

Visualising networks of artistic collaboration in AusStage – eResearchSA Summer Scholarship Report

Nathan Lambert and Jonathan Bollen
Drama Department, Flinders University
April 2009

Introduction

AusStage (<http://www.ausstage.edu.au>) is a relational database of performing arts events, with tables recording the relations between events, venues, organizations, people, resources and so on. Each event in AusStage is associated with a list of cast and crew.

Nathan Lambert was awarded an eResearch SA Summer Scholarship (Dec 2008-Jan 2009) to work with Jonathan Bollen on the development of new approaches to analysing data in AusStage by using network visualisation to address the question of ‘who works with whom?’

Lambert’s research explored software applications, data migration and techniques for visualizing AusStage network data relating to the performing arts in South Australia.

Aim

AusStage is leading the application of network visualisation and analysis in the performing arts by exploring the connections between innovation, creativity and collaboration, and mapping the topology of creative collaborations over time.

Consider biographies in theatre programs, where each member of the cast and crew recounts their relations with training institutions, theatre companies, leading artists and notable productions. A history of working where, with whom, and on what is how practitioners in the performing arts accumulate and articulate their experience, qualification and distinction.

What we want to explore are patterns of contributors working together and how these patterns change over time. We know, anecdotally, that social networks operate in the field of performing arts. The interactions between artists, as they train, rehearse and work together, have implications for the kinds of artists they become and the kind of performances they make.

With its relational design, the AusStage database records these networks of contact and collaboration and how they change over time. We set out to analyse these networks by visualising data from AusStage in network graphs.

Method

Bollen had developed a workflow for migrating and visualising AusStage datasets as network graphs using a combination of desktop software (Excel, Filemaker Pro) and open source software (Visone, SoNIA, BioLayout Express, MySQL) on Mac OS X Mac Mini and MacBook Pro. The workflow is effective at producing 2D images, 3D models and 2D animations of relatively small datasets focusing on particular companies, venues or individuals (<2,000 nodes, <20,000 links; see example).

Lambert extended this workflow by extracting new datasets from AusStage and exploring

visualization techniques using Visone and Cytoscape. Visone and Cytoscape require data from AusStage to be migrated two different formats.

Multiple layouts of network data were explored using both applications and exported to a variety of graphic formats, including PDF. A selection of some of the more successful visualisations are included. These samples include a brief analysis of the network visualisations.

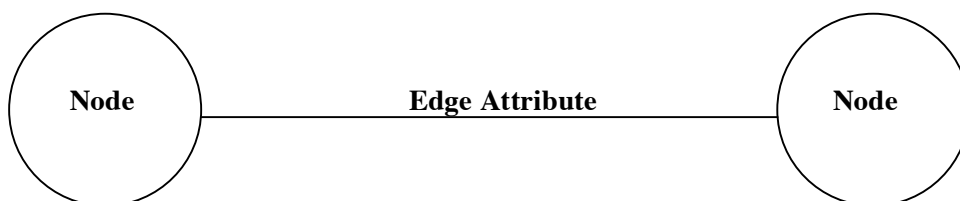
Data Visualisation: Visone

Visone is able to import a variety of formats for representing network data. To import AusStage data we used an edge list in the DL, a simple network format associated with the UCINET and Pajek network analysis packages.

The edgelist is formatted as follows

Node Data	Node Data	Edge Attribute
Darrellyn_Gunzburg_(536)	Darrellyn_Gunzburg_(536)	19880114
Catherine_Fitzgerald_(4774)	Darrellyn_Gunzburg_(536)	19880114
Darrellyn_Gunzburg_(536)	Catherine_Fitzgerald_(4774)	19880114
Catherine_Fitzgerald_(4774)	Catherine_Fitzgerald_(4774)	19880114

which is displayed as:



SQL queries were used to extract data from the event, contributor and event-contributor join tables in AusStage. We relied on the AusStage project manager to generate large exports using Oracle administrator software; the AusStage data entry interface provides an SQL facility which limits returns to 250 rows.

AusStage records associations between events and contributors (cast and crew); it does not directly record associations between contributions. The edge list format, however, requires associations between contributors that had worked together in any one event. A manipulation to transform event-contributor relations into a series of contributor-contributor relations based on event associations was scripted. This was performed subsequently on data exports, but could be scripted into the initial SQL query in future.

