Graphical Techniques for Exploring Social Network Data
by Linton C. Freeman

- Social Network scientist study structural patterns, specifically focusing on the following two:
  - patterns of emerging cohesive groups
  - patterns of emerging structuring equivalence
- Social Network scientists use data collection tools to examine the data on relationships/ties between actors
- Data is organized in NxN matrices where any \( i^{th} \) row and \( j^{th} \) column represent actors and entries corresponding to those actors represent presence or absence of the relationship between the actors or strength of those relationships
- Social Network analysts use statistical analysis procedures to examine relationships
  - useful only when testing hypothesis
  - not useful when trying to discover patterns
- The first social network scientist to use visual displays to portray the network data was Jacob Moreno
  - he used points for actors and link connecting two points for relationship between the actors
  - he showed the arranging points in various manners (i.e. actors’ attributes) can help uncover structural patterns
    - he provided many procedures for arranging the points however they were not standardized or generic and couldn’t be applied across all networks
- Several studies have been made to develop generic procedures or principled procedures for arranging the points
- All of the studies have an underlying assumption of preservation of social pattern
  - Nodes that are visually arranged should display the strength of the tie (i.e. visually by distance)
- Studies are looking to describe a systematic procedure to place a point on the picture with no more than three dimensions based on the actors’ social proximity
  - Therefore social proximity of the actors must be reflected in spatial proximity of the actors in the picture
Two approaches for developing such pictures:

- Multidimensional scaling, smallest space analysis or spring embedding
  - in a sense a search algorithm
  - requires dimensionality specification
  - uses search algorithm to find best/optimal placement for the points
    - placements/locations of those points can be either
      - those that are the closest to replicating the original pattern of social proximities
      - or, those that are the closest in replicating the order but not the magnitude of original social proximities
    - all procedures involve a search algorithm for those locations

- Singular Value Decomposition
  - algebraic approach
  - transforms original variables into new variables/dimensions that are arranged from largest to smallest based on the variance or patterning in the original data
  - this way all variance is contained in the first few dimensions
  - there are a number of ways of obtaining SVD solutions
    - correspondence analysis pre-processor
      - removing effects of differences in sizes among the totals of rows and columns
    - principal component analysis
      - removing effects of differences in means and variances

Before applying either MSD or SVD we need to know if there is an interesting pattern at all

- Experiment based on the data collected by Freeman, Freeman and Michaelson
  - 43 beach goers arranging name cards based on association (who is in social relationship with whom)
  - matrix was constructed where rows and columns were 43 beach goers and entries were the number of times two actors were grouped together
  - when data is input into MSD two dense clusters are recognized
  - in the center of each cluster there are a number of core actors
  - each core had its peripheral actors surrounding the core
  - there are also actors in between clusters that may serve as “bridges” between the clusters
  - MSD procedure calculates index of stress measure where high value of it means that data doesn’t closely correspond to original proximities
    - in the experiment index of stress = .17
  - Another data analysis was constructed on the same data by taking the data out of the matrix and then randomly putting the data back into the matrix but preserving symmetry
  - MSD was applied to this matrix and result suggested lack of patterns and high index of stress of .36
- SVD was then used on the non-random matrix and random matrix and the results were consistent with MSD results.
- Whether using MDS or SVD approach one can see if data has interesting patterns or not but looking at the shape it produces (i.e., disk/circular shape) is not an interesting result - no pattern.

- After uncovering structural patterns it is interesting to study the precedents and consequences of those patterns:
  - Labeling approach was introduced
    - Label nodes based on its attributes
    - Many software programs can now be used such as MAGE

- Next an analysis was conducted on the data provided by 25 employees of the cosmetics store providing information if one spends time with another outside of work:
  - MSD approach was used to find the patterns
    - Patterns were recognized in the data
  - Labeling approach was used to see why the data was clustered/patterned the way it was
  - Following are the findings of the researcher based on the labeling approach:
    - Ethnicity is not a factor in the clustering
    - Marital status is not a factor
    - Age is an important factor

- Next experiment was set up to test a priori hypothesis:
  - Hypothesis - athletes would confide in their teammates based on the type of sport team - individual (i.e., golf, tennis, etc.) or team sport (i.e., basketball, volleyball, etc.)
  - Hypothesis was rejected after visual clustering and patterns suggested that athletes belonging to both types of teams would confide in their teammates
  - Important finding however suggested that women’s sport teams served as “bridges” between the clusters

- Another example was presented that analyzed friendship relationships among Australian residential college students:
  - New method that was used has generated consistent results with SVD method thus was independently verified

- An experiment was conducted to analyze post hoc data:
  - The experiment studied the networks before and after the internet connection and internet communication
  - The network was analyzed prior to the internet communication using SVD method and post internet communication also using SVD method
  - Focus was to see how the network has changed
  - The change was observed as the nodes moved and became more clustered
Social Network Analysis: A Handbook
by John Scott

- Sociogram is widely used to portray the social network concepts
- There are limitations to sociograms as you are limited to a sheet of paper
  - Large graphs with numerous nodes end up not being interpretable
- There are techniques to overcome this limitation
  - Common technique is to arrange the nodes with their connections in a circle as patterns of connections may become visible
    - Minimizes overlapping between the ties
- Data is very sensitive to interpretation
  - Results may vary depending on how you present the data and arrange the nodes
- It was suggested that the physical distance between any two actors should closely correlate with the spatial distance on the graph between those two actors
  - This allows the preservation of mathematical properties of the graph and helps uncover new features

DISTANCE, SPACE AND METRICS

- Mathematical approach, multidimensional scaling (MDS) uses spatial proximity to map out relational data
- Metric Model - is defined by “any model of space and distance” with known relationships between the properties
- There are various concepts of distance -
  - In graph theory, distance is measured by the path length.
  - Metric concept of distance is closer to real life concept of physical distance
  - In Euclidean metric, distance is a measure of the length of the most direct route between two points.
    - More understood as, for example, atlas measure
- MDS tries to convert graph measures (i.e. path length) into metric measures which are closer to physical measures
- MDS uses proximity data to construct metric configurations of points
  - First step in MDS is to produce a proximity matrix from graph measures
    - Values in the proximity matrix show similarity and dissimilarity measure of each pair of actors
    - These proximity measures may be frequency of contacts, shareholder relationships etc.
    - Metric measures for those proximity measures are obscure therefore converted to correlation coefficients which conform to Euclidean metrics
      - If two points in the graph have the same pattern of connection then those two points will have a correlation factor of 1 therefore proximity measure of 1 as well
High proximity values indicate greater closeness if similarity matrix is used and vice versa if dissimilarity matrix is used where low values indicate closeness.

