Universität Konstanz

CENTRO RICERCHE ENRICO FERMI

Complexity Science Hub

Emergent Behaviors in LLM Populated Societies

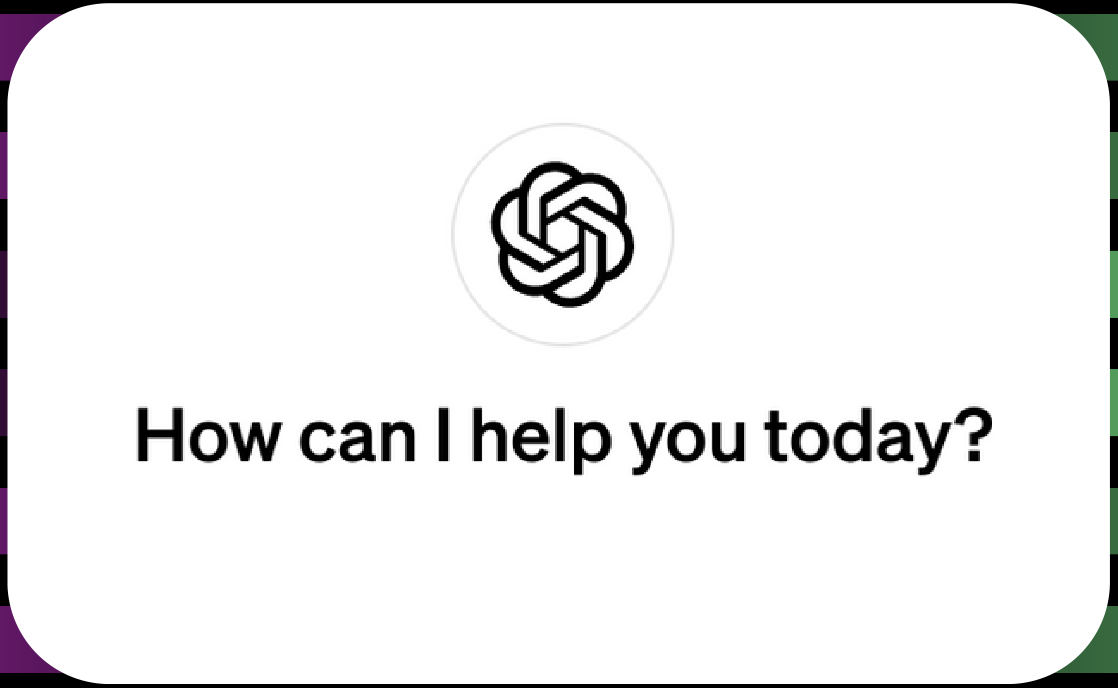
Presentation by

Giordano De Marzo

CSH Seminar

25 October 2024

ChatGPT and LLMs



ChatGPT impact has been huge

- Almost two years old
- Over 200 million weekly active users

There are countless applications

- Coding
- Text writing and editing
- Translation

But there is much more!

Page **2**

Generative Agents

Memory

Agents can be endowed with a memory stream that allows them to remember past actions

Autonomous Agents

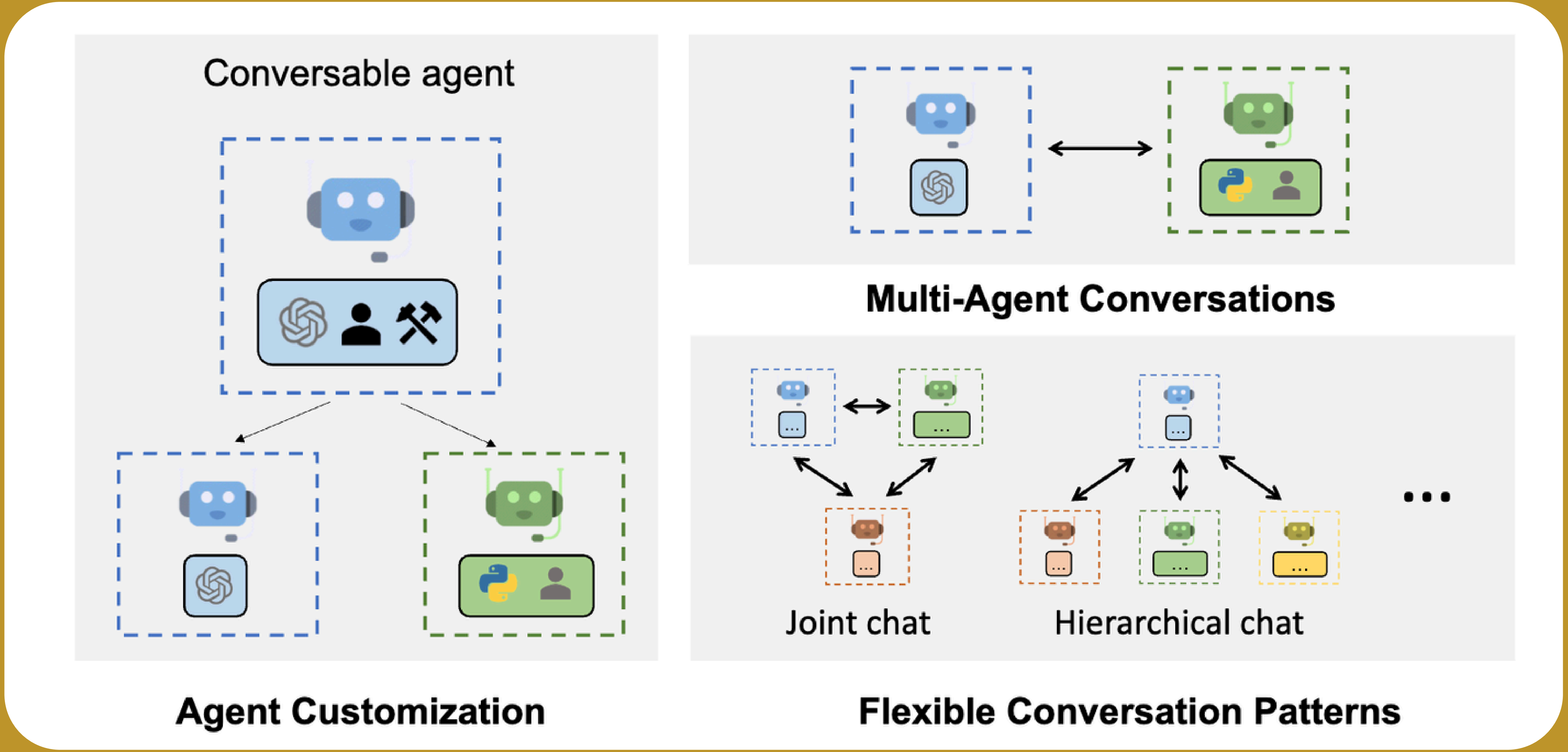
Agents reflect on what they experience and take decision autonomously

Park, Joon Sung, et al. "Generative agents: Interactive simulacra of human behavior." Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. 2023.

Teams of LLMs

Multiple LLMs can work together in a team to solve complex tasks
- each agent can have a different role
- they can access tools (python, search engine ...)
- examples include AutoGPT and AutoGen

Teams of simpler LLMs can outperform more advanced models



Wu, Qingyun, et al. "Autogen: Enabling next-gen llm applications via multi-agent conversation framework." arXiv preprint arXiv:2308.08155 (2023).

LLMs on Devices

AI Devices

Devices such as Rabbit R1 or Humane AI Pin are built around an LLM that can assists the user

AI Assistants

LLMs are revolutionizing AI assistants: Apple-OpenAI and Amazon-Anthropic agreements

LLMs Group Level Behavior

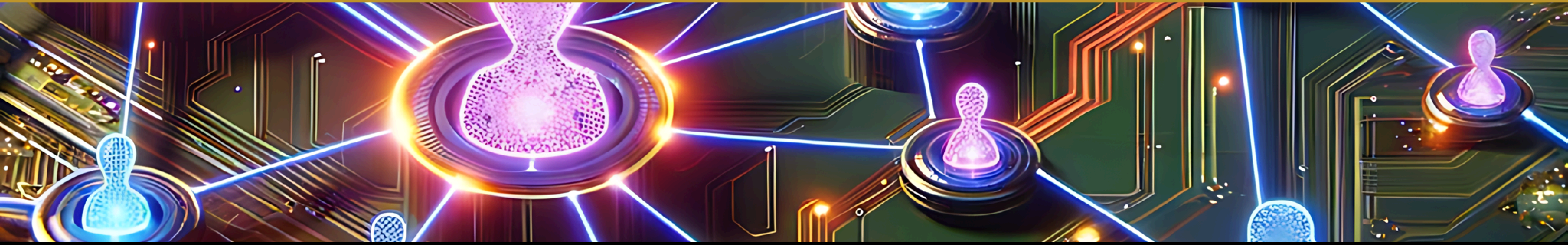
Understanding group behavior is crucial when studying humans:

- societies show emergent behavior that could hardly be derived from individuals' properties
- we approach most problems as a group, not as individuals

What about LLMs?

- do they show emergent group properties?
- can these properties be harmful?
- are groups of LLMs better in problem solving?
- can we use LLMs to simulate humans?

