Welcome to

The Session Will Begin Shortly

START

LLMs for Qualitative and Mixed-Methods Social Network Analysis

Session 6: Transition to Practice

Moses Boudourides

Faculty, Graduate Program on Data Science
Northwestern University

Moses.Boudourides@gmail.com

instats Seminar

Friday, January 30, 2026
7:30 PM – 9:00 PM UTC

Northwestern | SCHOOL OF PROFESSIONAL STUDIES

Outline

From seminar concepts to practical research application

Rigorous design, ethical safeguards, and continuous learning

Augmenting qualitative inquiry

What We Have Built So Far

Across five sessions, we established:

- *Networks as interpretive constructions:* relationships, boundaries, and categories are theorized and justified, not simply “read off" the data
- *LLMs as collaborators, not analysts:* LLMs can extract, suggest, and draft—but they do not determine meaning, validity, or claims
- *Design-centered mixed-methods workflows:* sampling, codebooks, prompts, validation, and audit trails link qualitative interpretation to scalable computation

In the last (sixth) session: we are going to translate these principles into *responsible research practice* in code and documentation

Time to Go from Seminar to Research

Bridging the Gap

- The seminar has provided the theoretical and methodological foundations of LLM-assisted qualitative SNA
- Now, the challenge is to translate this knowledge into concrete research practice

This involves:

- Translating concepts into designs
- Making practical tool choices
- Navigating ethical complexities
- Developing skills for rigorous and responsible research

Assessing Research Questions

The Guiding Star

- Research questions are paramount
- They should drive every decision about design, methods, and tools

- **Key questions for qualitative network analysis** to ask oneself:
  - What, exactly, is the network phenomenon to understand (relationships-ties, roles, boundaries, mechanisms)?
  - Is the primary interest *meaning*, *structure*, or their interplay?
  - What is the scope of inquiry (single case, multiple cases, or a comparative design)?
  - What counts as evidence for a *relationship-tie* in this setting (talk, documents, observation, interaction traces)?

Evaluating Research Tools

- Moving from method to agenda, the goal is not:
  - Tool adoption for its own sake
  - Automation as an end in itself
- The goal is:
  - New questions worth asking
  - New forms of evidence—and new ways to analyze them—that make those questions answerable
- What becomes possible now that LLMs make it feasible to:
  - Analyze large narrative corpora
  - Compare relational meanings across cases in a consistent way
  - Systematically support the integration of qual + quant (network) analysis
- Examples of new kinds of network questions:
  - How meanings of relationships-ties vary across positions?
  - When structural similarity masks interpretive difference?
  - How narratives stabilize or destabilize networks?

Choosing Your Data Sources

- **Textual data:** interviews, fieldnotes, organizational documents, email/webinar/film/series transcripts, social media, archival materials, literal text, encyclopedias etc.
- **Relational data:** name generators and rosters, observations of interaction, event logs, diaries, existing network datasets etc.
- **Key considerations:** access and permissions, feasibility at scale, interpretive richness, ethical implications (privacy, consent, context collapse)
- **Fit to the question:** do your sources capture *relationships*, *meanings of ties*, or both—and with what temporal resolution and multiplexity?

In practice, combining sources (triangulation) often yields the richest and most defensible insights!

Designing Your Research

Strategic Integration

Choose a mixed-methods design (sequential, parallel, embedded, or fully integrated) that matches your research questions, data sources, and inferential goals.

- **Specify the role of LLMs:** exploration, assisted coding, structured extraction, manuscript drafting, or cross-case comparison.
- **Operationalize concepts:** define relational evidence, categories, decision rules, and output formats (codebook + prompt protocol).
- **Plan validation:** held-out subsets, human checks, disagreement analysis, and iteration (revise prompts/codebook, then re-run).
- **Integrate qual + quant systematically:** map interpretive outputs to network representations and metrics, and return to texts for explanation and mechanism.

Building Your Team

- LLM-augmented quanlitative network analysis benefits from **interdisciplinary collaboration** (substance + methods + computation).
- Ensure coverage of **qualitative expertise** (coding, manuscript drafting, interpretation), **SNA and network science** (representation, measures, inference), and **computational expertise** (pipelines, APIs, reproducibility).
- Assign **clear roles** (e.g., on codebook, prompt/pipeline, validation, ethics/privacy), responsibilities, ommunication protocols, and set regular review checkpoints.
- Agree on **shared standards** for documentation, versioning, and decision logging (prompts, revisions, exclusions, known limitations).

Selecting Tools and Platforms

- **LLMs:**
  - Commercial APIs (such as OpenAI, Anthropic, Google, Cohere, AI21, Mistral API, xAI, Azure OpenAI, AWS Bedrock).
  - Open-source models (such as Llama, Mistral/Mixtral, Qwen, Falcon, OLMo, BLOOM, Gemma, Phi, DeepSeek, Yi).
- **Network Analysis:** Python libraries like NetworkX.
- **Visualization:** Matplotlib, Pyvis, Plotly etc.
- **Reproducibility:** Jupyter notebooks, GitHub.

Choose tools that align with your technical skills and research needs.

Addressing Privacy and Ethical Concerns

Proactive Ethics

Integrate ethical considerations from the outset of your research design (not as a final compliance step).