- It is important for a researcher to choose which data to use - similarity or dissimilarity as MDS procedure will be different based on the choice made.

- As an example of MDS we can take a number of towns and mileage distances between them and arrange them into proximity matrix and construct atlas like map.
  - The drawbacks of such atlas map of course include inaccurate mileage on the roads as the roads don’t necessary stretch across shortest distances between towns and height of the hills encountered on the roads is also not considered.
  - However, this is a useful approximation of MDS procedure.

- MDS can be explained by geometrical principals.
  - Let’s take three towns A, B, and C and scaled mileages between them and construct their location on the map.
    - First, we draw a line AB of a known length and then we draw a circle centering at A of a known radius equal to the distance of C from A as C can be located anywhere within AC distance from A. We then draw another circle with a center at B with radius equal to distance of C from A. Wherever, those circles intersect are the possible locations of C. There are only two places and we choose one of the triangles as they are both mirror image of each other.
    - Same procedure can also be used for more than three points.
    - This procedure provides two dimensional solution for point location in a metric space.

- MDS can also employ the concept of rotation which suggests the movement of the configuration of points for a better fit.
  - Goal of rotation is to discover meaningful dimensions.

- First MDS program was developed by Torgerson in 1952.
- The raw data for MDS is used as distance measures and MDS algorithms employ geometric principals for fitting of proximity data to final configuration on the map.

**PRINCIPAL COMPONENTS AND FACTORS**

- Similar approach to MDS.
- PCA (Principal Component Analysis) is sometimes referred to as “factor analysis”.
- Serves as a way to analyze case-by-variable attribute matrix to discover factors or components common to the variables.
  - Use raw data to plot scatter diagram with specific set of coordinates or axes.
  - For relational data contained in case-by-affiliation matrix scatter map can convey real information about relative positions of the points.

- Following is a PCA algorithm for case-by-variable matrix:
  - First - case-by-variable matrix is converted to variable-by-variable matrix.
Variable-by-variable matrix contains correlations between the variables (shows similarity vs. dissimilarity between the variables)

Then, we search for a set of highly correlated variables and replace them with artificial variable that is a measure of their correlation. Therefore, some set of variables with a correlation above a certain level will be replaced with constructed variable. This variable is then a first principal component.

Next, we search for a next set of highly correlated variables that are not correlated with previous set then the values again are replaced with constructed variable. This variable is then a second principal component.

The procedure is then repeated to identify all uncorrelated principal components. All together those components describe all of the variation in the data.

By the end of this procedure, when all principal components have been found, the original variable-by-variable correlation matrix is converted to variable-by-component correlation matrix.

First and second components are independent therefore second component is orthogonal to the first meaning it can be drawn the axis at the right angle to the first component.

It is difficult to visualize more than 3D space therefore PCA aims at finding the least number of principal components

Researcher considers a stopping point when searching for components

The rotation of the component aims at aligning first component with main axis

• Another method for PCA which involves scattering of the cases:
  o First we transpose original case-by-variable matrix followed by computing case-by-case correlation matrix.
  o Procedure is then the same as previous
  o Final result is case-by-component matrix
  o This method is sometimes referred to as Q-Mode method as opposed to R-Mode method used for variable-by-variable analysis

• There are differences in Q- vs R- mode methods -
  o If researcher is looking to investigate affiliations in the network then R- mode is used.
  o If researcher is looking to investigate structure of the relationships then Q- mode is necessary.

• One can use PCA to analyze adjacency matrix
  o If it is symmetrized and unidirectional then both row and column solutions (R- mode and Q- mode) will be identical.
  o If it is directional then one of the solutions will correspond to sending relationships and another to receiving relationships
NON-METRIC METHODS

- Most data is binary representing presence or absence of a relationship therefore when using MDS or PCA it must be first converted to proximity measures which leads to un-warranted assumptions by the researchers.
- Even valued relationships sometimes may not be converted to proximity measures.
- Non metric MDS techniques are referred to as smallest space analysis.
- Non metric MDS procedures uses symmetric adjacency matrices that are not converted into Euclidean distances but only uses rank order. Therefore, this procedure uses the same rank ordering for the output distances as for the original values.
- Smallest Space Analysis procedure entails:
  - Sorting original matrix in descending order
  - Then another matrix is constructed where values are replaced by the ranks based on their sorted order.
  - Then a matrix of Euclidean distances is constructed with the same rank ordering as original matrix
  - These distances are used to draw the scatter plot where their rank order corresponds to their proximity rank order.
  - In order to construct Euclidean distances trial and error approach is used
    - First the starting point or “guess” is chosen
    - Then rank order distances is compared to the proximities
    - Then trial and error refinements of initial estimates are conducted in order to achieve better configurations
    - To see when the best configuration has been achieved researcher plots original rank orderings and then plots the trial configuration. If points are scattered wildly then it is a bad fit if they are clustered close to 45 degree line then it is a better fit.
  - The number of dimensions plays its role and the same procedure is applied to each of the dimensions
  - The researcher must determine the number of dimensions needed to find the best fit to the original data
    - Statistic called “stress” is used to determine the number of dimensions needed
      - It represents an average spread of points around the line of good fit.
      - It can be plotted versus the number of dimensions to determine the appropriate number.
      - The breakdown of “stress” value is as follows:
        - “Good” fit - “stress” value of 5% and lower
        - “Fair” fit - “stress” value between 5% and 10%
        - “Bad” fit - “stress” value of 20% and above
Atkin has proposed dimensionality approach in Q-mode analysis but later it was rejected by Freeman who indicated the best approach would combine graph theory with geometrical dimensions discovered through MDS.

- The number of dimensions corresponds to the that that is necessary to embed the graph in a space with good fit (involves achievement of zero stress).

- Large number of dimensions is visually not useful therefore when required they can be plotted on the paper in two dimensions representing each slice of multidimensional space.