Network Growth



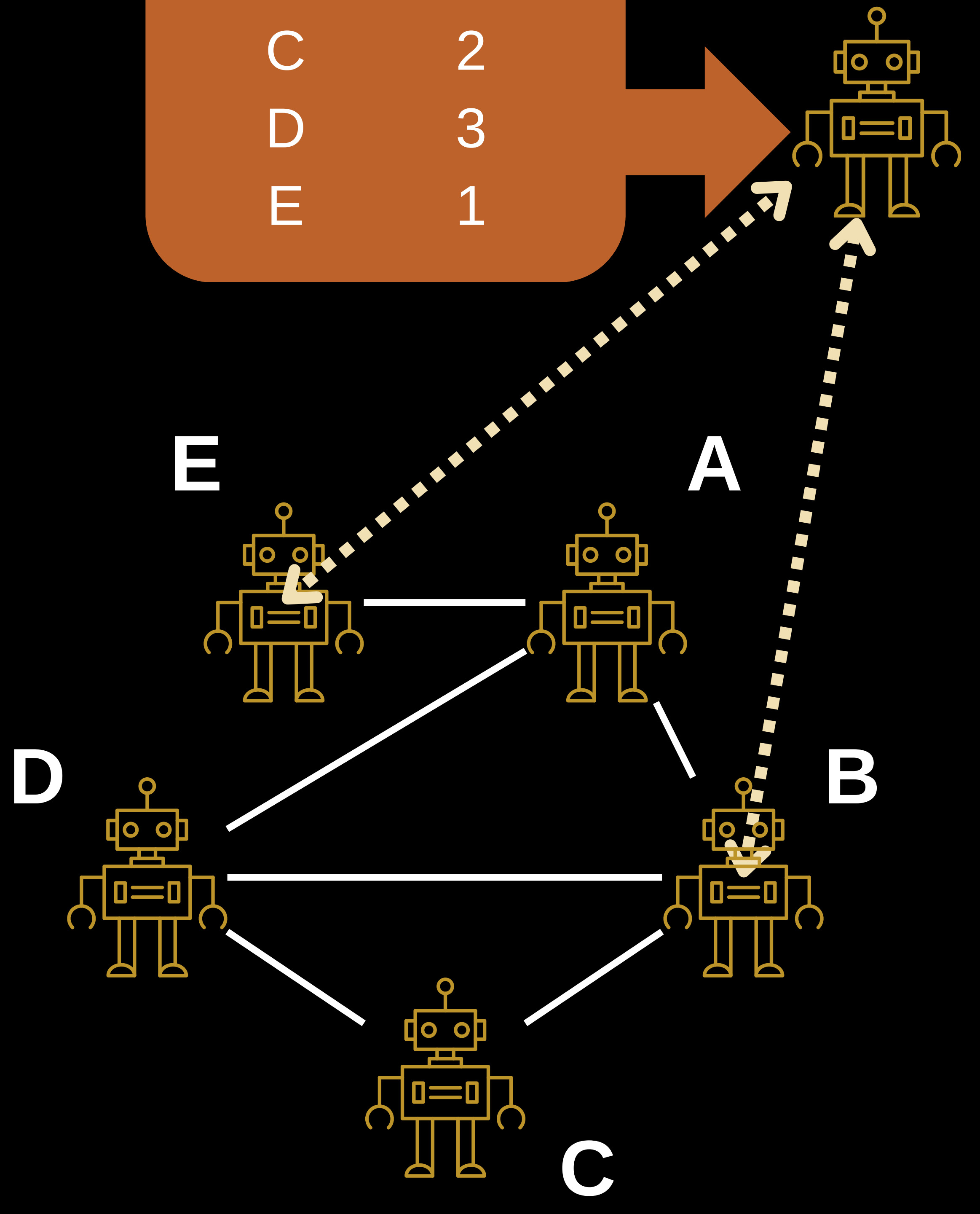
LLMs Social Networks

Barabasi-Albert like process:

- at each time step a new node is added
- it links to m already existing nodes
- the linking probability is not decided a priori
- a LLM decides which connections to establish

We exploit GPT3.5-Turbo as LLM

User	Friends
A	3
B	3
C	2
D	3
E	1



Prompt

- You've entered a virtual social network.
- You're tasked with connecting to exactly {m} individuals from the list below.
- Each individual is accompanied by their current number of connections.
- Please indicate your choices by replying with their names, separated by commas and enclosed within square brackets.

X7v 5

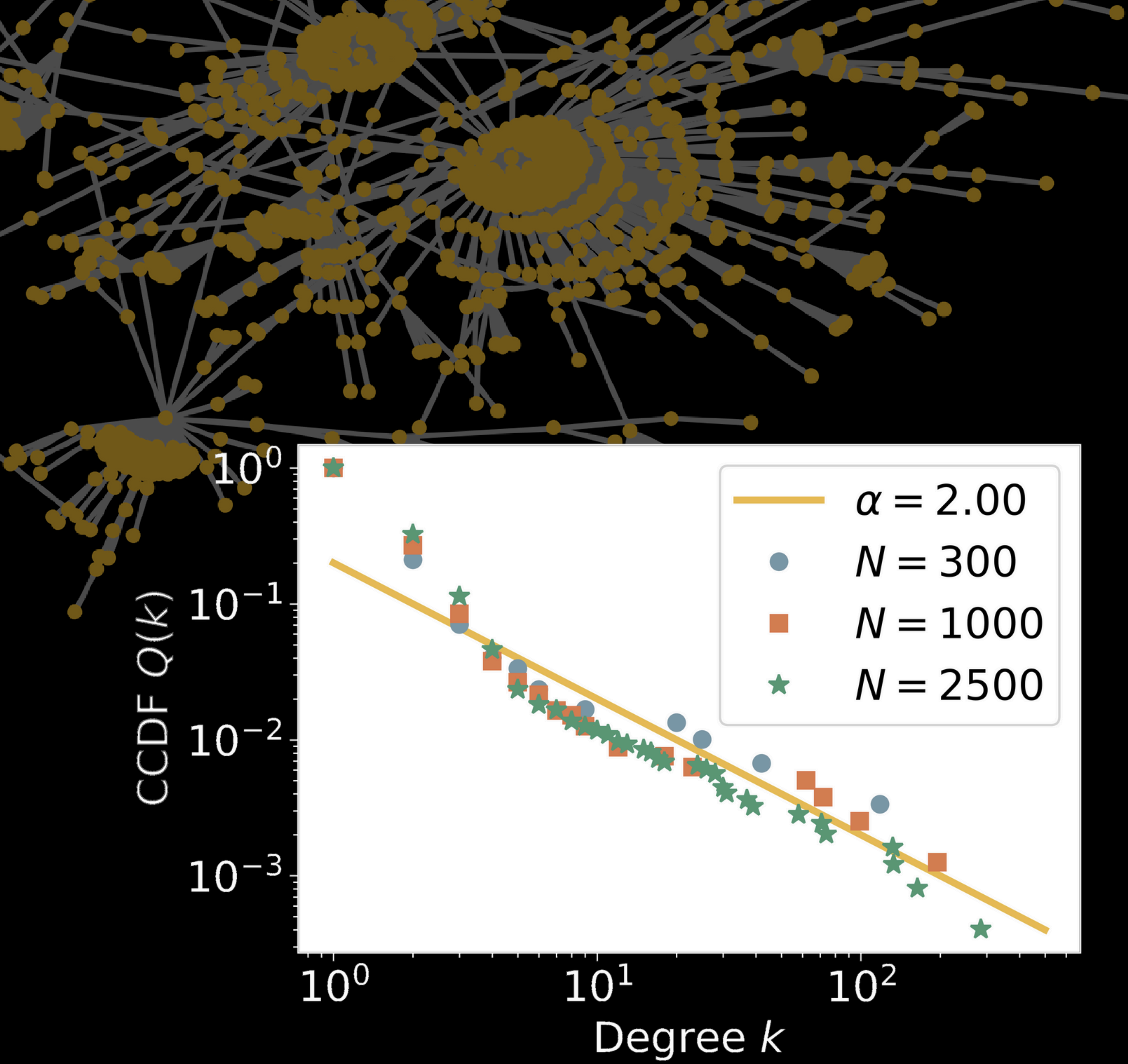
keY 1

91c 17

...

Page **8**

Scale-Free Networks



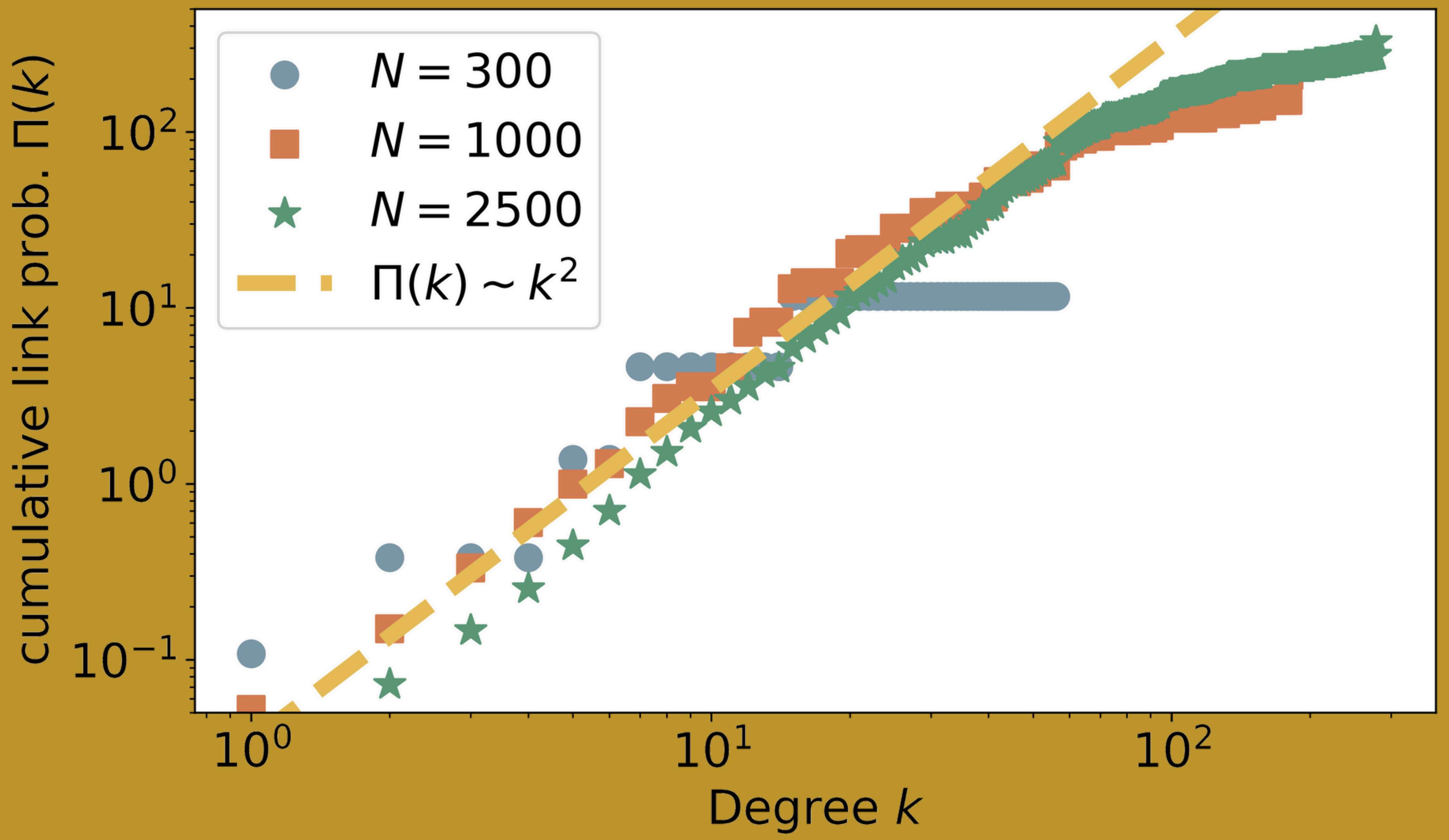
The resulting networks are similar to those formed by humans in social networks

