

# Temporal evolution of cognitive knowledge networks in AI-assisted conversations

Alexander Towell<sup>1\*</sup>, John Matta<sup>1</sup>

**1** Department of Computer Science, Southern Illinois University Edwardsville, Edwardsville, IL, USA

\* atowell@sieue.edu

## Abstract

How does a person's knowledge landscape evolve through sustained AI-assisted conversation? We previously introduced a *cognitive MRI* methodology that transforms linear conversation logs into semantic similarity networks, revealing knowledge communities and bridge conversations in a static snapshot. Here we extend this analysis to the temporal domain, tracking how the network grows over 29 months (December 2022 – April 2025) across 1,908 ChatGPT conversations. We construct cumulative monthly snapshots and discover several evolution patterns characteristic of real-world complex networks. The network exhibits *super-linear densification* ( $\gamma = 1.405$ ,  $R^2 = 0.993$ ), meaning knowledge exploration accelerates as the network grows. New conversations attach to existing topics via *sub-linear preferential attachment* ( $\beta = 0.763$ ,  $R^2 = 0.914$ ), indicating that popular topics attract disproportionate attention but less aggressively than in scale-free models. Community structure stabilizes early (modularity exceeding 0.6 by late 2023 and reaching  $\approx 0.75$  by the final snapshot) and persists through subsequent growth, with 40 distinct communities tracked through birth, continuation, and death events. Bridge conversations, critical cross-domain connectors, emerge at identifiable moments and maintain their structural role once established. Comparison across model eras (GPT-3.5 through GPT-4.5) reveals that AI model capabilities shape sub-network topology. These findings suggest that individual AI-assisted knowledge exploration self-organizes according to the same macroscopic laws

observed in collective knowledge systems such as citation and collaboration networks, but arising through a different mechanism: the progressive revelation of latent semantic structure rather than active social tie formation. This scale-bridging parallel offers both theoretical insight into distributed cognition and practical implications for knowledge management systems.

## Author summary

Every day, millions of people use AI chatbots to explore ideas, learn new topics, and solve problems. Over months of use, these conversations accumulate into a personal archive, but what structure, if any, lies hidden in that archive? We developed a method to find out. By connecting conversations that discuss similar topics into a network and watching how that network grows over time, we can see patterns that are invisible in a flat chat history. Surprisingly, the growth patterns we found in one person's 29-month ChatGPT archive mirror those seen in much larger knowledge systems like scientific citation networks, suggesting that individual curiosity may follow the same organizational principles as collective knowledge production. Our approach could help design smarter conversation archives that surface hidden connections between past and present inquiries, turning a passive chat log into an active tool for knowledge discovery.

## Introduction

The rise of conversational AI has created a new medium for knowledge exploration. Millions of users now interact daily with large language models (LLMs), generating conversation archives that constitute externalized records of cognitive activity [1]. These archives are typically presented as flat, chronological lists, a format that conceals the rich associative structure latent in accumulated inquiry.

In prior work, we introduced a *cognitive MRI* methodology that transforms such archives into semantic similarity networks, where conversations become nodes and edges connect semantically related pairs [2]. Analyzing the giant component of a single user's conversation archive (449 of 601 connected conversations at similarity threshold  $\theta = 0.9$ ), we identified 15 knowledge communities, heterogeneous network topology

(hub-and-spoke vs. tree-like structures), and three distinct bridge conversation types  
12 that facilitate cross-domain knowledge transfer. That analysis, however, treated the  
13 network as a static snapshot, collapsing 29 months of knowledge exploration into a  
14 single graph.  
15

This limitation is significant. The *temporal dimension* of knowledge exploration is  
16 precisely what distinguishes organic human inquiry from static knowledge bases.  
17

Questions build on previous questions; interests drift, deepen, and occasionally collide;  
18 new AI model capabilities reshape what is possible to explore. A static analysis cannot  
19 capture these dynamics.  
20

In this paper, we extend the cognitive MRI to the temporal domain, analyzing how  
21 the semantic similarity network evolves over 29 monthly snapshots spanning December  
22 2022 through April 2025. This temporal extension constitutes the core novel  
23 contribution beyond our previous work, addressing three new research questions:  
24

- Does the network densify over time, and if so, does it follow established  
25 densification laws?  
26
- How do new conversations attach to the existing network, whether randomly,  
27 preferentially, or through some intermediate mechanism?  
28
- Do knowledge communities emerge early and persist, or does the community  
29 structure remain in flux?  
30

Our analysis reveals that this single-user knowledge network exhibits evolution  
31 patterns remarkably consistent with those observed in much larger social, biological,  
32 and technological networks [3–5]. Specifically, we find:  
33

- Super-linear densification ( $\gamma = 1.405$ ,  $R^2 = 0.993$ ), where edges grow faster than  
34 nodes, meaning the network becomes proportionally denser over time.  
35
- Sub-linear preferential attachment ( $\beta = 0.763$ ,  $R^2 = 0.914$ ), where high-degree  
36 nodes attract new connections disproportionately, but less aggressively than in the  
37 Barabási-Albert model.  
38
- Early community stabilization, with modularity exceeding 0.6 by late 2023 and  
39 stabilizing near 0.75, persisting through continued growth despite the network  
40 tripling in size.  
41

- Bridge persistence, where once a conversation achieves bridge status (top-5% betweenness centrality), it maintains that role throughout the observation period. 42
- Model era effects, with sub-networks from different LLM eras exhibiting distinct 43 topological signatures. 44

These findings contribute to temporal network theory by providing a new empirical 46 case study (the evolution of a personal knowledge exploration network) and to the 47 emerging field of human-AI interaction by demonstrating that conversational AI usage 48 leaves structured, analyzable traces of cognitive activity. Our framing draws on the 49 distributed cognition tradition [6, 7], viewing the human-AI conversation archive as an 50 externalized cognitive system whose structure reveals patterns of knowledge 51 organization and exploration. 52