- Let’s take MSD output and optimal number of dimensions then we need to interpret it.
  - There are two issues with interpretation:
    - “Meaning of the dimensions”
    - “Significance of the spatial arrangements of the points”
  - Initial task in social network analysis is to interpret the dimensions.
    - Can be done by reflecting what each represents.
  - In order to provide significance to the spatial arrangement of the points, researcher first must apply cluster analysis on Euclidean distances and plot them on Euclidean MDS solution. The nested components will appear in the contour map. In such maps researcher can search for common characteristics.
  - All through the process researcher remains in control making his/her own interpretations.

**ADVANCES IN NETWORK VISUALIZATION**

- Integration of MDS techniques and structural analysis have been proposed to ease the visualization and exploration of the network.
- First solution proposed by Klovdahl used molecular modeling methods and simple 3D representations of points and lines. However, this solution still hasn’t provided interpretable visualization.
- Second solution proposed by Krempel used simple geometric shapes as a reference point to then organize the rest of the network. An example is a circle network. Circle becomes his reference shape and he uses graph theory to provide the best fit for the relational data to the circle. If more than one sub graph is present then multiple circles are employed.
- Freeman advocates that points should be color coded and arranged to at least portray their relational properties and organized around the most important structural features. He also advocates using molecular modeling techniques on social network data. He uses “MOVIEMOLD” to see the changes in the structure in animation mode.
- There is a more flexible modeling program - MAGE. It makes it possible to rotate the network model around multiple dimensions and explore it in detail by zooming in and out.
The first study employing the MDS techniques in sociology was constructed by Laumann.

- His research used framework of Parsonian systems theory to investigate community structures.
- He decided to use non-metric MDS approach to study the relational data.
- He employs the notion of social distance that refers to patterns of differential association. This distance measures how occupants of the social positions interact and associate with each other in a community life.
- Laumann’s goal is to obtain an “objective” measure of social distance and then employ MDS analysis to convert that measure to a social structural map.
- He obtained a sample of white males from two urban areas in Boston, MA.
- His unit of analysis was occupational positions rather than individuals themselves.
- Questions of interest were their friendship choices, kinship and neighboring which in combination produced measure for their positions.
- His initial analysis used five occupational categories which were used to construct position by attribute incidence matrices.
- His analysis of the matrices showed that friendship choices were highly correlated with occupational equals, while other social relationships were with people of various occupations.
- Laumann used actual patterns of association or social distance to study the hierarchy within the networks.
- The results of his work were as follows:
  - Three dimensional space provided best fit.
  - First dimension is prestige and it is evidently the most important one.
  - However, he failed to interpret the other two dimensions.
- Laumann’s later study explored friendship relationships between various social positions in Detroit.
  - A sample of 1013 white males was used.
  - Laumann initially looked at ethnicity and religion as separate relations however MDS solution suggested that the combining the two provides better results. He then calculated dissimilarity measures from their friendship choices.
  - He obtained three dimensional solution where:
    - First dimension separated Protestants, Catholics and Jews - religion.
    - Second dimension showed economic standing - income.
    - Third dimension showed - church attendance.
  - Another study that involved 16 occupational groups showed the two dimensional solution where one dimension represented status or prestige and another divided entrepreneurial from bureaucratic.
  - Another study will be described in the next paper. (SEE NEXT PAPER).
  - Levine studied the national economic elite using interlocking directorships in businesses.
Levine was first to use MDS techniques rather than graph theory for the study of social network analysis.

His study looked at a 100 top corporations and their relationships with 14 banks in three cities.

70 corporations had bank interlocks therefore he constructed 70 by 14 matrix where values showed the number of common directors and that value was taken as a measure of similarity between the enterprises.

The matrix was analyzed and joint space was created where both banks and corporations were present.

Results showed that three dimensional solution provide the best fit.

Two of these dimensions were interpreted:
  - First dimension - regional separation (NY, Pittsburg, Chicago)
  - Second dimension - not interpreted.
  - Third dimension - industrials vs. banks.

Levine provided spherical configuration where concentric circles formed around the joint center.

Center is occupied by corporations without interlocks.

Financial interests groups (those with a lot of interlocks) can be seen on the sphere as “wedges”.

“Wedges” represented spheres of influence. However, those spheres of influence intersect forming overall spherical configuration of the joint space.

Levine then switches the view point from remaining at the ego-centric spheres of particular banks level to seeing the socio-centric features of the overall structure itself.

Next task is to efficiently represent the 3D configuration on a flat piece of paper. In order to achieve that Levine uses cartographic projection methods as 2D MDS outputs are too distorted. He uses gnomonic projection as part of cartographic projection method.

Levine’s work combines ego-centric and socio-centric concerns into coherent model of embeddedness of networks in multidimensional space.
New Direction in the Study of Community Elites
by Edward Laumann and Franz Pappi

ABSTRACT

- A number of common theoretical methodologies came about from studying community decision making
- There are a number of weaknesses in the methodologies that have been developed
- A focus of research is on the structural analysis
- Some critical questions have been raised about identifying the aspects that influence the community and community decision making, specifically their attributes and ties
- A study of a small German town Altneustadt using new advances in graph theory and smallest space analysis was conducted to understand the influence system
- A new strategy is developed to study community issues and identify roots of formation of fractions and coalitions