- as the system grows, the degree probability distribution shows a power law tail
- this indicates a scale-free topology

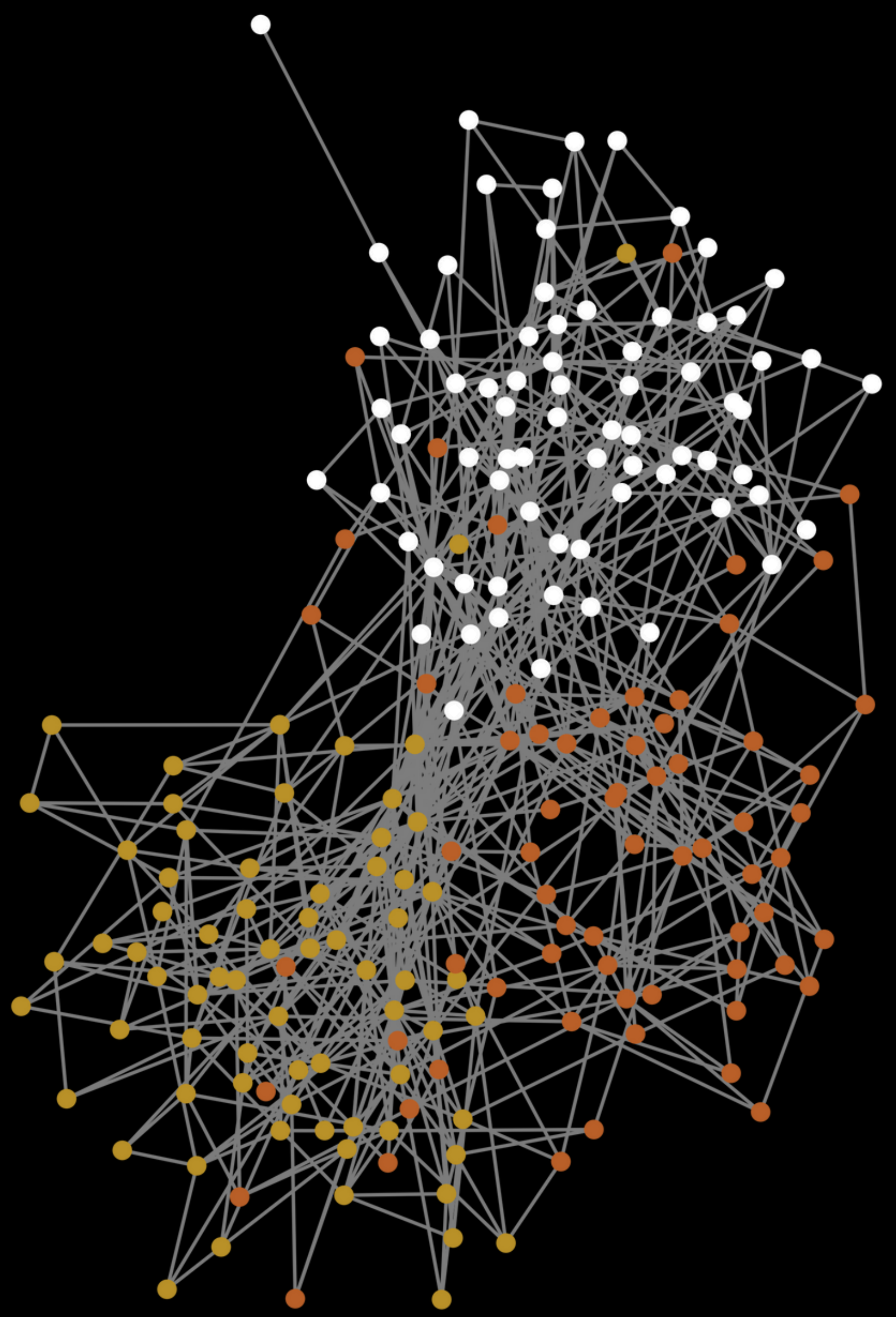
Preferential Attachment

In order to better understand the network growth process we can look at the (cumulative) linking probability.

LLM agents show linear preferential attachment!



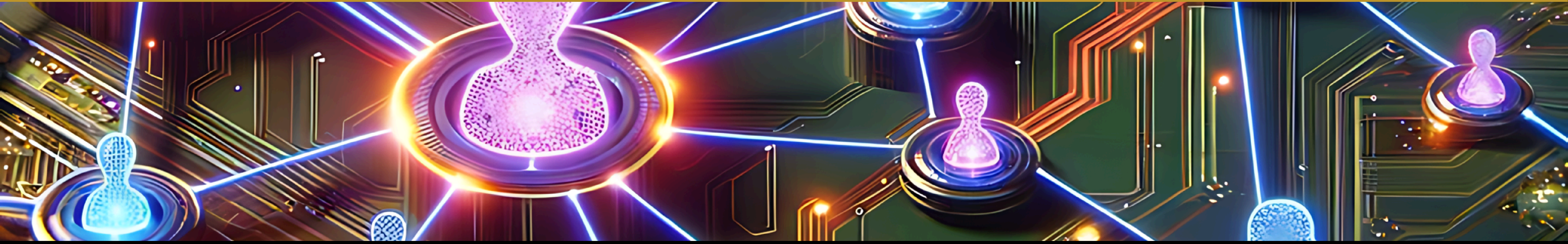
Homophily



Instead of specifying the number of connection we can show agents other features.

When ethnicity, gender or political leaning are shown, communities get formed.

Consensus Formation



The Social Brain Hypothesis

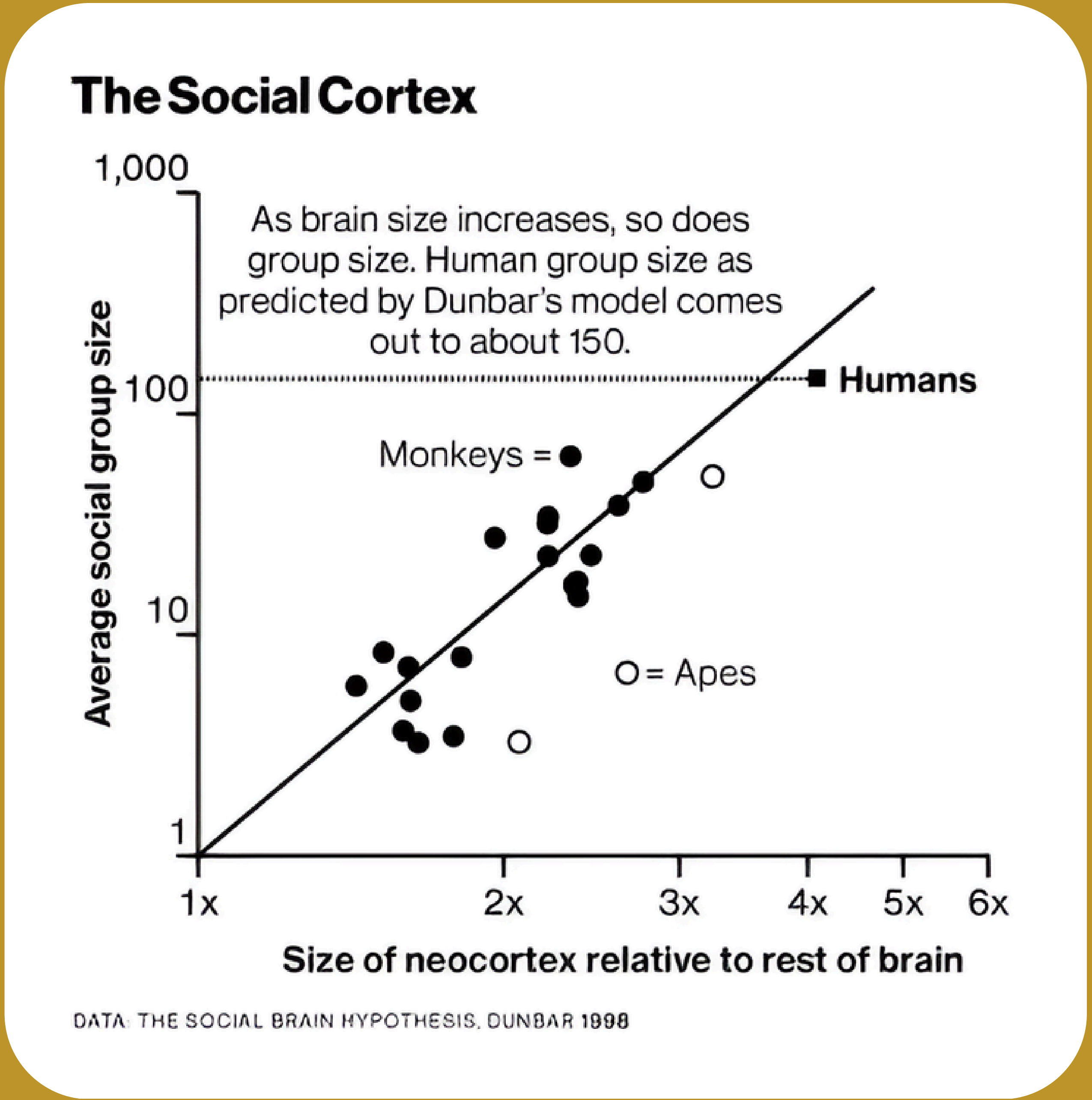
Humans and primates tend group into societies

- their size is intrinsically limited by the dimension of the neocortex
- for humans this leads to a maximal size of around 150 individuals (Dunbar's numbers)

What about LLMs?