In extracting data from Ausstage, it was necessary to concatenate some fields, such as dates, names and IDs, to use the data as a node or an edge attribute. For example, AusStage stores dates in day, month and year fields; these were combined into a single string, such as 19880114. We also concatenated the Event name with the Event ID and the contributor with the contributor ID. Concatenation was undertaken using FileMaker Pro; it would also be possible to script this into the initial SQL queries.

Data Visualisation: Cytoscape

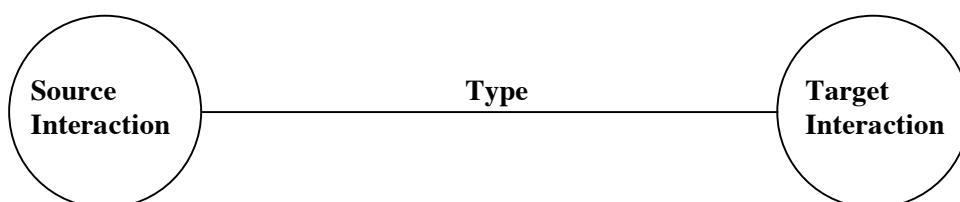
Cytoscape supported more flexible explorations of network relationships in AusStage data. It can import directly from Microsoft Excel (xls) or comma separated value (csv) formats. Interactions between nodes are defined on data import.

Cytoscape Import UI

Excel File that Cytoscape is able to import data from

EVENT_NAME	YYYYFIRST_D	UMBRELLA	VENUE_NAME	SUBURB	NAME	COUNTRYNAME
Birdbrain	2001	BAM 2001 Next	The Joyce Thea	New York, New	Australian Danc	America
The Age of Unb	2002		Drama Theatre	Sydney	Australian Danc	Australia
Ignition 3	2002		Wonderland Bal	Hawthorn	Australian Danc	Australia
Wildstars	1980		Opera Theatre	Adelaide	Australian Danc	Australia
Wildstars	1979		Opera Theatre	Adelaide	Australian Danc	Australia
Wildstars	1980		National Theatre	St Kilda	Australian Danc	Australia
Wildstars	1979		Canberra Theat	Canberra	Australian Danc	Australia
Wildstars	1980	Australian Danc	Royal Lyceum T	Edinburgh	Australian Danc	Scotland
Labyrinth / Brok	1982	ADT EuropeanT	Sadlers Wells T	London	Australian Danc	United Kingdom
Winter By Spring	1982	ADT EuropeanT	Sadlers Wells T	London	Australian Danc	United Kingdom
While We Watch	1982	ADT EuropeanT	Sadlers Wells T	London	Australian Danc	United Kingdom

Network data is displayed by Cytoscape as:



Project Outcomes

The outcomes of this research will be adopted in planned developments for AusStage as enhancements to the interface of the AusStage database and its analytical capabilities.

During the project, Lambert developed many skills, including:

- An increased knowledge of SQL databases (*Oracle, MYSQL*)
- An advanced knowledge of *Visone* and *Cytoscape*
- Analysis of network graphs
- Network theory
- PDF authoring
- Various data formats of GraphML, GML, DL, XML, NET.

- Data manipulation and data migration across various platforms/formats

Acquiring the skills for these languages and applications was challenging but rewarding. Lambert is now capable of constructing and analysing network visualisations. He has a series of graphs planned for his honours project this year already.

There are many possible uses for network graphs in humanities research. This research will feed into Ausstage development of network tools for researchers. More work is required to decrease the time spent when manipulating and migrating data by scripting data manipulation procedures into SQL queries.

It was also apparent that a time is needed to review the data that the graph is constructed from. If the data is not refined and made relevant to an outcome, little or no analysis can be made from it.

Lastly it is important to note how quickly network graphs can become confusing. The more data represented, the harder that data is to follow. It is therefore important to keep each graph's outcome focused and relatively narrow. For instance, if an SQL query returns more than say 8000 lines, the resulting graph can become difficult to follow.

