VISUALIZATION

UNIT - V

Dr. V.S. Felix Enigo, DCSE, SSNCE

Introduction

- Visualization powerful technique for exploring social relationships in social networks
- Earlier methods - hand drawn images, computational methods (Factor analysis, multidimensional scaling), lay out nodes in a 2D or 3D
- Evolved to machine-drawn images and screen-oriented graphics
- Displaying fine graph layouts, coloring, presenting node-edge relations, visualizing complex relations is still challenging
- Moreno – developed sociograms to represent social networks - to explore social relations in a formal study

Graph Theory

- Concepts and metrics in SNA are derived from graph theory

- Node degree: number of edges incident to the node, for loops counted twice
  - At most for undirected graph: N * (N-1) /2
  - At most for directed graph: N * (N-1), where ‘N’ no. of nodes

- Node density:
- Density of an undirected graph can be defined as (2 *E) / N * (N -1) where E is number of edges
- Density of a directed graph can be E / N * (N -1)

Contd...

- Path length: distances between pairs of nodes in a network graph
- Average path length - is the average of these distances between all pairs of nodes
- Component size: counted by the number of connected nodes in a graph
- if a graph is not connected, the graph can be partitioned into several connected subgraphs
- Component size of each subgraph can be calculated by the number of connected nodes in each subgraph.

Centrality

- Identify the most important or central nodes in the network
- HITS and PageRank algorithm are two most famous representatives using centrality for ranking
- HITS - analyzes important nodes based on calculating Authorities (indegrees) and Hubs (out-degrees)
- PageRank calculates node values based on out-degrees
- In social network analysis, three most popularly adopted methods to measure the centrality:
- “Degree”, “Betweenness”, and “Closeness”

Contd...

- Degree centrality: number of edges incident upon a node
  - Centrality is computed for the nodes that have direct connections
  - If the edges are directed, in-degree centrality is differentiated from out-degree centrality
- Betweenness centrality: Computes the extent to which a node lies between other nodes in the network
  - Gives higher value for nodes which bridge clusters
  - It also reflects the number of nodes which a node is connecting indirectly through the direct links

Contd...

- Closeness centrality: is a measure of the order of magnitude that a node is near to all other nodes
  - Calculating by finding mean shortest path for a node to all other nodes
  - Highly ranked with closeness centrality acts as information distributors in social network

Clustering

- In social networks, subsets of nodes highly connected within subset and have relatively few connections to nodes outside the subset
- Such subsets are likely to share some attributes and form their own communities
- Efficient and effective discover such community structures is important
- Important measure to help explore the grouping effects is clustering coefficient
- Clustering Coefficient: measure the degrees of nodes to decide which nodes in a graph tend to be clustered together
  - It is utilized for small world analysis

Visualization

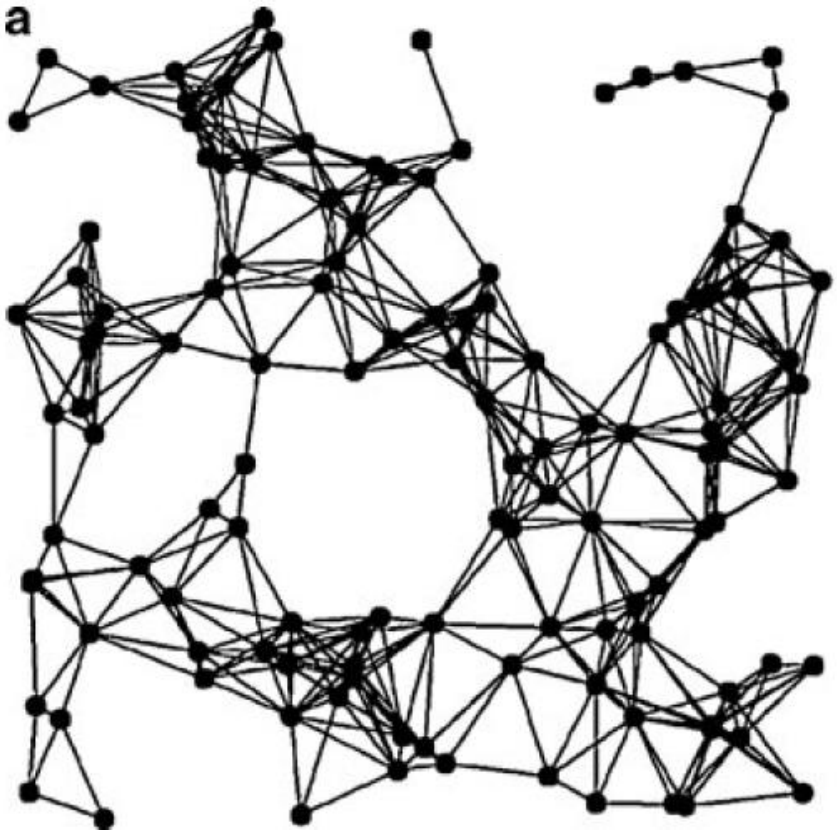
- Visualization plays crucial role of linking human vision and computer
- Helps to identify patterns, extracting insights from large amounts of information
- Help to understand different information structures, various visual representations and metaphors
- Single visualization method cannot fit all kinds of information structures
- Popularly employed visual representations – node-edge diagrams and matrix representations

Node – Edge Diagrams

- An intuitive way to visualize social networks
- It can better present for many network analysis tasks - component size calculation, centrality analysis, and pattern sketching
- Many node-edge layouts have been presented based on size, complexity, and structure of the social network
- Three kinds of layouts for Node-Edge Diagrams - random layout, force-directed layout, and tree layout

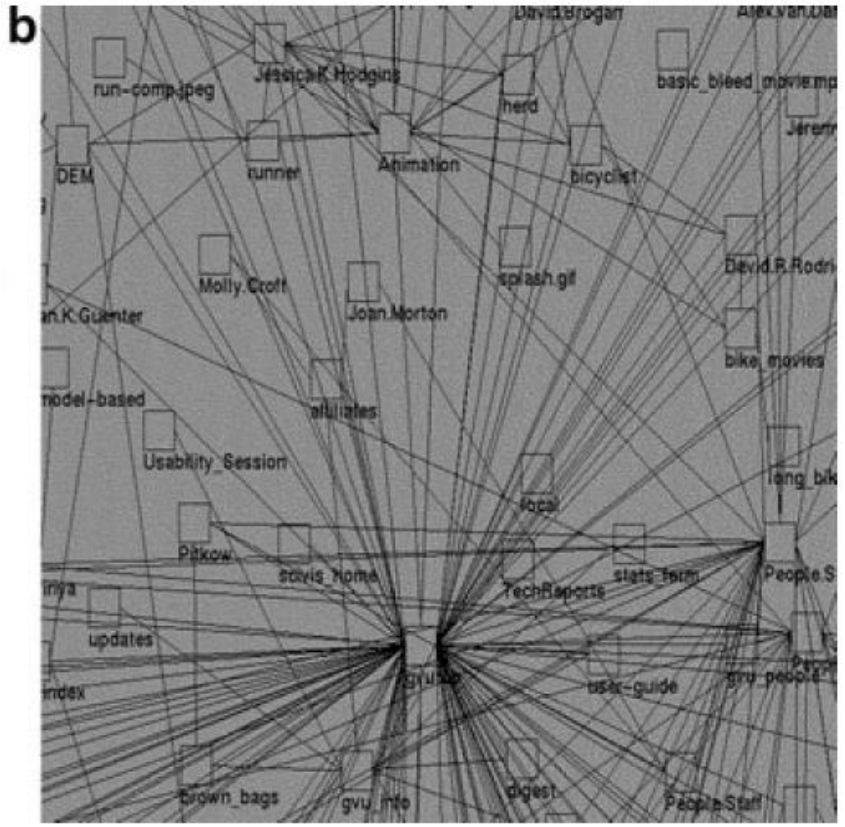
Random Layout

- Put the nodes at random geometric locations in the graph
- Don’t have clear visualization for more than thousands of nodes
- Can efficiently draw the social network graph in linear time O(N)
- Sometimes it can be usable to visualize very large network graphs



Force-Directed Layout

- Edges act as spring and the nodes act as repelling objects (Edge attraction and vertex repulsion)
- Mimics Hooke's law and the Coulomb's law
- Initial random layout yielded first
- force-directed algorithms run iteratively to adjust the positions of nodes
- Performed till all graph nodes and attractive forces between the adjacent nodes run to convergence



Force-Directed Layout Contd...

