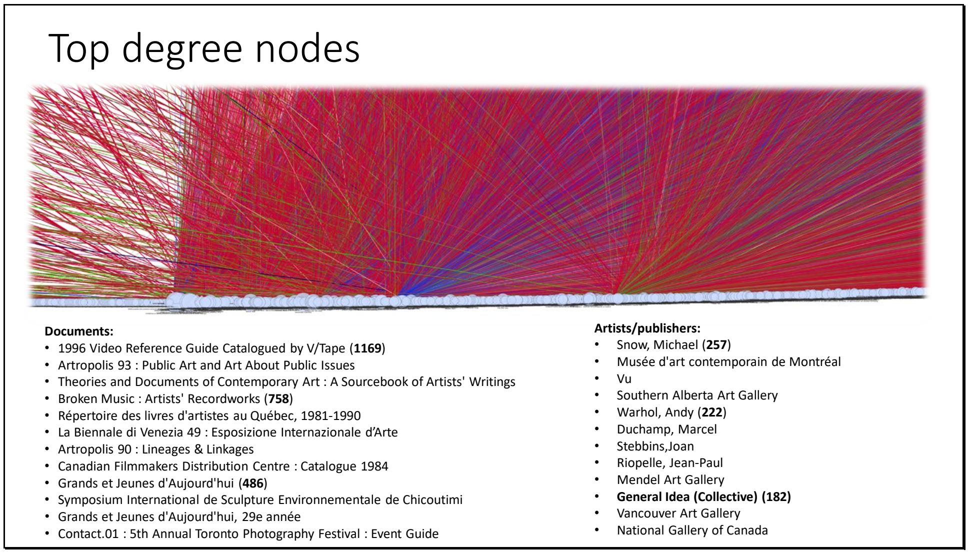
Part 2: grid layout
One of the computationally quickest layout algorithms in Cytoscape positions the nodes on a grid. And this shows what a bibliographic dataset consisting of approximately 200 e-artexte items looks like as a network with a grid layout.

Instead of a result set in the form of a list, you see a grid of nodes representing authors, artists, editors, etc, connected to documents by edges.



Circular (degree sorted) layout of e-artexte catalogue

The graph shows the relationship between node degree and the number of nodes in the network. Networks have been found to have a 'scale-free property' when a power law relationship between node degree and the number of nodes in the network follows this type of long-tailed power relation. This is visible on this layout of the graph, very few nodes have a particularly high degree (1000), near the bottom right of the circle, while the overwhelming majority have the degree of under 10.



Top degree nodes

Documents:

- 1996 Video Reference Guide Catalogued by V/Tape (**1169**)
- Artropolis 93 : Public Art and Art About Public Issues
- Theories and Documents of Contemporary Art : A Sourcebook of Artists' Writings
- Broken Music : Artists' Recordworks (**758**)
- Répertoire des livres d'artistes au Québec, 1981-1990
- La Biennale di Venezia 49 : Esposizione Internazionale d'Arte
- Artropolis 90 : Lineages & Linkages
- Canadian Filmmakers Distribution Centre : Catalogue 1984
- Grands et Jeunes d'Aujourd'hui (**486**)
- Symposium International de Sculpture Environnementale de Chicoutimi
- Grands et Jeunes d'Aujourd'hui, 29e année
- Contact.01 : 5th Annual Toronto Photography Festival : Event Guide

Artists/publishers:

- Snow, Michael (**257**)
- Musée d'art contemporain de Montréal
- Vu
- Southern Alberta Art Gallery
- Warhol, Andy (**222**)
- Duchamp, Marcel
- Stebbins,Joan
- Riopelle, Jean-Paul
- Mendel Art Gallery
- **General Idea (Collective) (182)**
- Vancouver Art Gallery
- National Gallery of Canada

“s.n” is node with second highest degree - publisher can not be determined, then enter s.n. (which stands for sine nomine) - we see that cataloguing practice is certainly a factor that can influence the results.

