
5-Weeks Course on Interactive Visual Network Exploration

Week 3: Data Shaping Techniques
and Challenges

Jan 26th, 2022



Visual+
Interactive
Data

Inria



THE UNIVERSITY
of EDINBURGH

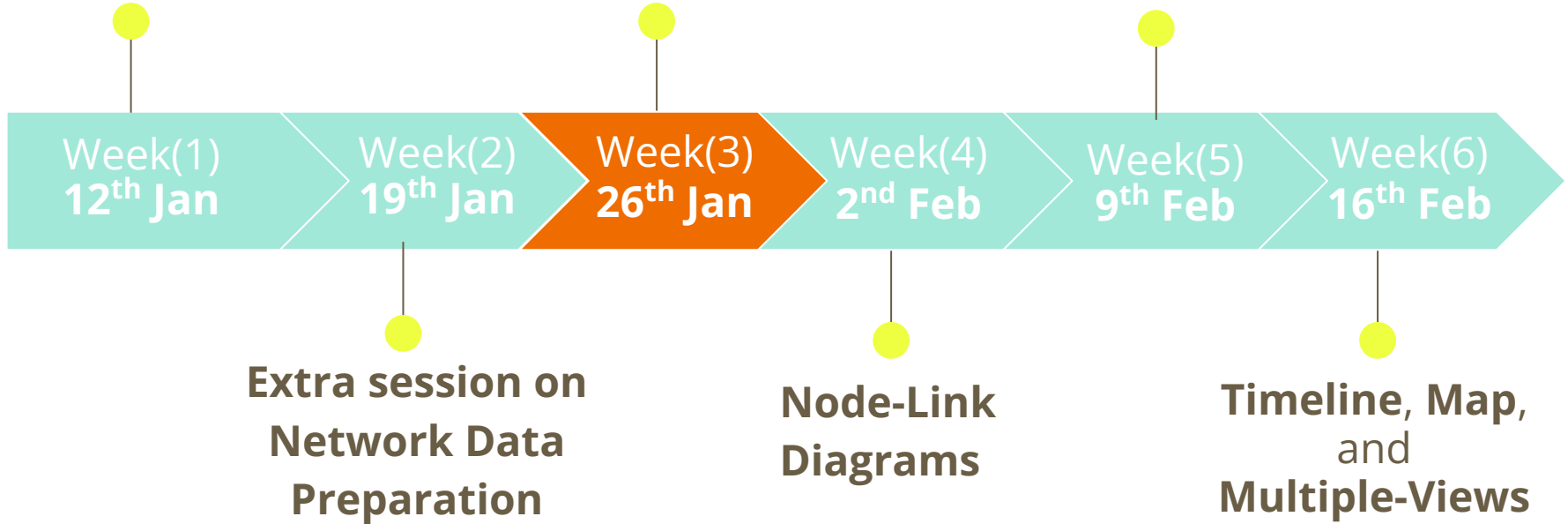
Course Goals

1. **Structure your network data and prepare it** for visualization with the Vistorian.
2. **Define goals of your exploration** and what you aim to learn about your network data using visualizations.
3. **Know a range of network visualizations**, through theory and hands-on use.
4. **Use different types of interactive visualizations** to explore your data.

Network Data Preparation

Data Shaping Techniques and Challenges

Adjacency Matrices



Session Outline

- Discussion on Challenges Faced from Concept Map and Network Tables
- Examples on Visual Network Exploration
- Ensuring Consistency of your Data
- Common Challenges in Network Visualizations
- Assistive Techniques in Shaping your Data

Let us hear from you

Types of Data you might dealt with:

- Temporal
- Geolocation
- Numerical
- Strings and text
- Categorical

Have you faced any obstacles in:

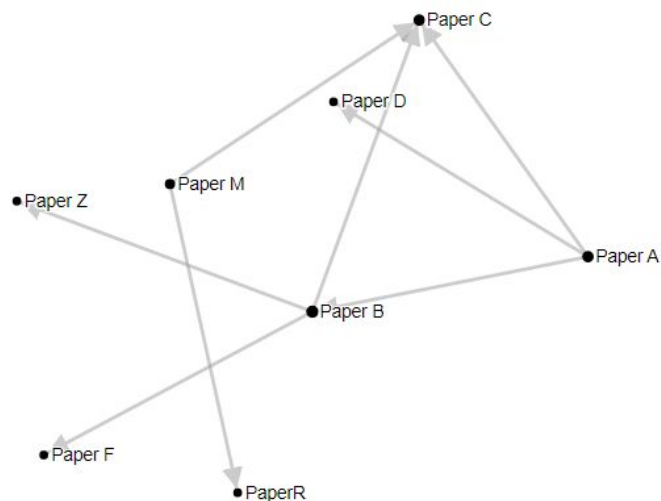
- Mapping your nodes and links
- Creating your tables
- Choosing the suitable tables
- Certain data types that you were