## Related work

### Temporal network evolution

The study of how networks evolve over time has revealed universal patterns across 55 diverse domains. Holme and Saramäki [8] provide a comprehensive review of temporal 56 network analysis, distinguishing between *contact sequences* (discrete events) and 57 *interval graphs* (persistent connections). Our cumulative snapshot approach falls into 58 the latter category: once a conversation enters the network, it persists permanently, 59 making the network monotonically non-decreasing. This models the reality that 60 knowledge, once explored, remains part of one’s cognitive landscape. 61

Dorogovtsev and Mendes [9] survey evolution models for growing networks, 62 identifying preferential attachment, fitness-based growth, and aging as key mechanisms. 63 Albert and Barabási [10] provide a broader statistical mechanics perspective on complex 64 network evolution, including the emergence of scale-free properties and small-world 65 characteristics during growth. 66

### Densification laws

Leskovec et al. [11] discovered that many real-world networks exhibit *densification*: the 68 number of edges grows super-linearly with the number of nodes, following a power law 69

$e(t) \propto n(t)^\gamma$  with  $\gamma > 1$ . They documented this pattern across citation networks  
70 (with  $\gamma = 1.69$ ), patent networks ( $\gamma = 1.26$ ), and autonomous systems graphs ( $\gamma = 1.18$ ),  
71 among others [3]. This contrasts with constant-density growth (Erdős-Rényi) or  
72 constant-degree growth (Barabási-Albert, where  $\gamma = 1$ ). Densification implies that as a  
73 network grows, participants increasingly find connections to existing content rather than  
74 remaining isolated. The exponent  $\gamma$  characterizes how aggressively this acceleration  
75 occurs.  
76

## Preferential attachment

77

Barabási and Albert [4] proposed preferential attachment as a mechanism for generating  
78 scale-free networks: new nodes connect preferentially to high-degree existing nodes with  
79 probability  $\Pi(k) \propto k^\beta$ . When  $\beta = 1$ , this produces power-law degree distributions.  
80 Jeong et al. [12] developed methods to measure preferential attachment empirically,  
81 finding linear ( $\beta \approx 1$ ) attachment in citation and collaboration networks. Subsequent  
82 work has shown that many real networks exhibit *sub-linear* attachment ( $\beta < 1$ ), where  
83 high-degree nodes attract connections but not as strongly as the pure model  
84 predicts [12]. Sub-linear attachment produces networks with more moderate degree  
85 heterogeneity than pure scale-free networks.  
86

## Community dynamics

87

Palla et al. [5] pioneered the study of community evolution in temporal networks,  
88 tracking overlapping communities in mobile phone and collaboration networks. They  
89 identified key lifecycle events: birth, growth, contraction, merging, splitting, and death.  
90 Greene et al. [13] proposed event-based frameworks for tracking non-overlapping  
91 communities across snapshots using set overlap measures. Mucha et al. [14] introduced  
92 multiscale community detection for time-dependent networks using generalized  
93 modularity optimization. Rossetti and Cazabet [15] survey the broader field of dynamic  
94 community discovery, cataloging approaches from incremental methods to tensor  
95 decompositions.  
96

A key finding across this literature is that community structure in growing networks  
97 tends to *stabilize*: after an initial transient period, the mesoscale organization persists  
98

even as the microscale (individual nodes and edges) continues to change [5].

99

## Knowledge and citation network evolution

100

Citation networks provide a natural comparison for knowledge exploration networks. de  
Solla Price [16] first studied their growth dynamics, and subsequent work has  
characterized their densification [3], preferential attachment [12], and community  
structure [17]. Shi et al. [18] model the evolution of scientific knowledge as a dynamic  
network, finding that the structure of science exhibits path-dependent growth with both  
conservative (within-field) and innovative (cross-field) exploration patterns. Our work  
extends this paradigm from collective scientific knowledge to individual knowledge  
exploration through AI conversation.

101

102

103

104

105

106

107

108

## Distributed cognition and the extended mind

109

Hutchins [6] introduced the framework of *distributed cognition*, arguing that cognitive  
processes are not confined to the individual brain but are distributed across people,  
artifacts, and the environment. His analysis of ship navigation demonstrated how a  
complex cognitive task is accomplished through the coordinated interaction of multiple  
agents and tools rather than any single mind. Clark and Chalmers [7] extended this  
reasoning with the *extended mind thesis*, arguing that when external resources play the  
right functional role in cognitive processing, they constitute genuine parts of the  
cognitive system. A notebook that reliably stores and retrieves beliefs, for instance,  
functions as an extension of memory. These frameworks are directly relevant to our  
analysis: the AI conversation archive functions as an external cognitive artifact whose  
structure reveals patterns of knowledge organization. We treat the archive not merely as  
a record of past inquiries but as an externalized cognitive system whose network  
structure is amenable to the same analytical tools used for other complex networks.

110

111

112

113

114

115

116

117

118

119

120

121

122

## AI conversation analysis

123

Research on AI conversational data has primarily focused on dialogue quality, user  
satisfaction, and topic modeling [19]. Network-based approaches to conversation  
analysis remain rare. Our conference paper [2] introduced the first complex network

124

125

126

**Fig 1. Static snapshot of the knowledge network at the final time step.** Nodes represent conversations, edges connect semantically similar pairs (cosine similarity  $\geq 0.9$ ). Colors indicate Louvain communities labeled by dominant topic. Layout by ForceAtlas2 in Gephi. Reproduced from [2].

analysis of a personal AI conversation archive, treating conversations as nodes in a semantic similarity network. The present work extends this to the temporal domain, a direction identified as key future work in the original paper.

## Methods

We describe the temporal analysis methodology that extends our previous static analysis [2]. The base network construction (embedding generation, similarity computation, threshold selection) is unchanged; we refer readers to our previous work for those details and summarize the essentials here.