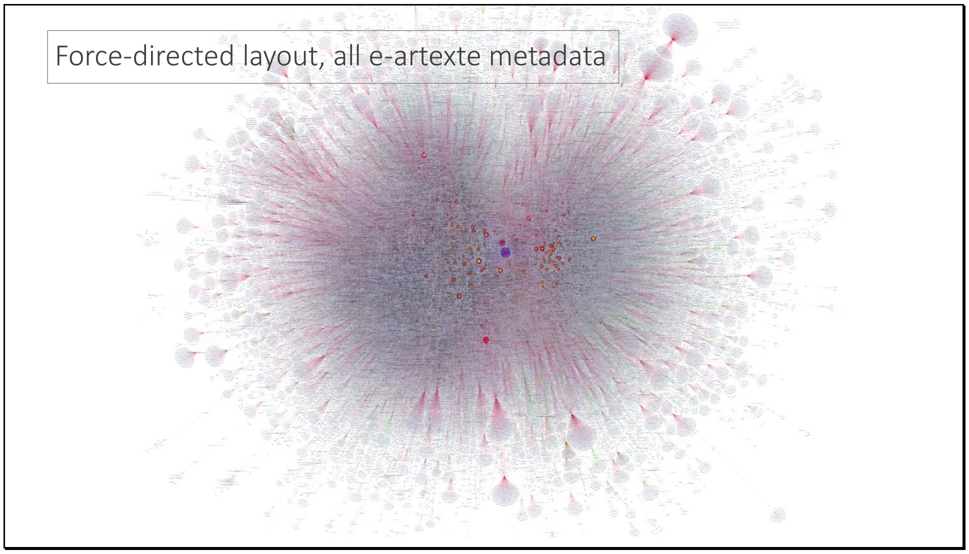
If we zoom in on that bottom of the circle, we can identify the top-degree nodes. These are the documents or artists/publishers with the most edges.

In terms of documents, large catalogues, like a Biennale, that include a lot of artists are at the top.

The top artist/publisher nodes include:

- well-known artist names, such as Michael Snow, Andy Warhol and Marcel Duchamp.
- General Idea (Collective).
- artist run centeres, such as the Quebec based “Vu”
- as well as large Canadian galleries: Montreal Museum of Contemporary Art, Vancouver Art Gallery, National Gallery of Canada.

Node degree is one measure of importance in the collection.



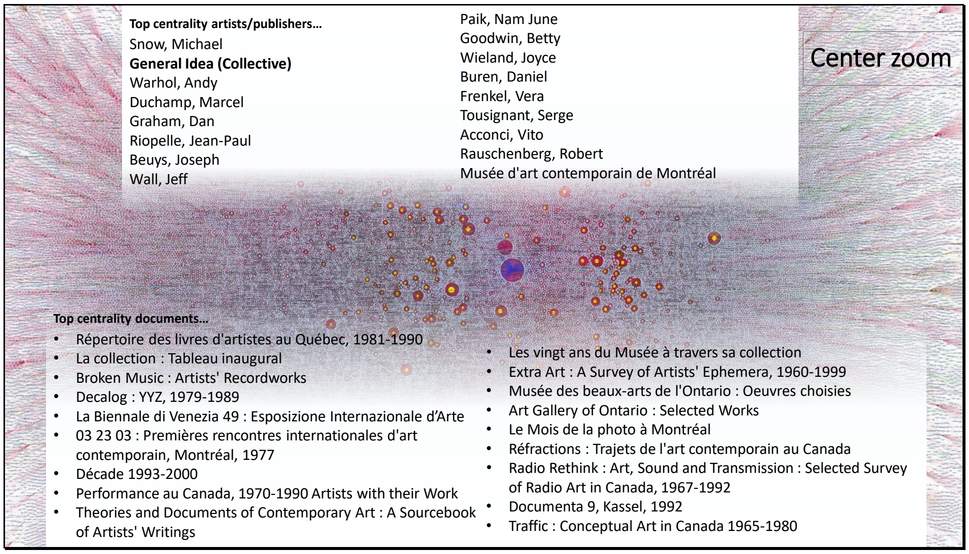
Force-directed layout, all e-artexte metadata

An image that lays out the entire e-artexte metadata set according to a force-directed layout, with some default circular gravity.