BODY

- There are a number of conflicting methodologies to study community decision making based on a number of pluralists and elite models
- New methodologies are now being assessed with a shifted foci from studying one individual community at a time to studying as many communities as possible
- Most of the emphasis is placed on empirical data and not theoretical issues
- An open-ended input-throughput-output theoretical model underlies many studies currently conducted
  - The model suggests that certain attributes of community (i.e. population size, age, location etc.) - “inputs” - together with its attributes of political institutions drive the decision making in the community
  - The decision making -“throughput” - in turn determines the issues that will be faced by community and decisions that will need to be made - “outputs”
  - The “throughput” or sometimes called “elite decision making” is central to the studies
- In this study the structure is one of the emphasis and it is carefully defined
- The absence of agreement regarding the conceptual definition of social structure makes it harder to study other problems such as structural change
- In this study of a small German town Altneustadt researchers are focusing on the “throughput” or “the elite/blackbox” of the community which turns “inputs” into “outputs”
- Central focus of the analysis is the fact that conflict acts as an epidemic where competing goals facilitate formation of fractions and coalitions
- Community Context - Altneustadt, West Germany
  - population of 20,000
  - no outside control by a larger town/city
- rich in farming activities
- town has a number of higher ranking public officials
- town also has a few manufacturing plants
  - local owners comprise the solid “middle class”
- Government built the largest National Science Research Center in Altneustadt
  - largest employer in the area
  - caused inflow of new people (Nebürger)
    - highly educated
    - highly salaried
    - urban origin
    - possess global and sophisticated perspective
    - mainly protestant
    - very different from local population (Altbürger)
      - Catholic
  - caused the formation of coalitions because of the differences in opinions between newcomers and locals
    - both newcomers and locals are relatively in the same socio economic structure layer but possess fundamentally different perceptions, behaviors and styles of life
      - facilitates the study of “status” vs. “class”
    - Nebürger joined with SPD (German Social Democratic Party) to voice their opinions and bring about change in Altneustadt
    - Altbürger were members of CDU (Christian Democratic Union) and CDU was predominant in the area but unstable
- Social Positions and Incumbents
  - Community influentials and their institutions sectors
    - unit of analysis is an individual
    - two approaches that try to identify persons in charge (“who governs?”)
      - reputational approach - measure reputation rather than power
      - issue approach - too conservative
    - structuralist approach, however, tries to find persons with general influence that affect decision making at multiple levels of institutional sectors
    - in this study, first researchers identified political officials of highest ranking that were responsible for organizations that were part of four institutional sectors
    - parsons view of hierarchical structure
      - Three levels of the structure
        - technical, managerial and institutional
    - since not all community subsystems are organized with clearly defined leaders the nominations for leader roles by well-informed community members were incorporated in addition to formal leaders already identified
since some of the leaders may occupy many leadership positions in the community researchers distinguish between primary and secondary positions of that leader depending on amount of time spent in each leadership role

- Rank order of influence
  - 55 leaders (influentials) were identified
  - 48 were interviewed
    - researchers asked if the leaders can hierarchically order themselves and other leaders based on the power
      - surprising consensus was reached when identifying top seven most influential leaders
    - in order to indirectly validate the list researchers compiled an average influence status of proponents or opponents of certain issues requiring decision making (see paper pg. 6 for calculation)
      - in doing so the people on the list were frequently mentioned to hold positions on proponents and opponents sides of various issues
      - the lower the value the higher is the influence status of opponents vs. proponents
      - based on that value five decisions on the five issues at hand were accurately predicted

- Social Relationships
  - Rationale for describing community influence structures
    - Definitions:
      - “Social Structure - persistent pattern of social relationships among social positions”
      - “Social Relationship - any link between incumbents of two social positions that involves mutual but not necessarily symmetric orientations and activities”
      - “Social Differentiation - the differing allocation of tasks and responsibilities among positions in social system”
      - “Differentiated social structure - one whose actors tend to confine their consensual relationships with others performing similar tasks”
        - Clustering occurs among people with similar positions
    - From the definitions it can be inferred that different models of structures will arise if different types of social relationships are used - i.e. professional vs. personal
      - need to define the dimensionality of the relationships
        - Researchers state underlying assumption:
          - “Similarities in social positions, interests, attitudes, beliefs, and behavior facilitate the formation of
consensual relationships among incumbents of social positions.”

- the more dissimilar any two positions are based on the above similarities less likely the relationship will form between them
  - this assumption affirms distance generating mechanism

The methodology of structural analysis: Graph Theory and Smallest Space Analysis

- Three focal relationships:
  - business-professional relationships
    - play critical role in decision making and formations of coalitions
    - task oriented relationships
    - obtained via interviews
  - “social” or expressive relationships
    - based on shared interests, values, beliefs
    - shows the differences between the Altbürger and Nebürger
  - “community affairs relationships”
    - formation of those types of relationships is said to be influenced by business-professional relationships as well as “social” relationships and centering around community affairs

Recent research has been trying to develop a technique for identifying certain cliques and the interrelationships within those cliques and between people who don’t belong to the cliques

Sociagram as a graphic representation was first developed to portray such relationships

- the drawback of sociograms is the size of the network, once the network becomes too large it is very difficult to distinguish relationships in the network
- large network analysis using sociograms sometimes leads people to different representations and conclusions

- a lot of computer generated techniques have been employed to generate the analysis of large networks
- researchers in this paper employ concepts and recent developments in graph theory and smallest space analysis to define the structure of the network

Mathematical theory concepts:

- Graph - set of points and lines connecting them, no directionality is assumed
- Three graph theoretical ideas:
  - adjacency matrix
  - reachability
    - a node a is reachable from node b if there exists a path between b and a
- path - set of directed links between b and a
  - path distance
    - minimum number of directed (if not symmetric) links that are used in order to reach a from b
- Researchers choose to symmetrize the data
- Reachability matrix portrayed the incumbents that were unreachable based on their choices and choices of others
  - In the given data - all respondents were reachable when social and community affairs relationships were assumed and five respondents were not reachable when business-professional relationships were assumed
- Maximum number of steps along the shortest path:
  - Business-professional and social relationships - 5
  - Community affairs relationships - 6
- Most influential incumbent could reach 91% of the network in 2 steps or less in community affairs structure, and 73% in business-professional and social structures
- Conclusions:
  - highly integrated influence structure across three networks where disconnected individuals represent less integrated structure
- Graphic Representations of Influence Structures
  - researchers used path distance matrix to run symmetric smallest space analysis program
    - euclidean two dimensional representations are produced
    - rank order of path distances is included in the analysis
      - the program arranges nodes in space based on the rank order
  - Researchers use two basic principles to interpret the space analysis:
    - principle of integrative centrality
      - central person will be located towards the center of their space
        - easier to reach all other nodes
        - depends on the relationship dimensionality
        - concentric circles in the space graph represent “zones” of importance
    - principle of sector differentiation
      - divides space into homogeneous regions closely based on the institutional sectors
      - these sectors represent natural coalition zones for community issues
less localized sector (less clustering of people based on institutional sector) means higher probability of dividing on community issues and vice versa

- Two additional hypotheses stated by researchers:
  - A person near the center of the space who belongs to a particular institutional sector may play an integral role as this sector’s representative.
  - Integrative centers are highly biased and may over represent some sectors while under representing others or even excluding them, but researchers state that their influence on decision making is therefore minimized.