- Force-based graph layout forms better layouts of the space
- Running cost is much higher than that of a random layout [O (N log N or O (E)]
- Not suitable for graphs larger than hundreds of nodes

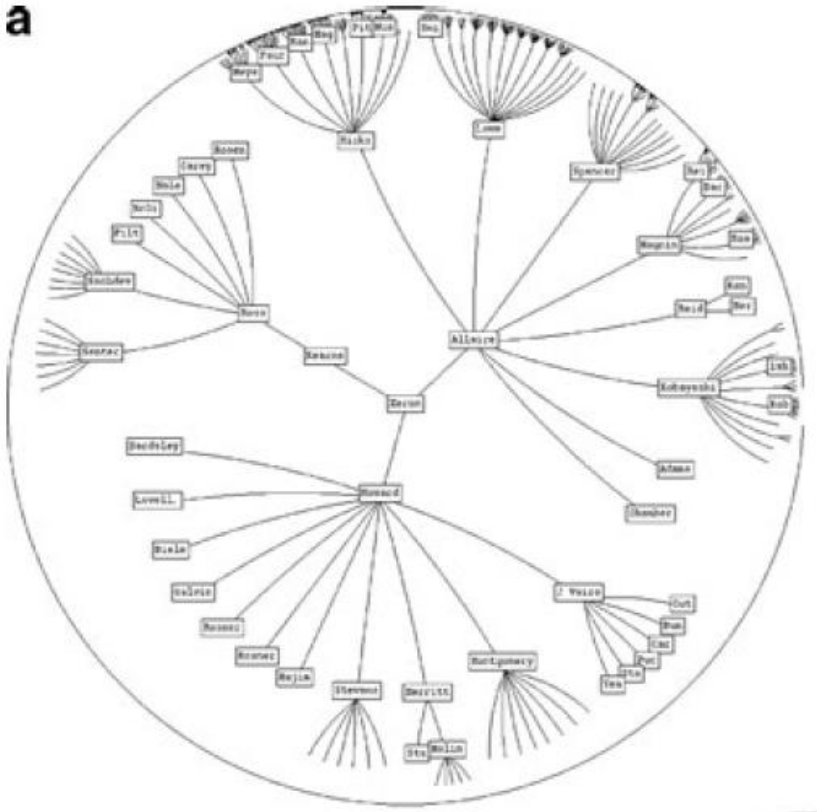
Tree Layout

- Basic Layout, chooses a node as root, nodes connected to it become children, nodes at more levels from root become grand-children so on
- Displays more structural layout than graph considering more contextual information
- Trees are more straightforward to grasp human eye than general graphs
- Special cases takes more constraints
- More contextual information can be extracted - facilitates network analysis

Tree Layout Contd...

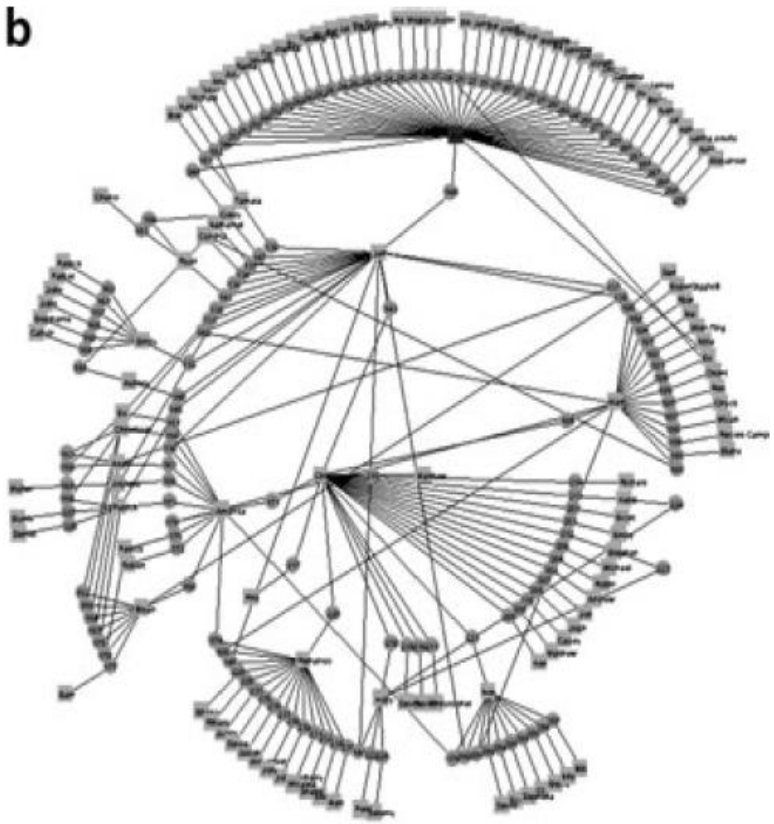
- Domain specific variants of tree layout - hyperbolic tree layout, radial tree layout
- tree visualizations utilize idea of focus + context + animation techniques for better visualization
- Help users to obtain both global and local views in 2D display

Hyperbolic Tree View

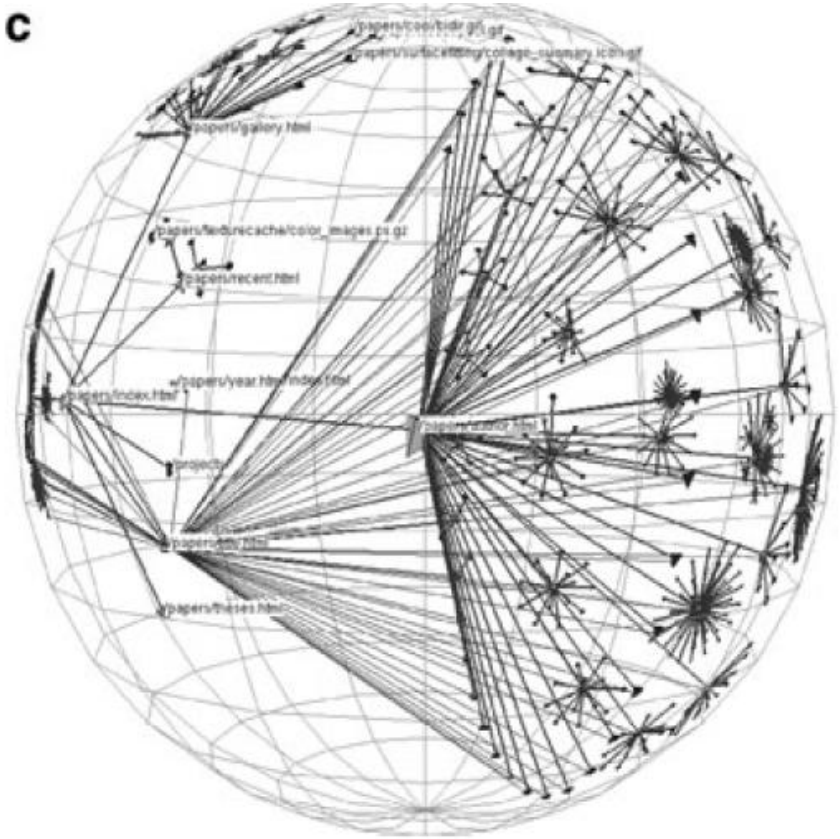


Contd...

Radial Tree Layout



H3 Visualization



Matrix Representation

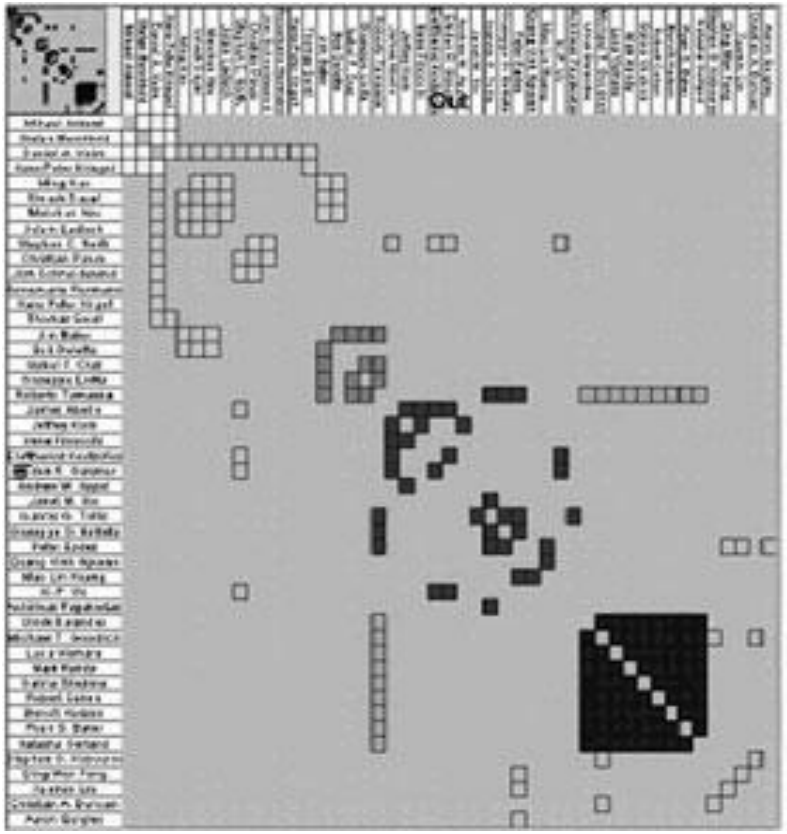
- Social network as graph with nodes and edges be represented as boolean matrix
- Boolean values in matrix can be replaced with valued attributes for more info about edges
- Minimizes occlusion problems caused by the node-edge diagram
- clusters and associations among the nodes can be better discovered
- For Complex relationship outperforms node-edge representations

