



基於多代理人 LLM 之 社群動態模擬系統

以人格化代理人、LLM 決策與動態斷交重組機制，模擬同溫層形成與社群兩極化

指導教授：李政德 教授

專題學生：陳峻霖

研究動機

現有模型的痛點

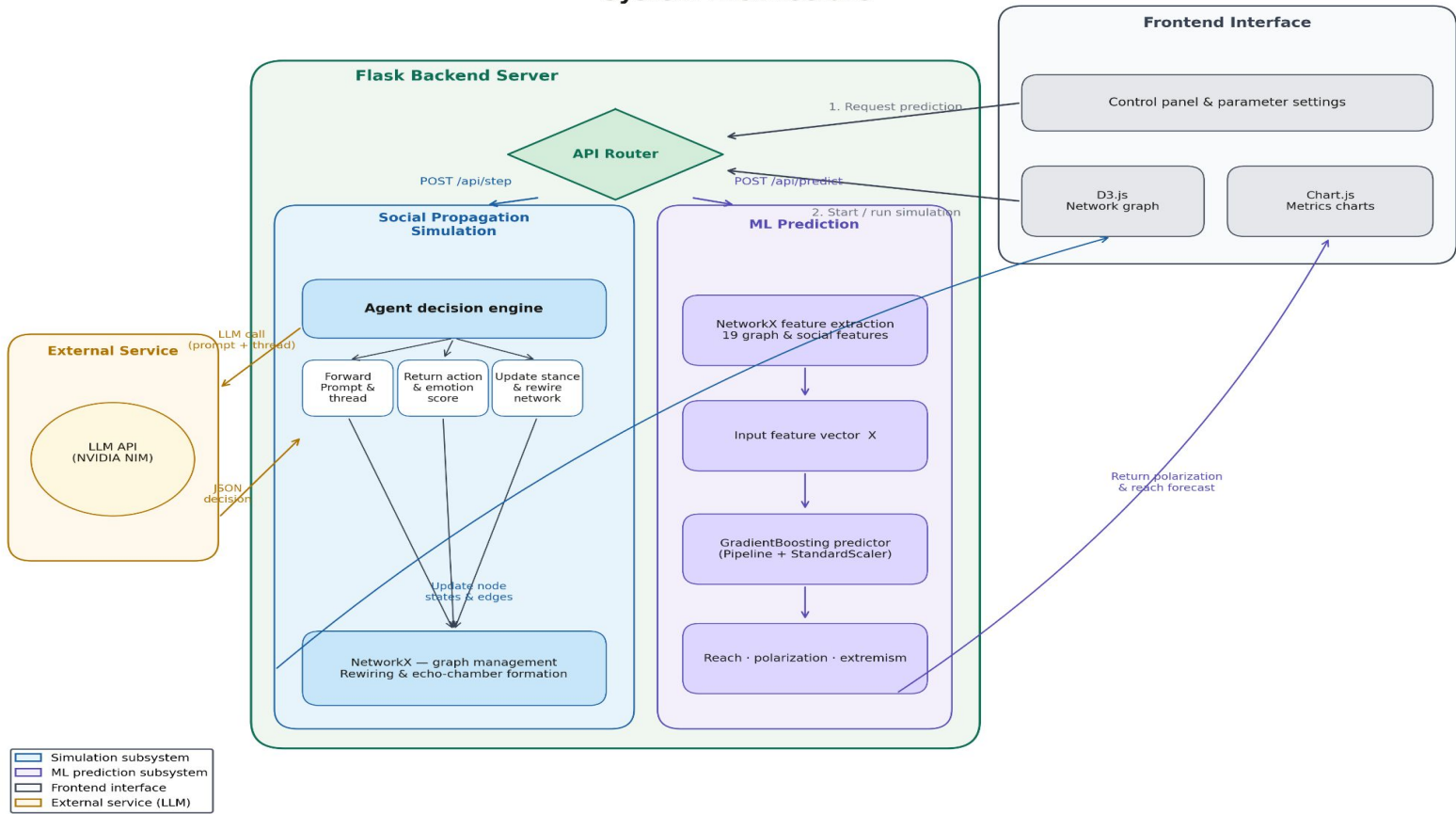
- IC Model / HT-ABM 缺乏語意感知
- 傳播決策為硬性指標，無個人差異
- 無法模擬情緒性反應（逆火效應）
- 缺乏社群斷交重組（同溫層形成）機制

本專題提案

- 導入 LLM 賦予 Agent 語意理解能力
- 加入固執度、易怒值等非理性人格特徵
- SHARE / REWRITE / IGNORE 三種決策
- 動態斷交重組（Network Rewiring）機制

