Finding Papers for Discussion Presentations

DSCC/LING 251/451 - Spring 2026

This is a brief supplement to the discussion presentation guidelines on the course website. See that page for format expectations, grading criteria, and signup details.

Your Goal

Find a research paper that:

1. **Relates to the week's topic** (e.g., self-supervised learning, active learning)
2. **Connects to your interests** - ideally in a domain you care about or want to learn more about
3. **Is accessible** - you should be able to understand the core ideas with effort

Timeline: Papers must be approved at least one week before your presentation and shared with the class by the Friday before.

Where to Find Papers

1. Semantic Scholar (Recommended starting point)

URL: semanticscholar.org

Why it's useful:

- Clean, readable interface focused on understanding research
- Shows "highly influential citations" to find seminal papers
- Provides citation graphs showing how papers relate
- Displays paper summaries and key figures
- Lists "Papers Citing This" and "References" for exploration
- Free and no account needed

How to use it:

1. Search for the topic (e.g., "active learning natural language processing")
2. Filter by year (usually want papers from last 5-10 years, though classics are fine too)
3. Sort by "Most Influential" or "Highly Cited"
4. Click on promising papers to see abstracts and citation graphs
5. Follow citation trails: look at what influential papers cite, and what cites them

Example search flow:

```
Search: "semi-supervised learning medical imaging"
→ Find a highly-cited survey/benchmark paper
→ Check "Papers Citing This" to find recent work
→ Look at the citation graph to identify key methods
```

2. Google Scholar

URL: scholar.google.com

Strengths:

- Comprehensive coverage (finds everything)
- Shows citation counts (useful proxy for impact)
- "Cited by" links help you find follow-up work
- Can set up alerts for new papers on topics

Limitations:

- Interface is less readable than Semantic Scholar
- Harder to browse and explore connections
- Can be overwhelming with results

Best used for:

- Finding papers when you have a specific author or title in mind
- Checking how influential a paper is (citation count)
- Following citation chains from a known starting point

3. ArXiv

URL: arxiv.org

What it is: Pre-print server where researchers post papers before or during peer review

Strengths:

- Access to very recent work (sometimes months before publication)
- All papers are freely available
- Many ML researchers post here first

Limitations:

- Papers aren't peer-reviewed yet (quality varies)
- Can be harder to assess importance
- Overwhelming volume (hundreds of ML papers per week)

Best approach:

- Use Semantic Scholar or Google Scholar to find papers first, then check if they're on ArXiv for free PDF access
- Browse specific categories like `cs.LG` (Machine Learning) or `cs.CL` (Computation and Language)

4. Venue Proceedings

Major ML conferences and journals where you can browse accepted papers:

Conferences:

- NeurIPS, ICML, ICLR (general ML)
- ACL, EMNLP, NAACL (NLP)
- CVPR, ICCV, ECCV (computer vision)
- AAAI, IJCAI (AI)

Journals:

- JMLR (Journal of Machine Learning Research)
- PAMI (Pattern Analysis and Machine Intelligence)
- Transactions on NLP, Speech, etc.

Why browse venues:

- Quality control (peer-reviewed and accepted)
- See what's trending in the field
- Find papers in your subfield's main venues

How to access:

- Most conferences have open-access proceedings online
- Search "[conference name] 2024 proceedings" or check the conference website
- Papers With Code (see below) aggregates papers by venue

5. Papers With Code

URL: paperswithcode.com

What it is: Database of papers with implementations and benchmarks

Strengths:

- Shows which papers have code available
- Organized by task (e.g., "image classification," "named entity recognition")
- Leaderboards show current state-of-the-art
- Great for finding papers with reproducible experiments

Best for: Finding papers on specific tasks where you want to see implementations

---

Search Strategies

Strategy 1: Start Broad, Then Narrow

1. Search the topic (e.g., "few-shot learning")
2. Find a recent survey paper or highly-cited method
3. Use its references and citations to zoom in on specific subtopics
4. Pick something at the intersection of the topic and your interests

Strategy 2: Domain + Technique

Search: `[ML technique] + [your domain]`

- "transfer learning robotics"
- "active learning medical diagnosis"
- "self-supervised learning audio"
- "semi-supervised time series"

Strategy 3: Follow Citation Chains

1. Start with a paper from lecture or a textbook
2. Look at what it cites (older foundational work)
3. Look at what cites it (newer applications and extensions)
4. Find a paper at the "right level" for discussion

Strategy 4: Browse Recent Conference Proceedings

1. Go to a recent conference (e.g., NeurIPS 2025, ICML 2024)
2. Search or filter by topic keywords
3. Read abstracts until something catches your interest
4. Use Semantic Scholar to check influence and connections

What Makes a Good Discussion Paper?