## Dataset and base network

The dataset comprises 1,908 ChatGPT conversations generated by one of the authors between December 2022 and April 2025. Conversations were conducted for authentic research, learning, and problem-solving purposes with no anticipation of future network analysis. Each conversation was embedded using `nomic-embed-text` [20] with a 2:1 user:AI message weighting ratio ( $\alpha = 2$ ), validated through a 63-configuration ablation study [2]. Pairwise cosine similarities were computed and filtered at threshold  $\theta = 0.9$ , yielding 601 connected nodes and 1,718 edges across 59 connected components, with a giant component of 453 nodes (1,307 conversations remain isolated at this threshold). Our previous work [2] analyzed the giant component (449 nodes at the time of that analysis); the present work analyzes all connected nodes. Fig 1 shows the resulting network at the final time step, illustrating the community structure that the temporal analysis below traces from emergence to stabilization.

Each conversation carries a creation timestamp, enabling temporal ordering. Conversations span five model eras based on the underlying LLM: GPT-3.5 (pre-GPT-4,  $n = 1,214$ ), GPT-4 ( $n = 44$ ), GPT-4o ( $n = 453$ ), Reasoning models (o1/o3 series,  $n = 181$ ), and GPT-4.5 ( $n = 16$ ).

## Cumulative temporal snapshots

152

We construct the network's temporal evolution through cumulative monthly snapshots. 153  
For each month  $t$  in the observation period, the snapshot  $G(t) = (V(t), E(t))$  contains: 154

$$V(t) = \{v \in V \mid \text{created}(v) \leq \text{end}(t)\} \quad (1)$$

$$E(t) = \{(u, v) \in E \mid u \in V(t) \wedge v \in V(t)\} \quad (2)$$

This cumulative construction models the irreversibility of knowledge exploration: 155  
conversations, once created, permanently enrich the knowledge landscape. It produces 156  
29 monthly snapshots (December 2022 through April 2025), growing from 1 node to the 157  
full 1,908. 158

For each snapshot, we compute a comprehensive set of network metrics on the 159  
connected subgraph: node count, edge count, density, number of connected components, 160  
giant component size and fraction, mean and maximum degree, average clustering 161  
coefficient, transitivity, average shortest path length (within the giant component), 162  
Louvain modularity and community count [21], average betweenness centrality, and 163  
degree assortativity. Community detection uses a fixed random seed for reproducibility. 164

For visualization and narrative purposes, we divide the 29-month observation period 165  
into five temporal phases based on usage intensity and model availability: *Early* 166  
(December 2022 – February 2023; 59 conversations, network bootstrapping), *Exploration* 167  
(March – July 2023; 635 conversations, rapid growth), *Established* (August 2023 – 168  
January 2024; 402 conversations, structural consolidation), *GPT-4o* (February – 169  
September 2024; 396 conversations, new model capabilities), and *Reasoning* (October 170  
2024 – April 2025; 416 conversations, reasoning model era). These phases appear as 171  
background shading in several figures. 172

## Community lifecycle tracking

173

To track community identity across snapshots, we apply Louvain community detection 174  
independently at each time step and align communities between consecutive snapshots 175  
using Jaccard similarity of node sets [13, 22]. 176

The tracking algorithm (Algorithm 1) proceeds in five passes per transition. First, 177

for each pair of consecutive communities  $(C_{t-1}^i, C_t^j)$ , we compute the Jaccard index  
 $J(C_{t-1}^i, C_t^j) = |C_{t-1}^i \cap C_t^j| / |C_{t-1}^i \cup C_t^j|$ . Communities are then classified:

- *Continuation*:  $C_t^j$  has a unique best match  $C_{t-1}^i$  with  $J \geq 0.3$ , and vice versa.

The tracked identity is preserved.

- *Birth*:  $C_t^j$  has no match  $\geq 0.3$  with any previous community. A new tracked identity is assigned.

- *Death*:  $C_{t-1}^i$  has no match  $\geq 0.3$  with any current community. The tracked identity is recorded as dissolved.

- *Merge*: Multiple previous communities' best match is the same current community.

- *Split*: One previous community maps to multiple current communities.

Each tracked community is labeled with a dominant topic based on keyword analysis of constituent conversation titles (e.g., ML/AI, Programming, Statistics, Philosophy, Health, Networks).

---

**Algorithm 1** Community lifecycle tracking via Jaccard alignment

---

**Require:** Community partitions  $\{P_1, P_2, \dots, P_T\}$ , threshold  $\tau = 0.3$

**Ensure:** Tracked communities with lifecycle events

```

1: next_id  $\leftarrow 0$ 
2: for  $t \leftarrow 2$  to  $T$  do
3:   Compute Jaccard matrix  $J[i, j] = |C_{t-1}^i \cap C_t^j| / |C_{t-1}^i \cup C_t^j|$ 
4:   Pass 1 (Continuations): Match pairs where  $\arg \max_j J[i, j] = j^*$  and  $\arg \max_i J[i, j^*] = i$  and  $J[i, j^*] \geq \tau$ 
5:   Pass 2 (Merges): Detect  $N:1$  mappings among unmatched communities
6:   Pass 3 (Splits): Detect  $1:N$  mappings among unmatched communities
7:   Pass 4 (Births): Assign new tracked_id to unmatched current communities
8:   Pass 5 (Deaths): Record unmatched previous communities as dissolved
9: end for

```

---

## Preferential attachment analysis

To test whether new conversations preferentially attach to high-degree existing nodes, we analyze each monthly transition. For month  $t$ , we identify new nodes  $V_{\text{new}}(t) = V(t) \setminus V(t-1)$  and new edges incident to these nodes in the existing network  $G(t-1)$ . For each existing node  $v \in V(t-1)$ , we compute the fraction of new nodes that connect to it and correlate this with  $v$ 's degree in  $G(t-1)$ .

To assess statistical significance, we compare the observed degree–attachment  
 correlation against a null model of uniform random attachment using 1,000 permutation  
 tests per month. We compute a  $z$ -score indicating how many standard deviations the  
 observed correlation exceeds the null expectation.