- **Force-directed** algorithms simulate the network as a physical system. The nodes are considered physical objects and edges are forces, think of it as springs, connecting those objects together. The algorithm then attempts to reach a "steady state" of minimized energy.
  - The algorithms position the nodes so that all the edges are of more or less equal length and there are as few crossing edges as possible.
  - Closely related items end up positioned closer together

Characteristic path length is 6 – that's how many 'hops' it typically takes to get from any one node to any other. The diameter is 20 – that is the maximum number of hops it takes to get from one node to another.

There is a handful of nodes that stand out near the center of the network that are larger than the others, because this visualization maps betweenness centrality to node size - these have higher centrality than others.



Center zoom

Top centrality artists/publishers…

Snow, Michael
General Idea (Collective)
Warhol, Andy
Duchamp, Marcel
Graham, Dan
Riopelle, Jean-Paul
Beuys, Joseph
Wall, Jeff
Paik, Nam June
Goodwin, Betty
Wieland, Joyce
Buren, Daniel
Frenkel, Vera
Tousignant, Serge
Acconci, Vito
Rauschenberg, Robert
Musée d'art contemporain de Montréal

Top centrality documents…

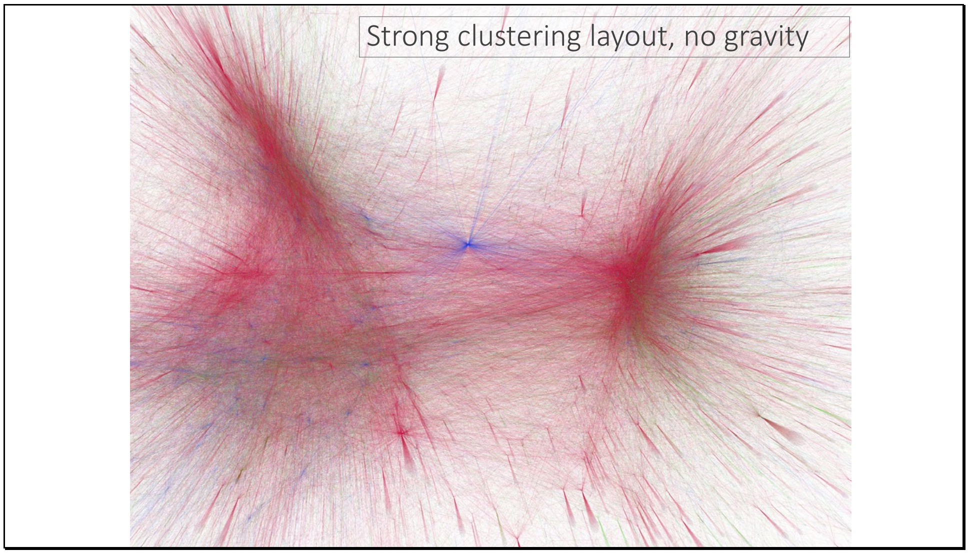
- Répertoire des livres d'artistes au Québec, 1981-1990
- La collection : Tableau inaugural
- Broken Music : Artists' Recordworks
- Decalog : YYZ, 1979-1989
- La Biennale di Venezia 49 : Esposizione Internazionale d'Arte
- 03 23 03 : Premières rencontres internationales d'art contemporain, Montréal, 1977
- Décade 1993-2000
- Performance au Canada, 1970-1990 Artists with their Work
- Theories and Documents of Contemporary Art : A Sourcebook of Artists' Writings
- Les vingt ans du Musée à travers sa collection
- Extra Art : A Survey of Artists' Ephemera, 1960-1999
- Musée des beaux-arts de l'Ontario : Oeuvres choisies
- Art Gallery of Ontario : Selected Works
- Le Mois de la photo à Montréal
- Réfractions : Trajets de l'art contemporain au Canada
- Radio Rethink : Art, Sound and Transmission : Selected Survey of Radio Art in Canada, 1967-1992
- Documenta 9, Kassel, 1992
- Traffic : Conceptual Art in Canada 1965-1980