Co-Authorship Network Example

Authors	Paper	References
John, Sara, Kyle	Paper A	Paper B, Paper C, Paper D
Jamal, Sara	Paper B	Paper C, Paper F, Paper Z
Rayn, Nicole	Paper M	Paper C, Paper R

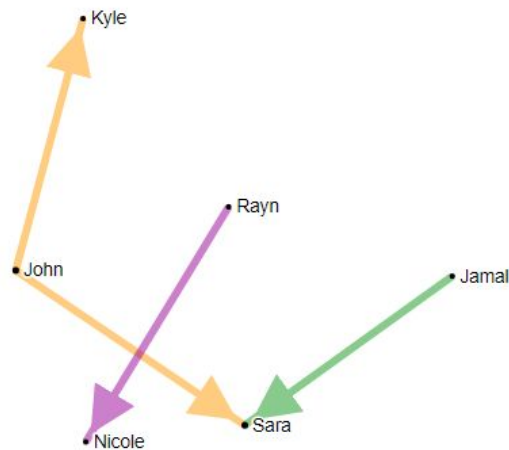
Co-Authorship Network Example (Paper-to-Paper)

Paper	References
Paper A	Paper B
Paper A	Paper C
Paper A	Paper D
Paper B	Paper C
Paper B	Paper F
Paper B	Paper Z
Paper M	Paper C
Paper M	PaperR



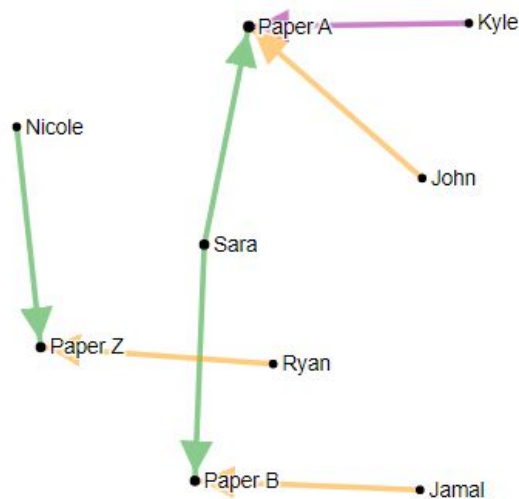
Co-Authorship Network Example (Author-to-Author)

Author 1 (Source)	Author 2 (Target)	Paper (Link Type)
John	Sara	Paper A
John	Kyle	Paper A
Jamal	Sara	Paper B
Rayn	Nicole	Paper M



Co-Authorship Network Example (Author-to-Paper)

Author	Paper	Link Type
John	Paper A	First
Sara	Paper A	Second
Kyle	Paper A	Third
Jamal	Paper B	First
Sara	Paper B	Second
Ryan	Paper Z	First
Nicole	Paper Z	Second



Ensuring Consistency of your Data

Ensuring Consistency of your Data (Data Types Format)

By checking your data have:

- Same spelling of similar nodes/locations/relation types
- Same date format for each single column.

Exercise (1) : Python Notebook- Checking Format (15 mins)

Google Colaboratory

- Visit <https://colab.research.google.com/>
- Upload your data file + Python notebook
- Play with your data

Ensuring Consistency of your Data (Duplicate Values)

- **Dealing with duplicate:**
 - Links
 - Nodes
 - Check your IDs are correct - if exist- .

Exercise (2) : Python Notebook- Checking Duplicates (10 mins)

Ensuring Consistency of your Data (LinkWeight)

Do you need to use the link weight? If yes, is it a binary link (0/1) or a nonbinary link weight where the weight denotes to a value (e.g. number of connections, amount of money spent, number of collaborations or patents)

- **Link Weight:**

- Quantifying link weight in case of categorical values
- Changing link weight based on temporal changes
- the range limit on numeric values - if exist -(ex. Link weight)

Exercise (3) : Python Notebook- Checking Link Weight (15 mins)

Common Challenges in Network Visualizations

Common Challenges in Network Visualizations

- **What if my data contains semi/unstructured data?**

Here are some examples of semi/unstructured data:

- Webpages
- Customer feedback: from where it was initiated? Website, online forms, social media accounts
- Emails are semi-structured:
 - From, to, category \Rightarrow structured
 - Subject and content \Rightarrow unstructured
- **Images, Audio, and Video** : can be tagged with title, category, subject, description, ..