To quantify the attachment kernel, we pool data across all months and bin existing  
 nodes by degree. For each bin, we compute the empirical attachment rate  $\Pi(k)$  (mean  
 number of new connections per existing node at degree  $k$ ). We then fit the power-law  
 kernel:

$$\Pi(k) \propto k^\beta \quad (3)$$

where  $\beta = 0$  corresponds to uniform random attachment,  $\beta = 1$  to linear preferential  
 attachment (Barabási-Albert model), and intermediate values indicate sub-linear  
 preferential attachment.

## Densification law analysis

Following Leskovec et al. [11], we test whether the network exhibits densification by  
 fitting a power law in the log-log space of connected nodes versus edges:

$$e(t) \propto n(t)^\gamma \quad (4)$$

where  $n(t) = |V_{\text{connected}}(t)|$  and  $e(t) = |E(t)|$  are the number of connected nodes and  
 edges at time  $t$ . The exponent  $\gamma > 1$  indicates super-linear densification (the network  
 becomes proportionally denser),  $\gamma = 1$  indicates constant average degree, and  $\gamma < 1$   
 indicates sparsification.

We fit this relationship using ordinary least squares (OLS) regression on the  
 log-transformed data, reporting the exponent  $\gamma$ , coefficient of determination  $R^2$ , and  
 $p$ -value. We note that OLS on log-transformed data is an approximation; more rigorous  
 power-law fitting methods exist [23], but OLS is standard practice for densification  
 analysis [3] and sufficient given our high  $R^2$  values. We exclude early months with fewer  
 than 4 connected nodes where metrics are unstable.

An important methodological consideration: because our edges derive from a  
 pre-computed similarity matrix, the densification we observe reflects the *progressive*

*revelation* of latent semantic structure rather than the *creation* of new connections (as  
223  
in social networks where people actively form ties). We discuss the implications of this  
224  
distinction in the Densification paradox subsection.  
225

## Bridge formation dynamics 226

Our previous work identified five bridge conversations with high betweenness centrality  
227  
that facilitate cross-domain knowledge transfer [2]. We track these bridges over time,  
228  
computing normalized betweenness centrality and the number of distinct neighbor  
229  
communities at each snapshot from each bridge’s creation month onward. A bridge is  
230  
considered to have achieved “bridge status” when its betweenness centrality enters the  
231  
top 5% of all nodes.  
232

## Model era sub-network comparison 233

To assess whether AI model capabilities influence network topology, we construct  
234  
separate sub-networks for each model era. Each sub-network contains only conversations  
235  
from that era and only edges between them. We compute standard network metrics for  
236  
each era’s sub-network and compare structural signatures across eras.  
237

# Results 238

## Network growth patterns 239

Fig 2 shows the network’s growth over 29 months. Total conversations grow from 1  
240  
(December 2022) to 1,908 (April 2025), with a rapid expansion phase from March  
241  
through July 2023 (averaging 135 new conversations per month) followed by steadier  
242  
growth (averaging 54 per month thereafter). Of the 1,908 total conversations, 601  
243  
(31.5%) appear in the connected network at  $\theta = 0.9$ .  
244

**Fig 2. Network growth over 29 months.** (a) Cumulative node count with monthly  
additions (bars). (b) Cumulative edge count with monthly additions. (c) Edges per  
connected node (connectivity ratio), which rises steeply through late 2023, then  
stabilizes near 3.0. Background shading indicates five temporal phases: Early (Dec 2022  
– Feb 2023), Exploration (Mar – Jul 2023), Established (Aug 2023 – Jan 2024), GPT-4o  
(Feb – Sep 2024), and Reasoning (Oct 2024 – Apr 2025).

Edge growth outpaces node growth throughout the observation period. The  
edges-per-node ratio increases from 0.5 (January 2023) to 2.86 (April 2025), confirming  
that the network becomes proportionally denser over time, because edge count grows  
faster than the number of nodes.

## Structural evolution

Fig 3 tracks four structural properties over time. Modularity rises rapidly from 0.0 to  
0.44 during the Exploration phase (March–July 2023), then undergoes a step increase to  
0.64 in November 2023 when the giant component undergoes significant restructuring.  
From late 2024, modularity stabilizes around 0.74 and reaches 0.750 at the final  
snapshot, matching our previous work’s static analysis exactly.

**Fig 3. Structural evolution of network properties.** (a) Modularity jumps to 0.64  
in November 2023, then gradually reaches 0.75. (b) Community count grows from 1 to  
15. (c) Clustering coefficient and transitivity remain stable after initial growth.  
(d) Giant component fraction shows a step increase in November 2023.

The number of detected communities grows from 1 to 15 over the observation period,  
with most community births occurring before mid-2024. Clustering coefficient and  
transitivity gradually converge to approximately 0.44 by the final snapshot, indicating  
that topically related conversations tend to share neighbors (if A is similar to B and B  
to C, then A and C are also likely similar), a pattern that establishes early and persists.

The giant component fraction exhibits interesting non-monotonic behavior. It grows  
to  $\sim 0.47$  during the Exploration phase, drops temporarily, then increases sharply to  
0.69 in November 2023 when previously isolated clusters merge. This step change  
coincides with the modularity jump, suggesting a phase transition in the network’s  
mesoscale organization. We note that this step change is partly mechanical: when  
previously isolated clusters merge into the giant component, modularity captures the  
separation between these formerly independent groups, inflating the apparent  
magnitude of the jump. The communities themselves are real, but the abruptness of the  
increase reflects the measurement context as much as the network dynamics.