We zoom in on the center, and it is interesting to see what nodes are among this "core", in part because right in the middle is: "General Idea (Collective)".

In relative terms, the betweenness centrality of General Idea node is higher than its degree. One could argue that this suggests that the structure, network shape of the metadata, elevates the importance of General Idea (Collective) above its degree standing.   It is certainly higher than that of AA Bronson.

Seen as a communication system, a node with high betweenness centrality has a large influence on the transfer of information through the network.

We have just answered the question: what are some of the most central nodes in the entire e-artexte collection.



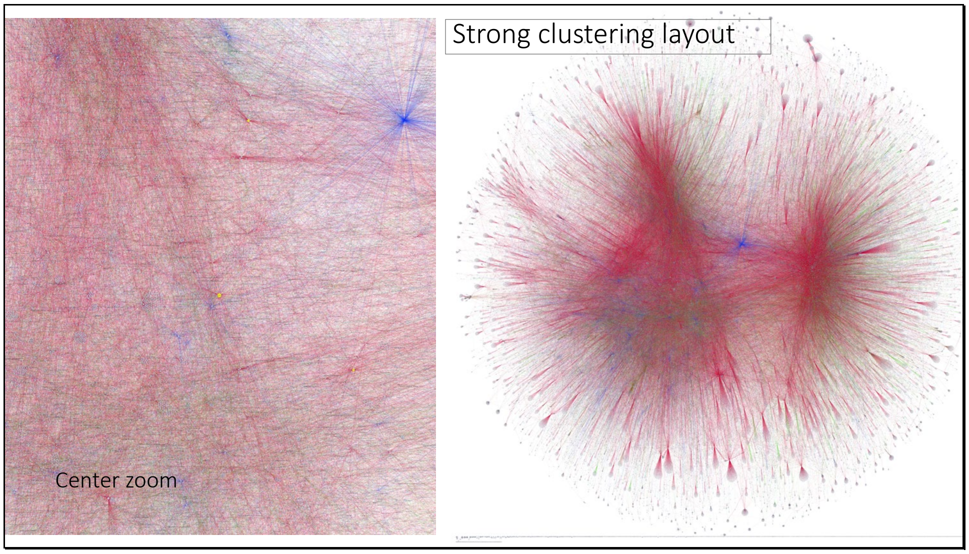
Zooming out again, this uses a “strong clustering” layout, but with no gravity, thus emphasizing the clusters even more.

Can see structure of publisher (blue) and authorship (green), as well as clustering of nodes.

What do these clusters represent? Why are they there? And is this a question that is worth asking?

Clearly, there is some division within the structure, is it based on language? Is it a division based on place (international vs Canadian)?

This is a new research question that emerges from this study.

