Visualizing Relational/Network Data
A Smörgåsbord of Approaches, Interpretations and Caveats

Carl Nordlund (AND Manu)
Post-doctoral Research Fellow
Center for Network Science, Department of Political Science
nordlundc@ceu.hu
Relational/network data?

- Not stuff – but stuff between stuff
- Data is relational: only makes sense in pairs of entities (actors)
- Entities depend on each other – not independent
- Sampling problematic – instead look at complete populations/networks. (With exceptions)
- Non-conventional approaches to deal with missing data
"Traditional" cross-comparative scientific inquiry

A  Age: 35
   Income: 40k
   Education: BSc
   Gender: Male

B  Age: 20
   Income: 30k
   Education: MA
   Gender: Female

C  Age: 41
   Income: 42k
   Education: -
   Gender: Female

D  Age: 52
   Income: 53k
   Education: -
   Gender: Male

E  Age: 21
   Income: 34k
   Education: MA
   Gender: Male

F  Age: 56
   Income: 28k
   Education: PhD
   Gender: Female
Cross-comparative approach: Modeling on internal properties

Comparing two or more parameters among the various entities

Identifying would-be relations

Assumed independency between entities → Sampling often possible
The network approach: Focusing on relations

Interpersonal relationships (aka friends)
Relational data types

Similar to ”traditional” statistics and its different data types, network data comes in many forms
The network approach: New analytical tools

Measuring actor centrality:
(compare between actors in a network)

Measuring network density:
(Compare between whole networks)

Number of actors (N) = 6
Number of ties (e) = 7
Number of possible ties (e_{max}) = N*(N-1)/2 = 15
Density (e/e_{max}) = 7/15 = 0.47

SO LET'S LOOK AT SOME COOL VISUALIZATIONS OF RELATIONAL DATA!!!

Degree centrality vs Income

To emphasize how utterly COOL VISUALIZATIONS ARE, this LABEL is ANIMATED!!!!
No. 1

A legislative network
No. 2
French political bloggosphere
No. 3
Love

Brad Pitt
Angelina Jolie
Mitt Romney
Barack Obama
No. 4
Wheat trade ties (with attributional data)
No. 5:
Gross wheat trade ties
Okay, so..... What do we know?
More love...
Five women (F1-F5) and five men (M1-M5)
Past and present relationships; general patterns

Representation 1: Sociograph

Representation 2: Sociomatrix
More love...

Five women (F1-F5) and five men (M1-M5)
Past and present relationships; general patterns
A social network of PolBeRG mail recipients
Date: 18 April 2012, 19:55

- Data mining and processing using a heuristic that is quite common in network-analytical approaches, particularly within the social sciences.

Data specifics:
- Using publicly available personal information of the PolBeRG mail recipients, subsequently transforming this data from N-to-M-mode, complementing these monomodal affiliation ties with label-based actor attributes.
Visualizing this network of PolBeRG mail announcement recipients

Apart from ocular inspection: Centrality, cliques, subgroups, regular equivalence etc. - loads of metrics to apply here!

Interpretations, interpretations! Social fabric of PolBeRG falling apart?
The pros and cons (mostly cons) of network visualizations

- Highly aesthetic and beautiful - beats a bar chart any day
- Major trickery – smokescreen for rubbish data
- Often assumes an almost superhuman ability of ocular interpretation
- Suggestive – certain features can be enhanced, others can be disguised and hidden (McGrath et al 1997; McGrath and Blythe 2004; etc)
The pros and cons (mostly pros) of network visualizations

- Very effective way to display large relational dataset
- Ocular inspection can reveal structural properties that would require a superhuman ability to see when looking at sociomatrices
  - Clusters, Bridges
  - Structural folds (Vedres and Stark 2010)
- Can display both relational and attributional data
- Different visual devices for different purposes
- Formal methods for checking the accuracy of a visualization
What follows: basic visualization techniques

- Look at various visual devices/heuristics to display relational data
  - i.e. Not look at specific software packages (Gephi, Pajek, Visone, Ceunet/Indra etc)
- Force-directed layouts (a.k.a spring-embedders)
- Multidimensional scaling
- Kruskal stress indices
- Target sociograms
- Block imaging
- Etc.
Force-directed layouts (spring-embedding)

- Simulates a physical system
  - Nodes: balls/rings
  - Edges/relations: springs
- Springs relax: optimal placements of nodes
- Various algorithms: Fruchterman-Reingold, Kamada-Kaway, home-brewed ones
- Binary/dichotomous data: spring have equal lengths
- Valued data: spring length differ
- Both for similarity and dissimilarity data
- Symmetrical data only

Fruchterman-Reingold spring-embedder of the PolBeRG mail (visualized using Ceunet/Indra)
Spring-embedder
Distances between Danish cities

Spatial distances are unique:
(Almost) ALWAYS perfect fit.