## Community lifecycles

269

Community tracking across 29 snapshots reveals 40 unique tracked communities, of  
270 which 15 survive to the final snapshot (Table 1). The lifecycle analysis recorded 189  
271 continuation events, 40 births, and 25 deaths. No merge or split events were detected at  
272 the Jaccard threshold of  $J = 0.3$ , suggesting that communities in this network grow and  
273 dissolve rather than recombining.  
274

**Table 1. Community lifecycle event summary.** Events tracked across 28 monthly transitions.

Event Type	Count
Continuations	189
Births	40
Deaths	25
Merges	0
Splits	0
Unique communities tracked	40
Surviving to final snapshot	15

Fig 4 shows the size trajectories of the eight largest communities. Several patterns  
275 emerge. First, the two earliest and largest communities (Stats & R Packages and Deep  
276 Learning) are born in the first months and grow throughout the observation period,  
277 consistent with the “first-mover advantage” observed in other temporal community  
278 studies [5]. Second, community births are distributed across the entire period rather  
279 than concentrated at the beginning: later communities such as Software Dev and Prog.  
280 Languages appear during the Reasoning phase, indicating ongoing diversification of  
281 knowledge exploration. Third, community deaths are concentrated among small,  
282 specialized communities that emerge briefly and dissolve (e.g., Casual / Philosophy),  
283 often subsumed by their larger neighbors as the network densifies.  
284

**Fig 4. Community size trajectories.** Line plot showing the eight largest communities (by maximum size) over time; each line is labeled by its dominant topic. Stars mark community birth months. The two earliest communities (Stats & R Packages, Deep Learning) grow steadily throughout the observation period, while later-born communities (Software Dev, Prog. Languages) emerge during the Reasoning phase.

The absence of merge and split events is notable. In social networks, community  
285 merging and splitting are common [5]. Their absence here may reflect the nature of  
286 knowledge domains: topics like “machine learning” and “statistics” are cognitively  
287

distinct categories that grow internally rather than fusing, unlike social groups whose  
288 membership boundaries are more fluid.  
289

## Densification law

The relationship between connected nodes and edges follows a power law with  
291 remarkable fidelity (Fig 5). Fitting  $\log e(t) = \gamma \log n(t) + c$  yields:  
292

$$\gamma = 1.405 \pm 0.022, \quad R^2 = 0.993, \quad p < 10^{-29} \quad (5)$$

**Fig 5. Densification law.** Log-log plot of edges versus connected nodes across 28 monthly snapshots. The fitted power law  $e(t) \propto n(t)^{1.405}$  (dashed line) fits the data with  $R^2 = 0.993$ , indicating super-linear densification.

This super-linear exponent ( $\gamma = 1.405$ ) places our knowledge network in the  
293 company of other densifying real-world networks, albeit with a moderate exponent.  
294 Table 2 compares our result with previously reported densification exponents.  
295

**Table 2. Densification exponents across network types.** Our knowledge network exhibits moderate super-linear densification comparable to patent and social networks.

Network	Exponent $\gamma$	Source
arXiv citations	1.69	[3]
Patent citations	1.26	[3]
Autonomous systems	1.18	[3]
<b>Knowledge network (this work)</b>	<b>1.405</b>	—

The exponent  $\gamma = 1.405$  indicates that for every doubling of the connected node  
296 count, the edge count increases by a factor of  $2^{1.405} \approx 2.65$ . This acceleration reflects  
297 the increasing semantic interconnectedness of conversations as the knowledge landscape  
298 fills in: later conversations are more likely to find existing semantic neighbors than early  
299 ones, because there are more potential neighbors in a richer knowledge base.  
300

## Preferential attachment

Fig 6 presents the preferential attachment analysis. The degree–attachment correlation  
302 is consistently positive and significantly exceeds the null model of random attachment  
303 across nearly all months. The median  $z$ -score is 8.4 (range: 0.5–13.2), indicating strong  
304 statistical evidence for preferential attachment.  
305

**Fig 6. Preferential attachment analysis.** (a) Attachment kernel  $\Pi(k)$  versus degree  $k$  on log-log axes, with fitted power law  $\Pi(k) \propto k^{0.763}$ . (b) Monthly degree–attachment correlation (blue) versus null model distribution (gray band shows mean  $\pm 2\sigma$ ). (c) Distribution of  $z$ -scores across months, consistently exceeding the significance threshold.

The pooled attachment kernel follows a power law  $\Pi(k) \propto k^\beta$  with:

$$\beta = 0.763 \pm 0.074, \quad R^2 = 0.914 \quad (6)$$

This sub-linear exponent ( $0 < \beta < 1$ ) indicates that high-degree nodes attract disproportionately many new connections, but less aggressively than pure preferential attachment ( $\beta = 1$ ). This is consistent with the observation from our previous work that our network exhibits “evolution beyond preferential attachment, reflecting cognitive exploration with hub formation limited by specialization and cross-domain constraints” [2]. The sub-linear kernel produces networks with moderate degree heterogeneity (broad-tailed degree distributions without extreme hubs), matching the empirical degree distribution observed in our network.

The month-by-month correlation shows temporal variation. The Exploration phase (March–July 2023) exhibits the strongest preferential attachment ( $z > 8$ ), possibly because rapid growth leads new conversations to cluster around established topics. Later months show reduced but still significant preferential attachment, consistent with diversification into new topics.

## Bridge formation dynamics

Our previous work identified three bridge types based on how conversations achieve cross-domain connectivity [2]. *Evolutionary bridges* are conversations that begin in one domain and organically evolve to span multiple topics through the natural progression of inquiry. *Integrative bridges* occupy a central position connecting several communities simultaneously, serving as multi-domain hubs. *Pure bridges* provide a minimal but critical link between exactly two communities, connecting domains that would otherwise be disconnected.

Fig 7 tracks the five bridge conversations identified in our previous work over time. The most striking pattern is the *dominance and persistence* of the primary bridge,

`geometric-mean-calculation`, which achieves bridge status immediately upon  
330 appearing in the connected network (November 2023) and maintains normalized  
331 betweenness centrality above 0.44 throughout the remaining 18 months. This  
332 conversation, which evolved from geometric means into probability theory and neural  
333 networks, exemplifies the “evolutionary bridge” type from our previous taxonomy.  
334

**Fig 7. Bridge formation dynamics.** Normalized betweenness centrality over time  
for the five bridge conversations identified in the conference paper. Vertical dashed lines  
mark creation dates. The “geometric-mean-calculation” bridge dominates throughout,  
while other bridges emerge later and stabilize at lower centrality levels.