Matrix Explorer

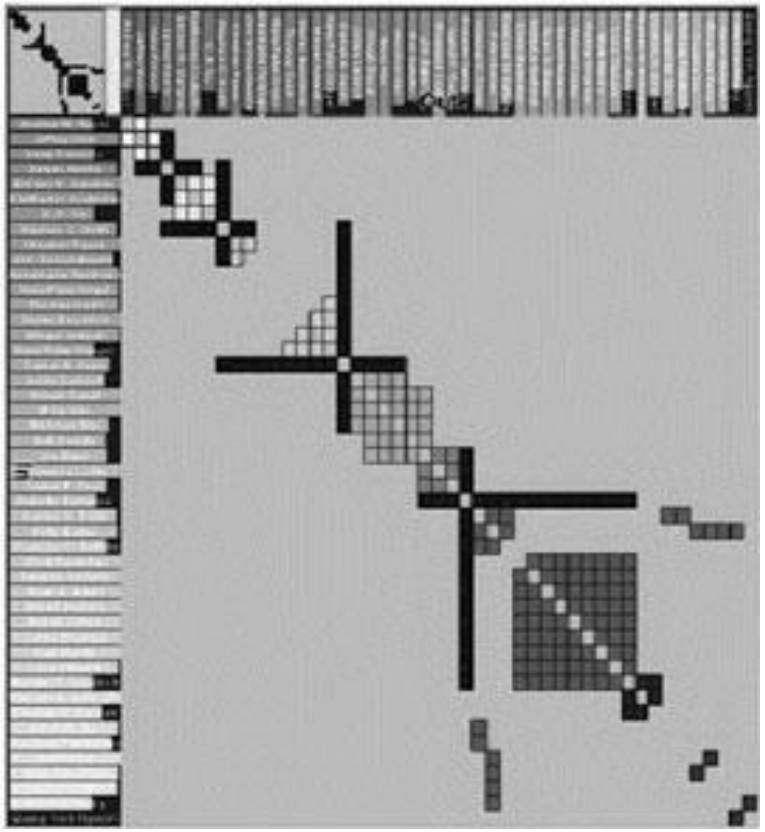
- MatrixExplorer - enhanced matrix-based representation
- Visualize social networks with a Dual-Representation - matrix and node-edge
- For social network with highly interlaced edges, help users quickly recognize the associations between nodes
- Reordered matrix can evidently help users find more clusters
- Not a replacement, but complement the shortcomings of a node-edge diagram

Matrix Explorer

Initial Order



TSP Order

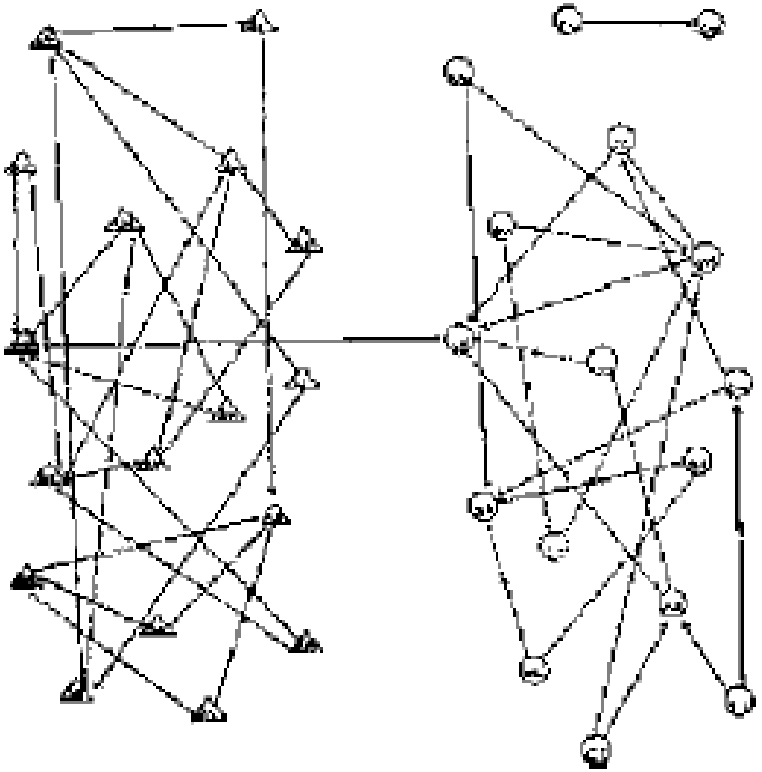


NODE LINK DIAGRAMS

Dr. V.S. Felix Enigo, DCSE, SSNCE

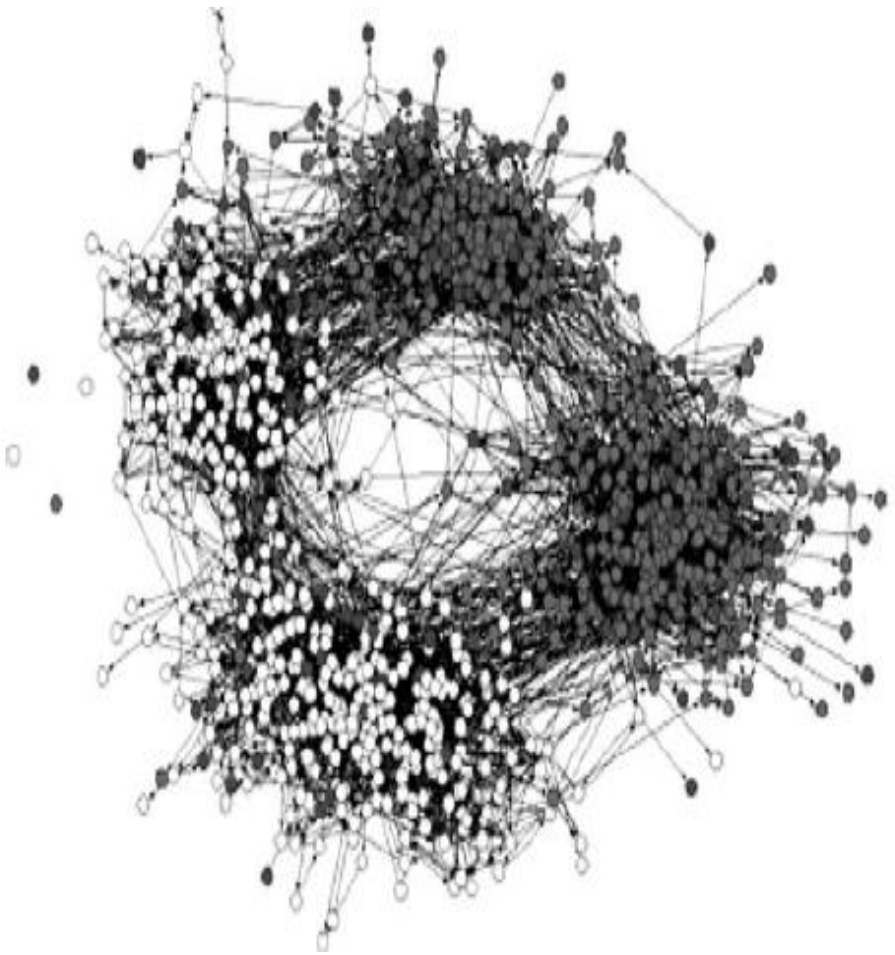
Introduction

- Jacob Moreno was first & pioneer of social network visualization
- To used visualization to support his findings about social friendship in schools
- He used node-link diagrams to represent actors by nodes and
- connections by links
- Different shapes used - for males and females - arrows connect them (direction of friendship relation)



Contd…

- visual representation can highlight central actors as in fig.
- Since 90, it's a Graph Drawing problem - generate algorithms to place nodes in the space based on certain criteria (Ex. Minimize cross links)
- More than 300 graph drawing algorithms to layout graphs in 2D space by researchers
- Difficult to identify a core set - create ideal layout algorithm.