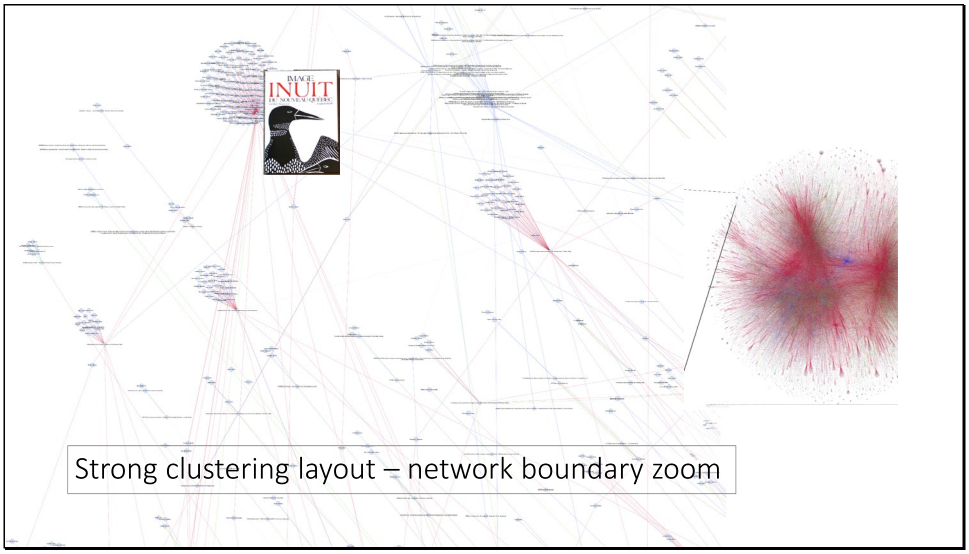


A strong clustering layout, with some circular gravity.

First question that emerges is: What lies in the middle, inbetween the clusters?

Where is Bronson AA in terms of these clusters?

With a high betweenness centrality, Bronson is in the middle, right outside of the large cluster on the right. This is consistent, since betweenness centrality favours those that are positioned to be intermediary nodes, and so lie on the boundaries of clusters.



Strong clustering layout – network boundary zoom

The second question that emerges is:

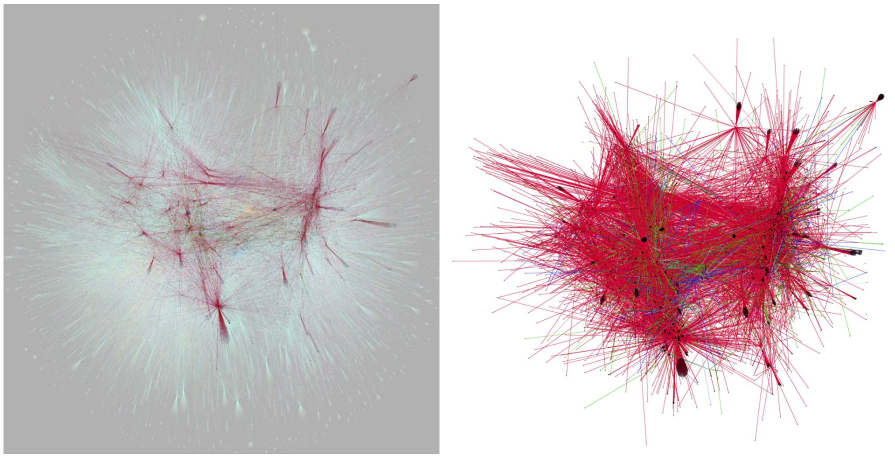
> What lies on the edges of the network? These would be works and names that are referenced in limited way.

When I zoom in on part of left edge of network, I found, for example:

> A book about an exhibition of Inuit artists.
>
> IMAGE INUIT DU NOUVEAU-QUÉBEC (left)

Perhaps examining these ‘boundary items’ would reveal a pattern?

Again, I don’t have an answer to this, but it is an avenue of research that is facilitated by this methodology.