Common Challenges in Network Visualizations

- **What if my data contains semi/unstructured data?** More examples such as
 - **What are you exploring?**
 - Relationship between participants?
 - Things or relations that are common in your network?
 - How did people feel toward a certain topic (sentimental Analysis) with two types of values (link type: positive and negative)
 - **How can I start?**
 - Use libraries to extract words of interest such as natural language processing (NLP) with keyword/entity extraction

Common Challenges in Network Visualizations

- Preprocessing data involves:
 - Join data from different tables
 - Reducing noise,
 - Eliminating irrelevant information (for example, stop words),
 - Slicing data into more manageable pieces of content

Common Challenges in Network Visualizations

- **What if my network** is too large to visualize?
 - **Breaking your data into smaller chunks:**
 - E.g. based on criteria
 - Network Sampling
 - **Filtering mechanisms to reduce size:**
 - Select nodes and links of specific types
 - Slice by time, choose only specific nodes or links

Assistive Techniques in Shaping your Data

Grouping/Aggregating Data Techniques :

- Aggregate rows to higher level (e.g. broader categories, semantic operations, .. etc)

User	Click	Time
A	Announcement 1	9:05
A	Announcement 2	9:08
A	Serve Client 1	9:10
A	Serve Client 2	9: 20
A	Start Task R1	9:30


User	Click	Time
B	Start Task B3	9:00
B	Start Task R1	12:00
B	Announcement 1	1:03
B	Serve Client 3	1:30
B	Start Task R2	2:00

User	Click	Time
A	Announcement 1	9:05
A	Announcement 2	9:08
A	Serve Client 1	9:10
A	Serve Client 2	9: 20
A	Task R1	9:30



User	From	To	Time

User	Click	Time
A	Announcement 1	9:05
A	Announcement 2	9:08
A	Serve Client 1	9:10
A	Serve Client 2	9: 20
A	Task R1	9:30

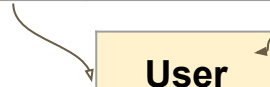


User	From	To	Time
A	Reading Announcements	Serving Client	9:03
A	Serving Client	Task R1	9:10

User	Click	Time
A	Announcement 1	9:05
A	Announcement 2	9:08
A	Serve Client 1	9:10
A	Serve Client 2	9: 20
A	Task R1	9:30

User	Click	Time
B	Task B3	9:00
B	Task R1	12:00
B	Announcement 1	1:03
B	Serve Client 3	1:30
B	Start Task R2	2:00

User	From	To	Time
A	Reading Announcements	Serving Client	9:03
A	Serving Client	Task R1	9:10
B	Task B3	Task R1	9:30
B	Task R1	Reading Announcements	12:00
B	Reading Announcements	Serving Client	1:03



Assistive Techniques in Shaping your Data

Grouping/Aggregating Data Techniques :

- Group multiple columns to a certain column(s) based on your network :
 - Applied for job + rejected \Rightarrow Unsuccessful and link weight =1
 - Applied for job + interview + rejected the offer \Rightarrow Unsuccessful and link weight =2
 - Applied for job + second interview \Rightarrow Successful and link weight =2

Name	Job Title	Initial Acceptance	1st Interview	2nd Interview	Accepted Offer	Link Type	Link Weight
John	Job A	Rejected	No			Unsuccessful	1
Sara	Job B	Accepted	Yes		No	Unsuccessful	2
Adel	Job A	Accepted	Yes	No	Yes	Successful	1
Lia	Job C	Accepted	Yes	Yes	Yes	Successful	2

Assistive Techniques in Shaping your Data

Grouping/Aggregating Data Techniques :

- Group multiple columns to a certain column(s) based on your network :
 - Applied for job + rejected \Rightarrow Unsuccessful and link weight =1
 - Applied for job + interview + rejected the offer \Rightarrow Unsuccessful and link weight =2
 - Applied for job + second interview \Rightarrow Successful and link weight =2

Source	Target	Initial Acceptance	1st Interview	2nd Interview	Accepted Offer	Link Type	Link Weight
John	Job A	Rejected	No			Unsuccessful	1
Sara	Job B	Accepted	Yes		No	Unsuccessful	2
Adel	Job A	Accepted	Yes	No	Yes	Successful	1
Lia	Job C	Accepted	Yes	Yes	Yes	Successful	2

Assistive Techniques in Shaping your Data

Grouping/Aggregating Data Techniques : (Advanced)

- Choose only highly connected nodes to upload
- Do some clustering and visualize clusters as nodes.