Contd…

- Information visualization different from social visualization
- Last 2 figs shows best actors and social groups
- Different representations may help discover different insights in the data
- information visualization - not aim on ideal representation
- It advocates use of multiple representations for multiple perspectives on the data + interactions to quickly explore them

Scaling to Larger Networks

- Node-link diagrams with large no. of nodes becomes hairball of nodes
- Difficult to transform either automatically or manually into a readable representation
- Solutions:
- Reducing the quantity of information by filtering or aggregating data
- Representing a subset of the network and exploring it incrementally
- Providing more visual space to represent the graph
- Using an alternative representation

Reducing the Quantity of Information

- An obvious technique to reduce size of a graph (remove some of its vertices and edges)
- Two approaches exist to filter networks:
- Filtering out elements, preserving a representative sample
- Filtering data that is not of current interest to the analyst
- Challenging for social networks due to small-world networks properties
- Filtering links results in disconnecting network or losing power-law distribution of connections

Contd…

- Alternatively, hierarchical decomposition of graphs
- each level is coarsened version of the previous one
- It speed-up the layout computation - provides several zooming level
- But due to small-world property, coarsening locally dense graph still produces a locally dense graph in smaller

Second Approach

- Filter nodes and edges according to value of a given measure
- Measure computed from:
- Structural properties of graph. Ex. filter by connected components
- Based on data properties Ex. filter data by year
- Ex. SocialAction - nodes and edges ranked on centrality or betweenness
- It controls sections of the network displayed also color and size

Different Approach

- To reduce quantity of information displayed - aggregate nodes and edges Cluster data using graph clustering techniques
- Then, to gain space, vertices of same cluster is aggregated into a single representative super-node
- Aggregating the network at multiple levels of details Ex. Ask-GraphView
- Aggregation results in loss of detailed information inside the super-node
- So data attributes of individual nodes is averaged
- Or other attributes is created (count of elements in the cluster, averages, min values, etc.)

Incremental Exploration

- Main challenge is to obtain a readable layout in a reasonable time
- Draw trees without crossings in linear time with the number of nodes exists
- Further possibilities is to draw networks as trees and “fix” them by adding additional links
- If no tree structure, the visual representations become less readable
- Solution is to show only a subset of the network + provide interaction to explore the remaining parts Ex. TreePlus
- Disadvantage of incremental exploration is the lack of overview
- Difficult to guide the analysis for users to explore whole network

Using More Visual Space

- Alternate approach to offer more display space minimizing number of link crossings
- Drawback of 3D representations is the occlusion
- Difficult for users to create a mental map of the whole network
- To solve these issues, provides multiple views to users
- Or offer navigation and interaction techniques to visualize network under multiple angles
- However these techniques disorient users, making visual exploration fruitless
- studies show that if 3D visualizations do not improve performances – sometime decreases for several tasks

Alternate Representations

- Treemaps - similar to Venn Diagrams where sub-trees are depicted with inclusion
- Treemaps + Links
- Bar charts and scatter plots – No visual overview of actual actors and connections – answering by query
- Adjacency Matrix Representation

Adjacency Matrix Representation

- Vertices represented in rows and columns – if connected marked at intersection
- node-link diagrams suffer from link crossings for large networks not in matrix representation
- Time to draw less, actors placed linearly
- 2 Factors need to be considered:
- Require reordering of rows and columns to get insights about data
- As it need quadratic space of number of nodes, effective navigation techniques needed to explore

Contd...

- Replace numerical values by visual indicators, reordering rows and columns - improves readability of tables and matrices
- Numerical table difficult to grasp higher-level organization of data
- So, transformed into graphical indicators - rows, columns manually reordered – can easily discover a number of insights
- Techniques to linearize graph, minimize bandwidth of a table are used to reorder adjacency matrices
- Techniques quality depends on data and task

Navigation

- Techniques exists to navigate in large spaces at different levels of details
- Bird’s eye views - miniature overviews of whole representation, users can move the position of their current view, faster navigation than standard scrollbars
- Fisheyes - allows visualizing multiple levels of details in a single view, acts as magnifying lenses increasing details on regions of interest
- Folding the space in 1D or 2D to provide both readable labels and context
- Techniques provide navigation in aggregated matrices

Visualization with Matrix or Node-Link

Dr.V.S.FelixEnigo

Advantages of Matrices

- Low create time, readable, good to initiate an exploration.
- Do not suffer from node overlapping (good to read actors label)
- No link crossing each other (good for dense network)
- Highlight the lack of connections (as shows all possible pairs)
- Shows directedness of the connections

Advantages of Node-Link

- Familiar representation to a wide audience
- For small or sparse networks, node-link diagrams are more effective than matrices
- For compact representation, node-link diagrams are a better choice.
- Analysis for path-related tasks, node-link diagrams are more appropriate.

Matrix + Node-Link Diagrams

- Combines the advantages of both representations
- Goal is to support the visual exploration of social networks. Example: MatrixExplorer

MatrixExplorer

- It consists of 4 stages:

1. Initiate the exploration
2. Explore interactively and iteratively
3. Find a consensus in the data or validate an hypothesis
4. Present the findings

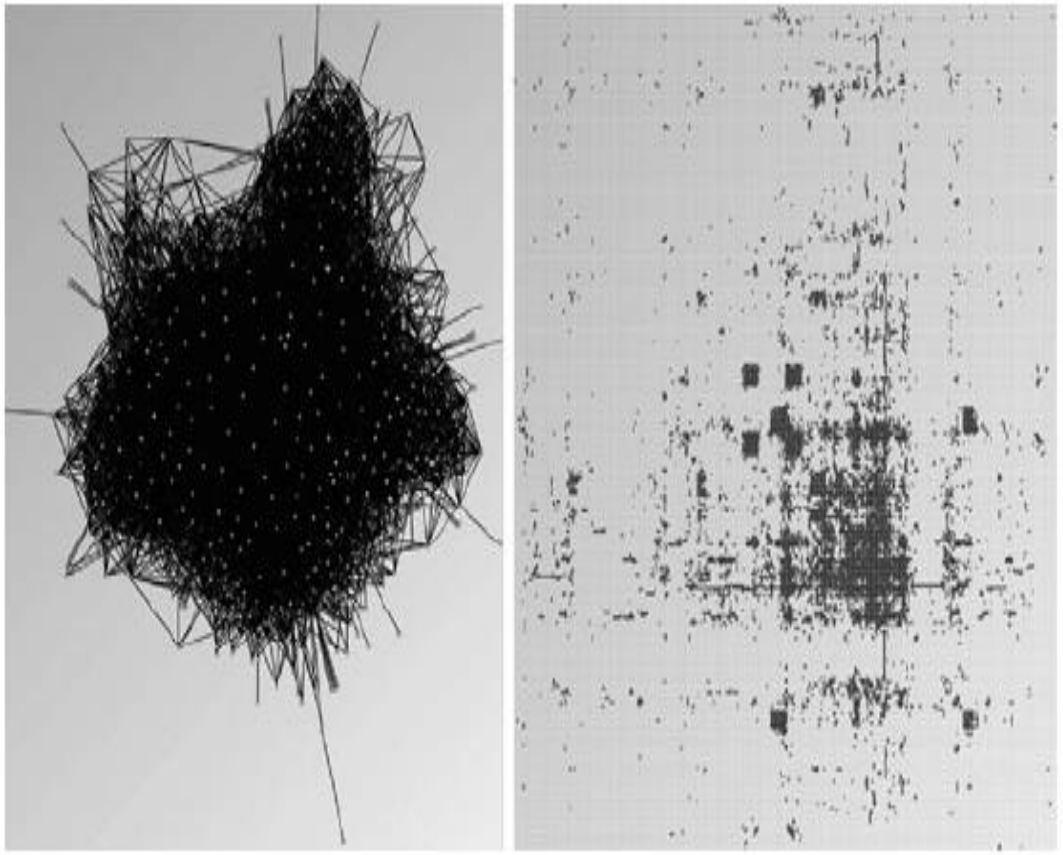
Initiate Exploration

- Matrices provide readable representation of a network (large), low rendering time – suitable to initiate the exploration

