Entropy and Heterogeneity in the AI Research-to-Product Knowledge Transfer Network

Anonymous Submission

We characterize the complexity of knowledge transfer from AI research to commercial products using information-theoretic measures applied to PeerGraph, an open bipartite network of 144 researchers, 91 products, and 216 builder-declared adoption links across 150 papers. The network exhibits high degree heterogeneity ($\mathrm{CV} = 3.7$), intermediate domain-flow entropy ($H = 4.48$ bits, $H/H_{\max} = 0.71$), and an effective channel count of 22 out of 78 possible domain pairs. These properties indicate an intermediate-entropy regime characteristic of structured complexity in knowledge transfer systems.

1. Network Construction and Degree Structure

PeerGraph is a bipartite graph $G = (R \cup B, E)$ with researcher nodes ($|R| = 144$), product nodes ($|B| = 91$), and edges ($|E| = 216$). An edge (r, b) indicates a builder declared product b uses a paper by researcher r. Edges carry domain annotations (12 research domains, 14 product domains, 78 observed pairs). Data derives from builder self-reports with public provenance; the dataset is a convenience sample with salience bias toward prominent papers. This construction parallels patent-to-paper citation networks [1] but captures declared deployment rather than inventive activity.

The researcher degree distribution has mean 1.5, median 1, max 58, and $\mathrm{CV} = 3.7$—well above the exponential threshold ($\mathrm{CV} = 1$), placing it in the heavy-tailed regime characteristic of heterogeneous knowledge networks [2]. Formal distributional fitting [3] requires larger samples than $N = 144$. The product partition shows lower heterogeneity (mean 2.4, max 11, $\mathrm{CV} \approx 1.6$).

2. Information-Theoretic Analysis

We treat the distribution of links over 78 domain pairs as a discrete distribution and compute three measures. (i) Shannon entropy $H = 4.48$ bits ($H/H_{\max} = 0.71$), yielding perplexity $2^H = 22.4$: the network concentrates flow through 29% of available pathways. (ii) KL divergence from uniform is $D_{\mathrm{KL}} = 1.80$ bits. (iii) Mutual information $I(\mathrm{src}; \mathrm{tgt}) = 0.18$ bits (normalized $\mathrm{MI} = 0.08$), indicating weak coupling: knowing the research domain provides little predictive information about the product domain, consistent with the cross-domain transfer ratio of 1.91 (547 cross vs. 287 intra-domain links).

The normalized entropy $H/H_{\max} = 0.71$ places this network between maximal order and maximal disorder. Complex systems theory associates such intermediate values with structured complexity [4]. Four dominant channels (NLP→GenAI, GenAI→GenAI, NLP→NLP, GenAI→NLP; 55% of links) coexist with 74 weaker channels.

3. Higher-Order Structure and Concentration

Products adopting k papers define $(k-1)$-simplices in the paper co-adoption complex. With mean product degree 2.4, multi-paper co-adoption is common; analysis of this simplicial structure (Betti numbers, persistence diagrams) is planned follow-up work. Paper adoption Gini is $G = 0.57$; domain flow Gini is $G = 0.72$. The most-adopted paper accounts for 26.9% of all links, comparable to inequality in citation distributions [5]. Salience bias in declarations may inflate this concentration.

Table 1: Complexity metrics for the PeerGraph network. Normalized values in parentheses.

Metric	Value	Interpretation
Degree CV (researchers)	3.7	Heavy-tailed regime
Shannon entropy H	4.48 bits (0.71)	Intermediate; structured
Effective channels	22.4 / 78 (29%)	Concentrated pathways
$D_{\mathrm{KL}}(\mathrm{obs} \| \mathrm{unif})$	1.80 bits	Departure from uniform
$I(\mathrm{src}; \mathrm{tgt})$	0.18 bits (0.08)	Weak domain coupling
Cross / intra ratio	1.91	Majority cross-domain
Paper adoption Gini	0.57	Moderate-high concentration
Domain flow Gini	0.72	High inequality

4. Limitations

Convenience sample ($N = 216$). Unverified declarations with salience bias. Entropy and Gini estimates have wide CIs at this scale. Future work: expanded coverage, bipartite null models [6], temporal entropy evolution, topological data analysis.

References

[1] A.B. Jaffe, M. Trajtenberg, *Patents, Citations, and Innovations*, MIT Press, 2002.

[2] M.E.J. Newman, "Scientific collaboration networks," *PNAS*, 98(2):404–409, 2001.

[3] A. Clauset et al., "Power-law distributions in empirical data," *SIAM Rev.*, 51(4):661–703, 2009.

[4] D.P. Feldman, J.P. Crutchfield, "Measures of statistical complexity," *Phys. Lett. A*, 238:244–252, 1998.

[5] S. Redner, "How popular is your paper?" *Eur. Phys. J. B*, 4(2):131–134, 1998.

[6] F. Saracco et al., "Randomizing bipartite networks: the case of the World Trade Web," *Sci. Rep.*, 5:10595, 2015.

NSIA Relevance Statement

(This section does not count toward the one-page abstract limit, per submission guidelines.)

This work directly addresses three topics in the NSIA 2026 Call for Abstracts.

Network Science for AI & AI for Network Science. We apply information-theoretic measures (Shannon entropy, KL divergence, mutual information) and network complexity metrics (Gini coefficient, coefficient of variation) to characterize the structural properties of AI's own knowledge transfer network. The bipartite research-to-product graph provides a new empirical setting for network science, while the findings inform our understanding of how AI research becomes technology.

Complex Network Intelligence. The intermediate normalized entropy ($H/H_{\max} = 0.71$) and effective channel concentration (22 of 78 pathways carry meaningful flow) reveal that research-to-product transfer operates as a structured complex system, not as a diffuse or random process. The coexistence of dominant transfer channels with a diverse tail of weaker channels exemplifies the kind of structured complexity studied in complex network intelligence.

Higher-Order Interactions and Learning. Multi-paper co-adoption by products defines higher-order interactions in the form of simplicial complexes in the paper co-adoption space. Products integrating research from multiple papers and domains represent higher-order knowledge integration events not reducible to pairwise relationships. Analysis of this simplicial structure is proposed as follow-up work with direct relevance to the NSIA community's interest in higher-order network analysis.

The underlying dataset is released under a CC0 license at `peergraph.ai`[1], providing the NSIA community with a new open empirical substrate for studying knowledge diffusion dynamics at the intersection of network science and artificial intelligence.

[1] URL provided for data access and review purposes. Authorship will be disclosed upon acceptance.