- Business and professional space
  - Most integrated persons are high-ranking administrative official members that control much of economic and financial interests, therefore integral to decision making in economic matters; they are also located close to each other implying common interests in those matters.
  - Others such as small businessmen, religious, educational, and research center persons are located in periphery and are in considerable distance from each other, meaning less common interests.

- Social Relations space
  - Core of the space is comprised of mostly long time residents and members of CDU who are Catholic.
  - The order of economic sector is reverse of business and professional space’s economic sector:
    - Small businessmen and merchants at the center and larger manufacturing and financial enterprises in periphery.
  - Research personnel is also located at a considerable distance from the center and clustered away from other sectors.
    - Only one member, the most influential member in the town, is located close to the center.
      - The member is very assimilated to the local long time residents.

- Community Affairs Space
  - Central core includes high density of personnel.
  - Center personnel is highly homogeneous.

- There is significant correlation between the business/professional and community affairs space, and insignificant correlation with social relationships space.
  - This result is consistent with the principle of integrative centrality.

- Integrative status may differ in context based on the relationship dimensionality.
• it is evident that integral person is more relevant in the context of community affairs and business/professional relationships but not so much in the contest of social relationships

■ Researchers make two general statements regarding the findings:
  • pattern maintenance personnel are divided into two groups / clusters at the opposite ends of the axis that runs through the center of the space and Research Center personnel is clustered and located at the considerable distance from the center (weaker influence on the decision making) and the traditional religious and educational leaders are located at the other end of the axis with closer proximity to the center
  • the economic interests seem be less differentiated with division mostly between the local merchants and small businessmen and large corporations

○ Community Issues
  ■ Two types of issues:
    • Instrumental issues - allocation of scarce resources
      o labeled as “class politics”
      o conflict is moderate and characterized by bargaining and compromising
      o outcome is shaped by the influence factor/power of opposing parties
    • Expressive issues - concerned with community affairs and beliefs and values that shall drive them
      o labeled as “status politics”
      o high conflict with high emotions with “all or none” nature
      o difficult to reach compromise

■ Two favored strategies for identifying community issues:
  • Researcher discovered recent issues from newspapers or community informants
  • Researcher has a vested interest in an issue already and wants to know the factors that affect the outcome
  • Both strategies have drawbacks in the areas of comparative analysis and historical particularities bias

■ Researchers in this paper state that frame for reference is necessary to define the community issues

■ Two frames of reference have been identified:
  • Interest group approach
    o this approach identifies issues directly related to the interest groups identified in the community
  • Functional approach
- this approach is selected by the researchers and therefore is used to “define universe of content of possible issues”

- In order to construct a theoretical scheme to define community issues following points have to be addressed:
  - “definition of universe of content of possible community issues”
  - defining the biases in the set of community issues
  - translation of historical individuality into meaningful categories in order to conduct a comparative analysis
  - stating the hypothesis that connects the issues to the structural analysis

- Researchers in this paper state that those points have been addressed when functional approach was chosen

- Researchers claim that the outcome of the particular functional issue can be predicted by locating a functional sector that is activated by that issue

- They state that they can also predict the likelihood of the outcome when the sector is divided on the issue by analyzing the clustering of the sector and the locations of the clusters relative to the integral center

- Five issues were selected by the researchers based on the following two criteria:
  - issue must have an impact on the community for at least past 3 to 4 years
  - issues must be distributed across the functional areas

- The notion for studying five functionally different issues was to identify if the elite (decision makers) tended to be differentiated into coalitions structurally and functionally or elite stayed undifferentiated as a unity to make the decisions for all functional areas

- Graphic Representation of the cleavage structure
  - in the space graph “fault lines” representing five issues have been drawn to divide proponents and opponents
    - in order to construct these lines intensive interviews were conducted
  - It is found that the composition of the opponents and proponents differs based on the issue at hand
  - It is also evident that the most integral/central people tend to be involved in more than one issue
  - It is shown by the researchers that structural differentiation is necessary to generate stable coalitions that activate differentially based on the functional issue
The Effect of Spatial Arrangement on Judgments and Errors in Interpreting Graphs
by Cathleen McGrath, Jim Blythe, David Krackhardt

1. INTRODUCTION AND PROBLEM STATEMENT

- "It is commonly assumed that graphs communicate important characteristics of network data."
  - Graphs show actors within a network as nodes, and diagram the relationships between each pair of nodes.
  - Graphs can also use “spatial relations” (i.e., the physical positioning of the nodes as drawn on the paper) to highlight relationships among actors (such as to indicate geographical distance, etc.).
- However, people tend to make assumptions about the data based not just on the actual connections between nodes (the "graph theoretical properties"), but also on the spatial relationships between nodes (the "Euclidean spatial factors"), regardless of whether the nodes’ positioning in space carries any actual information or is purely for aesthetic purposes.
- This paper looks at how individuals’ perceptions of common social network measures can be influenced by the spatial arrangement of the nodes in the network visualization, specifically:
  - How proximity to the center of the spatial arrangement affects perceptions of “prominence”;
  - How positioning between clusters of nodes in spatial arrangement affects perception of “bridging”; and
  - How spatial clustering of groups of nodes in spatial arrangement affects perception of grouping.

2. EXPERIMENTAL DESIGN

The study was conducted using the graph of a network with 12 actors and 48 ties. This graph was then drawn with 5 different spatial arrangements: The nodes and connections between them were identical for all 5 arrangements (in other words, the graphs were isomorphic), but their placement on the physical page was varied (to be closer together, more spread out, etc.).

Questionnaire Design and Administration Methods

- A questionnaire was given to 80 graduate students who had just completed a course in organizational theory (so the test subjects understood what the graphs represented, and the concepts they were being asked about):
  - Questionnaire contained 3 of the 5 graph images, and stated that the images represented three different M&A teams of an investment banking firm.
A connection between nodes meant those two actors had discussed work-related matters with each other. If no connection was shown on the graph, then they did not discuss work-related matters with each other.