- are there intrinsic limits to the size of an LLM populated society?

We answer to this by simulating opinion dynamics and studying if and how consensus emerges

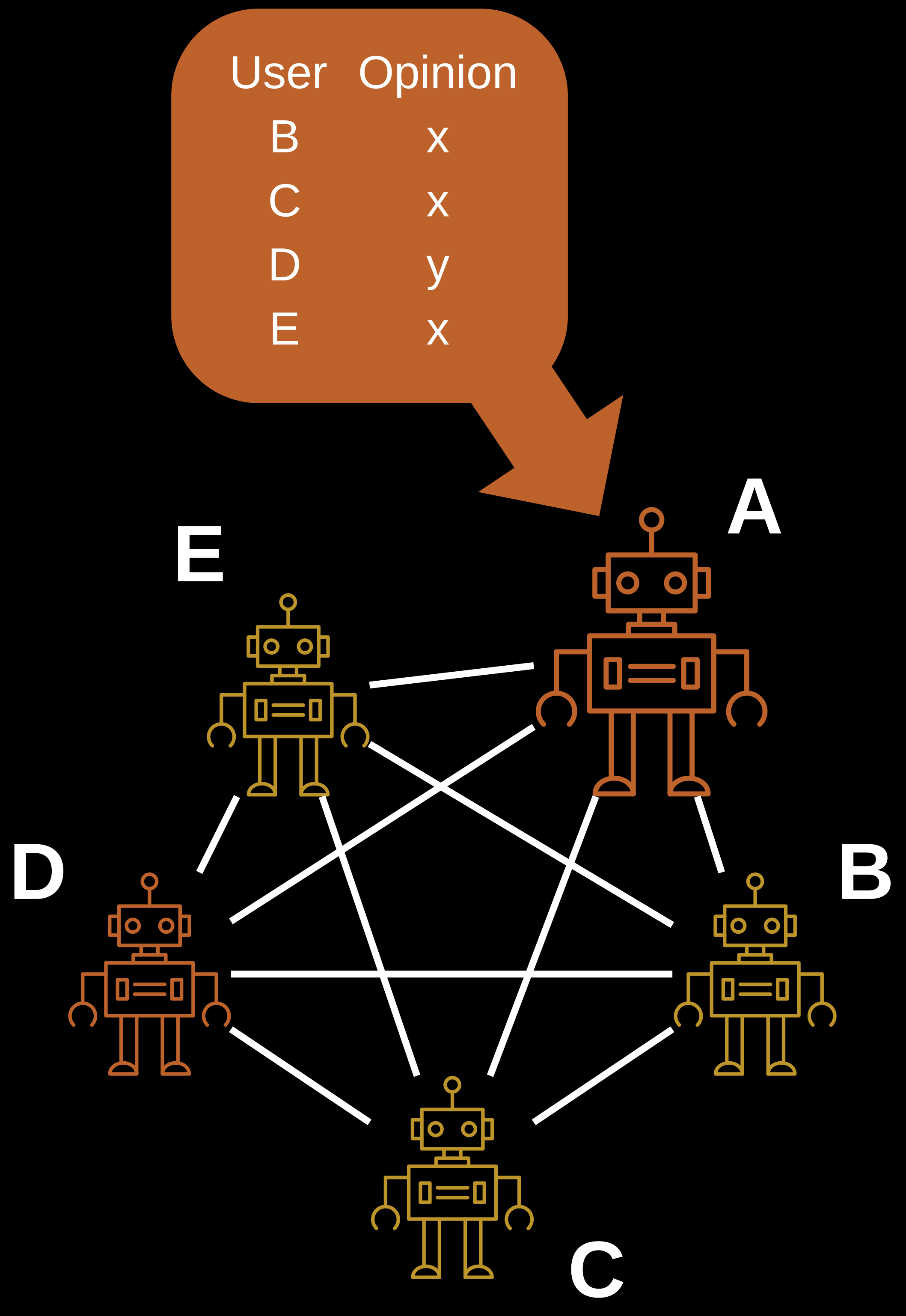


LLMs Opinion Dynamics

Binary Opinion Dynamics Process

- at each time step we select an agent on the network
- we provide it the list of its connections with the opinion they support
- an LLM autonomously decides which opinion to align to

We exploit several different LLMs and we consider a fully connected network.

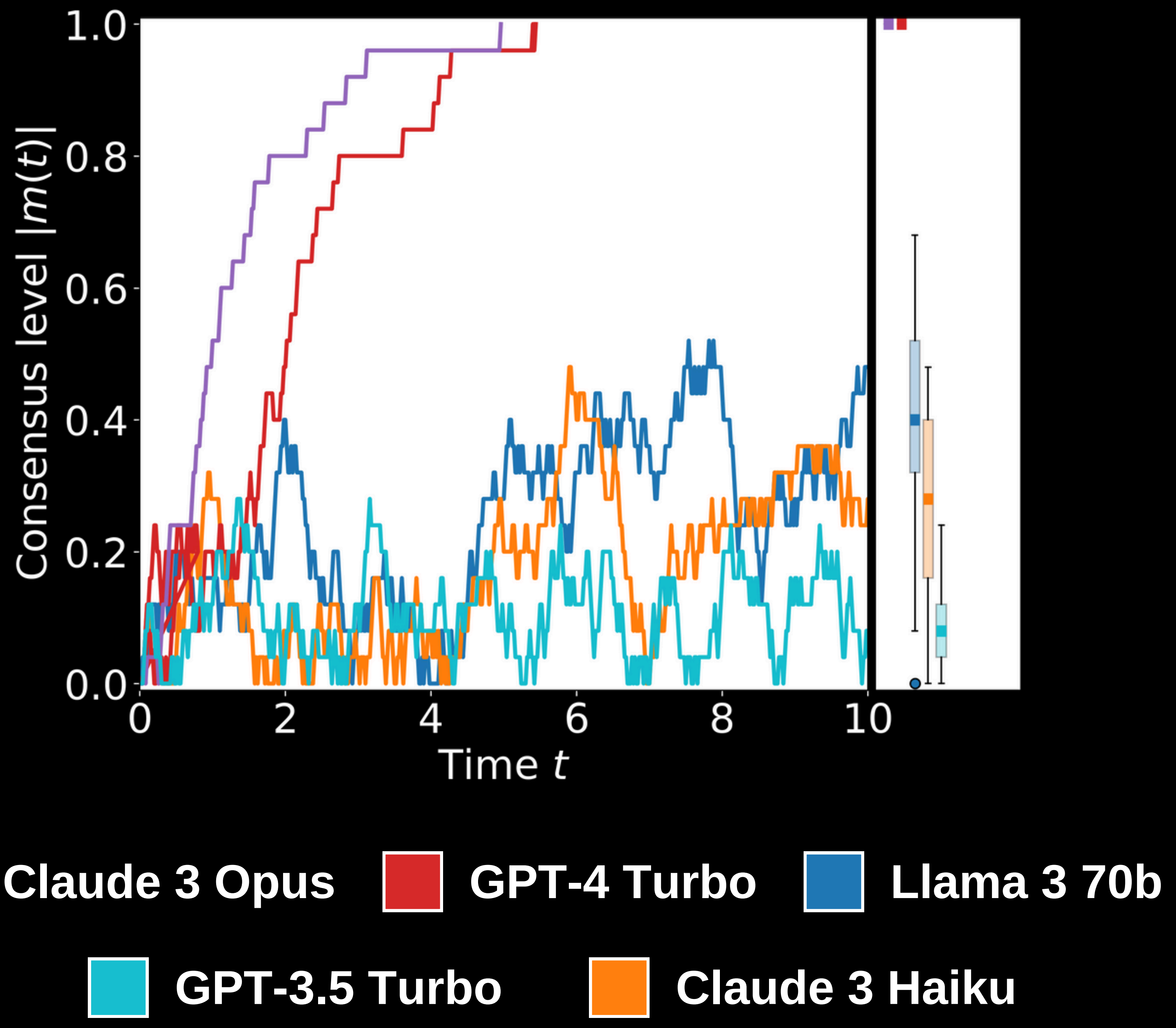


Prompt

- Below you can see the list of all your friends together with the opinion they support.
- You must reply with the opinion you want to support.
- The opinion must be reported between square brakets.

X7v x
keY x
91c y
gew x
41O y
...