Bronson AA related sub-network within context of e-artexte

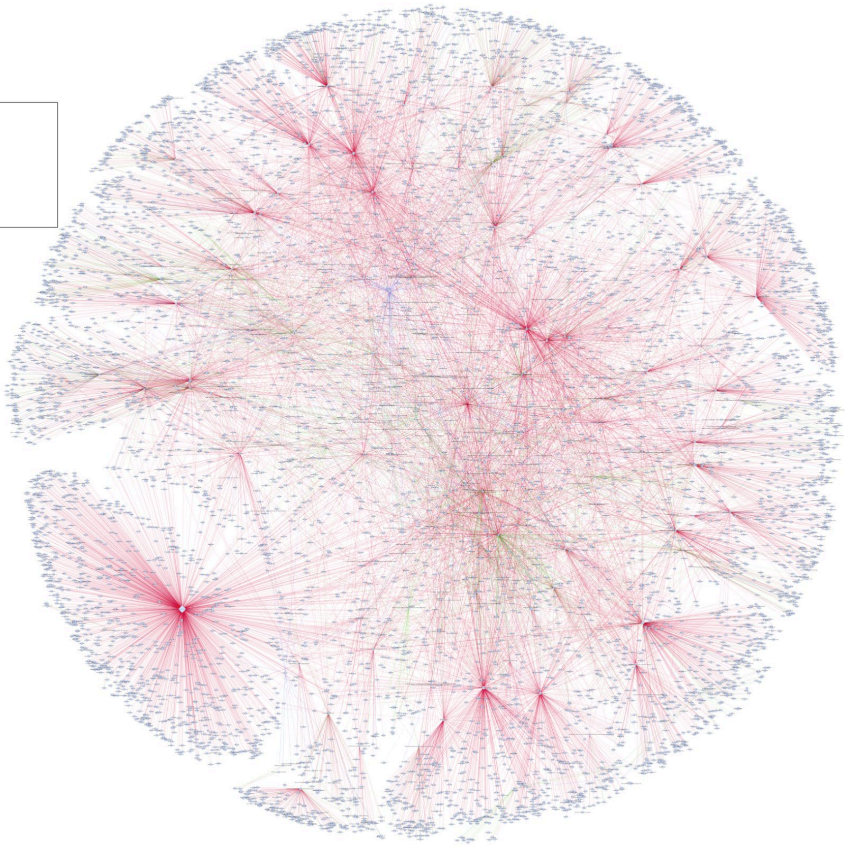
Returning to the case study of AA Bronson within the e-artexte collection.

The following shows the ‘sub-network’ of nodes that make up the 296 e-artexte items that we identified as closely related to AA Bronson.

296 items.
7351 nodes
13576 edges

On the left, shows where the AA Bronson nodes are positioned in the context of the whole e-artexte metadata set (using Allegro Strong Clustering layout). The white background is the whole network. Interestingly, the shape of this mirrors that of the whole collection, just under 300 out of over 20,000 items, yet the network they form spans across it.

