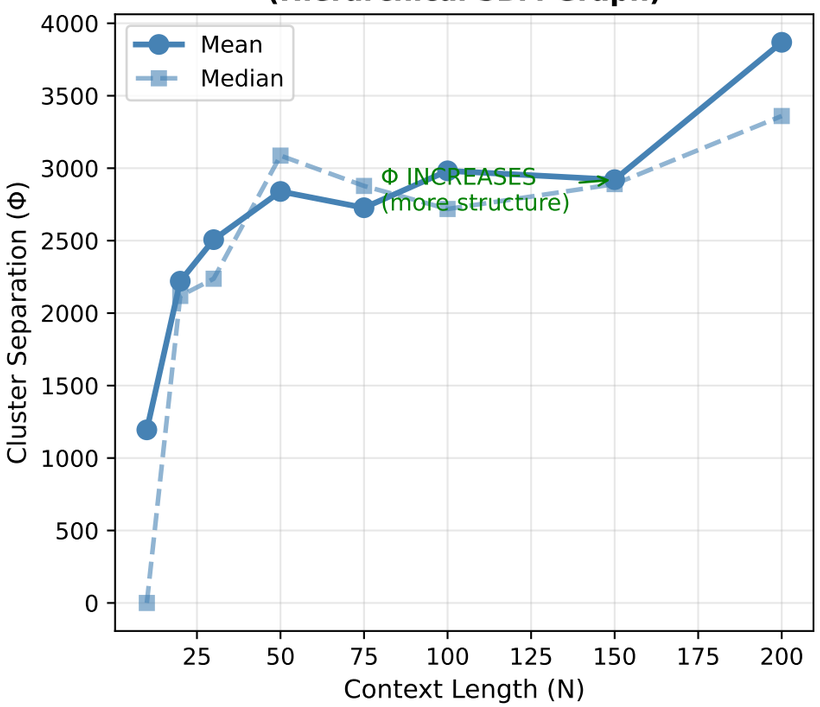
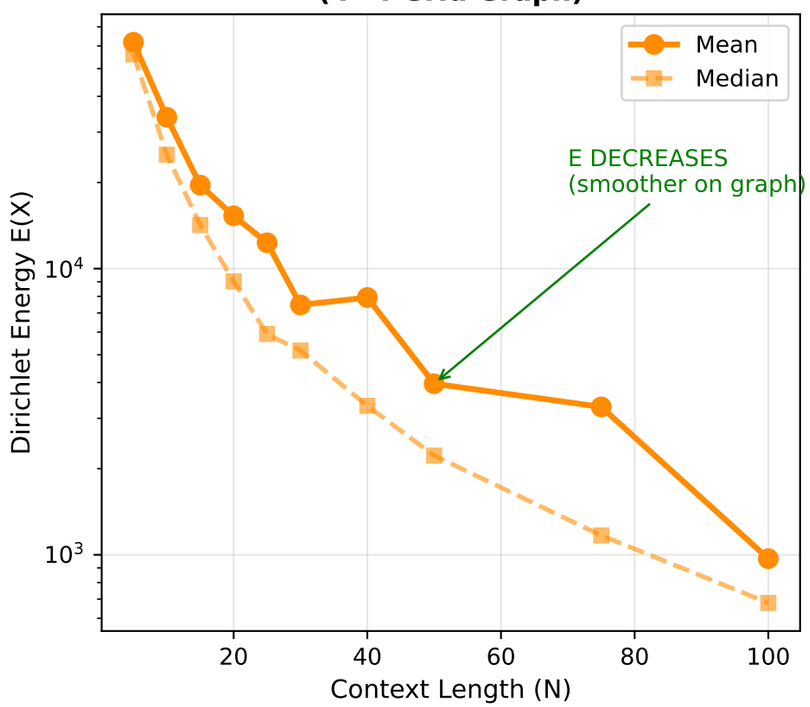
Comparison: Our Hierarchical Experiment vs Park et al. Reproduction
Model: LLaMA-3.1-8B