Emergence of Consensus



The state of the system is given by the collective opinion m

- we can follow the evolution by looking at the consensus level |m|
- the most advanced models reach consensus in all the runs
- the less advanced models never reach consensus

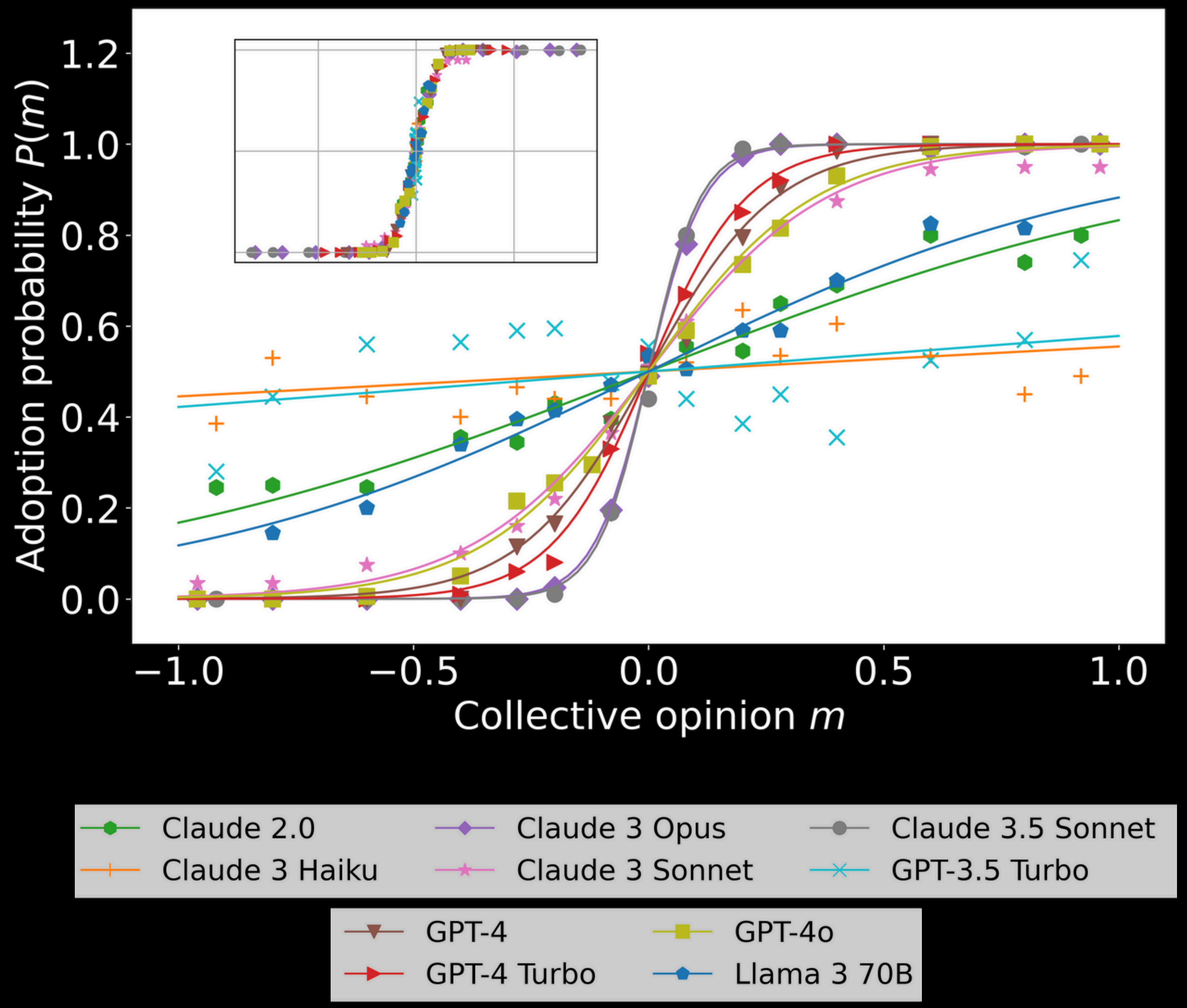
Some LLMs are able to coordinate and reach consensus others are not

Adoption Probability

We can understand the opinion dynamics process looking at the adoption probability

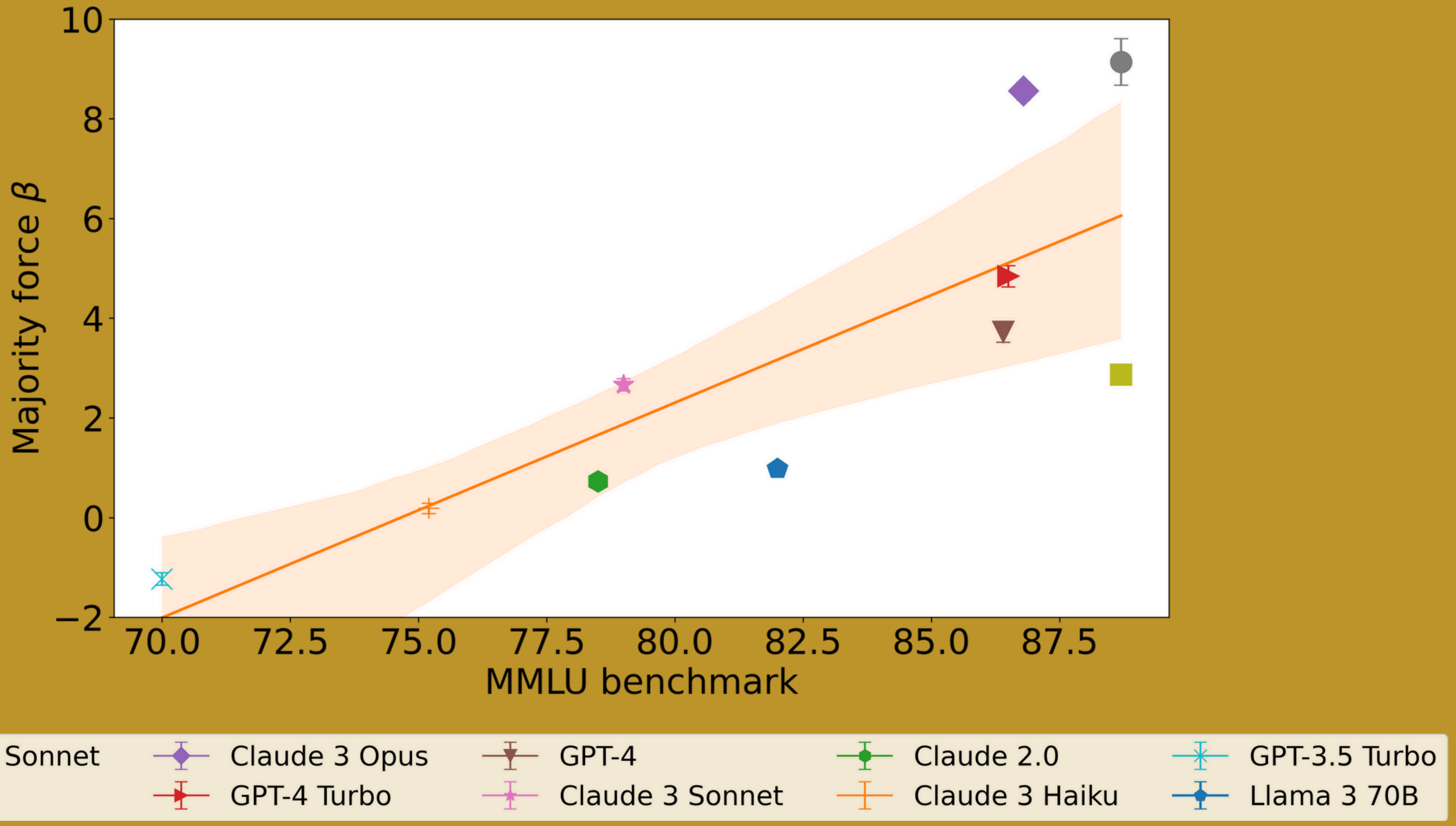
- probability P(m) to choose the first opinion as function of m
- we observe an universal behavior $P(m)=0.5+0.5\cdot\tanh(\beta m)$
- the only difference is in the majority force β

This is the same probability of the Curie-Weiss model!



