Cognitive MRI of ChatGPT Conversations

Revealing the Hidden Structure of AI Conversations

Alex Towell
Southern Illinois University-Edwardsville
atowell@siue.edu
lex@metafunctor.com

April 29, 2025

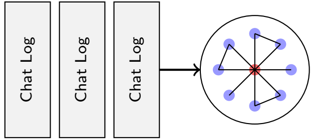
Code: github.com/queelius/chatgpt-complex-net

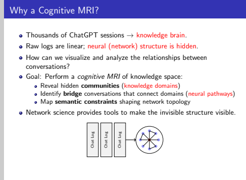
Cognitive MRI of ChatGPT Conversations
Revealing the Hidden Structure of AI Conversations
Alex Towell
Southern Illinois University-Edwardsville
atowell@siue.edu
lex@metafunctor.com
April 29, 2025
Code: github.com/queelius/chatgpt-complex-net

- Introduce: Research on analyzing ChatGPT conversations with network methods
- Key concept: "Cognitive MRI" - revealing hidden structures in conversation data
- This combines network science with modern language models
- (Transition) Let me explain why this matters...

Why a Cognitive MRI?

- Thousands of ChatGPT sessions → knowledge brain.
- Raw logs are linear; neural (network) structure is hidden.
- How can we visualize and analyze the relationships between conversations?
- Goal: Perform a *cognitive MRI* of knowledge space:
  - Reveal hidden **communities** (knowledge domains)
  - Identify **bridge** conversations that connect domains (neural pathways)
  - Map **semantic constraints** shaping network topology
- Network science provides tools to make the invisible structure visible.





- Personal context: Generated 1000+ ChatGPT conversations over 18 months
- Problem: Linear archive with no organization or visible connections
- Goal: Transform flat data into structured knowledge map
- MRI analogy: Just as medical imaging reveals hidden structures in tissue
- Looking for: communities (domains), bridges (pathways), constraints
- (Point to diagram) From linear logs to organized network
- (Transition) First step: convert text to vectors...

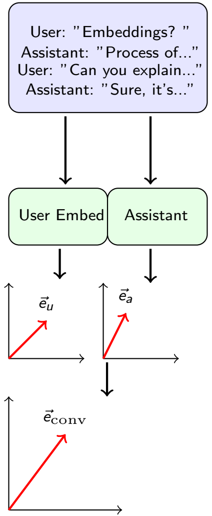
Representing Conversations as Vectors

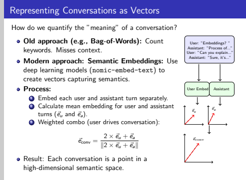
How do we quantify the "meaning" of a conversation?

- **Old approach (e.g., Bag-of-Words):** Count keywords. Misses context.
- **Modern approach: Semantic Embeddings:** Use deep learning models (`nomic-embed-text`) to create vectors capturing semantics.
- **Process:**
  1. Embed each user and assistant turn separately.
  2. Calculate mean embedding for user and assistant turns ($\vec{e}_u$ and $\vec{e}_a$).
  3. Weighted combo (user drives conversation):

$$\vec{e}_{\text{conv}} = \frac{2 \times \vec{e}_u + \vec{e}_a}{\|2 \times \vec{e}_u + \vec{e}_a\|}$$

- Result: Each conversation is a point in a high-dimensional semantic space.





- Challenge: Making "meaning" mathematically analyzable
- Traditional methods (bag-of-words): Just keyword counting, loses context
- Modern approach: Neural embeddings capture semantic meaning
- Process: (point to diagram)
  - Split by speaker role
  - User turns weighted 2x (they drive conversation direction)
  - Result: Each conversation = point in high-dim semantic space
- (Transition) Now we build connections between these points...

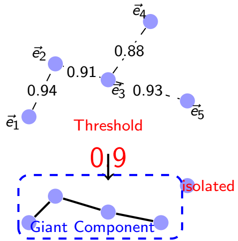
Building the Semantic Network

From vectors to a network:

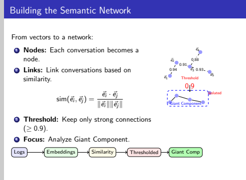
1. **Nodes:** Each conversation becomes a node.
2. **Links:** Link conversations based on similarity.

$$\text{sim}(\vec{e}_i, \vec{e}_j) = \frac{\vec{e}_i \cdot \vec{e}_j}{\|\vec{e}_i\| \|\vec{e}_j\|}$$

3. **Threshold:** Keep only strong connections (≥ 0.9).
4. **Focus:** Analyze Giant Component.



