Welcome to

The Session Will Begin Shortly

START

LLMs for Qualitative and Mixed-Methods Social Network Analysis

Session 5: Computational Practice and Ethical Responsibility

Moses Boudourides

Faculty, Graduate Program on Data Science
Northwestern University

Moses.Boudourides@gmail.com

instats Seminar

Friday, January 30, 2026
6:00 PM – 7:30 PM UTC

Northwestern | SCHOOL OF PROFESSIONAL STUDIES

Outline

A practical workflow for LLM-augmented SNA in Jupyter notebooks

Validation, documentation, and ethical responsibility in qualitative and network analysis

From Design to Practice

Previous session(s) addressed:

- *Interpretation-first networks:* ties as meanings and practices, not just edges
- *LLM capabilities and limits:* what models can assist with—and what they cannot decide
- *Research design patterns:* where LLMs fit in qualitative research, coding, and comparative analysis

In this session: we move from principles to implementation: *what does responsible practice look like in code?*

Why Computational Practice Is Not Neutral

Computational workflows:

- *Encode assumptions:* definitions of a “tie," unit of analysis, sampling choices, thresholds, and prompts
- *Stabilize interpretations:* once operationalized, choices become defaults that shape what is seen, counted, and reported
- *Create path dependence:* early decisions propagate through extraction, cleaning, modeling, and visualization

Practice is theory made operational!

The Jupyter Notebook as Research Environment

What Jupyter Notebooks Enable

- A de facto standard for computational work in Python (and increasingly beyond)
- A single place to combine text, code, figures and visualizations, results, narrative, and interpretation
- A transparent, inspectable workflow that supports replication, revision, and sharing

What Jupyter Notebooks Cannot Do

- Eliminate interpretive judgment
- Guarantee validity or correctness of conclusions

A notebook supports reasoning and documentation; it does not replace them.

A Typical LLM-Augmented Typical Workflow in a Jupyter Notebook

- Data collection
- Preprocessing
- LLM API calls for text analysis
- Validation of LLM output
- Network construction
- Network analysis and visualization.
- Interpretation and narrative write-up.

Data Collection and Preprocessing

- Data can come from APIs, web scraping, or document ingestion
- Preprocessing is crucial: cleaning text, handling missing data, and structuring data for analysis
- Document all preprocessing steps for reproducibility

Using LLM APIs in Python

Programmatic Access

Most LLM providers offer Python libraries to interact with their models programmatically

- Craft a well-defined function to send prompts and receive responses
- Handle potential errors and rate limits
- Be mindful of cost, as API calls can be expensive

Prompt Engineering for Qualitative and Network Analysis

The Prompt is Key

The quality of an LLM output is highly dependent on the quality of the prompt

- Be specific about the entities and relationships you want to extract
- Provide examples in the prompt (few-shot prompting)
- Specify the desired output format (e.g., JSON) to make parsing easier
- Iterate and refine prompts based on test results

Validation and Quality Control

Trust, but Verify

Never trust LLM output blindly–validation is non-negotiable

- **Human-in-the-loop:** Have a human coder review a sample of the LLM's output
- **Inter-rater reliability:** Calculate agreement between the LLM and human coders (e.g., using Cohen's Kappa)
- If accuracy is low, refine prompts or methods

Constructing Networks from LLM Output

- Once validated, the extracted relationships can be used to construct a network
- Python libraries like **NetworkX** are excellent for this
- Create an edge list from the LLM output and load it into a NetworkX graph object

Network Visualization and Exploration

- Visualization is a powerful tool for exploring network structure
- Use libraries like **Matplotlib**, **Pyvis**, or **Plotly** to create visualizations
- Remember that visualization is an interpretive act; the layout and aesthetics can shape the story one tells

Quantitative and Statistical Network Analysis

Once the network is constructed, one can compute a wide range of metrics:

- **Centrality measures:** Degree, betweenness, closeness, PageRank, eigenvector etc.
- **Community detection:** Identify clusters or subgroups
- **Network-level metrics:** Density, diameter, average path length, clustering coefficient, small-world-ness, scale-free-ness, robustness and resilience etc.
- **Network Statistics:** ERGM, R(H)EM, prediction, machine learming

Documentation and Reproducibility

The Core of Scientific Practice

Jupyter notebook should serve as a complete record of the entire research process

- Use Markdown cells (supporting LaTeX) to explain reasoning
- Comment code clearly
- Make data and code available and replicable so that others can reproduce your work

Data Ethics and Informed Consent

A Critical Responsibility

Using LLMs to analyze qualitative data raises significant ethical challenges.

- **Data Privacy:** If you are using a third-party API, how is your data being used and stored? *Consider using local models for sensitive data!*
- **Informed Consent:** Have participants consented to their data being analyzed by a computational tool?
  - Do data involve third parties?
  - Do they encode relationships, not just individuals?
  - *Consent is relational, not atomic!*
- **Context Collapse:** Is information created for one audience or setting gets pulled into a different one, so the original meaning might be distorted or the speaker's expectations might be violated?
  - *Public availability does not imply ethical permission!*

Bias and Fairness in LLM Analysis

Extra care is required for vulnerable populations when:

- Stakes are high
- Harm is asymmetric
- LLMs can amplify harm at scale

AI bias challenges:

- LLMs can reproduce and amplify societal biases present in their training data
- Be alert to how these biases might affect your analysis
- Test for bias and be transparent about potential limitations in the findings

Inherited bias amplification in LLMs:

- Cultural bias
- Linguistic bias
- *Scaling interpretation scales bias!*

Hallucination and Factual Accuracy

The Risk of Fabrication

- LLMs can "hallucinate," creating plausible but false information
- This is why rigorous validation is essential
- Hallucination as ethical risk can:
  - Fabricate entities
  - Invent relationships (ties) and categories
  - Misattribute motives
- This is not just a technical issue: *it is an ethical imperative!*

Responsible Use of Research Data and Findings

- Researchers must be mindful of the data's origins and the model's provenance:
  - What data (content), and whose (ownership)?
  - Which model was used?
  - With what assumptions?
- *Opacity undermines accountability!*
- Consider how their findings might be used and by whom
- Take steps to prevent their work from being used to harm or stigmatize individuals or groups
- Engage with the communities they study to ensure their research is beneficial and respectful

Transparency and Accountability

- Researchers must be transparent about their use of LLMs and maintain thorough documentation records of:
  - Prompts (and versions)
  - Validation procedures
  - Corrections and post-processing
  - Exclusions and filtering rules
  - Known limitations
- Responsible workflows should include:
  - Human review
  - Reflexive pauses
- Researchers should disclose:
  - Where LLMs intervened
  - How outputs were filtered
- Students must learn:
  - What LLMs do well
  - Where they fail

Reflexivity in Computational Research

- Researchers should ask continuously:
  - Why this workflow?
  - Who benefits?
  - Who might be harmed?
- Reflexivity is just as important in computational research as it is in traditional qualitative research
- Critically reflect on how choices of tools and methods is shaping findings
- Be aware of biases (including self-biases) and underlining assumptions

Iterative Collaborative Development and Learning

A Cycle of Discovery

- LLM-augmented research is an iterative process of analysis, validation, and refinement
- Researchers should embrace this cycle as a core part of the research process

Building Collaborative Workflows

- This type of research often requires collaboration between researchers with different skills (e.g., qualitative, computational, domain expertise)
- Use tools like Git and shared notebooks to facilitate collaboration
- Clear communication and documentation are essential

Session 5 Summary and Looking Forward

- We have covered the practical workflow of LLM-augmented qualitative research and SNA in Jupyter notebooks.
- We have emphasized the critical importance of validation, documentation, and ethical responsibility.

Next Session

In our final session, we will discuss how to transition from this seminar to your own research practice.

Questions and Discussion

Thank you!

Questions?

STOP