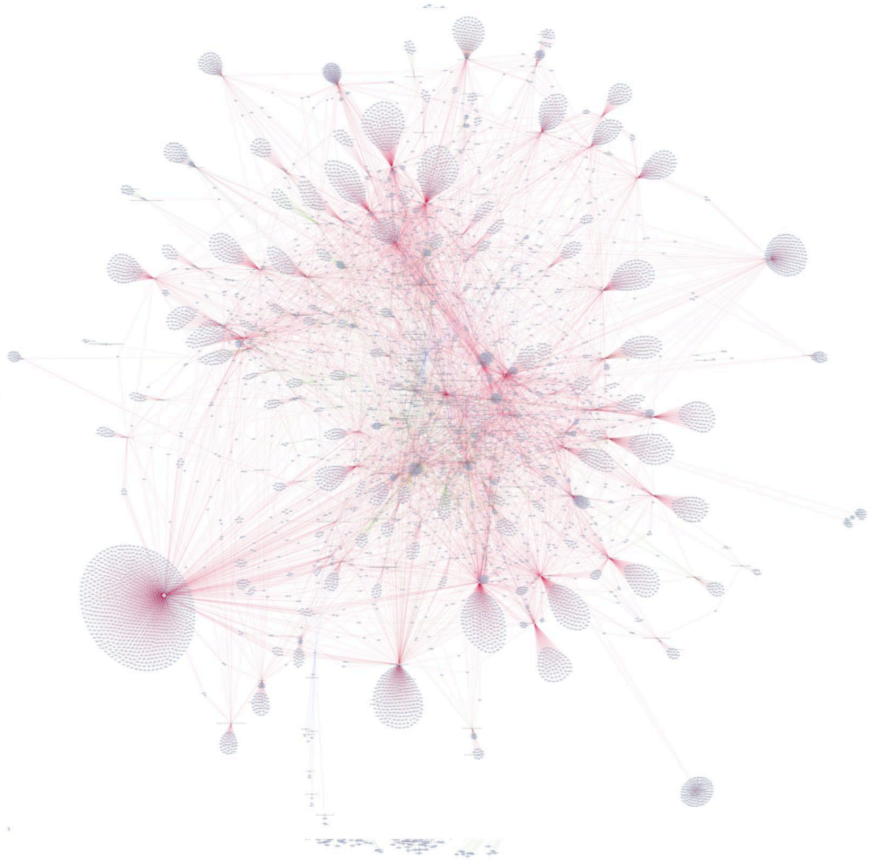
Bronson AA sub-network



Part 1: The Bronson AA network with the force-directed layout

The layout, for the exact graph will be different depending on its context. This is the layout based only on the 7000 nodes and 13000 edges that make up the Bronson AA sub-network.