The second bridge, `loss-in-l1m-training`, enters the connected network in  
335 November 2023, the same month as the primary bridge, but with low initial  
336 betweenness centrality (0.006). It achieves bridge status only in September 2024, when  
337 its centrality jumps to 0.045 as it begins connecting multiple communities. The  
338 remaining three bridges (`mcts-code-analysis-suggestions`,  
339 `algotree-generate-unit-tests-flattree`, `compile-cuda-program-linux`) enter in  
340 October 2024, coinciding with the expansion of the giant component during the  
341 Reasoning model era. Once established, each maintains a stable betweenness centrality  
342 level: `mcts-code-analysis-suggestions` and `loss-in-l1m-training` stabilize  
343 around 0.35, `algotree-generate-unit-tests-flattree` around 0.30, and  
344 `compile-cuda-program-linux` around 0.09–0.12.  
345

The stability of bridge centrality over time (once established, bridges maintain their  
346 structural role) suggests that the cross-domain connections they provide are not  
347 incidental but reflect genuine semantic bridging between knowledge communities.  
348

The bridge conversations span 2–5 neighbor communities each. The  
349 `geometric-mean-calculation` bridge consistently connects 4 communities, confirming  
350 its role as a multi-domain integrator. In contrast, `compile-cuda-program-linux`  
351 connects exactly 2 communities throughout its tracked lifetime, exemplifying the “pure  
352 bridge” type: a minimal but critical link between domains.  
353

## Model era effects

As an exploratory analysis, we examined sub-network metrics for each model era  
355 (Table 3). The two eras with sufficient data for meaningful comparison are GPT-3.5  
356

(360 connected conversations) and GPT-4o (112 connected). Both produce well-structured sub-networks with comparable modularity (0.675 and 0.670, respectively), suggesting that the knowledge organization patterns we observe are robust across model generations. The Reasoning era (32 connected conversations) shows lower modularity (0.353), which may reflect the nature of reasoning model usage (focused, cross-domain technical problems) or simply the smaller sample size. The GPT-4 and GPT-4.5 eras contain too few connected conversations for meaningful analysis (15 and 2, respectively); the small GPT-4 sample is partly a metadata artifact, as ChatGPT’s export format did not record the model field until approximately March 2024.

**Table 3. Sub-network metrics by model era.** Each era’s sub-network contains only conversations from that era and edges between them.

Era	Connected	Edges	Avg. Degree	Clustering	Modularity
GPT-3.5	360	1,102	6.12	0.381	0.675
GPT-4	15	10	1.33	0.156	0.001
GPT-4o	112	148	2.64	0.290	0.670
Reasoning	32	27	1.69	0.168	0.353
GPT-4.5	2	1	—	—	—

## Discussion

### Cognitive interpretation

The temporal evolution patterns we observe can be interpreted through the lens of distributed cognition [6] and the extended mind thesis [7]. The conversation archive functions as an externalized cognitive system, and its growth dynamics reveal how knowledge exploration self-organizes over time.

The super-linear densification ( $\gamma = 1.405$ ) has a natural cognitive interpretation: as one’s knowledge base grows, new inquiries increasingly connect to existing knowledge rather than standing alone. This reflects the cumulative nature of learning, as later conversations benefit from a richer contextual landscape, making semantic connections more likely. The densification exponent quantifies the *rate* at which knowledge becomes interconnected.

Sub-linear preferential attachment ( $\beta = 0.763$ ) suggests a balanced exploration strategy. Popular topics (high-degree nodes) do attract follow-up conversations, but the

sub-linear kernel indicates that the user also explores less-established topics rather than  
380 exclusively deepening existing interests. This balances *exploitation* (deepening known  
381 topics) with *exploration* (investigating new ones), a trade-off recognized as fundamental  
382 in organizational learning [24] and cognitive search [25]. The same caveat from the  
383 densification paradox (Section ) applies here: high-degree nodes may attract new  
384 connections partly because they occupy broad regions of semantic space where new  
385 conversations are more likely to land, rather than through a generative rich-get-richer  
386 mechanism. The permutation null model tests against random attachment but does not  
387 account for this geometric confound.  
388

The early stabilization of community structure (modularity exceeding 0.6 by late  
389 2023 and stabilizing near 0.75) suggests that a user’s knowledge domains crystallize  
390 relatively quickly. Once the major thematic communities are established, subsequent  
391 growth fills in rather than restructures. This is consistent with schema theory in  
392 cognitive psychology [26]: once mental frameworks are established, new information is  
393 assimilated into existing schemas rather than triggering wholesale reorganization.  
394

## Comparison with known network evolution patterns

395

Our findings place this personal knowledge network squarely within the family of  
396 densifying, preferentially attaching real-world networks documented by Leskovec et  
397 al. [3] and Barabási and Albert [4]. The densification exponent ( $\gamma = 1.405$ ) falls  
398 between those of patent citation networks ( $\gamma = 1.26$ ) and arXiv citation networks  
399 ( $\gamma = 1.69$ ), suggesting comparable but not identical growth dynamics. The sub-linear  
400 preferential attachment ( $\beta = 0.763$ ) is lower than the near-linear values reported for  
401 citation networks [12], consistent with our previous observation that hub formation is  
402 limited by cognitive specialization constraints.  
403

The community lifecycle patterns (persistent major communities, gradual births, no  
404 merges or splits) differ from social networks where community fusion and fission are  
405 common [5]. This distinction likely reflects the fundamental difference between social  
406 identity (fluid, negotiated) and knowledge domain identity (more stable, ontologically  
407 grounded). A “machine learning” community and a “statistics” community may grow  
408 closer as related conversations accumulate, but they do not merge in the way social  
409

groups do.