Logs → Embeddings → Similarity → Thresholded → Giant Comp



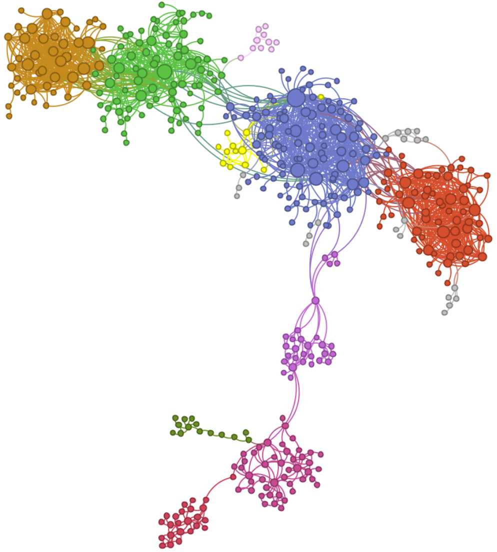
- Network construction process: [point to each step]
  1. Nodes = conversations
  2. Links = similarity above threshold (cosine similarity)
  3. High threshold (0.9) keeps only strong connections
  4. Focus on giant component for analysis
- (Point to visualization) Shows filtering process
- Similar to contrast enhancement in medical imaging
- 449 conversations in giant component (about half of dataset)
- (Transition) Let's see what this network looks like...

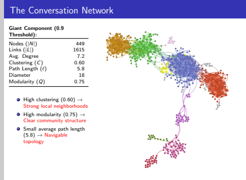
The Conversation Network

Giant Component (0.9 Threshold):

Nodes ($\|N\|$)	449
Links ($\|L\|$)	1615
Avg. Degree	7.2
Clustering (C)	0.60
Path Length (ℓ)	5.8
Diameter	18
Modularity (Q)	0.75

- High clustering (0.60) → Strong local neighborhoods
- High modularity (0.75) → Clear community structure
- Small average path length (5.8) → Navigable topology





- (Point to visualization) Clear clustering visible in different colors
- Key metrics tell us about structure: (point to metrics)
  - High clustering (0.60): Strong local neighborhoods
  - High modularity (0.75): Clear community boundaries
  - Relatively short paths (5.8 steps): Knowledge is navigable
- Notice varying density across regions - different knowledge organization
- (Transition) Let's identify these communities...

Community Detection:

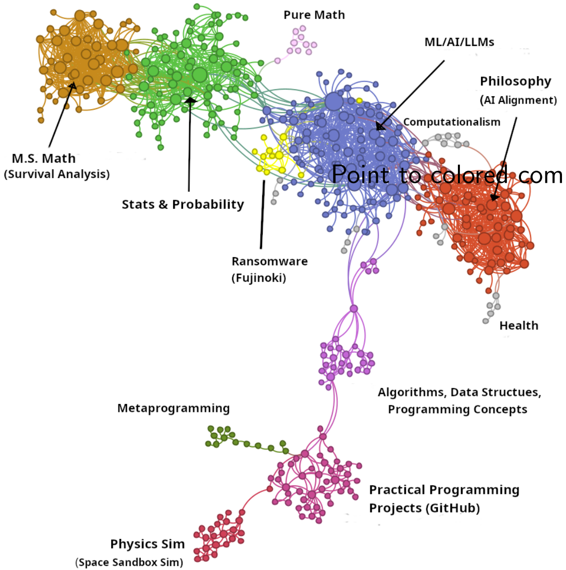
- ~ 15 communities
- (High) modularity score: 0.75

Major Knowledge Domains:

1. ML/AI/LLM (23%)
2. Statistics & Probability (18%)
3. M.S. Math/Analysis (14%)
4. Philosophy & AI Ethics (10%)
5. Programming Projects (10%)
6. Programming/Algorithms (7%)

Distinct structural patterns across communities:

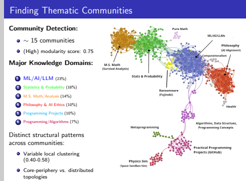
- Variable local clustering (0.40-0.58)
- Core-periphery vs. distributed topologies



2025-04-29

ChatGPT Conversation Graph

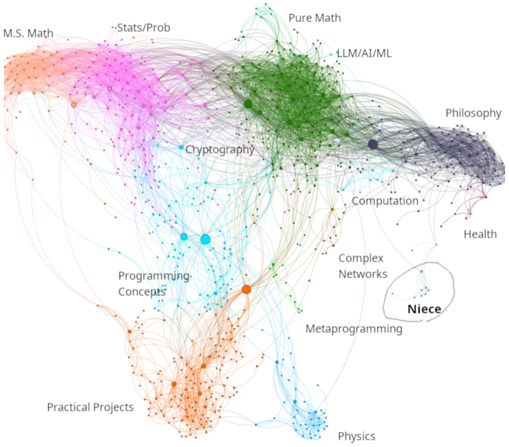
└Finding Thematic Communities



- Used Louvain algorithm - found 15 communities
  - Point to colored communities list
  - Blue: ML/AI/LLM (23%) - largest domain
  - Green: Statistics (18%)
  - Orange: Advanced math (14%)
  - Others: Philosophy, programming, etc.
- Different internal structures:
  - AI/ML has hub-spoke pattern
  - Programming projects more distributed
- (Point to visualization) Colors match communities list
- (Transition) How robust are these findings?