A. Our Experiment: ClusterSeparation (Hierarchical SBM Graph)



B. Park et al. Reproduction: Dirichlet Energy (4×4 Grid Graph)



C. Interpretation

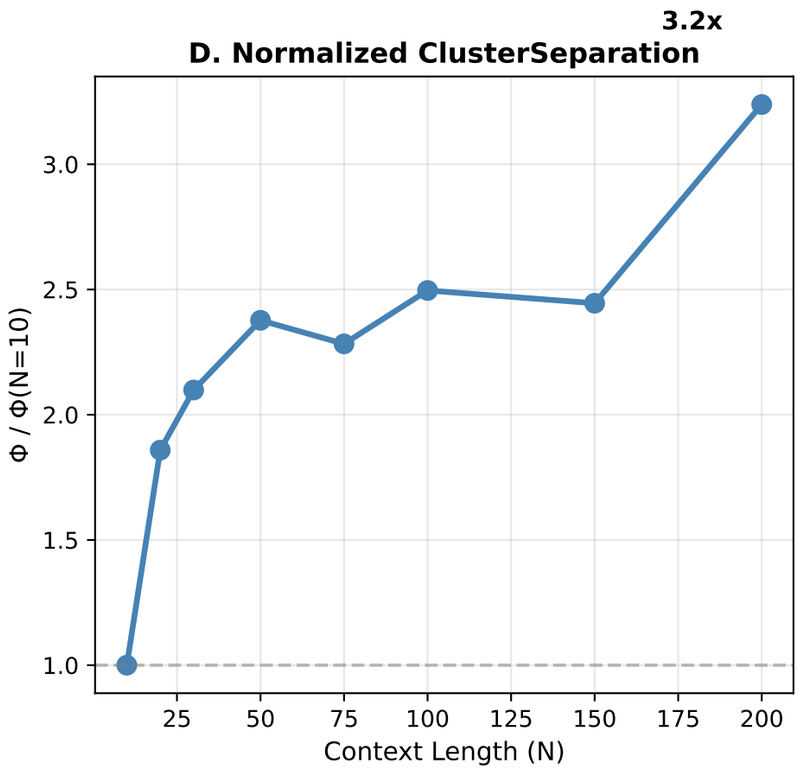
KEY INSIGHT: Both Metrics Show Structure Learning!

ClusterSeparation (Φ) ↑
- Measures: between/within cluster variance
- Higher = clusters more distinct
- INCREASES = learning cluster structure

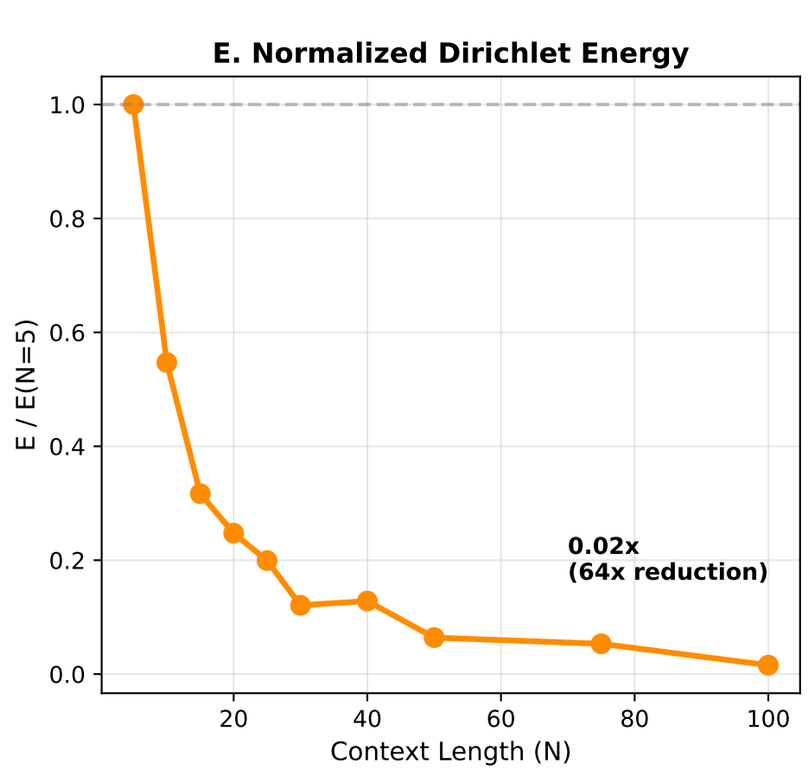
Dirichlet Energy (E) ↓
- Measures: smoothness on graph
- Lower = adjacent nodes similar
- DECREASES = learning graph topology

BOTH confirm Park et al.'s finding: "Representations reorganize to match context-specified structure"

D. Normalized ClusterSeparation



E. Normalized Dirichlet Energy



F. Comparison Summary

Metric	Our Experiment	Park Reproduction
Graph Type	Hierarchical SBM	Simple Grid
Metric Name	ClusterSeparation (Φ)	Dirichlet Energy (E)
Direction	INCREASES ↑	DECREASES ↓
Interpretation	Clusters more distinct	Graph smoother
N=10 → N=100	2.5x increase	64x decrease
Phase Transition	N≈20 (after bugfix)	N≈5-15
Conclusion	✓ Structure learned	✓ Structure learned