

MCN Research Log — Technical Progress Report (Complete)

Mycelial Council Network v0.1 | Runs 1-10 + Ablations | February 2026

Central Finding (Phase 3): Temperature is the dominant variable. Single-agent T=0.3 achieves 91% — best overall. MCN-LinUCB (homogeneous T=0.3) achieves 86% (-5pp). MCN-GNN (heterogeneous T=0.1/0.5/0.9) achieves 88% = T=0.9 single agent. T_2 achieves 95% in isolation (within-run), giving a -7pp routing cost vs an oracle. No specialisation in any run (chi-sq p>0.73).

Project	Mycelial Council Network (MCN) v0.1
Hardware	Single GPU workstation · Docker Compose stack
Model	Qwen/Qwen2.5-Coder-7B-Instruct-AWQ (4-bit AWQ)
Phase 1	Runs 1-6: GNN/LinUCB on 12-task benchmark
Phase 2	Runs 7-8: Expanded to 16-task benchmark (exploratory)
Phase 3	Runs 9-10 + 3 ablations: Controlled temperature comparison
Central finding	T=0.3 single agent (91%) beats all MCN variants (86-88%). Temperature dominates routing.
Date	2026-02-23

All Experimental Runs (1-10)

Ru n	Router	Task set	Pass%	fibonacci	Routing T0/T1/T2	Key finding
1	GNN	12	83%	0% (false neg)	10/69/21	Infra bugs: unbounded Hypothesis, sys.maxsize, PATCH_MIN=2
2	GNN	12	90%	0% (false neg)	33/48/19	Partial fix; reference_solution channel inactive
3	GNN	12	91%	0% (false neg)	33/48/19	Partial fix; Hypothesis still unbounded
4	GNN	12	98%	75%	39/13/48	All infra defects fixed; GNN explores all tribes
5	GNN*	12	98%	75%	10/8/82	Inherited GNN state; eps near floor; T2 dominant (82%)
6	LinUCB	12	97%	62%	100/0/0	Cold-start routing collapse; degenerate single-tribe system
7-8	LinUCB	16	84.5%	n/m	mixed	Exploratory; epsilon-restoration bug; ref-bugs for 3 new tasks
9	LinUCB	16	86%	40%	71/13/16	Post-fix; epsilon warm-up working; 5pp BELOW single-agent T=0.3
10	GNN	16	88%	60%	66/13/21	Hetero T=0.1/0.5/0.9; MCN=T=0.9 single; T2 isolation=95%; no specialisation (p=0.76)

Phase 3: Five-Condition Comparison

Condition	Temp	Pass%	fibonacci	search_insert	unique_paths	vs MCN-LinUCB	vs MCN-GNN
T=0.1 single agent	0.1	86%	40%	25%	n/m	=0pp	-2pp
T=0.3 single agent	0.3	91%	80%	100%	100%	+5pp	+3pp
T=0.9 single agent	0.9	88%	40%	n/m	100%	+2pp	=0pp

Condition	Temp	Pass%	fibonacci	search_insert	unique_paths	vs MCN-LinUCB	vs MCN-GNN
MCN-LinUCB (Run 9)	0.3x3	86%	40%	n/m	n/m	—	-2pp
MCN-GNN (Run 10)	0.1/0.5/0.9	88%	60%	n/m	n/m	+2pp	—

MCN Variants vs Best Single Agent

Metric	T=0.3 Solo	MCN-LinUCB	MCN-GNN
Pass rate	91%	86% (-5pp)	88% (-3pp)
fibonacci	80%	40% (-40pp)	60% (-20pp)
Routing dist.	—	71/13/16	66/13/21
Chi-sq. p	—	0.737	0.762
Specialisation?	—	No	No
T_2 isolation	88% (T=0.9)	n/a	95% (within-run)
Oracle gap	—	n/a	-7pp vs T_2 solo
Verdict	BEST	-5pp vs best	=T=0.9 solo

Defect Log (All Phases)

Defect	Phase	Impact
sys.maxsize in adversarial test B6	1 (Run 1)	fibonacci hang; masked true cause
st.integers() unbounded in Hypothesis	1 (Runs 1-3)	fibonacci 0%; corrupted rewards for 3 runs
PATCH_MIN_ATTEMPTS=2 hardcoded	1 (Run 1)	0 patches stored; hints inactive
permutations reference returns tuples	2 (Runs 7-8)	permutations 0%; tests expect list[list[int]]
unique_paths math.comb crash on m=0/n=0	2 (Runs 7-8)	overseer edge case raises ValueError
epsilon restoration overwrites config warm-up	2 (Runs 7-8)	97/1/2 routing collapse after restart
invert_dict spec-oracle contradiction	1 (Run 1)	8% pass-rate loss from bad task spec
has_cycle: model capability limit	2-3 (all)	0% across all runs; Kahn's reference ineffective for 7B

Current State and Next Steps

Overall finding	Temperature dominates routing. T=0.3 optimal (91%). Routing adds no benefit with same-model tribes.
MCN-LinUCB (R9)	86%; 71/13/16; chi-sq p=0.74; 5pp BELOW single-agent T=0.3
MCN-GNN (R10)	88%; 66/13/21; chi-sq p=0.76; T_2 isolation=95%; oracle gap=-7pp
T=0.3 ablation	91% — best overall; fibonacci 80%; search_insert 100%; unique_paths 100%
T=0.9 ablation	88% — matches MCN-GNN; fibonacci 40%; unique_paths 100%
T=0.1 ablation	86% — matches MCN-LinUCB; search_insert 25% (deterministic failure)
has_cycle	0% in ALL conditions — 7B model capability limit
fibonacci	T=0.3 best (80%); T_2 in MCN-GNN improved to 60% via high-variance search
Patch registry	GNN runs: 279-377 patches (ChromaDB); LinUCB runs: 86-97 patches (in-memory)
Next step A	True model diversity: Route between qualitatively different models
Next step B	Scale: n=500+ tasks, 20+ task types for GNN specialisation signal
Next step C	Adaptive temperature: Continuous router output, not fixed per-tribe

Complete edition — Runs 1-10 + 3 ablations | MLflow tracked | Raw data: mcn-results:/results/runs.jsonl | Feb 2026