Threshold Sensitivity: 0.9 vs 0.875

How robust is the structure to the similarity threshold?



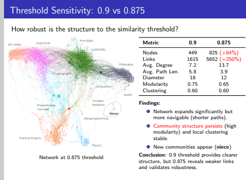
Network at 0.875 threshold

Metric	0.9	0.875
Nodes	449	825 (+84%)
Links	1615	5652 (+250%)
Avg. Degree	7.2	13.7
Avg. Path Len.	5.8	3.9
Diameter	18	12
Modularity	0.75	0.65
Clustering	0.60	0.60

Findings:

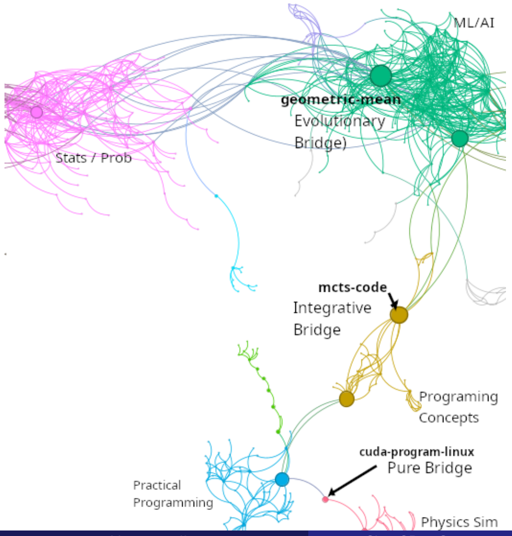
- Network expands significantly but more navigable (shorter paths).
- Community structure persists (high modularity) and local clustering stable.
- New communities appear (**niece**)

Conclusion: 0.9 threshold provides clearer structure, but 0.875 reveals weaker links and validates robustness.



- Testing robustness: Slightly lowered threshold (0.9 -¿ 0.875)
- (Point to numbers) Dramatic expansion: +84% nodes, +250% links
- (Point to key findings)
  - Community structure persists despite expansion
  - Clustering coefficient stays exactly 0.60
  - New distinct communities appear (e.g., niece conversations)
- Confirms findings aren't artifacts of arbitrary threshold
- (Transition) How do these knowledge domains connect?

How do different knowledge areas connect? → Betweenness Centrality (C_B)

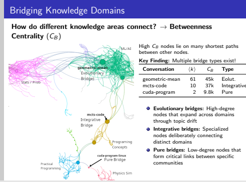


High C_B nodes lie on many shortest paths between other nodes.

Key Finding: Multiple bridge types exist!

Conversation	$\langle k \rangle$	C_B	**Type**
geometric-mean	61	45k	Eolut.
mcts-code	10	37k	Integrative
cuda-program	2	9.8k	Pure

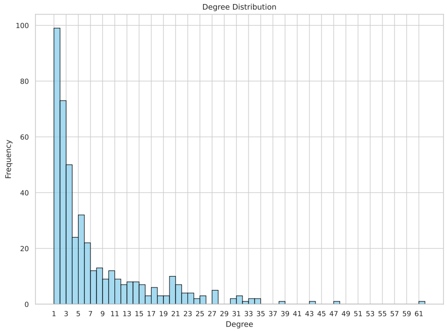
- **Evolutionary bridges:** High-degree nodes that expand across domains through topic drift
- **Integrative bridges:** Specialized nodes deliberately connecting distinct domains
- **Pure bridges:** Low-degree nodes that form critical links between specific communities



- Betweenness centrality reveals bridges between communities
- (Point to visualization) Zoomed view of key bridges
- Discovered three bridge types: (point to table)
  - Evolutionary bridges (geometric-mean): High degree & betweenness
  - Integrative bridges (mcts-code): Deliberate synthesis
  - Pure bridges (cuda-program): Minimal connections but critical links
- Different mechanisms for knowledge integration
- (Transition) Let's look at connectivity patterns...

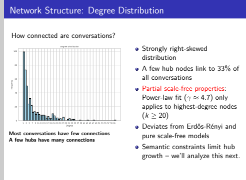
Network Structure: Degree Distribution

How connected are conversations?