Software Used

Network analysis software

- Visone – Analysis and visualisation of social networks (<http://visone.info/>)
- BioLayout Express 3D – visualisation of large network graphs (<http://www.biolayout.org/>)
- Cytoscape – network analysis and visualisation (<http://www.cytoscape.org>)

Other software and languages

- Oracle
- MySQL
- PHP
- FilemakerPro
- Microsoft Excel
- Acrobat Reader Professional

Appendix – attached files

Figure 1 1966 Adelaide Festival contributor-contributor links

Figure 2 South Australia, 1970-1980, contributor-function-event organic

Figure 3 Australian Dance Theatre event-venue degree-sorted circle

Figure 4 Adelaide Festival and Fringe, selected years, contributor-contributor curves vertical

Figure 5 Adelaide Festival and Fringe, selected years, event-event random layout

Figure 6 South Australia, 2008, event-venue-organisation circle

Contact

Dr Jonathan Bollen

Drama Department, Flinders University, GPO Box 2100, Adelaide SA 5001, Australia

jonathan.bollen@flinders.edu.au

+61 8 8201 5874

Figure 1 1966 Adelaide Festival contributor-contributor links

Software Visone <http://visone.info/>

Austage Data Used

Event name, Contributor name, Event first date

Relationship used (Node-Link-Edge Value)

ContributorID/Name – ContributorID/Name – Event First DDMMYYYY Links between two people Established through a production both contributors took some part in. Network Layout This layout was one of the first ones experimented with. A query on %Adelaide%Festival where first year was equal to 1966 was run, then a MySQL query was used to connect any contributor to each member of the cast/crew of each production in the 1966 Adelaide Festival

SQL Query

```
SELECT DISTINCT coneblink.conevlinkid, contributor.contributorid, contributor.last_name,
contributor.first_name, coneblink.function, contributorfunctpreferred.preferredterm, events.eventid,
events.umbrella, events.event_name, events.ddfirst_datell'/'ll
events.mmfirst_datell'/'llevents.yyyyfirst_date
FROM coneblink, contributor, contributorfunctpreferred, events
WHERE contributor.contributorid=conevlink.contributorid
AND events.eventid=conevlink.eventid
AND contributorfunctpreferred.contributorfunctpreferredid=conevlink.function
AND events.umbrella like '%adelaide %festival'
AND events.yyyyfirst_date=1966
AND coneblink.conevlinkid>0
ORDER BY coneblink.conevlinkid
```

Network Layout

Uniform, node links and colours are defined by degree

Explanation

This graph is not very exciting, but it is a good example of a simple graph to Explain very simply how Visone works. Each of the eclipse represents a node. Each of these nodes represents a contributor. You should be able to notice that some nodes are bigger than others but also have a different colour. The node's colour/size represents the degree of connectivity that contributor had in 1966. Each black line represents the link between each contributor. By following these links it is possible to see which contributor linked other contributors together. Not surprisingly, the linking contributors are usually Directors, Playwrights or Designers.

Advantages

This graph shows clearly where the links occurred in the 1966 Adelaide Festival, as well as loosely highlighting the major connected artists at the time.

Disadvantages

Whilst offering a reasonably straight forward view of a network, this view is limited. It is not possible to see what event links the contributors or see what part they played in each event.

Possible Uses

This method of visualization could be used to establish clustering on the Austage database and in each state's community network.

Figure 2 South Australia, 1970-1980, contributor-function-event organic

Software Cytoscape

Austage Data Used

Contributor, Contributor Function, Event Name, Event Date, Event State

Relationship used (Node-Attribute-Edge)

Contributor – Contributor Function - Event

SQL Query

```
SELECT DISTINCT events.eventid, events.event_name, events.yyyyfirst_date, events.mmmfirst_date,
events.ddfirst_date, events.umbrella, venue.venue_name, venue.street, venue.suburb,
orgevlink.organisationid, organisation.name, venue.countryid, country.countryname
FROM conevlink, contributor, contributorfunctpreferred, events, venue, orgevlink, organisation,
orgfunctmenu, country
WHERE venue.countryid=country.countryid
AND contributor.contributorid=conevlink.contributorid AND events.eventid=conevlink.eventid AND
events.eventid=orgevlink.eventid AND orgevlink.organisationid=organisation.organisationid AND
contributorfunctpreferred.contributorfunctpreferredid=conevlink.function
AND events.yyyyfirst_date=2008 AND events.yyyyfirst_date>1970
AND events.yyyyfirst_date<1980 AND venue.venueid=events.venueid
AND venue.state=1
AND events.eventid>0
ORDER BY events.eventid
```

Network Layout

Organic, Spring embedded

Explanation

This graph is complicated to say the least. The graph includes any contributor that was involved in an event between 1970 and 1980 in South Australia and links each contributor to each event they are associated with.

