

---

# VISGRAPHVAR: A BENCHMARK GENERATOR FOR ASSESSING VARIABILITY IN GRAPH ANALYSIS USING LARGE VISION-LANGUAGE MODELS (SUPPLEMENTARY MATERIAL)

---

**Camilo Chacón Sartori**

Artificial Intelligence Research Institute (IIIA-CSIC)  
Bellaterra, Spain  
`cchacon@iiaa.csic.es`

**Christian Blum**

Artificial Intelligence Research Institute (IIIA-CSIC)  
Bellaterra, Spain  
`christian.blum@iiaa.csic.es`

**Filippo Bistaffa**

Artificial Intelligence Research Institute (IIIA-CSIC)  
Bellaterra, Spain  
`filippo.bistaffa@iiaa.csic.es`

Table 1: Task 1 (Detection)

(a)

LVLM	Layout									
	random		spiral		circular		shell		spring	
	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot
claude-3.5-sonnet	80.27	79.20	<b>78.49</b>	74.54	68.88	66.96	72.14	69.37	69.12	71.39
geminipro-1.5	<b>81.69</b>	<b>82.16</b>	78.31	<b>75.83</b>	<b>72.12</b>	<b>74.77</b>	<b>75.73</b>	<b>76.37</b>	80.87	<b>80.89</b>
gpt-4o-2024-08-06	67.87	68.30	66.01	61.70	61.16	58.50	63.20	56.38	<b>81.55</b>	80.85
llama-3.2-90b-vision-instruct	62.80	62.37	67.45	67.07	70.06	65.27	63.50	70.40	50.77	50.62
pixtral-12b	61.17	68.27	67.35	58.45	52.81	47.81	52.98	49.38	65.17	70.57
qwen-2-vl-72b-instruct	66.01	68.20	61.45	53.35	55.84	44.57	47.48	48.89	67.91	69.78
									73.57	72.29
									44.71	55.78
									50.92	54.56

(b)

LVLM	Edge type		
	directed		undirected
	CoT	0-shot	CoT 0-shot
claude-3.5-sonnet	75.27	74.41	64.63
geminipro-1.5	<b>79.40</b>	<b>79.52</b>	<b>71.97</b>
gpt-4o-2024-08-06	74.21	70.57	65.39
llama-3.2-90b-vision-instruct	62.92	63.52	54.31
pixtral-12b	62.79	66.94	54.81
qwen-2-vl-72b-instruct	66.92	64.66	55.37

(c)

LVLM	Node color		
	one color		random colors
	CoT	0-shot	CoT 0-shot
claude-3.5-sonnet	69.58	68.92	70.33
geminipro-1.5	<b>74.92</b>	<b>75.45</b>	<b>76.47</b>
gpt-4o-2024-08-06	69.08	67.12	70.52
llama-3.2-90b-vision-instruct	60.06	59.53	57.18
pixtral-12b	62.22	60.05	55.38
qwen-2-vl-72b-instruct	62.02	59.33	60.27

(d)

LVLM	Node labels		
	present		not present
	CoT	0-shot	CoT 0-shot
claude-3.5-sonnet	71.63	70.84	68.28
geminipro-1.5	<b>76.33</b>	<b>76.27</b>	<b>75.05</b>
gpt-4o-2024-08-06	73.41	73.45	66.19
llama-3.2-90b-vision-instruct	60.27	60.20	56.97
pixtral-12b	59.65	60.33	57.95
qwen-2-vl-72b-instruct	64.05	62.55	58.25
			55.06

Table 2: Task 2 (Classification)

LVLM	Graph type													
	acyclic		cyclic		bipartite		complete		mesh		planar		tree	
	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot
claude-3.5-sonnet	<b>70.00</b>	50.00	90.00	90.00	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	90.00	<b>100.00</b>	<b>100.00</b>	40.00	0.00	<b>90.00</b>	90.00
gemini-pro-1.5	40.00	40.00	10.00	80.00	<b>100.00</b>	<b>100.00</b>	0.00	40.00	90.00	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>90.00</b>	<b>100.00</b>
gpt-4o-2024-08-06	50.00	30.00	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	10.00	30.00	<b>90.00</b>	90.00
llama-3.2-90b-vision-instruct	10.00	10.00	10.00	40.00	0.00	10.00	20.00	20.00	20.00	10.00	30.00	40.00	50.00	50.00
pixtral-12b	0.00	10.00	60.00	50.00	20.00	40.00	50.00	80.00	90.00	50.00	10.00	10.00	60.00	70.00
qwen-2-vl-72b-instruct	30.00	<b>60.00</b>	70.00	50.00	<b>100.00</b>	90.00	90.00	90.00	<b>100.00</b>	<b>100.00</b>	20.00	20.00	70.00	40.00

Table 3: Task 3 (Segmentation)

