Sentence-Level Information Analysis of Narratives

Muhammad Abraar Abhirama[1], Siew Ann Cheong[2], and Yunfan Lai[3]

[1]Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore
[2]Physics Department, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
[3]Linguistics Department, School of Humanities, Nanyang Technological University, Singapore

Keywords: sentence information, transitional information, sentence-level analysis

Extended Abstract

Motivation. Previous works have shown that sequential comic-based visualizations can communicate temporal changes in dynamic networks [1], and that hypergraph derivatives can quantify mesoscopic relationships between linguistic units to characterize authorial style and language complexity [2]. We extend these frameworks by asking: do sentence-level character interactions carry measurable information-theoretic signatures that reflect narrative complexity and structural surprise?

Approach and Methodology. We propose a quantitative measure (information bits) for narrative data interpreted by a sequence of graphs. Consider a narrative of T sentences with character set $C = \{a, b, c \ldots, n\}$. Each sentence t is represented as a directed graph $G_t = (V_t, E_t)$, where $V_t \subseteq C$ are the characters mentioned and E_t are directed edges encoding verb-based interactions. For each graph $G_t = (V_t, E_t)$, we compute the likelihood $P(G_t)$ as the product over all characters (their presence and absence) in C, and we also compute the presence and absence of links between all characters. We then apply the negative logarithmic value to obtain the information value $I(G_t)$ in bits.

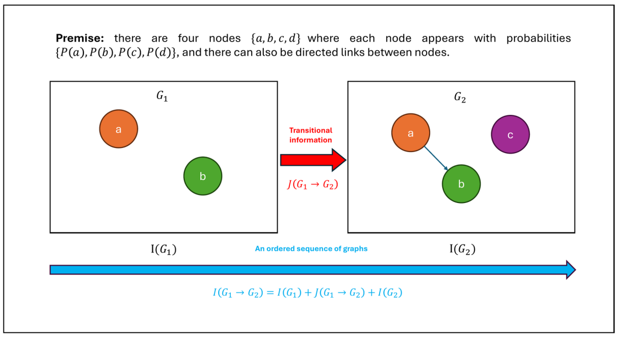


Figure 1: **A sketch of the transitional information within an ordered sequence of graphs.**

Now that in a sequence of graphs, there occurs a transition from one graph G_t to the next graph G_{t+1}. Within this graph transition, there appears information change which we define as TRANSITIONAL INFORMATION. The transitional information $J(G_t \rightarrow G_{t+1})$ captures how surprising the transition from one graph to the next is, given the observed sequence. We classify each graph G_t into a state $S_t = (V_t, E_t)$, grouping graphs that share exactly the same set of nodes and edges. The conditional probability and transitional information are then $R(S_{t+1}|S_t) = \frac{\#(S_{t+1}|S_t)}{\#(S_t)}$ and $J(G_t \rightarrow G_{t+1}) = -\log_2 R(S_{t+1}|S_t)$. See Figure 1 for the sketch.

We apply the framework across three narratives: *The Three Little Pigs* (16 sentences, 4 characters), Hawthorne's *Dr. Heidegger's Experiment* (132 sentences, 6 characters), and Rowling's *Harry Potter and the Philosopher's Stone* 16 chapters (3,980 sentences, 25+ characters).

Results. At the short-story scale, no transitional information is observed since the narrative is too brief for any graph state to recur. At the medium scale, transitional information emerges, and the framework identifies informational spikes. At the long-narrative scale, transitional information is likewise present, though the static information per sentence is lower due to differences in writing style (like how many characters appear per sentence and the density of their interactions). Notably, J exhibits elevated values *preceding* critical events that likely serves as an early warning signal, and remains elevated *following* such events, analogous to aftershock activity. This suggests that the storyteller must normalize the narrative again after the critical events.

Conclusions and Outlook. Our framework quantifies narrative structure across different scales: short, medium, and long. Future work includes a scaling collapse analysis across narratives to identify whether a universal scaling governs narrative information dynamics.

References

[1] B. Bach et al., "Telling stories about dynamic networks with graph comics," in *ACM CHI*, 2016.

[2] Á. Criado-Alonso et al., "Derivative of a hypergraph as a tool for linguistic pattern analysis," *Chaos Solitons Fractals*, vol. 163, 112604, 2022.

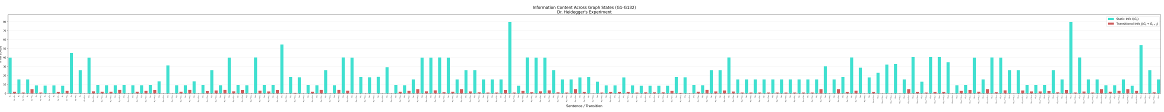


Figure 2: **Dr. Heidegger's Experiment** (132 sentences, 6 characters). $I(G_t)$ (teal) and $J(G_t \to G_{t+1})$ (red) across sentences.

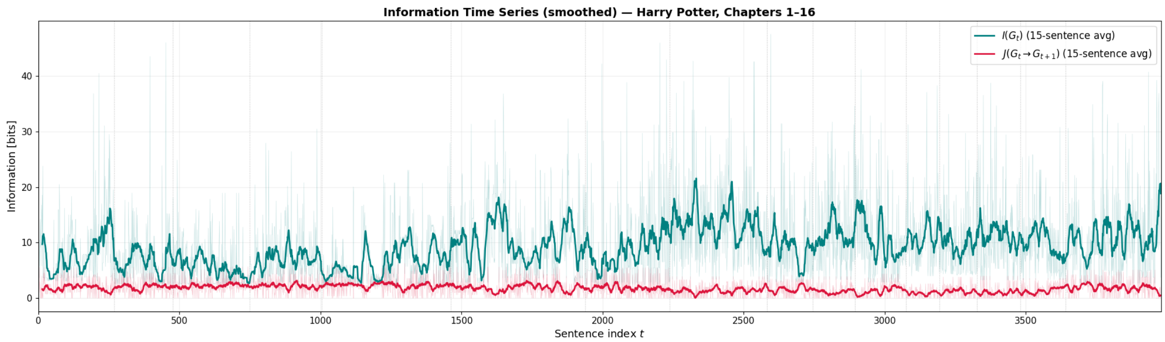


Figure 3: **Harry Potter** (3,980 sentences, 16 chapters): information time series with 15-sentence moving average. Chapter boundaries shown as vertical lines.

