

# Влияние графовых представлений на коллективное рассуждение LLM-агентов в игре «Мафия»

Пеганов Никита, Назарько Михаил

2025

## Аннотация

В работе исследуется влияние включения графовых представлений отношений между игроками на успешность коллективных рассуждений LLM-агентов в игре «Мафия». Мы сравниваем различные схемы инъекции графа в промпт и фиксируем статистику побед для разных режимов и LLM-моделей. Результаты свидетельствуют о выигрыше стратегий с графом, основанным на накопленной истории взаимодействий.

## 1 Введение

Современные языковые модели (LLM) демонстрируют способность к рассуждению и коллективному мышлению в игровых и социальных задачах. Одна из ключевых проблем — сохранение и использование сложного контекста взаимодействий. Мы выдвигаем гипотезу, что графовое представление отношений между игроками, построенное из истории обсуждений, может существенно повысить эффективность кооперации и качество принятых решений в полукооперативных играх, таких как «Мафия».

## 2 Методология

Проект основан на форке репозитория `llm-mafia-game`, адаптированного для запуска с локальными моделями (`vllm`). Все агенты в эксперименте — клоны одной LLM. Эксперименты симулируются батчами по 80 партий на различных моделях и с разными видами графовых подсказок. Используется автоматизированная система логирования и сбора результатов на платформе Firebase, а для анализа предусмотрен веб-дэшборд.

## 3 Модели и алгоритмы

Для экспериментов использовались различные open-source LLM: `gryphe/mythomax-12-13b`, `mistralai/mistral-small-24b-instruct-2501`, `deepseek/deepseek-llm-7b-chat` и др. Графы строит отдельная аналитическая LLM, анализируя диалоги и действия игроков. Рассматривались следующие типы графов:

- **Коммуникационный граф** — связи между игроками на основе явных взаимодействий (обвинения, защита, голосования)
- **Граф текущего раунда** — только связи в рамках одного круга
- **Глобальный граф** — агрегация всей истории взаимодействий и событий

Recent Games				
Game ID	Winner	Participants	Language	Timestamp
<a href="#">05337dc6-749e-4980-8472-c14d5aeded97</a>	Villagers	mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Doctor), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager)	English	6/26/2025, 6:13:22 AM
<a href="#">7d9aef1a-da2f-44fc7-b599-35099e36e54d</a>	Villagers	mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Doctor), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia)	English	6/26/2025, 6:11:08 AM
<a href="#">0e569b3c-0e30-4b1f-a3f0-af1c0d554c75</a>	Villagers	mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Doctor), mythomax-l2-13b (Villager)	English	6/26/2025, 6:07:48 AM
<a href="#">4ddff938-bca4-4e90-9d7f-bf2a9bba1265</a>	Mafia	mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Doctor), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia)	English	6/26/2025, 6:02:41 AM
<a href="#">4cf20fc3-ca73-44f3-96af-3aa89514a5a8</a>	Mafia	mythomax-l2-13b (Villager), mythomax-l2-13b (Doctor), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia)	English	6/26/2025, 5:56:41 AM
<a href="#">59ec3f8f-2dd9-49af-82a8-af0232201b7a</a>	Mafia	mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Doctor), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager)	English	6/26/2025, 5:53:00 AM
<a href="#">c222674f-0f11-4f8b-8639-797d080afd90</a>	Mafia	mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Doctor), mythomax-l2-13b (Villager)	English	6/26/2025, 5:47:02 AM
<a href="#">cbd60304-47e5-48cd-8f11-0052a91a6734</a>	Villagers	mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Doctor), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager)	English	6/26/2025, 5:42:47 AM
<a href="#">ae4968c7-1b46-4013-a717-6ed01db51dc6</a>	Mafia	mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Villager), mythomax-l2-13b (Mafia), mythomax-l2-13b (Mafia), mythomax-l2-13b (Villager), mythomax-l2-13b (Doctor)	English	6/26/2025, 5:40:54 AM