LVLM	10 nodes		20 nodes		30 nodes	
	CoT	0-shot	CoT	0-shot	CoT	0-shot
claude-3.5-sonnet	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	40.00	30.00
gemini-pro-1.5	<b>100.00</b>	90.00	<b>100.00</b>	<b>100.00</b>	<b>70.00</b>	<b>60.00</b>
gpt-4o-2024-08-06	<b>100.00</b>	<b>100.00</b>	70.00	80.00	50.00	50.00
llama-3.2-90b-vision-instruct	50.00	50.00	30.00	30.00	30.00	30.00
pixtral-12b	50.00	60.00	60.00	80.00	30.00	30.00
qwen-2-vl-72b-instruct	60.00	50.00	60.00	50.00	10.00	20.00

Table 4: Task 4 (Pattern recognition)

LVLM	Pattern type																	
	chain						clique						star					
	2		3		4		2		3		4		2		3		4	
	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot
claude-3.5-sonnet	96.67	96.67	<b>96.67</b>	<b>96.67</b>	88.33	86.67	<b>100.00</b>	<b>100.00</b>	95.00	93.33	<b>100.00</b>	<b>98.33</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
gemini-pro-1.5	<b>98.33</b>	<b>98.33</b>	93.33	93.33	<b>90.00</b>	<b>88.33</b>	<b>100.00</b>	<b>100.00</b>	96.67	96.67	96.67	96.67	<b>100.00</b>	95.00	<b>100.00</b>	<b>100.00</b>	90.00	85.00
gpt-4o-2024-08-06	96.67	95.00	95.00	95.00	83.33	78.33	<b>100.00</b>	<b>100.00</b>	96.67	<b>100.00</b>	91.67	86.67	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	95.00	75.00	85.00
llama-3.2-90b-vision-instruct	55.00	51.67	61.67	55.00	53.33	48.33	56.67	50.00	70.00	60.00	65.00	58.33	40.00	55.00	55.00	40.00	55.00	50.00
pixtral-12b	91.67	83.33	83.33	78.33	80.00	71.67	65.00	66.67	66.67	56.67	70.00	66.67	<b>100.00</b>	95.00	90.00	85.00	90.00	95.00
qwen-2-vl-72b-instruct	<b>98.33</b>	90.00	88.33	88.33	75.00	86.67	91.67	95.00	<b>98.33</b>	98.33	80.00	81.67	90.00	80.00	60.00	85.00	55.00	45.00

Table 5: Task 5 (Link Prediction)

LVLM	4 nodes		5 nodes		6 nodes	
	CoT	0-shot	CoT	0-shot	CoT	0-shot
claude-3.5-sonnet	90.00	<b>95.00</b>	<b>90.00</b>	<b>90.00</b>	65.00	<b>65.00</b>
gemini-pro-1.5	<b>100.00</b>	<b>95.00</b>	<b>90.00</b>	75.00	66.67	50.00
gpt-4o-2024-08-06	95.00	<b>95.00</b>	85.00	70.00	60.00	60.00
llama-3.2-90b-vision-instruct	55.00	65.00	60.00	75.00	60.00	55.56
pixtral-12b	70.00	55.00	75.00	60.00	<b>75.00</b>	55.00
qwen-2-vl-72b-instruct	65.00	75.00	75.00	60.00	60.00	60.00

Table 6: Task 6 (Reasoning)

LVLM	Shortest path (Jaccard similarity)					
	5 nodes		6 nodes		7 nodes	
	CoT	0-shot	CoT	0-shot	CoT	0-shot
claude-3.5-sonnet	67.83	63.83	66.83	<b>67.00</b>	68.50	<b>72.00</b>
gemini-pro-1.5	66.67	59.17	59.17	63.50	<b>75.64</b>	67.71
gpt-4o-2024-08-06	68.17	<b>72.17</b>	<b>68.67</b>	63.67	66.83	58.67
llama-3.2-90b-vision-instruct	58.50	31.50	54.00	43.67	44.50	50.83
pixtral-12b	65.17	53.83	59.00	58.67	61.00	59.00
qwen-2-vl-72b-instruct	<b>73.50</b>	64.67	52.33	50.17	74.00	57.50

Table 7: Task 7 (Matching)

LVLM	match						no match					
	4 (nodes)		5 (nodes)		6 (nodes)		4 (nodes)		5 (nodes)		6 (nodes)	
	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot	CoT	0-shot
claude-3.5-sonnet	40.00	50.00	40.00	50.00	<b>90.00</b>	90.00	<b>100.00</b>	<b>100.00</b>	80.00	80.00	80.00	80.00
gemini-pro-1.5	50.00	30.00	50.00	50.00	70.00	70.00	80.00	90.00	60.00	70.00	60.00	60.00
gpt-4o-2024-08-06	<b>60.00</b>	<b>80.00</b>	<b>60.00</b>	<b>70.00</b>	80.00	<b>100.00</b>	90.00	80.00	50.00	50.00	70.00	60.00
llama-3.2-90b-vision-instruct	30.00	10.00	30.00	20.00	30.00	10.00	90.00	40.00	<b>100.00</b>	90.00	<b>100.00</b>	80.00
pixtral-12b	0.00	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
qwen-2-vl-72b-instruct	0.00	0.00	20.00	10.00	20.00	0.00	80.00	90.00	<b>100.00</b>	90.00	<b>100.00</b>	<b>100.00</b>