系統架構

System Architecture



社群網路的建立 - BA 無尺度網路模型

BA 網路示意圖



Agent 決策機制

Agent 特徵

- 政治立場 (0-10)
- 固執度 (0-1)
- 易怒值 (0-1)
- 職業 / 背景

Prompt
+ Thread

LLM Agent

JSON 輸出
action · new_stance ·
emotion

SHARE

認同轉發
立場微幅收斂

REWRITE

反嗆改寫
可能觸發斷交重組

IGNORE

忽略訊息
立場不變

量化指標

1. **Polarization (兩極化指數)**: 參考 Hegselmann & Krause 經典論文, 直接抓整個網路立場的 **Standard Deviation**來衡量社會撕裂度。
2. **Group Consensus (群體共識度)**: 與極化成反比。

$$Polarization(\sigma) = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - \mu)^2}$$

N 為節點總數, S_i 為節點 i 的當前立場, μ 為全網平均立場

$$Consensus = \frac{1}{1 + \sigma}$$

量化指標



3. 同溫層係數：用來量化 Echo Chamber 效應的強弱，我們參考 Newman 提出的 Assortativity coefficient。

$$r = \frac{\sum_{xy} x y (e_{xy} - a_x b_y)}{\sigma_a \sigma_b}$$

意義：數值越高，代表網路上相連的 nodes 立場越像，也就是「同溫層」越嚴重。

量化指標



4. **Extremist Ratio (極端份子比例)**: 算落入極端區間的節點佔比。

$$ExtremistRatio = \frac{|\{i \in N \mid S_i \leq 2.0 \vee S_i \geq 8.0\}|}{N}$$

5. **Average Attitude Shift**: 量化演算法造成的「洗腦/說服」程度。

$$\Delta S_{avg} = \frac{1}{N} \sum_{i=1}^N |S_i^{(t)} - S_i^{(0)}|$$

量化指標



6. Average Emotion Score: 追蹤網路炎上的指數，由llm自我評估1-10 分憤怒值之平均。

$$Emotion_{avg} = \frac{1}{K} \sum_{j=1}^K E_j$$

7.Active Participants Count: 統計做出 SHARE 或 REWRITE 的獨立節點總數，衡量觸及率。

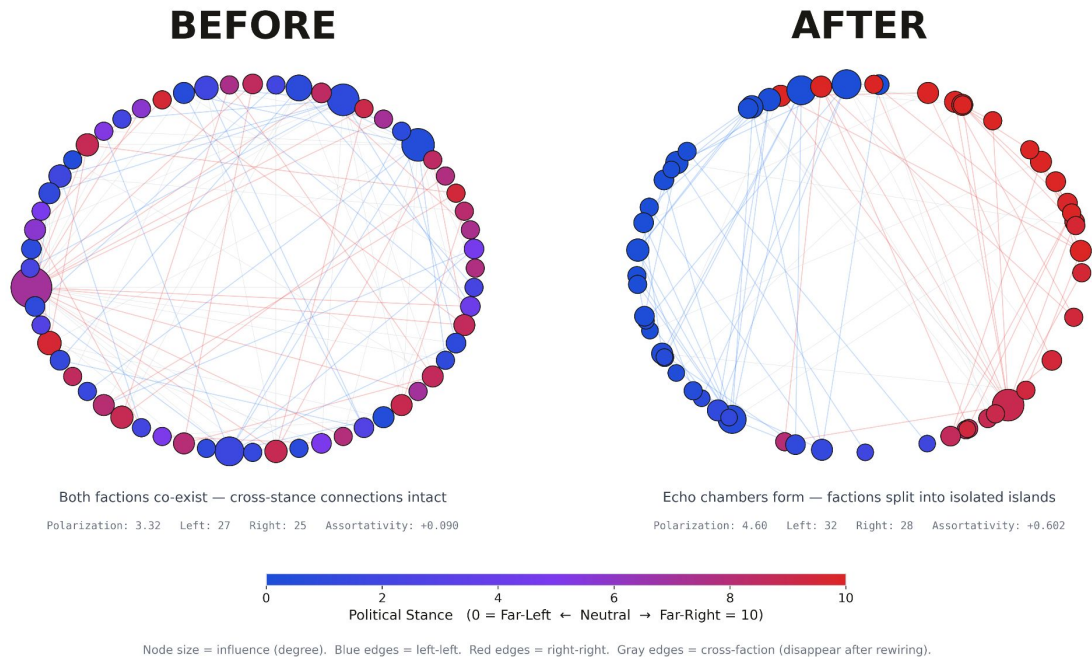
動態斷交重組 (Network Rewiring)