Example: Email exchange of more than 450 persons during a year

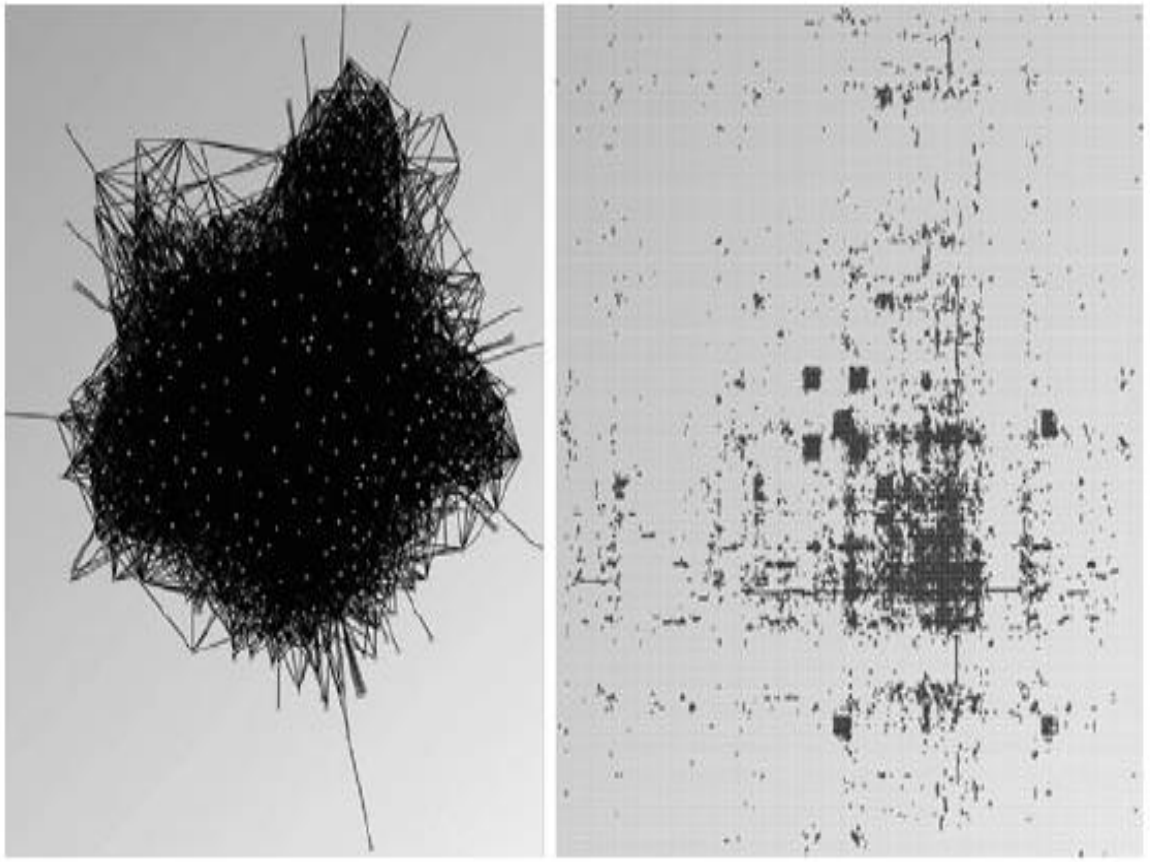
Inferences:

- Email exchanges between two persons are represented by a link or a cell filled with black in the matrix.
- Using traditional force-directed layout, makes it difficult to identify specific nodes or links
- Dense network represents everyone have been exchanging emails
- Few nodes on the periphery, indicating that a few persons did not communicate



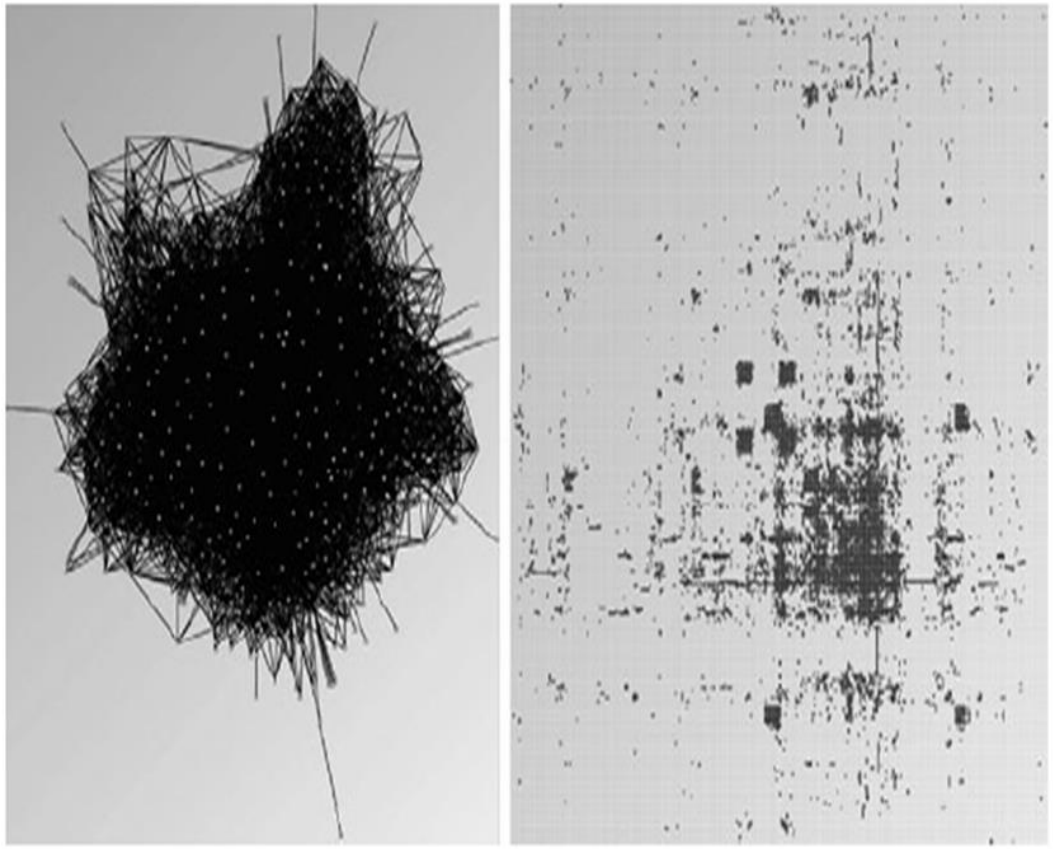
Contd…

- Matrix representation conveys far more information
- Black dot represents a connection between a row and a column (i.e. an email exchangebetween two persons)
- The gray background shows the lack of connection
- Allows an analyst can quickly assess the network
- Majority of gray in the matrix showing that many actors did not exchange email with each other
- Clusters of black dots represents groups or research teams



Contd…

- Cross pattern: vertical and horizontal lines constituted of black dots- administrative service, dealing with travels of the whole institutions and thus, communicating with many persons in the network
- Shows power of matrices
- When correctly reordered, matrices highlight salient patterns of a network such as clusters or central actors
- Need expertise to decode and interpret these visual patterns



Explore Interactively

- Exploration process itself is iterative and requires the creation of multiple visualizations
- Interaction on these representations includes the configuration of the visualization
- Example: adjust its layout and its graphical attributes, the filtering, grouping, aggregation of some of its elements
- Both the matrix and node-link representations support the analysis of the network at different levels of details.
- Overview of the network to identify its main communities, the matrix is the best option
- More detailed analysis to identify actors bridging two communities node-link diagrams good

Contd…

- MatrixExplorer provides multiple views of the network and number of tools to interactively manipulate matrix and node-link representations
- Matrix and node-link representations are synchronized to ease the identification of visual patterns
- Selecting a row or column in the matrix highlights the corresponding node in the other representation.
- Visual variables such as size or color can be shared by both visualizations
- Use matrices for some tasks and node-link diagrams for other
- Selecting a visual pattern in the matrix and visualizing its equivalent in the node-link diagram makes easy understanding to less expert users

Contd…

- To interactively manipulate matrix and node-link representations set of tools:

1. Interactive specification of visual attributes.

- The user controls the mapping data-visual encoding by entering values in a text field or selecting a value in a list
- Visual attributes of nodes label, color, transparency or size etc

2. Interactive layout and reordering

- Users may directly move a node or a row/column in both representations to change its position or order

Contd…

3. *Automatic layout and reordering techniques*.