Most conversations have few connections
A few hubs have many connections

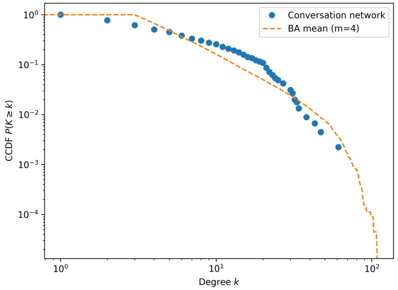
- Strongly right-skewed distribution
- A few hub nodes link to 33% of all conversations
- Partial scale-free properties: Power-law fit ($\gamma \approx 4.7$) only applies to highest-degree nodes ($k \geq 20$)
- Deviates from Erdős-Rényi and pure scale-free models
- Semantic constraints limit hub growth – we'll analyze this next.

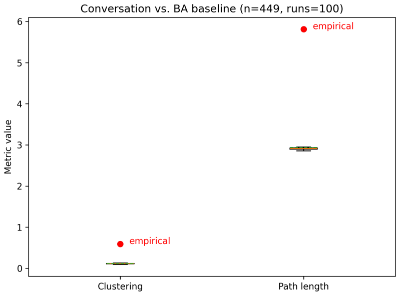


- (Point to histogram) Heavily skewed distribution
- Most conversations have few connections
- Few hubs connect to many others (one links to 14% of network)
- Not purely scale-free: Power law only in tail (k¿=20)
- Much steeper falloff than typical scale-free (gamma 4.7 vs 2-3)
- Suggests semantic constraints limit hub growth
- (Transition) Is this just preferential attachment?

Beyond Simple Network Growth Models

Is this just preferential attachment (rich get richer)?





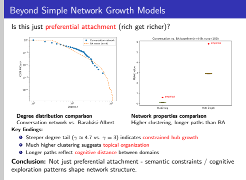
Degree distribution comparison
Conversation network vs. Barabási-Albert

Network properties comparison
Higher clustering, longer paths than BA

Key findings:

- Steeper degree tail ($\gamma \approx 4.7$ vs. $\gamma = 3$) indicates constrained hub growth
- Much higher clustering suggests topical organization
- Longer paths reflect cognitive distance between domains

Conclusion: Not just preferential attachment - semantic constraints / cognitive exploration patterns shape network structure.



- Compared with Barabási-Albert preferential attachment model
- (Point to left graph) Much steeper degree falloff
- (Point to right graph) 6x higher clustering than BA model
- Three key differences:
  - Constrained hub growth (semantic limitations)
  - Strong topical organization (high clustering)
  - Meaningful distances between domains (longer paths)
- Not just "rich get richer" - semantic constraints matter
- (Transition) Let's wrap up what we've learned...

Application: Network-Aware Recommendations

If time permits - otherwise skip to conclusions

Score past conversations by balancing:

$$\mathrm{score}(\mathrm{c_i}) = \alpha \times \underbrace{\mathrm{sim}(\tilde{\mathrm{e}}_{\mathrm{cur}}, \tilde{\mathrm{e}}_{\mathrm{i}})}_{\text{Familiarity}} + \beta \times \underbrace{\widehat{C_B}(c_i)}_{\text{Bridging}}$$

- $\alpha \gg \beta$: Favor similar conversations
- $\alpha \approx \beta$: Balance familiar with bridging
- $\alpha \ll \beta$: Prioritize knowledge-connecting bridges

Example ($\alpha = 2, \beta = 1$):

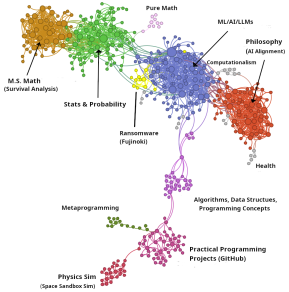
- High-sim conversations dominate top slots
- Bridge nodes (e.g., "geometric-mean") still appear mid-list
- Single parameter to control exploration vs. exploitation

- Practical application: Recommendation system using network structure
- Formula balances: similarity vs. bridging potential
- Can tune parameters for exploration vs. exploitation
- (Skip if running behind schedule)

Conclusions

Key Findings:

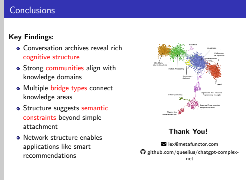
- Conversation archives reveal rich cognitive structure
- Strong communities align with knowledge domains
- Multiple bridge types connect knowledge areas
- Structure suggests semantic constraints beyond simple attachment
- Network structure enables applications like smart recommendations



Thank You!

lex@metafunctor.com

github.com/queelius/chatgpt-complex-net



- Key contributions:
  - Network approach reveals cognitive structure in conversations
  - Communities align with knowledge domains
  - Multiple bridge types connect domains (evolutionary, integrative, pure)
  - Evidence of semantic constraints beyond simple growth models
- "Cognitive MRI" makes visible what was hidden in linear logs
- Code available on GitHub for reproducing analysis
- Thank you! Questions?