Рис. 1: Веб-дэшборд для мониторинга партий, статистики и истории

```

# Communication Graph
Reese -> suspicion -> Kennedy
Quinn -> goal -> win
Winter -> disagreement -> Reese
Oakley -> action -> vote
Ellis -> action -> vote
Gray -> alliance -> Winter
Winter -> alliance -> Gray
Kennedy -> suspicion -> Quinn

# Current Round Graph
Oakley -> vote -> Kennedy
Cameron -> vote -> Kennedy
Kennedy -> suspicion -> Quinn
Winter -> question -> Reese
Reese -> vote -> Kennedy
Winter -> vote -> Kennedy
Ellis -> accusation -> Kennedy
Gray -> support -> Winter
Quinn -> action -> vote

# Global Graph
Quinn -> total_votes -> Kennedy [count: 3]
Kennedy -> total_suspicious -> Multiple [count: 4]
Gray -> support_sequence -> Winter [days: 1,2,3]
Cameron -> accused -> Kennedy [rounds: 1, 2]
Ellis -> changed_vote -> Reese -> Kennedy [on round 2]
Winter [status: alive, activity: high]
Kennedy [status: eliminated, role: Mafia]
Reese -> pattern -> "frequently hesitates"
Quinn -> trust -> Oakley [established: round 1]

```

Рис. 2: Примеры различных типов графов в игре

## 4 Результаты

```
[VILLAGER GRAPH for Parker]:
Noel -> protecting -> Finley
Gray -> suspicious -> Quinn
└ Parker (Villager)
  *Thinks* I agree with Noel that Quinn and Finley seem suspicious. However, I'm not sure about protecting Finley. It's possible he's just playing it safe
  and not saying much. I think we should focus on eliminating Quinn tonight. If we're wrong, we can always protect Finley next time. VOTE: Quinn

ACTION: Parker (Villager) - Vote Quinn

[VILLAGER GRAPH for Finley]:
Noel -> suspicious -> Quinn
Noel -> trust -> Finley
Noel -> protect -> Finley
Gray -> suspicious -> Quinn
Parker -> suspicious -> Quinn
Parker -> vote -> Quinn
└ Finley (Villager)
  *Thinks* Well, this is interesting. I'm not sure what to make of all this. I trust Noel and think he's trying to help us. But I also agree that Quinn see
  ms suspicious. I guess we'll just have to see how this plays out. VOTE: Quinn

ACTION: Finley (Villager) - Vote Quinn
└ Frankie (Villager)
  *Thinks* I agree with Parker. Quinn seems like the more suspicious of the two. We can always protect Finley if needed. VOTE: Quinn

ACTION: Frankie (Villager) - Vote Quinn
└ Kai (Villager)
  *Thinks* Hmm, interesting. I agree that Quinn seems suspicious, but I'm not sure about Finley. It's possible he's just playing it safe and not saying muc
  h. I think we should focus on gathering more information before making any rash decisions. Let's not forget that the Mafia might be trying to throw us off b
  y making us suspicious of each other. VOTE: None

ACTION: Kai (Villager) - Vote Parker (auto-selected)
└ Taylor (Mafia)
  *Thinks* I agree with Kai. We need to gather more information before making any rash decisions. Let's not forget that the Mafia might be trying to manipu
  late us into suspecting each other. I suggest we focus on observing everyone's behavior and gathering more clues during the day phase. VOTE: None

ACTION: Taylor (Mafia) - Vote Gray (auto-selected)
```

Рис. 3: Демонстрация графа отношений, внедрённого в промпт для LLM-агента

Таблица 1 демонстрирует число побед мирных игроков (из 80 партий) для разных моделей и режимов графовой инъекции:

Модель	Без графа	Ком- муник.	Комм. история	Раунд	Раунд история	Глобал. история
mythomax-l2-13b	41	37	48	39	50	55
mistral-small-24b-instruct-2501	39	36	47	38	49	56
deepseek-llm-7b-chat	40	35	45	38	47	54
deepseek-r1-distill-llama-70b	42	38	49	39	51	58
hermes-3-llama-3.1-70b	41	36	46	40	48	55
DeepSeek-R1-Distill-Qwen-32B	40	35	46	39	48	54

Таблица 1: Число побед мирных игроков из 80 партий для разных моделей и режимов графов

## 5 Обсуждение

Результаты показали, что графы без накопления истории малоэффективны: агенты теряют контекст и не могут скоординировать действия. Наилучших результатов удалось достичь при использовании общего глобального графа с накоплением истории — winrate мирных превышает случайный. Граф текущего раунда показал минимальное влияние. Повторение экспериментов автора на локальных моделях подтверждает воспроизводимость подхода и даёт простор для масштабных исследований.

## 6 Заключение

Включение графовых представлений, особенно основанных на всей истории игры, позволяет языковым моделям эффективнее коллективно рассуждать и добиваться более высоких

победных стратегий в социальных играх. Использование локальных моделей и автоматизация пайплайна открывают путь к дальнейшим экспериментам, расширению анализа и исследованию новых сценариев кооперации LLM.