Algorithms to Automate layouts and reorderings to ease users computation time and quality

4. *Computer-assisted layoutand reordering techniques*:

To apply layout and reordering algorithms to specific subsets of the network

5. *Interactive filtering*.

Filtering actors or connections according to a selection or by selecting a specific value of a data attribute from a list using dynamic queries

Contd…

6. *Interactive clustering*.

- Groups of actors mark them and associate them to a visual attributed such as the color or shape of the nodes.

7. *Overview + Detail techniques to navigate in both representations*.

- To support navigation in large visual spaces, focus + detail techniques Bird’s eye view to nagivate and a fisheye lens to magnify regions of interest for details
- A Treemap to represent the macrostructure of the network
- A fast filtering mechanism to isolate each connected component of the network.

Find a Consensus in the Data

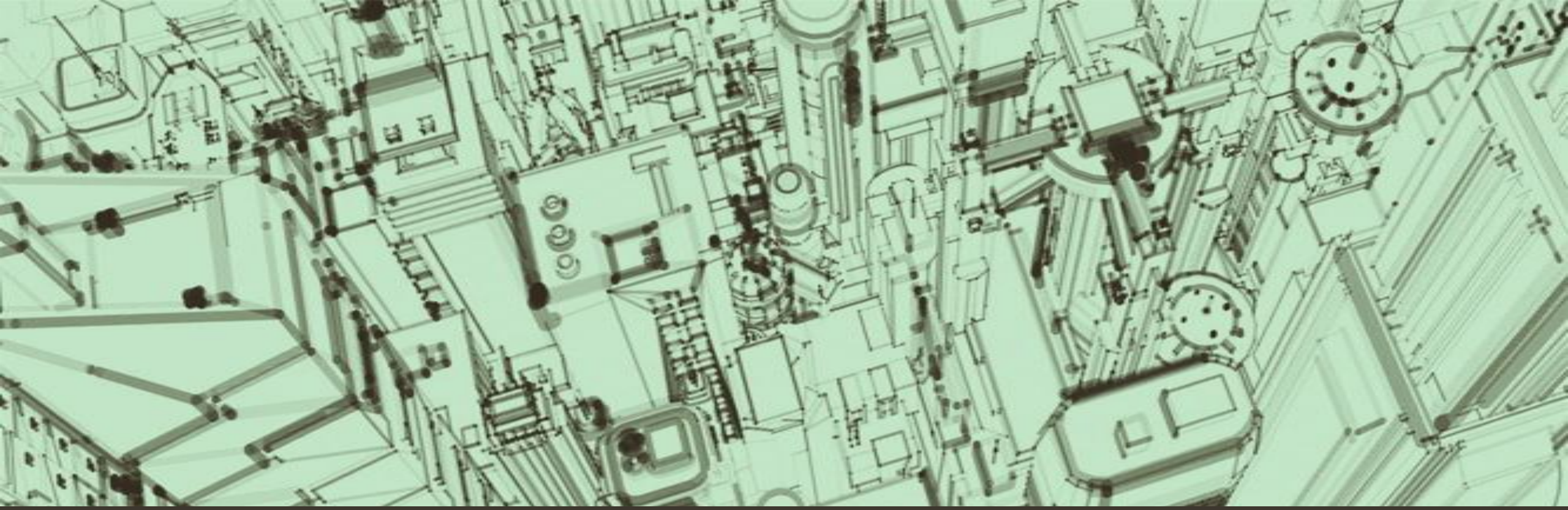
- Each visualization may lead to the discovery of different insights. While in many cases, these i
- Insights confirmed by searching them using different representations, layouts or order during the analysis
- Different techniques to reorder the matrix may lead to different cluster sets.
- To help analysts find a consensus and validate hypotheses, some support is needed.

Contd…

- MatrixExplorer allows analysts to find consensus in the data through simple interactions.
- Reordering the matrix several times, analysts can identify clusters appearing clearly in multiple orders as more valid.
- To mark the uncertainty of attribution of an actor to a given cluster
- Degree of membership of the element to a given cluster (less likely to belong to a cluster with a lighter color)
- Supports overlapping clusters and multiple sets of clusters: elements may belong to multiple clusters at the same time.

Present Findings

- Matrix representations may prove effective when exploring large networks
- Node-link diagrams are essential to communicate findings to a wide audience
- Many node-link diagrams may be created for presenting results with different filters and possibly different aggregations
- MatrixExplorer allows users to generate pictures while performing the exploration

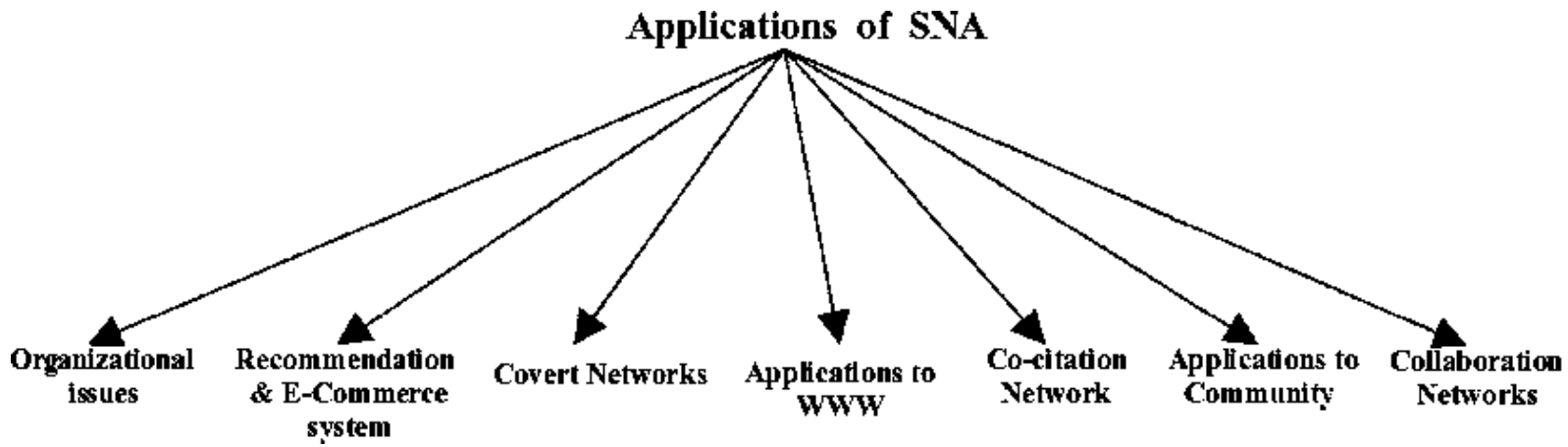


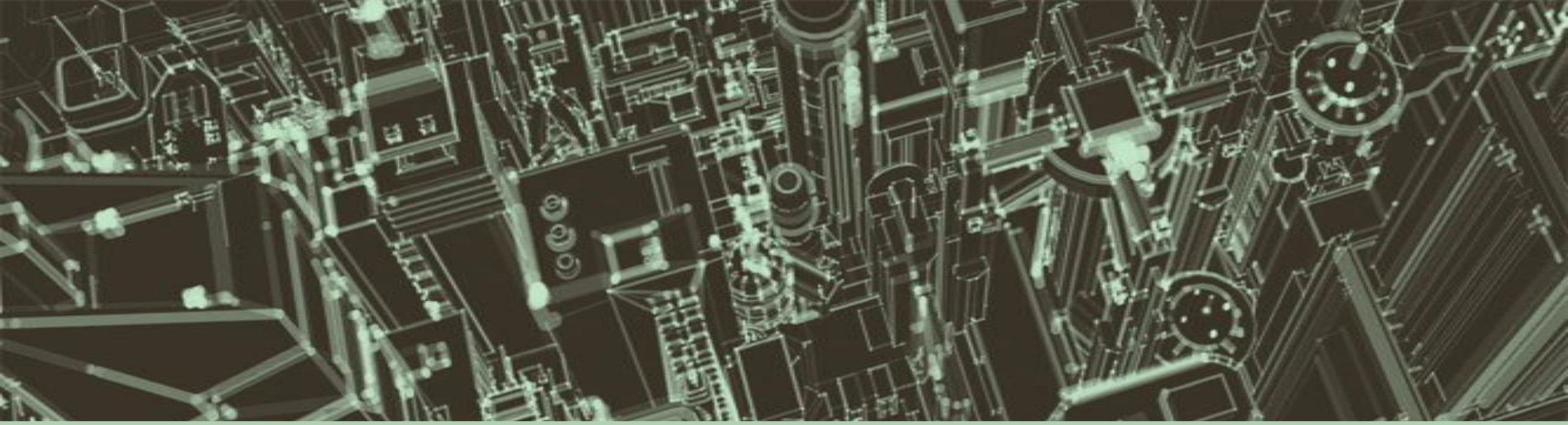
Applications of Social Networks

Introduction

- Covert Networks
- Community Welfare
- Collaboration Networks
- Co-Citation Networks

Focus is on these four!!