The nodes had different labels (the names of people who supposedly worked on each team) in each of the graphs, but the graphs themselves were isomorphic.

Five particular nodes of interest were identified in each graph on the questionnaire. These nodes actually had the same position (in graph theoretic terms) in every graph; only their names and spatial relationship to the other nodes was changed (for the analysis in this paper, these nodes have been designated as A, B, C, D, and E).

The respondents were asked two questions about these five nodes in each graph. First, they were asked to rate the following perceived relative attributes for each node on a Likert scale (1-7):

- Some individuals have a more prominent role in their team than other individuals. Please rate the following people according to how prominent within their team they appear to you [(circle 1-7) for each node].
- Some individuals are important because they form a bridge between subgroups. Please rate the following people according to how important they appear to be as bridges between subgroups [(circle 1-7) for each node].

Respondents were also asked to report the number of subgroups in the graph.

To administer the questionnaire in the least biased way possible, the researchers used several techniques:

- Each questionnaire contained 3 of the 5 isomorphic graphs (so there were five different versions of the questionnaire).
- Respondents only viewed one graph and its corresponding set of questions at a time.
- Researchers used an incomplete balanced Latin square design to control for order in the presentation of the graphs (so they displayed each spatial arrangement first, second, and third exactly once).

**Features of the Network**

- The network used for all 5 representations is the same: The data is symmetric, and involves 12 actors with 48 directed ties.
  - The overall density is 36%.
  - The (degree centrality, betweenness centrality) for each node of interest is:
    - A: (5, 8.67)
    - B: (5, 8.67)
    - C: (4, 4.67)
    - D: (3, 0.00)
    - E: (3, 0.00)
  - A and B have the highest scores for prominence and bridging, followed by C. D and E have the lowest scores for these measures.
  - A and B are automorphically equivalent, as are D and E (meaning we could exchange their positions in the graph without affecting the distances between all the other actors in the graph).
  - The network has four cliques.
- The visual representations were chosen because they conformed to commonly accepted aesthetic standards, such as regular spacing of nodes and minimal edge crossings. The circle
representation was chosen because of its general acceptance as a means for presenting social network data, and also because there is no variation in the spatial information provided when the nodes are arranged this way.

3. RESULTS

Order Effects

- Researchers were concerned that the order in which respondents viewed the three graphs might affect their perceptions of the measures being asked about.
  - Poulton (1982): Research subjects can learn a strategy in one condition and apply it to another, so what respondents observed in one arrangement might influence what they report observing in the next.
- To test for order effects, researchers ran two regressions (using prominence or bridging as the dependent variable, respectively). They controlled for individual respondents, node, and layout. When order of appearance was included, it had no significant effect on the independent variable for either prominence or bridging.
  - No significant overall mean effect. So when considered in aggregate, the order of presentation did not make a significant difference in respondents’ perceptions.
  - However, when each spatial arrangement was considered separately, Arrangement 3 showed some difference depending on the order in which it appeared.
    - Researchers believe this might mean that Arrangement 3 is particularly difficult to interpret, and that respondents’ perceptions were influenced by the preceding arrangement.

Predicting Judgment

Researchers used OLS (ordinary least squares) regression to compare Arrangements 1-4 with Arrangement 5 (which did not contain any spatial relationship data because it was a circle).

- Prominence
  - Likert scale data was converted to a mean centered prominence score (by calculating each respondent’s average prominence score across all nodes and arrangements, and subtracting that value from the prominence score assigned to each node) to control for individual tendencies to rate high or low in general.
  - In general, the relative ranking order was stable across all arrangements: A and B were ranked above C, which was ranked above D and E. Exception was Arrangement 1, where C was ranked below D and E.
  - Some results are as expected:
    - Node C: When it is pictured on the periphery (Arrangements 1 and 2), respondents perceived it as less prominent than when it is in the circle (A5). When it is pictured closer to the center of the graph (A3 and A4), respondents perceived it as more prominent than in the circle (A5).
    - Nodes D and E: When pictured closer to the center of the graph (A1 and A4), respondents perceived them as more prominent than in the circle (A5). When
 pictured in the periphery (A2 and A3), respondents perceived them as less prominent than in the circle (A5).

- Some results were unexpected:
  - Nodes A and B: These nodes are not perceived as significantly more prominent in A1 (where they are members of a central cluster in a central spatial location of the graph).
    - Researchers suggest that perhaps nodes with high structural prominence are not as affected by spatial positioning as nodes with low structural prominence (i.e., maybe it is easier to see from the graph that they are highly connected, regardless of where they are positioned).
  - Overall, positioning a node toward the center of the graph seemed to increase perception of its prominence.

- Importance as a Bridge
  - Researchers applied the same mean centering conversion.
  - The order of reported bridging is consistent for all nodes in all arrangements: A and B are always greater than C, which is always greater than D and E.
  - Results are generally as expected:
    - Nodes A and B: Both received higher bridging scores in A1 and A2, when they are spatially positioned between two clusters, than in A3 and A4.
    - Node C: Received higher bridging scores in A3 and A4, where it is positioned between clusters, than in A1 and A2, where it appears closer to the middle.
    - Nodes D and E: Received lower bridging scores in all four arrangements than they did in the circle, although E’s coefficients were not statistically significant.
  - Overall, positioning a node between groups seemed to increase perception of its importance as a bridge.

- Grouping
  - Spatial clustering of nodes may influence perception about the existence of sub-groups within the network, independent of the actual structural ties between the nodes.
  - The mean number of sub-groups reported for each arrangement varied, even though the actual number is the same for all graphs:
    - A1: 5.04 - highest reported
    - A2: 4.27
    - A3: 3.55
    - A4: 4.75
    - A5: 3.45 - lowest reported
  - Researchers also looked at the distribution of the suggested number of subgroups for each arrangement.
    - A5 (the circle) had a very flat distribution, suggesting that respondents had a hard time figuring out how many subgroups it contained (i.e., there was much disagreement).
    - Researchers used Kolmogorov-Smirnov test to determine if the difference between each pair of distributions was statistically significant. They found that, aside from A5 (which was significant), the only other significant difference was A1-A2 and A1-A3.
Finding the “Best” Representation