Advantages

This graph highlights main contributors in SA but also shows the vast network that exists across the performing arts community in SA.

Disadvantages

It is difficult to read many of the event names or contributor names. It is also hard to draw any data out of this graph as there are too many connections, especially towards the central cluster of the graph. It is not really possible to gain much insight without zooming right in, which limits a wide zoom analysis of this data.

Possible Uses

This graph could be used to analyse a certain artist's connections to other artist's within a state. This graph would work much better with a smaller data set.

Figure 3 Australian Dance Theatre event-venue degree-sorted circle

Software Cytoscape

Austage Data Used

Event Name, Venue Name, Organization Name, Event Country ID Relationship used (Node-Attribute-Edge) Event Country – Venue - Event

SQL Query

```
SELECT DISTINCT coneblink.conevlinkid, contributor.contributorid, contributor.last_name,
contributor.first_name, coneblink.function, contributorfunctpreferred.preferredterm, events.eventid,
events.event_name, events.yyyyfirst_date, events.mmfirst_date, events.ddfirst_date, events.umbrella,
venue.venue_name, venue.street, venue.suburb, orgevlink.organisationid, organisation.name,
venue.countryid, country.countryname
FROM coneblink, contributor, contributorfunctpreferred, events, venue, orgevlink, organisation,
country
WHERE venue.countryid=country.countryid
AND contributor.contributorid=conevlink.contributorid AND events.eventid=conevlink.eventid AND
events.eventid=orgevlink.eventid AND orgevlink.organisationid=organisation.organisationid AND
contributorfunctpreferred.contributorfunctpreferredid=conevlink.function
AND venue.venueid=events.venueid
AND orgevlink.organisationid=102
AND coneblink.conevlinkid>0
ORDER BY coneblink.conevlinkid
```

Network Layout

Degree sorted circle layout

Explanation

This view offers an interesting way to analyse the Australian Dance Theatre's work internationally. The nodes and edges represent the Event name and Event Country ID. The Edge Attribute is the Venue name. This graph allows an interesting view of the Australian Dance Theatre's toured work.

Advantages

This graph gives a very clear view of the company's wide coverage of both Australia and the World.

Disadvantages

It may have been more advantageous to include the date field as the edge attribute, rather than the Venue name.

Possible Uses

This sort of graph and dataset would be great for analysing a group or contributor's work internationally or even nationally.

Figure 4 Adelaide Festival and Fringe, selected years, contributor-contributor curves vertical

Software Visone <http://visone.info/>

Austage Data Used

Event name, Contributor name, Event first date, Umbrella.

The data on the Austage database for the Adelaide Fringe and Festival of Arts lacks consistency. It was necessary to select particular years where there was consistent data. The Festival/Fringe data used in the graph were from the following years : 1978, 1980, 1992, 1994, 1996, 2002, 2004.

SQL Query

```
SELECT DISTINCT coneblink.coneblinkid, contributor.contributorid, contributor.last_name,
contributor.first_name, coneblink.function, contributorfunctpreferred.preferredterm, events.eventid,
events.umbrella, events.event_name, events.ddfirst_datell'/'ll
events.mmfirst_datell'/'llevents.yyyyfirst_date
FROM coneblink, contributor, contributorfunctpreferred, event
WHERE contributor.contributorid=coneblink.contributorid
AND events.eventid=coneblink.eventid
AND contributorfunctpreferred.contributorfunctpreferredid=coneblink.function
AND events.umbrella like '%adelaide %festival'
AND events.yyyyfirst_date=[Year]
AND coneblink.coneblinkid>0
ORDER BY coneblink.coneblinkid
```

Relationship used (Node-Link-Edge Value) ContributorID/Name – ContributorID/Name – Event First DDMMYYYY Links between two people Established through a production both contributors took some part in.

Network Graph Name Adelaide Festival/Fringe Contrib/Contrib Curves (vertical)

Network Layout Uniform, node links and colours are defined by degree, links as curved lines

Explanation This network graph shows the links between contributors that have participated in Adelaide Festival/Fringe events during the selected years.