Good characteristics:

- **Clear contribution**: Has a specific method, result, or insight to discuss
- **Right scope**: Not so broad you can't cover it, not so narrow it's hard to discuss
- **Recent enough**: Usually last 5-10 years (unless it's a classic foundational paper)
- **Accessible**: Core ideas are understandable even if some technical details are complex
- **Interesting**: Connects to something you or your classmates care about

Watch out for:

- Papers that are just incremental improvements (hard to discuss)
- Very theoretical papers if you want to focus on applications (or vice versa)
- Papers that require deep domain knowledge to understand
- Papers with limited or questionable experiments

Approval Process

Remember: Papers must be approved one week before your presentation.

How to get approval:

1. Find 1-3 candidate papers
2. Email me with:
    - Paper title and link
    - 1-2 sentences on why you chose it
    - Any questions about whether it's a good fit

3. I'll confirm or suggest alternatives
4. Once approved, share the PDF with the class by Friday

Pro tip: Give yourself time! If your first choice doesn't work out, you'll want time to find an alternative.

Preparing Your Presentation

Visual materials are required. You should prepare slides or a handout for the class to follow. This helps everyone stay engaged and makes your presentation more effective.

What to include:

- **Key points from the paper**: Problem, method, main results
- **Figures or diagrams**: Pull important visuals from the paper (with proper attribution)
- **Connections to course concepts**: How does this relate to what we've learned?
- **Your critical analysis**: Strengths, limitations, open questions
- **Discussion questions**: The 2-3 questions you want the class to engage with

Format options:

- **Slides** (PDF, PowerPoint, Google Slides, Beamer, etc.) - most common
- **Handout** (1-2 pages summarizing the paper and discussion points)
- **Combination** (brief slides + detailed handout for reference)

Tips:

- Don't try to fit everything on slides - focus on what helps understanding
- Include page/figure numbers so you can reference the original paper
- Have your discussion questions visible so the class can think about them
- Test your slides beforehand (font size, readability, any videos/animations)

Example: Finding a Paper on Self-Supervised Learning

Let's walk through finding a paper for the self-supervised learning week:

1. **Go to Semantic Scholar**, search: "self-supervised learning"
2. **Filter for recent years** (e.g., 2020-2025) to see modern approaches
3. **Sort by "Most Influential"** to find important papers
4. **Browse abstracts**:
    - "SimCLR: A Simple Framework for Contrastive Learning" - computer vision
    - "BERT: Pre-training of Deep Bidirectional Transformers" - NLP
    - "Momentum Contrast (MoCo)" - another vision approach
5. **Pick based on your interest**:
    - If you care about NLP → explore BERT and related papers
    - If you care about vision → explore SimCLR or MoCo

    - If you care about audio/video → check "Papers Citing This" from a vision paper to find multimodal extensions

6. **Follow citations**: Click on a promising paper, see what it cites and what cites it
7. **Check the PDF**: Make sure you can access it and understand enough to present
8. **Submit for approval**: Email with your choice and reasoning

---

Tips for Exploration

- **Don't commit to the first paper you find** - browse 5-10 abstracts before deciding
- **Follow your curiosity** - if a paper mentions something interesting in passing, search for that
- **Check multiple sources** - cross-reference on Semantic Scholar, Google Scholar, and ArXiv
- **Read the abstract carefully** - if it's confusing or boring now, it might be hard to present
- **Look at figures** - papers with clear figures/diagrams are often easier to present
- **Consider your audience** - pick papers that will be interesting to discuss, not just to read

---

Questions?

If you're having trouble finding papers, come to office hours or post on the discussion board. I'm happy to help you narrow down options or suggest search directions!