觸發條件

- 決策為 REWRITE 且情緒分數 ≥ 7
- 與鄰居立場差距 > 4.0

執行步驟

- ① 刪除意見不和鄰居的連結
- ② 在立場差距 ≤ 2 的節點中新增連結
- ③ Assortativity (同溫層係數) 上升



推薦演算法



由於現代社群系統會觀察用戶的互動行為——按讚、留言、分享、停留秒數——判斷用戶偏好，然後優先推送「你可能認同」的內容。

推送對象的選擇根據**立場相似度**加權：

$$w_i = \frac{1}{1 + |s_{\text{src}} - s_i|}$$

立場越接近，權重越高，被推送到的機率越大。這模擬了平台「你可能喜歡」的推薦邏輯。

傳播流程



1 訊息注入網路

選定種子節點，注入帶立場的初始訊息



2 鄰居接收

訊息沿邊傳至鄰居，組成各自 Prompt



3 LLM 決策

輸出 SHARE / REWRITE / IGNORE



4 狀態更新

回寫立場與情緒，記錄觸達



5 斷交重組

依立場差異觸發 Rewire



6 指標累計

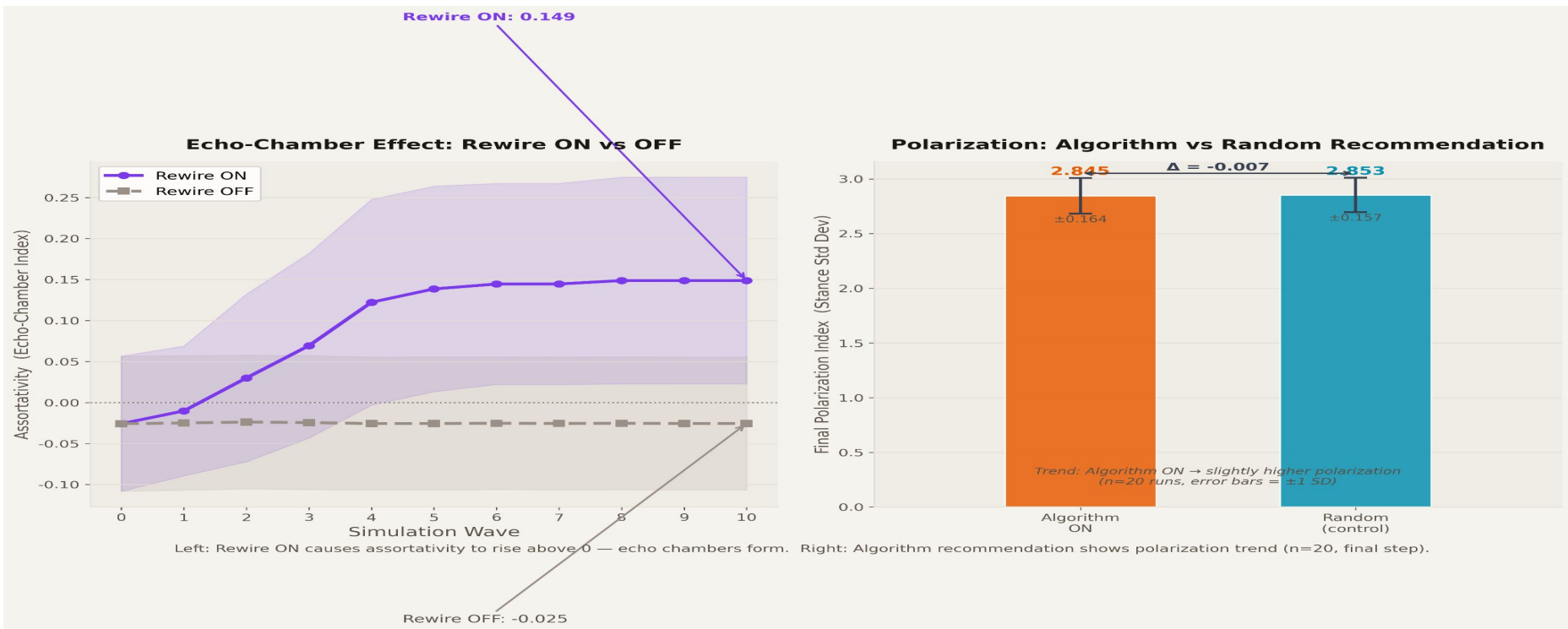
更新觸達率、極化指數、情緒等

↳ 多回合迭代直至收斂或達回合上限

實驗結果 I: 斷交重組的影響

同溫層係數增加率: +126%

Algorithm ON vs OFF · 真實模擬數據

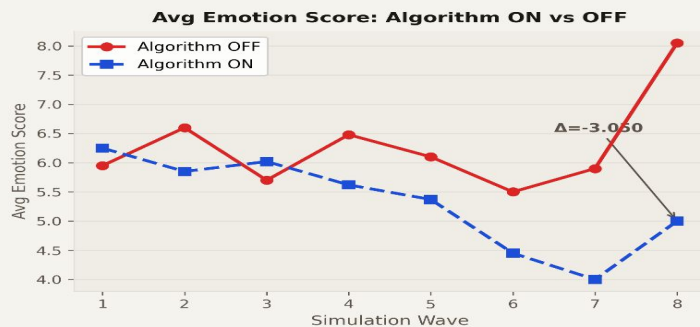
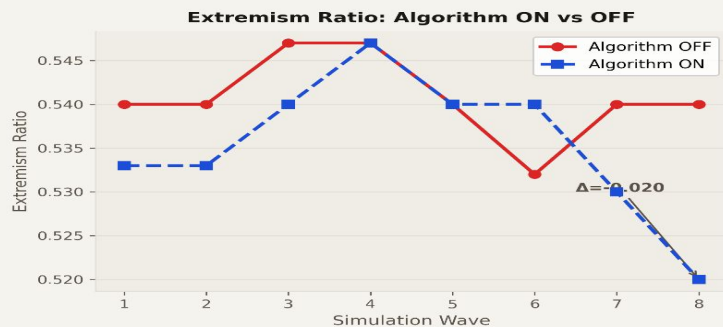
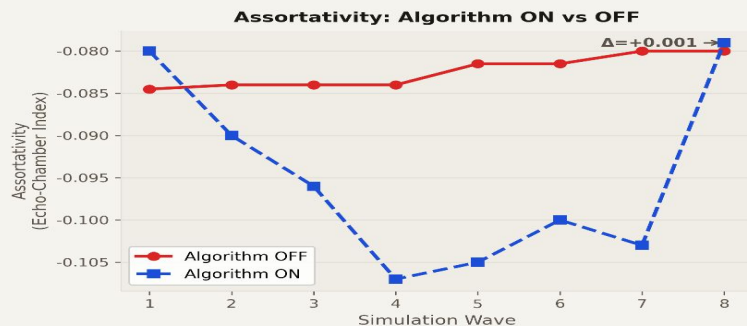
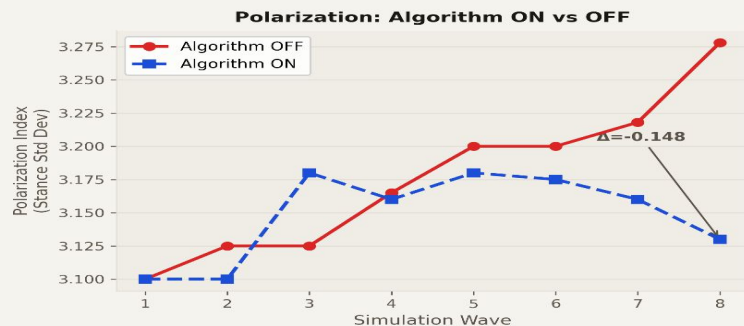