Assistive Techniques in Shaping your Data

Refactoring Data Techniques :

- Projection of your table

Project	Employee
Project A	John
Project A	Sara
Project A	Ryan
Project A	Elisa



Source	Target	Link Type
John	Sara	Project A
John	Ryan	Project A
John	Elisa	Project A
Sara	Ryan	Project A
Sara	Elisa	Project A
Ryan	Elisa	Project A

Exercise (5) : Python Notebook- Projection of Node Table(15 mins)

Assistive Techniques in Shaping your Data

Refactoring Data Techniques :

- **Lasting relation:** are relationships that last over a longer period but, the link weight changes over time.

ID	Sender	Receiver	Money	Year
0	Anton	Bob	100	1801
0	Anton	Bob	30	1803
1	Anton	Charles	10	1801
1	Anton	Charles	20	1802
1	Anton	Charles	30	1803
1	Anton	Charles	100	1804

Exercise (6) : Importing your Data

- Open the Vistorian: Our new version on <http://129.215.193.74/>
- Upload your data through the wizard:
 - Choose basic info: source and target
 - Add optional info: ex. Data, locations, link type and weight
- Create any of the following visualizations: Nodelink, Matrix, Timeline, Map
- Explore:
 - one at a time as you go.
 - Play with controls offered to adjust the view of your visualization

General Questions?

Next Week

- Need your tables
- Start Exploring with Node-link Diagrams

Homework

- Check your data structure and consistency.
- Run a short exploration session with your exploration/research questions in mind.
- Please share your feedback <https://forms.office.com/r/YKiWbzP9EZ>

Additional Resources on Preparing Data

- How to infer missing node attributes? Node Classification in Social Networks <https://www.youtube.com/watch?v=yVItc5gFps8>
- Graph Representation Learning <https://www.youtube.com/watch?v=YrhBZUtG4E>

References

- Associations within school-based same-sex friendship networks of children's physical activity and sedentary behaviours: a cross-sectional social network analysis by Salway et al.
<https://link.springer.com/article/10.1186/s12966-018-0653-9>
- CollaborationViz: Interactive Visual Exploration of Biomedical Research Collaboration Networks

Tools to assist in Checking Data Consistency

Tools to assist in checking data consistency :

1. Excel
2. Google Sheets
3. Python Scripts
4. KNIME <https://www.knime.com/knime-courses>
(<https://www.youtube.com/watch?v=l-UxKNTDl2Q>)

Additional Example: on Visual Network Exploration

- 1136 participants aged 8-9 years with same-sex friends in the study
- Children wore a waist-worn accelerometer for five days.
- The average number of MVPA and sedentary minutes per day were derived for each child.
- Children were asked to name up to four of their closest friends within their school and year group ⇒ Friendship Network

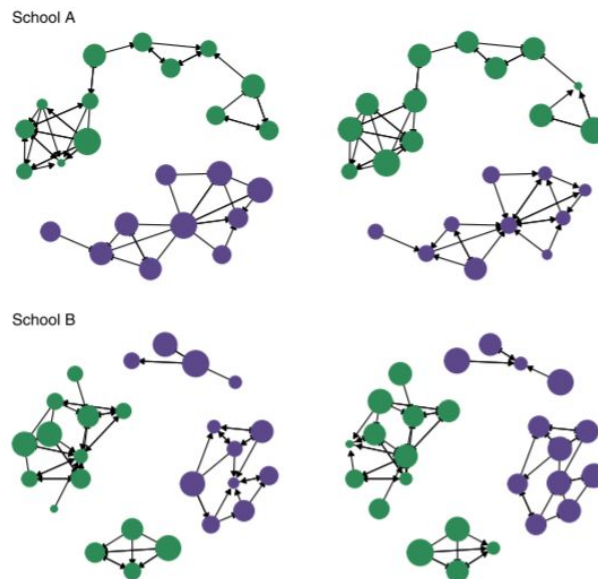


Fig. 1 Network plots of average MVPA (left) and sedentary time (right) for two typical schools (top and bottom). *Legend:* Nodes represent individual children, and are sized by average minutes of MVPA (left) and average minutes of sedentary time (right) and coloured purple for boys and green for girls. The same child is in the same position in both plots

Source: Associations within school-based same-sex friendship networks of children's physical activity and sedentary behaviours: a cross-sectional social network analysis by Salway et al.