Task: reduce distance between Århus and CPH, ceteris paribus

Geometrically impossible!
Stress indices
Measuring the goodness-of-fit for a particular visualization

- Calculates the discrepancy between geometrical distances in a visualization with actual relational data
- Various ways – Shepard diagrams, Kruskal stress indices (1 and 2)
- For Danish cities (assuming a flat earth): always a perfect fit, i.e. Kruskal stress index = 0
- Although primarily used in multi-dimensional scaling, these stress indices can be used to estimate how accurate a specific layout is
  - Kruskal stress index < .2 = acceptable
  - Kruskal stress index > .4 = not acceptable
Most social network data contains pairs of non-connected actors (i.e. density < 1)

This often leads to non-connected actors stacking on top of each other – which is not very intuitive (and often impossible to notice)

Thus, spring-embedded algorithms implement repelling of non-connected actors

Many social network dataset contains actors that are complete isolates, i.e. not connected to any other actors

Thus, many spring-embedders complement node repelling with gravity
Multidimensional scaling

- Principally similar to spring-embedding
  - Transform data (proximities) to Euclidean distances
- 2- or 3- (or more) dimensions
- Symmetrical data – binary or valued, similarity or dissimilarity data
- Metric or non-metric
  - Metric: proportional
  - Non-metric: rank-order
- Kruskal stress indices useful to validate a MDS
Multidimensional scaling II

- More formal and generally generates better (lower) stress indices
- Typically used in role-analysis on equivalence matrices
- (Which means: role-analysis of network data generates matrices containing continuous measures of role-similarity – thus no non-connected in this data)
- Nearness: role-similar

Role-similarity of countries: wheat trade, 1993-1995
2-dimensional **closed-space** spring-embedding
Data: domestic migration in Brazil

Surface of a sphere – no implicit gravity effect due to mean-coordinate clustering (above a certain threshold scaling)

For certain datasets (esp those without a given core): lower stress, better visualization!
Target sociograms

Distance from center reflects indegree (prestige)

More central: higher indegree

(Northway 1952)
Trivia: How to establish an optimal target sociogram?

Either a nice little computer algorithm...

```
//Bertma Katzess
int _t=(currlen-optlen) * (currlen-optlen);
int _n=optlen*option;
}
else if (repelunconnected.getState() as row!=col)
{
    from=node[row];
    to=node[col];
    dx=from.coords.x-to.coords.x;
    dy=from.coords.y-to.coords.y;
    currlen=MAKESQRT((dx*dx + dy*dy));
    rx=dx/currlen;
    ry=dy/currlen;
    from.displacement.x=x*rx/currlen;
    from.displacement.y=y*ry/currlen;
    to.displacement.x=x*rx/currlen;
    to.displacement.y=y*ry/currlen;
}
```

Or this steampunkish little contraption from 1950 (Chapin)

...pegs and rubberbands on a perforated board...
Exemplifying with a modified Galtung core-periphery typology

- Blockmodeling – typically used in role-analysis
- ...which is a collection of methods for establishing subsets of actors that share similar roles in the network
- Structural or Regular role equivalence
- Indirect (REGE, Concor etc) or direct methods (iterations, optimal fitting)
- Results in partition(s) of the actors into subsets

Modified core-periphery/imperialist Structure à la Galtung (1971)
Blockmodeling
Block images/Reduced graphs

Binary (dichotomous), symmetrical, no self-ties
Blockmodeling
Block images/Reduced graphs/Image graphs

Block image of modified Galtung typology

A
[c, e, f, j]

B
[The rest]

Complete/full block

Null-block

Regular

Null-block

Complete/full block
Interpreting image graphs

Ideal role-relational structures

Core-periphery:

Hierarchy:

Transitivity:

Mutual regular groups:

Functional anatomy:
intl trade in edible agricultural goods 1995-1999
“Recursive graphs”

Similar to image graphs: Each blue node represents a set of actors.

Ties between blue nodes represent all ties between the actors.

Blue nodes visualized using Fruchterman-Reingold.