The temporal analysis also provides a developmental account of the heterogeneous topology described in our previous work [2], where theoretical domains (ML/AI, Statistics, Philosophy) exhibited dense hub-and-spoke structures while practical domains (Programming) showed sparser, tree-like hierarchies. Community tracking reveals that this heterogeneity has a temporal origin: the theoretical communities are among the earliest born and have the longest growth histories, accumulating dense internal connections through sustained exploration over 20+ months. In contrast, practical communities tend to emerge later and grow through independent, focused conversations that branch rather than cluster. The sub-linear preferential attachment we measure ( $\beta = 0.763$ ) quantifies the mechanism observed qualitatively in our previous work, where hubs form but their growth is limited by cognitive specialization, producing the moderate degree heterogeneity rather than extreme scale-free structure.

## The densification paradox

An important caveat applies to the densification finding. In social networks, densification reflects the active formation of new ties. In our network, edges derive from a pre-computed similarity matrix: all potential connections exist latently from the moment both endpoints are created. The “densification” we observe is actually the *progressive revelation* of latent structure as the network fills in.

This distinction matters for interpretation but does not invalidate the finding. The densification law still accurately describes the growth trajectory and has predictive value: it tells us that later additions to the conversation archive will be proportionally more connected than earlier ones. The mechanism, however, is not active tie formation but rather the increasing density of the semantic space being sampled. As more conversations are added to more topics, new conversations are more likely to land near existing ones in semantic space.

We note that a similar argument applies to any threshold-based similarity network, including many co-occurrence and co-citation networks. The densification patterns reported by Leskovec et al. [3] for citation networks also involve retrospectively computed relationships rather than active social ties. Our case makes this mechanism

particularly transparent.

440

## Implications

441

Our findings have practical implications for the design of knowledge management and  
conversation archival systems:

442

443

- The early stabilization of community structure suggests that knowledge domains  
can be identified relatively early in a user's conversation history and used to  
organize archives thematically.
- The persistence of bridge conversations once established suggests that identifying  
bridges early could help users discover cross-domain connections.
- The densification law implies that retrieval systems should expect increasing  
connectivity over time, potentially enabling richer recommendation strategies as  
the archive grows.
- The distinct topological signatures across model eras suggest that AI model  
capabilities influence knowledge structure, which may be relevant for  
understanding how AI tools shape thinking patterns.

444

445

446

447

448

449

450

451

452

453

454

## Limitations

455

Several limitations constrain the generalizability of our findings.

456

*Single user.* This remains a case study of one user's conversation archive. While we  
demonstrate that the methodology produces meaningful and consistent results, the  
specific parameter values ( $\gamma$ ,  $\beta$ , community count, etc.) may differ for other users.  
Multi-user studies are needed to establish which patterns generalize.

457

458

459

460

461

462

463

464

*Pre-computed similarities.* As discussed in the Densification paradox subsection,  
edges are determined by semantic similarity rather than active social choice. The  
densification we observe is structurally real but mechanistically different from that in  
social or collaboration networks.

*Embedding model consistency.* The same embedding model (**nomic-embed-text**)  
was used for all conversations regardless of their creation date. In practice, embedding

465

466

models may evolve, and conversations from different periods might be better  
467  
represented by different models.  
468

*Louvain non-determinism.* Although we use a fixed random seed, Louvain  
469  
community detection is inherently non-deterministic across different implementations  
470  
and platforms. Community event counts may vary slightly across runs.  
471

*Single threshold.* We analyze the network at a single similarity threshold ( $\theta = 0.9$ ),  
472  
validated by the ablation study in our previous work. The temporal patterns at other  
473  
thresholds remain unexplored.  
474

*Model era confounds.* The consistency of modularity between GPT-3.5 and GPT-40  
475  
eras is encouraging, but differences between model era sub-networks may also reflect  
476  
temporal trends (topic interests change over time) rather than model capabilities alone.  
477  
Furthermore, model era classification relies on export metadata that was absent before  
478  
March 2024; the “GPT-3.5” era likely contains an unknown number of GPT-4  
479  
conversations. Larger multi-era datasets will be needed to disentangle model effects  
480  
from temporal trends.  
481

## Conclusion

We have extended the cognitive MRI methodology from static to temporal network  
482  
analysis, tracking how a personal knowledge network grows over 29 months of  
483  
AI-assisted conversation. The network exhibits evolution patterns characteristic of  
484  
real-world complex systems: super-linear densification ( $\gamma = 1.405$ ), sub-linear  
485  
preferential attachment ( $\beta = 0.763$ ), early community stabilization, and persistent  
486  
bridge formation.  
487

These findings suggest that personal knowledge exploration through conversational  
488  
AI is not random accumulation but produces network growth patterns quantitatively  
489  
similar to those observed in collective knowledge systems such as citation and  
490  
collaboration networks. Densification quantifies how knowledge becomes increasingly  
491  
interconnected; preferential attachment reveals the balance between deepening existing  
492  
interests and exploring new ones; community tracking shows how knowledge domains  
493  
crystallize early and persist; bridge dynamics illuminate how cross-domain connections  
494  
form and stabilize. That these patterns emerge at the scale of a single individual,  
495

mirroring dynamics previously documented only in multi-agent systems like scientific  
497 citation networks and online collaboration platforms, suggests that the organizational  
498 principles of knowledge exploration may be scale-bridging, operating whether the  
499 exploring agent is a community of scientists or a single person conversing with an AI.  
500

The mechanism differs: in our network, densification reflects the progressive  
501 revelation of latent semantic structure rather than active social tie formation. Yet the  
502 resulting growth laws are strikingly similar, raising the question of whether individual  
503 and collective knowledge exploration are governed by common underlying dynamics.  
504 Multi-user studies will be needed to test this hypothesis. Different users may exhibit  
505 different densification exponents depending on whether their exploration strategy is  
506 focused (deep specialist) or broad (wide generalist). The balance between exploitation  
507 and exploration captured by the attachment exponent  $\beta$  likely varies with individual  
508 cognitive style and professional domain. Multi-user comparison would reveal which  
509 parameters are universal properties of knowledge exploration through AI and which  
510 reflect individual differences in how people organize inquiry. Our methodology provides  
511 a template for such investigations, and our findings offer a first empirical  
512 characterization of how one person's knowledge landscape evolves through sustained AI  
513 interaction.  
514

The analysis pipeline is available at [27] and the temporal analysis code, data, and  
515 paper source at [28].  
516

## Data availability statement

All derived data needed to reproduce the analyses presented in this paper are included  
518 in the research compendium at  
519  
<https://github.com/queelius/cognitive-mri-conversations> (DOI:  
520 10.5281/zenodo.18778377), including conversation embeddings (with message text  
521 removed for privacy), the primary edge list, and all temporal analysis outputs. A  
522 reproduction script (`data/reproduce.sh`) regenerates all figures and data files from the  
523 curated data directory. The analysis pipeline code is available separately at  
524  
<https://github.com/queelius/chatgpt-complex-net> (DOI:  
525 10.5281/zenodo.15314235). Raw conversation content is not shared to protect the  
526

privacy of the human participant.