實驗結果 II: 推薦演算法的影響

Algorithm ON vs OFF · 真實模擬數據

推薦演算法使訊息集中在認同者之間傳播，雖短期降低情緒對立 ($\Delta \text{emotion} = -3.06$)，但同步加深社群同質化，形成隱性同溫層效應。

Algorithm Recommendation: Impact on Social Dynamics
(Real simulation data — single run per condition)



模擬前測設定

模型設計

- 演算法：Gradient Boosting Regressor (梯度提升迴歸樹)。
- 輸入：19 項模擬特徵；輸出：觸達率、極化指數 (分別建模)。
- 驗證：5-fold 交叉驗證，避免過度擬合並估計泛化能力。
- 可解釋性：以 Feature Importance 解析主要驅動因子。

$R^2 \approx 0.85$ (5-fold CV)

建模流程

- 300 次模擬資料集
- 特徵擷取 (19 維)
- GradientBoosting 訓練
- 5-fold 交叉驗證
- Feature Importance 分析

系統demo



結論



1. 導入 LLM 建構具語意理解與情緒感知的生成式代理人，解決傳統模型 缺乏個人化互動的問題
2. 結合動態斷交重組機制，量化實證同溫層聚集、演算法隔離與社會極化等複雜輿論現象
3. 整合為具互動性 Web 平台，提供即時可調參數的社群傳播研究工具

未來展望



1. 部署本地開源模型 (Llama 3) 挑戰更大型社群拓樸
2. 加入圖文多模態輸入與五大性格特質 (Big Five)
3. 爬取 PTT / Threads 真實數據進行交叉驗證