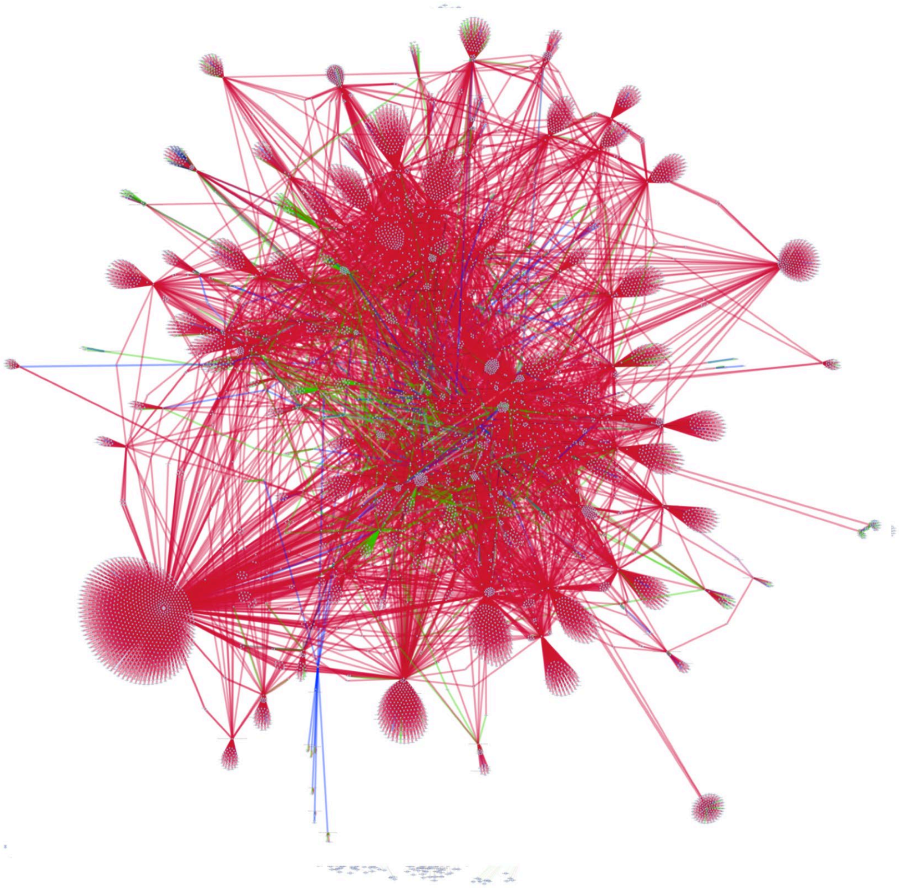
Bronson AA sub-network



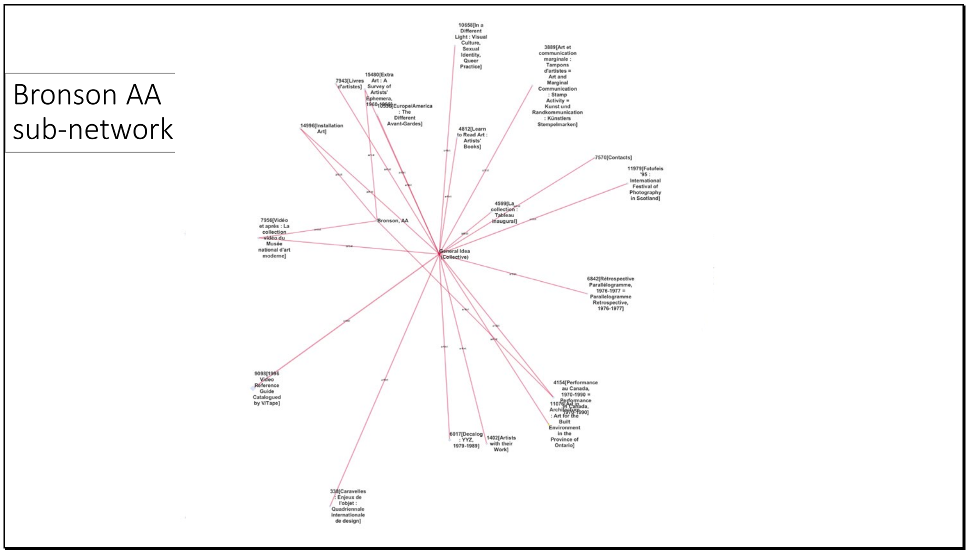
Part 2: Bronson AA network with the clustering layout

This layout positions the nodes related to each other closer together.

Bronson AA sub-network



Part 3: Bronson AA network with the clustering layout, and edge betweenness emphasized. Just like nodes, edges also have a betweenness centrality. This visualization emphasizes visually the roles structure with edge colour. The stronger the colour, the thicker the width of the edge.


Bronson AA sub-network
10658[In a Different Light : Visual Culture, Sexual Identity, Queer Practice]
3889[Art et communication marginale : Tampons d'artistes = Art and Marginal Communication : Stamp Activity = Kunst und Randkommunication : Künstlers Stempelmarken]
7943[Livres d'artistes]
15480[Extra Art : A Survey of Artists' Ephemera, 1960-1999]
10596[Europe/America : The Different Avant-Gardes]
14996[Installation Art]
4812[Learn to Read Art : Artists' Books]
7570[Contacts]
11979[Fotofeis '95 : International Festival of Photography in Scotland]
4599[La collection : Tableau inaugural]
Bronson, AA
7956[Vidéo et après : La collection vidéo du Musée national d'art moderne]
General Idea (Collective)
6842[Rétrospective Parallélogramme, 1976-1977 = Parallelogramme Retrospective, 1976-1977]
9098[1996 Video Reference Guide Catalogued by V/Tape]
4154[Performance au Canada, 1970-1990 = Performance in Canada, 1970-1990]
11076[Art in Architecture : Art for the Built Environment in the Province of Ontario]
6017[Decalog : YYZ, 1979-1989]
1402[Artists with their Work]
338[Caravelles : Enjeux de l'objet : Quadriennale internationale de design]

Part 4: Bronson AA, top 20 betweenness centrality nodes only.

Lastly, we can focus in on just the most central nodes, as a way of narrowing down/sorting our results. We decided to narrow it down to the top 20 most central ones. This would be the first page of result after applying a "centrality" sort to a list of results in a database. We can now ask a subject expert, Felicity, what her thoughts/reactions were about using centrality measures to narrow down the results in this way.