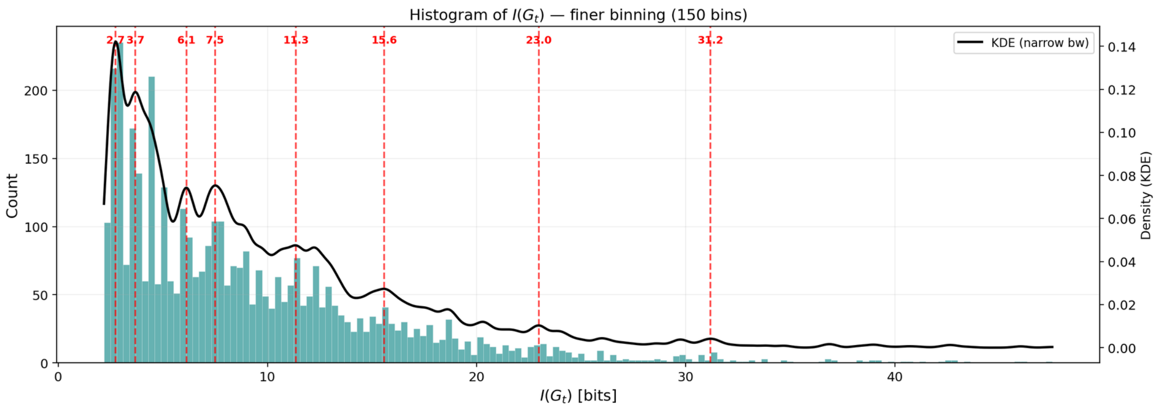


Figure 4: **Distribution of** $I(G_t)$ across all Harry Potter sentences (150 bins). KDE overlay with dashed lines marking prominent peaks.

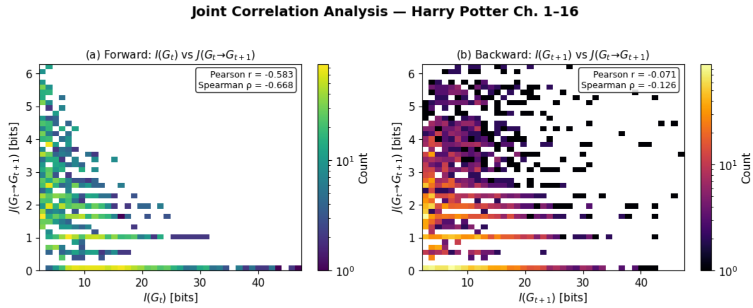
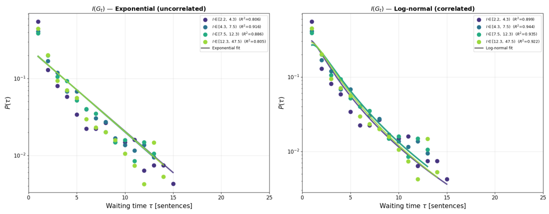
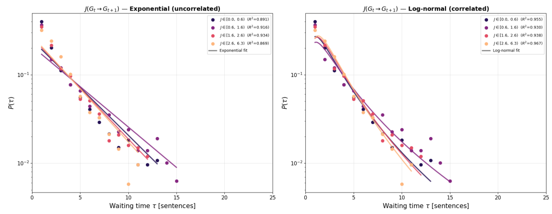


Figure 5: **Joint correlation heatmaps.** (a) Forward: $I(G_t)$ vs $J(G_t \to G_{t+1})$. (b) Backward: $I(G_{t+1})$ vs $J(G_t \to G_{t+1})$.

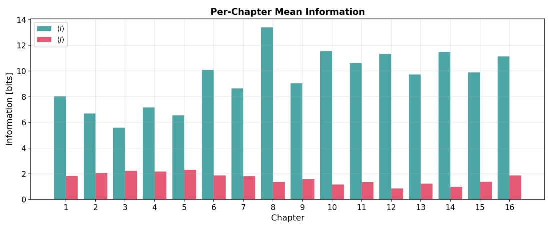


(a) $P(\tau)$ for $I(G_t)$: exponential (left) vs log-normal (right).

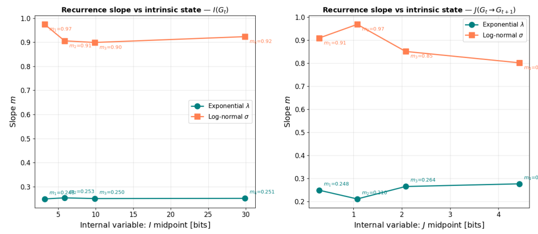


(b) $P(\tau)$ for $J(G_t \to G_{t+1})$: exponential vs log-normal.

Figure 6: **Recurrence time distributions** grouped by quartile bins. R^2 in log-space; log-normal consistently outperforms exponential.



(a) Per-chapter mean $\langle I \rangle$ and $\langle J \rangle$.



(b) Recurrence slope m vs intrinsic state.

Figure 7: **Chapter-level and state-dependent analysis.** (a) Mean information per chapter. (b) For I, slopes are flat; for J, λ increases and σ decreases with transition size.