527

## Author contributions

528

**Alexander Towell:** Conceptualization, Methodology, Software, Formal Analysis, Data Curation, Writing – Original Draft, Visualization. **John Matta:** Supervision, Writing – Review & Editing.

529

530

531

## Funding

532

This research received no specific funding from any agency in the public, commercial, or not-for-profit sectors.

533

534

## Competing interests

535

The authors declare no competing interests.

536

## Ethics statement

537

This study analyzes one author's own ChatGPT conversation archive. No human subjects were recruited, and no personally identifiable information about third parties is included in the dataset or analysis. The conversation data was generated through the author's routine use of a commercially available AI service.

538

539

540

541

## Acknowledgments

542

The authors thank the organizers of the PLOS Complex Systems special issue on Complex Networks 2025 for the invitation to submit this extended work.

543

544

## References

1. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. arXiv preprint arXiv:230318223. 2023.

2. Towell A, Matta J. Cognitive MRI of AI Conversations: Analyzing AI Interactions through Semantic Embedding Networks. In: Proceedings of the International Conference on Complex Networks and Their Applications (Complex Networks 2025). Studies in Computational Intelligence. Springer; 2025. In press.
3. Leskovec J, Kleinberg J, Faloutsos C. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*. 2007;1(1):2-es. doi:10.1145/1217299.1217301.
4. Barabási AL, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509-12. doi:10.1126/science.286.5439.509.
5. Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. *Nature*. 2007;446(7136):664-7. doi:10.1038/nature05670.
6. Hutchins E. *Cognition in the Wild*. MIT Press; 1995.
7. Clark A, Chalmers D. The extended mind. *Analysis*. 1998;58(1):7-19.
8. Holme P, Saramäki J. Temporal networks. *Physics Reports*. 2012;519(3):97-125. doi:10.1016/j.physrep.2012.03.001.
9. Dorogovtsev SN, Mendes JFF. Evolution of networks. *Advances in Physics*. 2002;51(4):1079-187. doi:10.1080/00018730110112519.
10. Albert R, Barabási AL. Statistical mechanics of complex networks. *Reviews of Modern Physics*. 2002;74(1):47-97. doi:10.1103/RevModPhys.74.47.
11. Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2005. p. 177-87. doi:10.1145/1081870.1081893.
12. Jeong H, Néda Z, Barabási AL. Measuring preferential attachment in evolving networks. *Europhysics Letters*. 2003;61(4):567-72. doi:10.1209/epl/i2003-00166-9.
13. Greene D, Doyle D, Cunningham P. Tracking the evolution of communities in dynamic social networks. In: Proceedings of the International Conference on

Advances in Social Networks Analysis and Mining; 2010. p. 176-83.  
doi:10.1109/ASONAM.2010.17.

14. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP. Community structure in time-dependent, multiplex, and other multiscale networks. *Science*. 2010;328(5980):876-8. doi:10.1126/science.1184819.
15. Rossetti G, Cazabet R. Community discovery in dynamic networks: a survey. *ACM Computing Surveys*. 2018;51(2):1-37. doi:10.1145/3172867.
16. de Solla Price DJ. Networks of scientific papers. *Science*. 1965;149(3683):510-5. doi:10.1126/science.149.3683.510.
17. Chen C, Ibekwe-SanJuan F, Hou J. The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology*. 2010;61(7):1386-409. doi:10.1002/asi.21309.
18. Shi F, Foster JG, Evans JA. Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks*. 2015;43:73-85. doi:10.1016/j.socnet.2015.02.006.
19. Serban IV, Lowe R, Henderson P, Charlin L, Pineau J. A survey of available corpora for building data-driven dialogue systems. arXiv preprint arXiv:151205742. 2016.
20. Nussbaum Z, Morris JX, Duderstadt B, Mulyar A. Nomic Embed: Training a Reproducible Long Context Text Embedder. arXiv preprint arXiv:240201613. 2024.
21. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;2008(10):P10008. doi:10.1088/1742-5468/2008/10/P10008.
22. Jaccard P. The distribution of the flora in the alpine zone. *New Phytologist*. 1912;11(2):37-50.

23. Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. SIAM Review. 2009;51(4):661-703. doi:10.1137/070710111.
24. March JG. Exploration and exploitation in organizational learning. Organization Science. 1991;2(1):71-87. doi:10.1287/orsc.2.1.71.
25. Hills TT, Todd PM, Lazer D, Redish AD, Coles IR, the Cognitive Search Research Group. Exploration versus exploitation in space, mind, and society. Trends in Cognitive Sciences. 2015;19(1):46-54. doi:10.1016/j.tics.2014.10.004.
26. Bartlett FC. Remembering: A Study in Experimental and Social Psychology. Cambridge University Press; 1932.
27. Towell A. chatgpt-complex-net: Complex network analysis of ChatGPT conversations. Zenodo; 2025. Available from: <https://github.com/queelius/chatgpt-complex-net>. doi:10.5281/zenodo.15314235.
28. Towell A, Matta J. Cognitive MRI of AI Conversations: Research Compendium. Zenodo; 2026. Available from: <https://github.com/queelius/cognitive-mri-conversations>.