- There is generally not one “best” visual representation of a network, but a “better” representation is one that highlights the characteristics that are being discussed.
  - For this network, we want to choose an arrangement that clearly communicates the relative order of \((A = B) > C > (D = E)\) in terms of prominence and bridging.
    - A2 has the highest proportion of correctly ordered responses for these measures, followed by A1, A4, A3, and A5.
    - However, A1 has the highest proportion of correct responses for the number of groups (4) in the network.
  - To analyze the relative effectiveness of each arrangement, researchers compared the relative ordering of each pair of nodes for each measure, and classified them according to three types of errors:
    - Alpha error: When a pair of nodes are the same, but reported as different;
    - Beta error: When a pair of nodes are different, but reported as the same;
    - Gamma error: When a pair of nodes are correctly reported as different, but have their orders reversed.
  - Some general observations emerged from these errors:
    - Overall, respondents had lower error rates for bridging than prominence.
    - The circle arrangement (A5) had a relatively high Beta error for bridging (i.e., everyone looked equally bridge-important even though that wasn’t really true).
    - There is much less Beta error for bridging than for prominence.
    - A1 had a high Gamma error rate for prominence (i.e., many people reversed the order of prominence in A1).
    - The low Gamma error rate for A5 is because so many respondents reported all nodes as equally prominent or equally bridge-important. As a result, the Gamma error rate is not especially meaningful.

4. CONCLUSIONS

- Spatial arrangement can impact viewer’s judgments of prominence, bridging, and grouping.
- There are some limits to this conclusion, particularly that perceptions of the most prominent nodes are not significantly affected by spatial arrangements (at least not based on the particulars of this experiment).
- The “best” spatial arrangement for a network varies depending on the information that the arrangement is trying to convey. In this study, the arrangement that led to the “best” perception on number of groups was different from the one that led to the “best” perception of relative prominence. However, the circle arrangement did not perform well for any measure, because it tends to hide structural differences between nodes.
- Further study is needed, but the hope is that this information can be used to understand how biases in perception can occur as a result of spatial positioning, and hopefully how those biases can be avoided.
Multidimensional scaling (MDS) is a procedure that places data points in physical space for the purpose of visual representation.

- The simplest example of MDS is to take data of the distances between pairs of cities, and use the data to construct a 2-dimensional map. However, most MDS problems are not this straightforward:
  - In the example, there is no ambiguity about what is meant by “distance,” since it refers to the actual physical distance between the pairs of cities. But in a more typical MDS problem, the assignment of a distance value is usually more arbitrary, and may be based on subjective assessments rather than actual measurements.
  - In the example, we know the data can easily be arranged in 2-dimensional space and can be represented as a map. But in a more typical MDS problem, we might not know how many dimensions will be needed to represent the data, and they might not produce simple 2- or 3-dimensional visualizations.
- In general, MDS is used to determine whether the distance matrix can be represented by a map that approximately reproduces the original distances. In particular, the objective of MDS is to obtain a good approximate representation of these distances in a small number of dimensions (to make it easy to understand and to visualize).
- There are two types of scaling:
  - Classical: Distances on the map are in the same scale of measurement as the original values from the distance matrix.
  - Ordinal (or non-metrical): Distances on the map reflect a ranking order. This is the method most frequently used for social science data, since much of the collected data is subjective ratings rather than cardinal measurements.

2. EXAMPLES

Two different examples are provided to illustrate the difference between Classical and Ordinal MDS.

- Classical MDS: Reproducing a 2-D map from air distances between pairs of cities
  - The distance matrix consists of pairwise distances between ten cities in Europe and Asia.
  - The MDS maps points in 2-D space (i.e., on a flat plane) such that the “straight line” distances between the points on the graph match the distances as reported by the distance matrix (although it is rescaled to make the graph size more reasonable).
The configuration may need to be rotated and/or reflected in order to more easily interpret the resulting graph. While this might not be as significant of an issue for Euclidean distance data, it is likely significant when dealing with subjective data.

There are three significant points to remember when interpreting MDS solutions as represented by a graph:

- The configuration can be reflected without changing the distances between points.
- The distance between points is not affected if we change the graph’s defined origin by adding or subtracting a constant from the row or column coordinates (as long as this is done uniformly)
- The set of points can be rotated without affecting the distance between points (in other words, the axes can be rotated).

- **Ordinal MDS**: An attempt to determine the dimensions underlying similarity judgments for pairs of 12 countries.
  - In 1968, a group of 18 students were asked to rate the degree of similarity between each pair of 12 countries on a Likert scale from 1 (very different) to 9 (very similar).
  - The study calculated the mean similarity ratings for each pair of cities, and applied ordinal MDS to create a graph.
  - Researchers were interested to know what factors the students weighted most heavily in determining which countries were most similar to, and most different from, one another. They used the visualization as a guide, and rotated the axes (since the original assigned orientation is arbitrary until they found meaningful variables. Specifically, they found variation along two seemingly logical axes:
    - Pro-Western to Pro-Communist (bottom-left to top-right); and
    - Developed to Developing (top-left to bottom-right).
  - Two cautions about these results:
    - Since the results were established by averaging the subjective assessments of 18 students and analyzing the MDS-created graph, it is assumed that all students are using the same two dimensions (defined above) to make their similarity determinations, and that all students are giving each of these factors the same relative weight. However, the reality may be more complicated than that.
    - This approach of mapping the results and then fitting an interpretation to the resulting graph is not always the most effective way of discerning interesting patterns in the data.

### 3. **CLASSICAL, ORDINAL AND METRICAL MULTIDIMENSIONAL SCALING**

- Classical MDS treats the values in the distance matrix as Euclidean distances (i.e., distances in space), and seeks to find a configuration in a low number of dimensions such that the distances between points in the visualization are close in value to the distances in the distance matrix.
  - We sometimes need to determine the values in the underlying data matrix based on the distances in the distance matrix; we can use MDS for this, and then check for fit (particularly to confirm how many dimensions are needed to get a good enough fit).
  - Calculate stress (also known as the normalized sum of squares). The closer the stress value is to zero, the better the MDS solution is a good fit to the original values.

- Ordinal (non-metric) MDS deals with mapping pairs of objects in relation to one another, rather than as an absolute measure. This is important when the underlying data is a subjective
assessment (as social science data often is) rather than a physical value (such as a Euclidean distance).