Treat each subset of (green) actors as an individual graph to be visualized!
Dendrograms – Tree diagrams

- Typically used with symmetrical, valued data
- Dendrograms: visualizing results from a hierarchical clustering scheme

Random dendrogram found using Google
No idea what it depicts...
Fixed-layout network flows

Gross value of inter-state commodity flows (SCTG 2,15 & 30) within USA in 1997 (Nordlund 2003)
Transforming into economic space

Gower-MDS of the previous inter-state commodity flow data (Nordlund 2003)
### Visualizing 2-mode data

#### Women attending events

1. Visualize 2-mode data directly as a bipartite network
2. Convert to 1-mode (with its implications) and visualize it as relational data

<table>
<thead>
<tr>
<th>Names of Participants of Group I</th>
<th>Code Numbers and Dates of Social Events Reported in Old City Herald</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) 6/21</td>
</tr>
<tr>
<td>1. Mrs. Evelyn Jefferson</td>
<td>X</td>
</tr>
<tr>
<td>2. Miss Laura Mandeville</td>
<td></td>
</tr>
<tr>
<td>3. Miss Theresa Anderson</td>
<td>X</td>
</tr>
<tr>
<td>4. Miss Brenda Rogers</td>
<td>X</td>
</tr>
<tr>
<td>5. Miss Charlotte McDowd</td>
<td>X</td>
</tr>
<tr>
<td>6. Miss Frances Anderson</td>
<td>X</td>
</tr>
<tr>
<td>7. Miss Eleanor Nye</td>
<td>X</td>
</tr>
<tr>
<td>8. Miss Pearl Oglethorpe</td>
<td>X</td>
</tr>
<tr>
<td>9. Miss Ruth DeSand</td>
<td>X</td>
</tr>
<tr>
<td>10. Miss Verne Sanderson</td>
<td>X</td>
</tr>
<tr>
<td>11. Miss Myra Liddell</td>
<td>X</td>
</tr>
<tr>
<td>12. Miss Katherine Rogers</td>
<td>X</td>
</tr>
<tr>
<td>13. Mrs. Sylvia Avondale</td>
<td>X</td>
</tr>
<tr>
<td>14. Mrs. Nora Fayette</td>
<td>X</td>
</tr>
<tr>
<td>15. Mrs. Helen Lloyd</td>
<td>X</td>
</tr>
<tr>
<td>16. Mrs. Dorothy Murchison</td>
<td>X</td>
</tr>
<tr>
<td>17. Mrs. Olivia Carleton</td>
<td>X</td>
</tr>
<tr>
<td>18. Mrs. Flora Price</td>
<td>X</td>
</tr>
</tbody>
</table>

Woman-vs-events (Davis et al 1941)
Visualizing 2-mode data
Women attending events

Two types of nodes:
Actors
Events

A.k.a. Bipartite network (n-partite etc)

Force-directed layout works fine (incl stress indices)

Of course: relations only between different types of nodes

Women and events share the same “interpretational space”
2-mode to 1-mode
i.e. going from two actors to one actor

Actors A-K are affiliated to “events“ 1-4 in various ways

The conversion thus assumes that actors have a tie based on the fact that they participate in the same event

E.g. B and G attends event 2, thus a tie between B and G.

G and I attends event 4, thus a tie between G and I

B and I don’t share any event, thus no tie

Note: B and D participate in 1 and 2, thus (in a valued graph) the tie strength between B and D is 2.
Inter-personal network visualization of the energy sector in France: 2-mode to 1-mode gone bad

Affiliation data

Badly done!!

Based on mix of affiliation data: Web searches Being in same school Org. membership No sources

Affiliation data creates artificial clusters (cliques/total subgraphs)

...which is quite okay if the data is contextually relevant

Sharing name initials are NOT typically relevant

Source: http://greenpeace.fr/facenuke/
Summary: on visualizing relational/network data

- Visualization an integral part of network analysis
  - From Moreno's sociograms to Nvidia graphics hardware: huge improvement
- Network visualizations is not (or very seldom) analysis per se
  - Visualizations is not formal analysis
- A wide set of alternatives
  - Different visual devices for different data and, particularly, purposes
- The utility of network visualizations
  - Assist in interpreting data: identifying certain general properties
  - Formal validation of visual representation
  - In certain cases: ocular inspection of visualization can replace formal analysis
PERG session next week
Network analysis of intl trade in wheat
Exchange-structural changes vs Arab spring

- Thursday, 3rd of May, 17:30 (room TBA)
- No fancy-schmancy visualization stuff though