Advantages It is possible to gain a very wide view of contributor links in this graph and establish the major connecting collaborators.

Disadvantages This graph is very large and difficult to follow. While the graph gives an idea of the major artists, it is not possible to see what role that artist played in a production or even what production they took part in.

Possible Uses This network graph would be suitable for establishing the main contributors for a certain year or over a period of time.

Figure 5 Adelaide Festival and Fringe, selected years, event-event random layout

Software Visone <http://visone.info/>

Austage Data Used

Event name, Contributor name, Event first date, Umbrella.

The data on the Austage database for the Adelaide Fringe and Festival of Arts lacks consistency. It was necessary to select particular years where there was consistent data. The Festival/Fringe data used in the graph were from the following years : 1978, 1980, 1992, 1994, 1996, 2002, 2004.

SQL Query

```
SELECT DISTINCT coneblink.coneblinkid, contributor.contributorid, contributor.last_name,
contributor.first_name, coneblink.function, contributorfuncpreferred.preferredterm, events.eventid,
events.umbrella, events.event_name, events.ddfirst_datell/'||
events.mmfirst_datell/'||events.yyyyfirst_date
FROM coneblink, contributor, contributorfuncpreferred, event
WHERE contributor.contributorid=coneblink.contributorid
AND events.eventid=coneblink.eventid
AND contributorfuncpreferred.contributorfuncpreferredid=coneblink.function
AND events.umbrella like '%adelaide %festival'
AND events.yyyyfirst_date=[Year]
AND coneblink.coneblinkid>0
ORDER BY coneblink.coneblinkid
```

Relationship used (Node-Link-Edge Value) ContributorID/Name – ContributorID/Name – Event First DDMMYYYY Links between two people Established through a production both contributors took some part in.

Network Graph Name Adelaide Festival/Fringe Event-Event curves

Network Layout Uniform, node links and colours are defined by degree, links as curved lines with arrow indicating direction of relationship.

Explanation This network graph shows the links between contributors that have participated in Adelaide Festival/Fringe events during the selected years.

Advantages It is possible to gain a very wide view of contributor links in this graph and establish the major connecting collaborators.

Disadvantages This graph is very large and difficult to follow. While the graph gives an idea of the major artists, it is not possible to see what role that artist played in a production or even what production they took part in.

Possible Uses This network graph would be suitable for establishing the main contributors for a certain year or over a period of time.

Figure 6 South Australia, 2008, event-venue-organisation circle

Software Cytoscape

Austage Data Used

Event Name, Venue Name, Organization Name, Event Name, Event First Date, State Id

Relationship used (Node-Attribute-Edge)

Venue – Event - Organisation

SQL Query

```
SELECT DISTINCT coneblink.coneblinkid, contributor.contributorid, contributor.last_name,
contributor.first_name, coneblink.function, contributorfunctpreferred.preferredterm, events.eventid,
events.event_name, events.yyyyfirst_date, events.mmfirst_date, events.ddfirst_date, events.umbrella,
venue.venue_name, venue.street, venue.suburb, orgevlink.organisationid, organisation.name FROM
coneblink, contributor, contributorfunctpreferred, events, venue, orgevlink, organisation WHERE
contributor.contributorid=coneblink.contributorid AND events.eventid=coneblink.eventid AND
events.eventid=orgevlink.eventid AND orgevlink.organisationid=organisation.organisationid AND
contributorfunctpreferred.contributorfunctpreferredid=coneblink.function
AND events.yyyyfirst_date=2008 AND venue.venueid=events.venueid
AND venue.state=1
AND coneblink.coneblinkid>0 ORDER BY coneblink.coneblinkid
```

Network Layout

Organic (Spring Embedded)

Explanation

This network graph visualizes the relationship between organisations, venues and events, in 2008, in South Australia. Unfortunately the data set is not complete so some venues are missing due to an error in an SQL statement.

Advantages

This visualization gives a good view of which venues where most active over the year. It is also possible to get an idea of which organisations used which venues in 2008 for various performances.

Disadvantages

As in some of the other graphs, it is not possible to distinguish between edges and nodes.

Possible Uses

This view could be used to find the major hubs in different performance hubs in different states.