- We construct fitted distances (“disparities”) that are in the same rank order as the original distances, using least-squares monotonic regression. The goal is to fit a monotonic curve to the to the original and relative distance data, while making the sum of squared vertical deviations as small as possible. The point on the monotonic curve is the fitted (or predicted) value of $d_{ij}$ from the monotonic regression.

- In determining how good of a fit the MDS is, we are now concerned with the closeness of the distances to the disparities, rather than the observed distances, because we only need to reproduce the rank order of the observed distances, not the actual distances themselves. This is also called Kruskal’s stress, type I, and the best configuration is obtained by minimizing this measure of stress.

- Metrical MDS involves non-metrical data, but assumes that the data can be transformed into Euclidean distances by some parametric transformation other than monotonic transformation:
  - Interval scaling assumes that a linear transformation can be used to turn the observed values into Euclidean distances. The disparities become points on the regression line, and the formula for stress remains the same.
  - Ratio scaling involves a situation where the regression passes through the origin, and often gives similar results to classical scaling.

### 4. COMMENTS ON COMPUTATIONAL PROCEDURES

- The aim of MDS is to find a configuration such that the stress value is minimized. Most computer packages accomplish this by starting within an initial configuration in $k$ dimensions, and iteratively improving it by moving the points short distances such as to slightly reduce the stress in each iteration. When these changes fail to affect the stress by some specified amount, the program assumes this is the MDS solution.

- However, because of this method, it is possible that the program will find a local minimum, rather than a global minimum, as a result of the chosen starting point for the iterative solution.

- Therefore, the MDS solution depends on:
  - The choice of initial configuration; and
  - The stress criterion used.

- Because of this, different software packages may give different solutions. If the solutions are very different, it could be that (a) the data contains no strongly identifiable structure, (b) at least one of the solutions is a local rather than global optimum, or (c) complete convergence has not been achieved for one (or both) solutions.

### 5. ASSESSING FIT AND CHOOSING THE NUMBER OF DIMENSIONS

- There are several possible methods to use when assessing whether the MDS solution is a good fit, and whether the appropriate number of dimensions has been chosen:
  - Calculate the stress values for the MDS solution, and compare them with the guidelines established empirically by Kruskal (1964).
  - Create a scree plot, where stress is plotted against the number of dimensions used in the MDS. Remember there is a trade-off: More dimensions means better fit, but harder to understand the visualization. Look for the “elbow” in the scree plot, where increasing
the number of dimensions has little effect on the stress values – this point will represent the best “compromise” between fitness and understandability.

- There are also several diagnostic plots that can assist in this determination:
  - Plot of the distance between points in the MDS versus the fitted values obtained from the monotonic regression. If MDS solution is a good fit, this plot should be a linear relationship with a 45-degree slope and little scatter.
  - Plot of the distance between points in the MDS versus the observed distances. If MDS solution is a good fit, these should have approximately the same rank order and plot should show a monotonic curve (increasing or decreasing).
  - Plot of the fitted value of distance between points versus the observed distances. The fitted values are “smoothed” to have the same rank order as the observed distances; if a large amount of smoothing is required to achieve a monotonic curve, this plot will show a large number of horizontal steps, and we will know the MDS is not a good fit.

6. A WORKED EXAMPLE: DIMENSIONS OF COLOR VISION

An experiment was conducted where subjects were shown several pairs of colors and asked to rate their “qualitative similarity” on a scale of 1 to 5.

- One might guess that a 1-dimensional solution is sufficient, since the difference between colors can be represented as a continuous metric of how far apart their wavelengths are.
- However, a scree plot showed a significant reduction in stress between 1 and 2 dimensions, and that the best MDS configuration was actually a 2-D plot.
  - The plot showed a relatively common “horseshoe” pattern, indicating that subjects were rating reds/violets as being more similar to each other than reds/greens, even though reds/greens are closer to each other in terms of wavelength.
  - This unusual result showed researchers that there were other elements at play in the perception of color similarity besides what could be captured in the wavelength data.

7. FURTHER EXAMPLES AND SUGGESTIONS FOR FURTHER WORK

The following are some topics for further research suggested by the authors.

- Economic and demographic indicators for 25 countries
  - Underlying data consisted of five economic and demographic indicators for 25 different countries. Data needed to be converted to a distance matrix; researchers used radio MDS, but suggest one could try ordinal scaling to compare the results.
  - First tried a 1-D solution, to see if the countries could be placed on a scale of development based on these 5 indicators. The countries lay approximately where one would expect, but Kruskal type I stress indicators determined it was a poor fit.
  - 2-D solution produced much better fit. Graph still showed a fairly straight line, but singled out Romania and Croatia, which had most of the characteristics of developed countries, except with a much lower GDP. So this second dimension is largely a function of GDP.
- Persian archers
• Underlying data consists of similarities between bas-relief carvings of archers in Persia. Researchers want to know whether they were all carved by the same sculptor, several independent sculptors, or team(s) of sculptors.
  o Scree plot suggests that 2-D is adequate, but 3 or 4-D might create a more meaningful representation.
  o Plots in 2-D space show clusters of carvings with high degree of similarity, allowing researchers to draw conclusions about which carvings were likely to have been the work of the same sculptor.

• Dialect words in 25 English villages
  o Underlying data consists of similarities between dialects (specifically, 60 words) in 25 English villages and their relative (Euclidean) distance to one another.
  o Researchers are interested to know how the map of dialects compares to the geographical village map. Topographical features such as rivers or roadways can have a more significant effect on communication than straight-line distance between villages.
  o Researchers suggest implementing ordinal MDS on the similarities in 2-D. Although 2-D does not have the best stress measurement, it is the most logical representation to compare to a geographical map of the region.

• Acoustic confusion of letters of the alphabet
  o Underlying data consists of an experiment involving 300 post office employees who wrote down the letters they thought they heard when letters were spoken against a background noise at a rate of one every 5 seconds. (Conrad 1964)
  o Similarity data was calculated based on the number of times one letter was confused for another. (Morgan 1973)
  o A scree plot shows no clear elbow, and stress remains above 0.1 (i.e., not very good) until you reach a 4-D solution.
  o However, the 2-D plot is easy to understand, and still conveys some valuable information about which letters are likely to be confused.