We compare the majority force with the MMLU benchmark

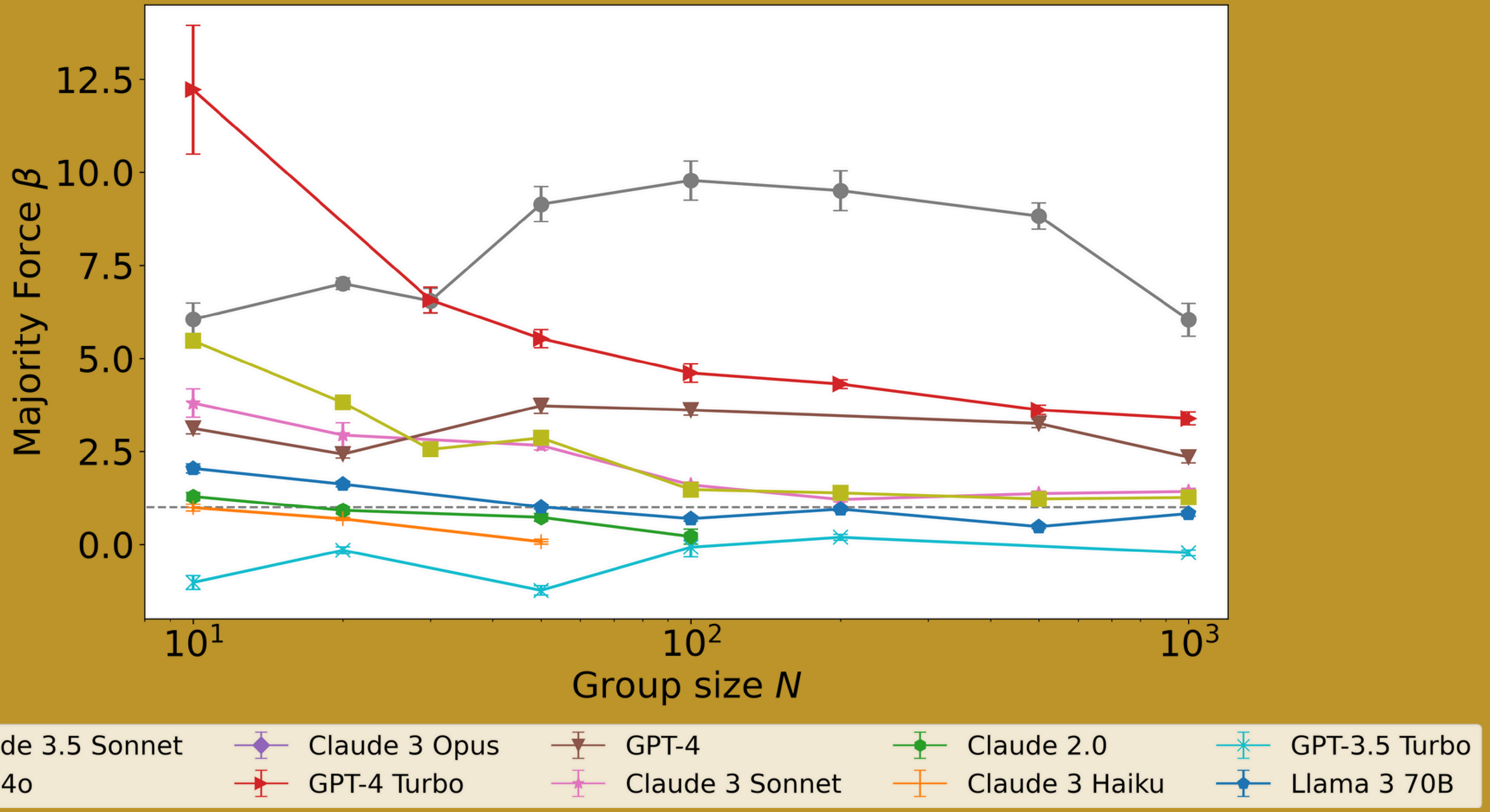
- β is strongly correlated with the language understanding and cognitive capabilities
- advanced models have a stronger majority following tendency

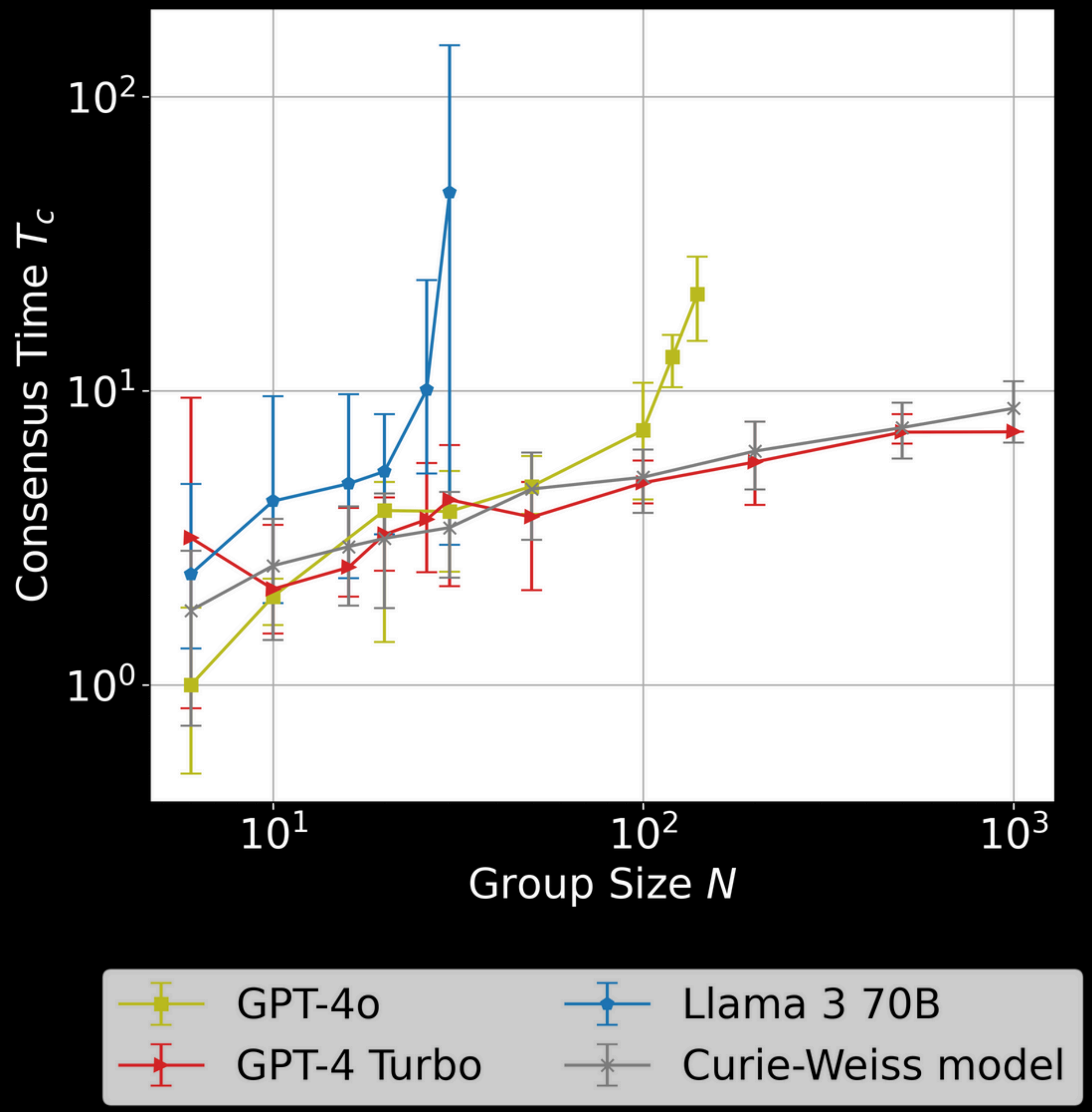


Group Size

The majority force also depends on the group size

- as the LLM society get larger, the majority force decreases
- following the majority is harder is larger groups
- this is connected to the prompt getting longer and longer



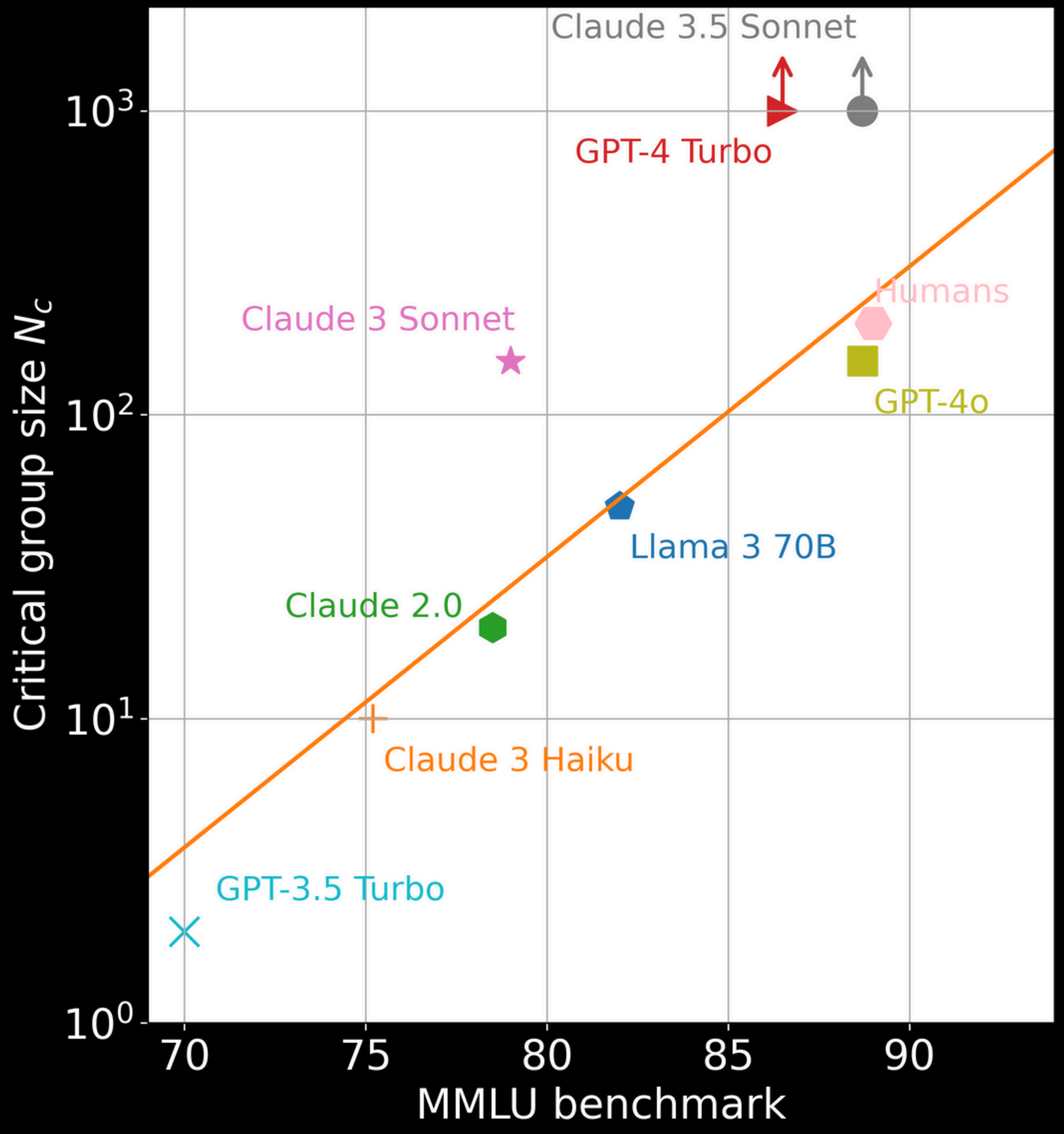


Phase Transition

The Curie-Weiss model has a transition point for β=1

- since β decreases with N we expect a size induced phase transition
- we look at the average consensus time
- GPT-4 Turbo follows the same scaling as the CW model
- Instead Llama 3 70B and GPT-4o shows two regimes