1. Covert Networks

What are Covert Networks?

- The covert networks are **hidden -** the actors of such network **does not disclose their information** to the external world.
- Covert groups have **cellular networks structure** which is different from hierarchical organizations.
- Ex. **The terrorist and criminal networks**.
- SNA has been successfully applied to such domains to **understand covert cell operations** and their organization.
- Thereby, you can **combat terrorism**.

SNA -Terrorist cells and database

- SNA has been used to **understand** the **communication** and **structure of terrorist cells**.
- SNA is applied on terrorism database for
  - predicting node and links
  - Discovering interesting patterns and actors involved in an event.
- In this context, SNA discovers
  - who is central within organizations
  - which individuals-removal would most effectively disrupt the network
  - what roles individuals are playing
  - Which relationships are vital to monitor.

SNA – Predict Terrorism

- Another vital **application of SNA** for terrorist database is to **predict terrorism activities**.
- **Terrorist organizations** have **special structures** on
  - Recruitment
  - Evolution
  - Ideas diffusion in network.
- Studies have shown that **these** types of **networks** can be **well understood** by **mapping them**.

One example = 9/11 attack

- The **Valdis Krebs** [9] has used **social network analysis** to **map the terrorist network** that attacked **on 9/11**.
- In spite of unavailability of complete and proper knowledge of all actors and connections in between them, his analysis has disclosed network which is almost near to real network.

Sources of data – to build terrorist database ??

1) The data to build and complete such networks is gathered from **publicly available resources** such as **news papers**
2) Now a days **Web resources** such as **blogs, emails** etc. are also used for hidden communication.
3) Hence, **various data mining** and social network analysis techniques are employed to **extract necessary information** to detect terror.

Problem of Connecting Dots

- SNA considers terrorists networks analysis as a problem of connecting dots.
- Connecting multiple pairs of dots exposes the total network.
- Centrality is the most important and widely used measure in SNA.
- The various other factors are:
  - Betweenness centrality
  - Degree centrality measures
  - Cohesion factors
  - Closeness

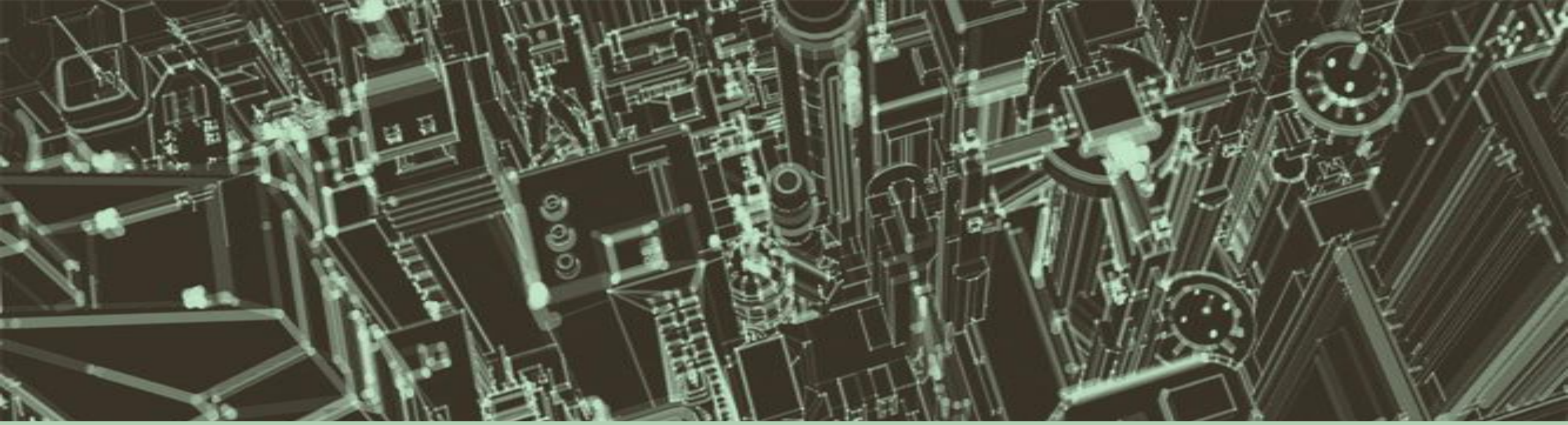
Steps involved in TNA

a. **Identify key players** in terrorist network using the problem of connecting dots.

b. Identify the **actors linked to these key players** – By doing so, the whole network is found out.

c. The **regular day-to-day activities** of the key players are **monitored**.

d. Use Structural cohesion to find **connectors among group of actors** - This measure is used to **identify sub-groups in an organization** having similar features skills and involvement in particular event.



2. Community Welfare

1. Spread of Disease

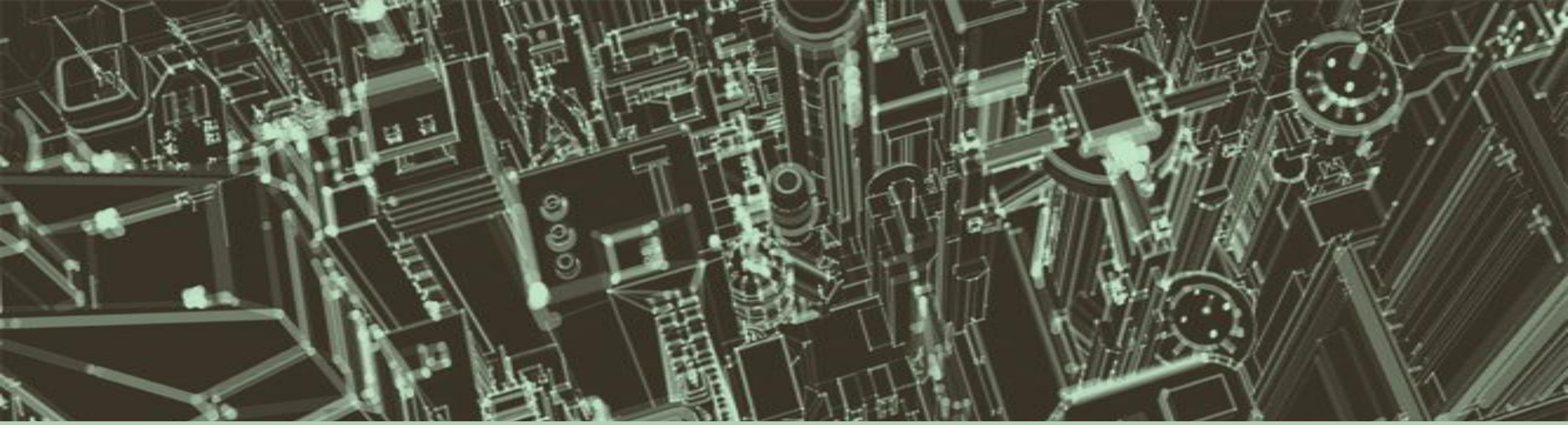
- The SNA techniques can also be used to improve the **community welfare**.
- SNA is used to analyze different types of relations such as
  - Communication patterns Physical contacts Sexual relationship etc.
- The **SNA** may **reveal** the **patterns of human contact** which may lead to **spread of disease** such as HIV in population.
- It has been employed in **epidemiology** and has shown considerable results for community improvement.
- **Another** interesting **application** is to use SNA to **examine and observe farm animal network** to identify **patterns of disease spread** from one animal to another.

2. Monitor Suspected People Behavior

- **Mass surveillance** practice is undertaken by some organizations and governments to **monitor the behavior of suspected people** of population.
- This is done with the **purpose** of **protecting people from criminals**, terrorists or political subversives to maintain social control.
- In US, the Total Information Awareness program of the Information Awareness Office **designed numerous technologies** to be used to **perform mass surveillance** which made **use of SNA tools.**

3. Strengthening Community resilience

- **Social Networks** which are made for **strengthening community resilience against disasters** (natural or human-made) can **reveal vulnerabilities** within a network [13].
- These networks are analyzed using SNA tools to **study the changes** that occur during disaster and further to **improve disaster preparedness** strategies.
- **Knowledge Sharing:** The **SNA tools** have also been used to **assess the communities of practices - This information** can further be used to **improve knowledge sharing** in community.