- **Data privacy:** data minimization, anonymization/pseudonymization, secure storage, controlled access, and (when needed) local/on-prem LLM deployment.
- **Informed consent:** disclose *what* data will be processed by LLMs, *why*, *where* it is processed, and what safeguards/retention policies apply.
- **Potential for harm:** anticipate re-identification and context collapse, protect vulnerable groups, and avoid releasing traceable excerpts or sensitive network relationships.

Starting Small and Iterating

Pilot and Refine

- Start with a small, diverse sample to develop and stress-test the workflow.
- Iterate on prompts, validation procedures, and analytic choices before scaling up.

Iteration helps you:

- **Find failure modes early:** hallucinations, category drift, prompt noncompliance, and brittle extraction.
- **Learn the corpus:** local vocabularies, cases, attributes, and categories, and the contextual cues that carry relational meaning.
- **Tighten the design:** refine the research question, scope conditions, and what counts as evidence for a relationship.

Developing Your Coding Scheme

- **Ground in data:** build initial codes from close reading of a purposive sample (typical and deviant cases).
- **Make it explicit:** write a notebook with definitions, inclusion/exclusion rules, and short anchor examples.
- **Keep it theory-aware:** connect codes to concepts and mechanisms while preserving participants' categories and meanings.
- **Pilot before scaling:** test on new material, resolve ambiguities, and version the scheme before applying it via LLMs to the full dataset.

Validating LLM Output

Rigorous Verification

Treat LLM-generated outputs as provisional and validate them systematically.

- **Human review:** double-code a representative (or stratified) subset and compare to model outputs.
- **Reliability + error analysis:** assess agreement and inspect where/why the model fails (false relationships-ties, wrong direction/type, missed context).
- **Document and report:** prompts, revisions, exclusions, and accuracy metrics appropriate to the task (and their limits).

Integrating Quantitative and Qualitative Findings

Synthesizing Insights

Use each method to discipline the other: patterns guide interpretation, and interpretation explains patterns.

- Use quantitative patterns (centralities, clusters, brokerage, mixing) to **target qualitative deep dives** and "critical cases."
- Use qualitative evidence (semantics, narratives, signification) to **interpret mechanisms** behind metrics and to refine their meaning and modes of operation.
- Move iteratively between **graphs and texts**: explain outliers, revisit all relational definitions, and update claims accordingly.
- Aim for an integrated account where **structure, meaning, and process** cohere.

Writing and Presenting Your Research

- Be transparent about methods, including the *specific role of LLMs* (extraction, coding support, manuscript drafting, comparison).
- Report *design choices and validation*: sampling, codebook, prompts, human checks, disagreement/error analysis.
- Present findings with *paired evidence*: network patterns *and* illustrative qualitative excerpts (with ethical care).
- Acknowledge *limitations*: construct/interpretive validity, model sensitivity, missingness, boundary choices, and generalizability.
- Share *artifacts*: Jupyter notebooks, prompt protocols, codebook versions, and reproducible pipelines (as permitted).

Engaging with Research Participants

Member Validation

When appropriate, share interpretations with participants and invite feedback to strengthen interpretive validity and ethical accountability.

- Do interpretations *resonate* with participants' lived experience and local categories?
- Are important *nuances, exceptions, or power dynamics* missing from the analytic story?
- Use feedback to identify *misread relationships*, misplaced emphases, or context that the corpus does not capture.

Managing Bias and Ensuring Fairness

- Treat bias as *multi-source*: data collection, coding categories, prompts, model behavior, and analytic interpretation.
- Run *checks for differential error*: who is misclassified, omitted, or over-attributed, and under what conditions?
- Use *counterfactual prompts* and alternative framings to probe sensitivity (and avoid single-prompt lock-in).
- Document mitigation steps and state *residual risks and limits* explicitly.

Contributing to Knowledge and Practice

Beyond the Data

Ask what your study changes in how we understand relational patterns (networks etc.)—and what it enables others to do.

- **Theory:** what new concepts or mechanisms about relationships, meanings, and dynamics does the study propose or refine?
- **Method:** what reusable workflows, codebooks, prompt protocols, or validation strategies does it contribute?
- **Practice:** what actionable insights follow for organizations/communities, and what are the ethical trade-offs?
- **Equity:** how might the work reduce harm, widen participation, or challenge exclusionary relational structures (networks etc.)?

Continuing to Learn and Develop

- LLM-augmented social science is evolving rapidly in *capabilities and norms*.
- Track changes in *models and tooling* (APIs, open models, privacy options, evaluation methods).
- Keep up with *methods*: mixed-methods designs, qualitative network analysis, and computational validation practices.
- Participate in community venues (workshops, conferences, preprints) and maintain a *living workflow* you can revise and develop.

Toward Responsible and Rigorous Research

The Path Forward

LLMs can expand what is feasible in qualitative and mixed-methods SNA, but only under disciplined, transparent, and ethical practice.

- Keep *human control* over meaning, categories, and claims.
- Validate outputs with *systematic checks* and explicit error analysis.
- Be *transparent and reflexive*: prompts, revisions, exclusions, limitations, and ethical safeguards.
- Aim for *nuanced accounts* of complex social realities: structure, meaning, and process together.

Session 6 Summary and Closing

- We have guided the transition from seminar concepts to practical research application.
- Emphasized the importance of careful design, ethical considerations, and continuous learning.
- The goal is to augment, not replace, human qualitative inquiry.

Questions and Discussion

Thank you!

Questions?

STOP