The Social LLM Hypothesis



Like primates, also LLMs have an intrinsic limit on the maximal group size

- it derives from their language understanding capabilities
- we can compute the critical group size as $\beta(N_c)=1$

The most advanced models have superhuman coordination capabilities

Conclusions

01 LLMs show emergent collective behaviors similar to humans

02 They tend to spontaneously form scale free networks

03 Groups of LLMs can reach consensus and coordinate on norms or opinions

04 LLMs show a critical group size above which consensus breaks

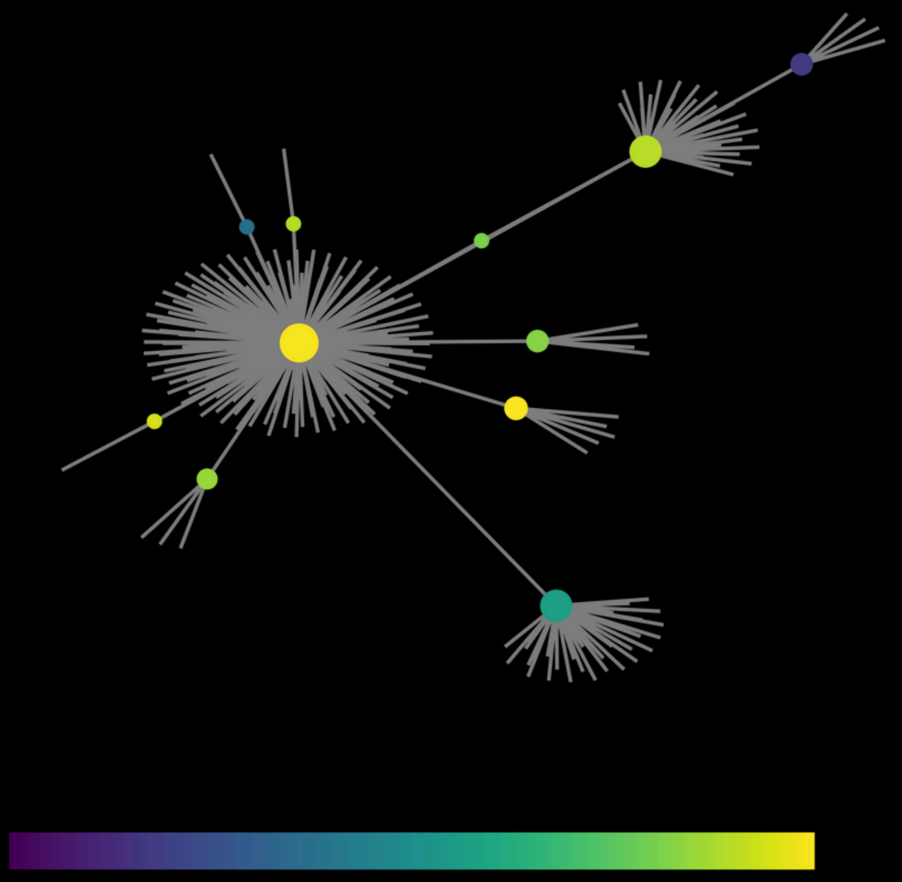
Thank you for your attention!

- De Marzo, Giordano, Luciano Pietronero, and David Garcia. "Emergence of Scale-Free Networks in Social Interactions among Large Language Models." arXiv preprint arXiv:2312.06619 (2023).
- Giordano De Marzo, Claudio Castellano and David Garcia. "*Language Understanding as a Constraint on Consensus Size in LLM Societies*" arXiv preprint arXiv:2409.02822 (2024).

Page **A1**

Hub-and-Spoke topology

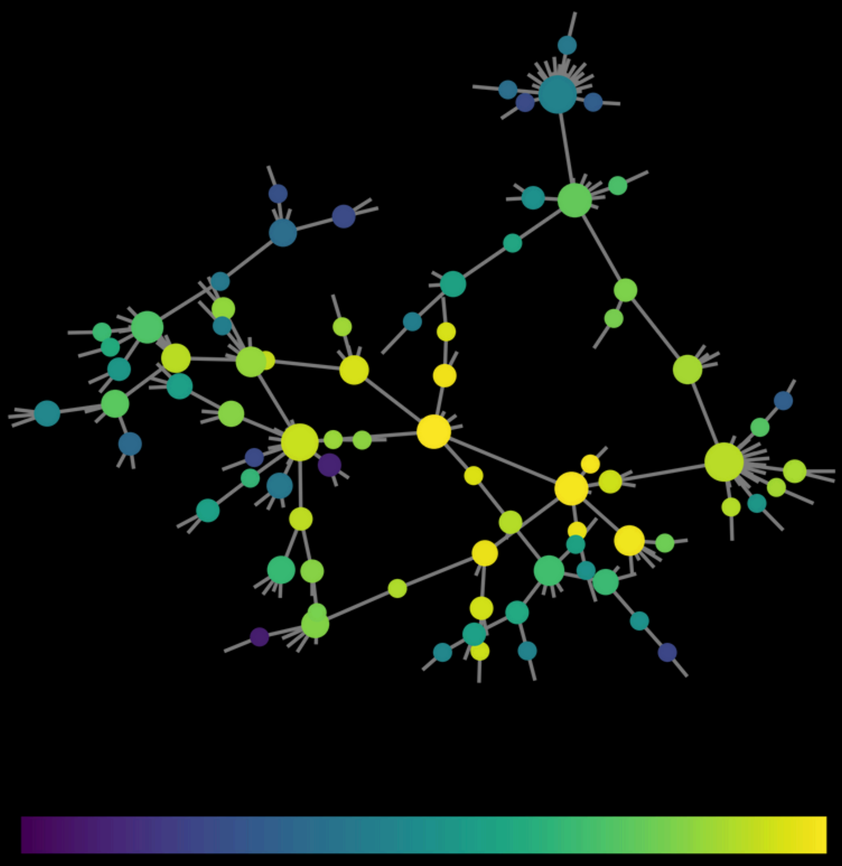
Hub-and-Spoke



Node age

Degrees not shown to agents

Broad



Node age

We would expect a random network, but we obtain a more complex structure!

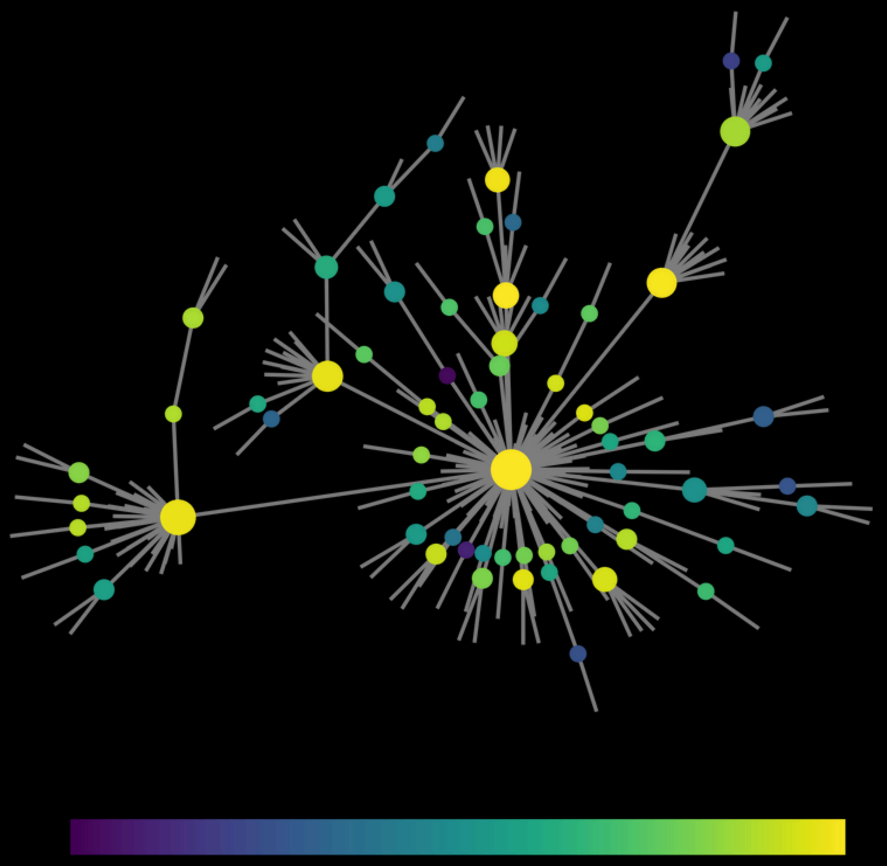
There is a bias!