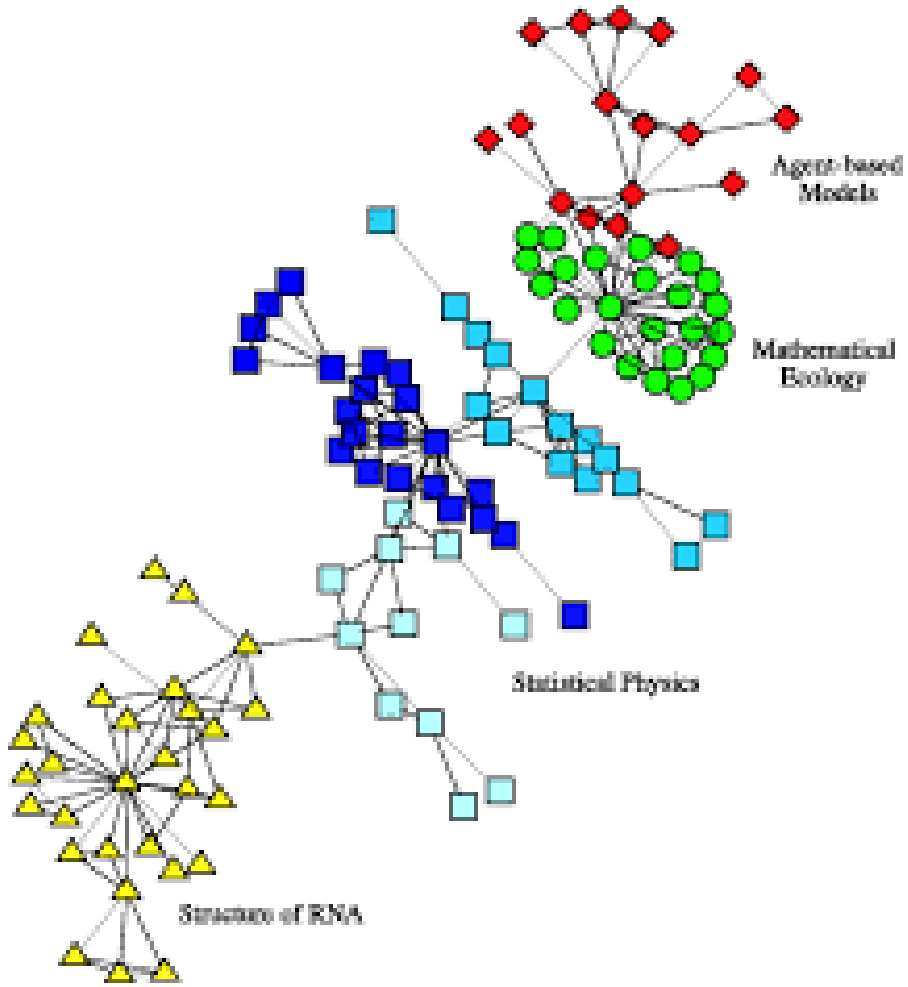
3. Collaboration Networks

What are Collaboration Networks?

- **Collaboration network** consists **groups of persons working together** to perform particular activity
- Studying human collaboration is an important topic in sociology.
- The various types of collaboration networks are:
  1. Co-authorship networks
  2. Movie actor network
  3. Knowledge collaboration network

1. Co-authorship Networks

- **Co-authorship** of a paper can be thought of as **documenting a collaboration between two or more authors**, and these collaborations form a “co-authorship network”



Advantages of Co-Authorship Networks

1. **B**etter way to **improve** the **interdisciplinary research** is by **identifying such current interactions** and engaging involved institutions and researchers for future research.
2. **R**eveals the **ego networks of** prominent **key-players** in the network.
3. **U**nderstand the **influence of individual researchers**.
4. **S**tudy **dynamics** in **patterns of interactions** between educational entities or communities.
5. **Strategic planning** of research and development.
6. **S**cope of research discipline at particular location so that further new inventions in same can be promoted.

Examples of Co-authorship Networks

- Wikipedia article authors
- Network of the pacific Asia Conference on Information Systems
- Network of European Conference on Information Systems (ECIS) etc.

Required Datasets

- The required datasets for co-authorship network analysis is mostly extracted from sources
  - Scientific journals
  - Bibliographic records
  - Digital libraries.

Measures

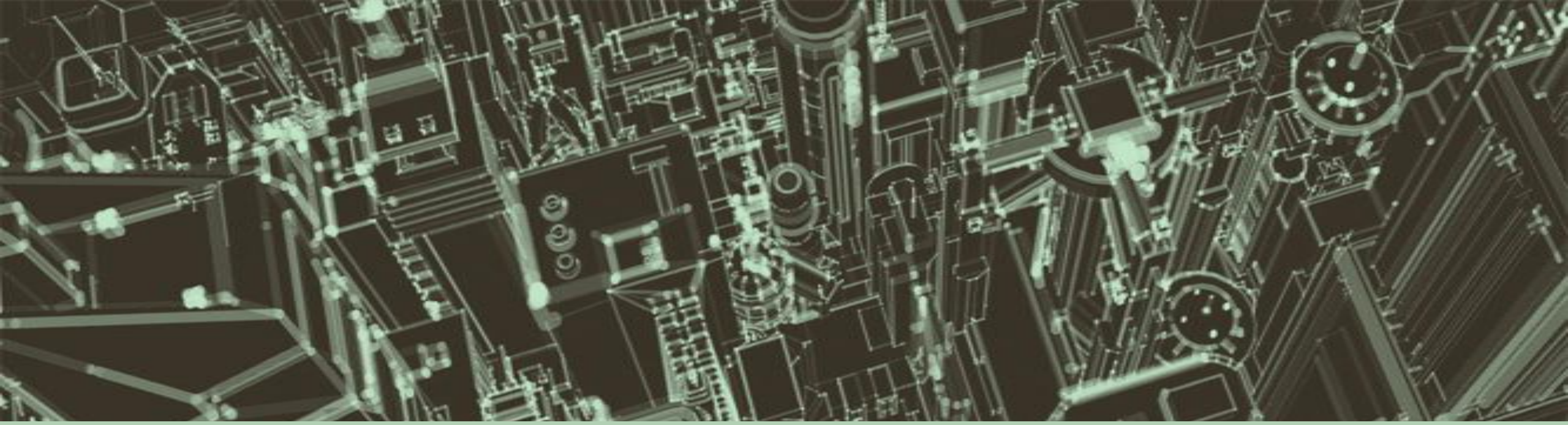
- The **important SNA measures** used **for co-authorship network** includes - **cohesion, network density and centrality.**
- The **cohesion** is used to **identify the subgroups** within network with respect to each research subject.
- The **node similarity** measure in this context **represents extend of similar subject skills**.

2. Movie Actor Network

- **Movie actor network** is analyzed to **study the interaction** amongst themselves, to **discover closely related actors**.
- It is **built based** on **Internet movie database** (www.imdb.com) consisting of all movies and their casts.
- In this network,
  - **Nodes** represents the **actors**
  - **Ties** represent two connected **nodes acted together in some movie**.

3. Knowledge Collaboration Network

- The **information** about **Open Source Software** needs to be **distributed** amongst community or users - because not all members have required knowledge or skills for such software usage and development.
- Hence, **success** of such software highly depends on **distribution of knowledge** using **tools** such as emails, discussion forums, web blogs etc.



4. Co – Citation Networks

What is co-citation?

- **Co-citation** is used as a **measure of similarity** between **two objects**.
- Co-citation analysis helps to understand the status and structure of scientific research.
- The **Co-citation network** can be viewed as a **bipartite graph** showing **linkage between two different groups of documents**.
- Basic **two approaches of co-citation** are
  - Author co-citation
  - Document co-citation

Example 1 – Research network

- In the field of **methodological evaluation**, **co-citation analysis** has been employed to **search for invisible colleges**.
- This **reveals the research network consisting of different institutions** linked to each other informally by having indicators to each others documents/papers.
- This can be used to **get group of institutes having similar ongoing research**.
- This may help to **promote further research** in respective area in those institutions.

Example 2 – Finding Journal Importance

- **SNA** has been also studied as an **approach to understand journals importance** or prestige.
- It also helps to **figure out how** does any **journal - influence or get influenced by - the other journals/papers** in same or different discipline.
- The **node similarity** measure is used **to find similarity between two articles** or publications.
- **Nodes** represent **papers**.
- **Existence of link** shows that **two articles were cited in other articles**.