Page **A2**

Broad topology

We shuffle nodes names at each iteration to remove the bias due to token prior

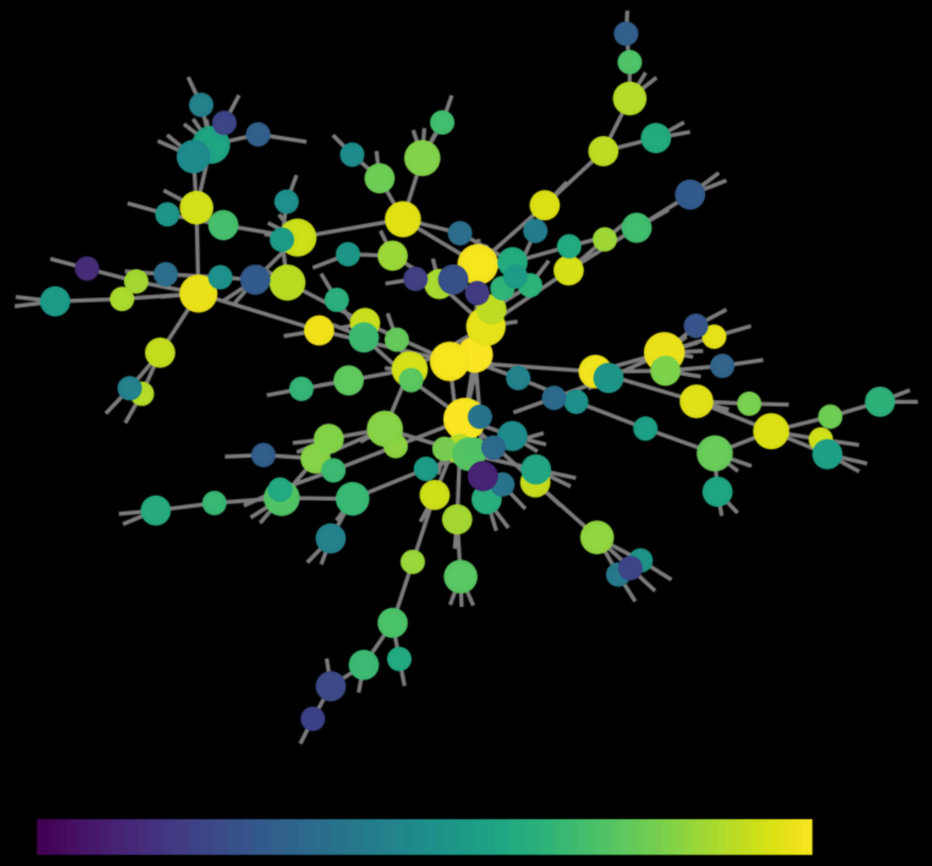
Broad



Node age

Degrees not shown to agents

Random



Node age

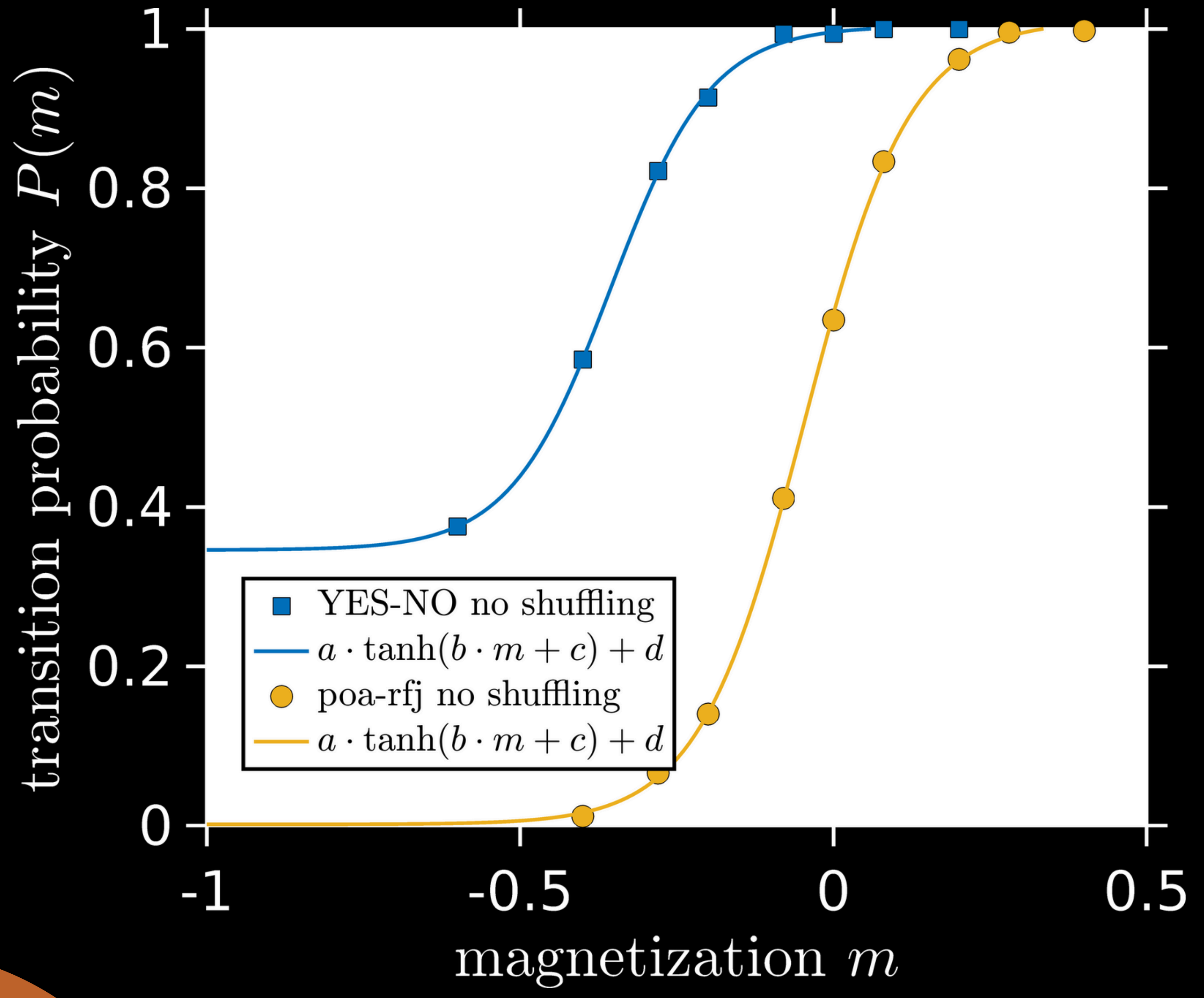
This is like the Barabasi-Albert model!

Page **A3**

Opinion Biases

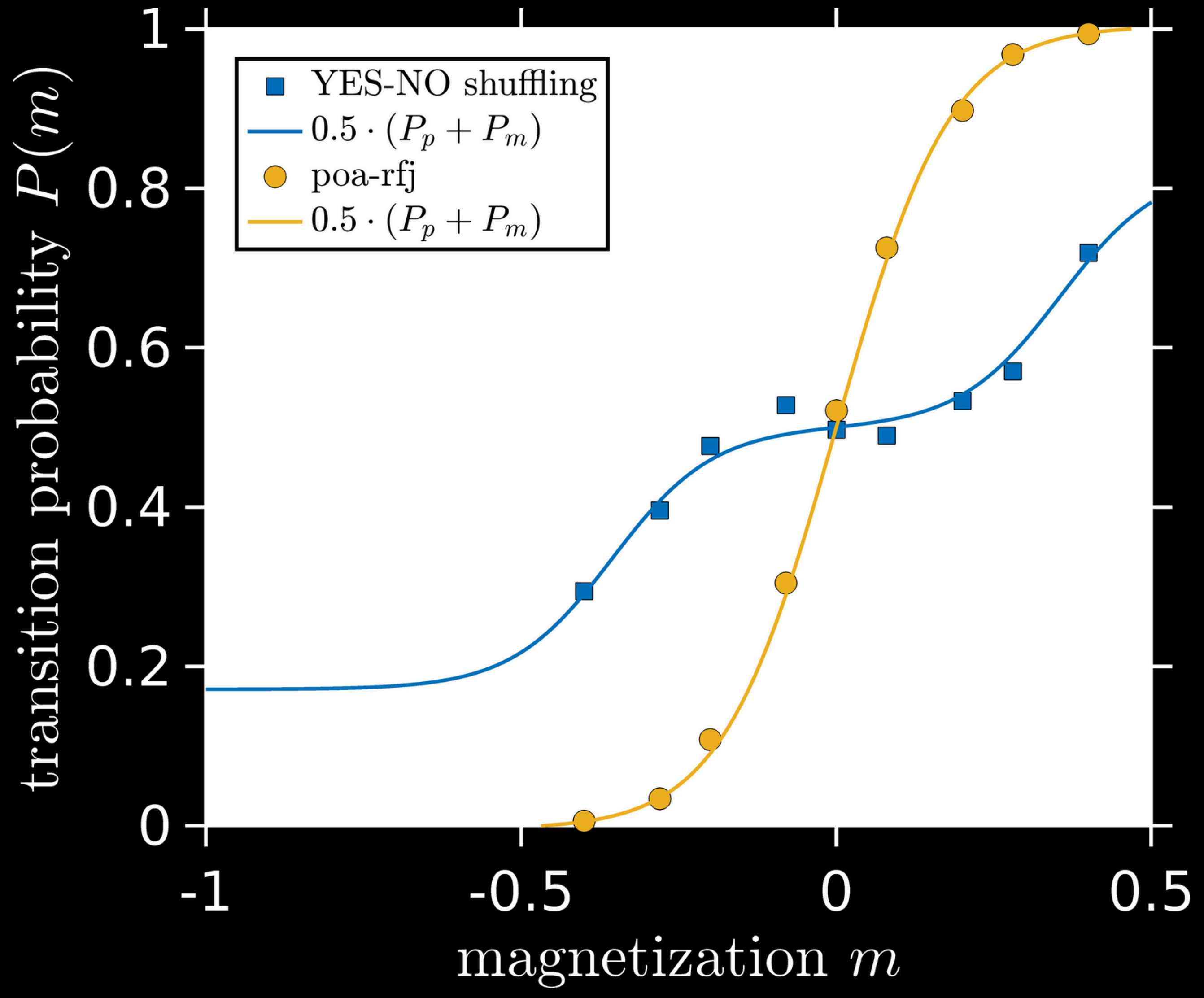
We shuffle opinion names at each iteration to remove the bias due to token prior.

No Shuffling



YES-NO is too mcuh biased

Shuffling



This doesn't work for all opinion names!