

Modelování rodinných interakcí pomocí generativních agentů

Petra Vidnerová, Roman Neruda
Ústav informatiky AV ČR



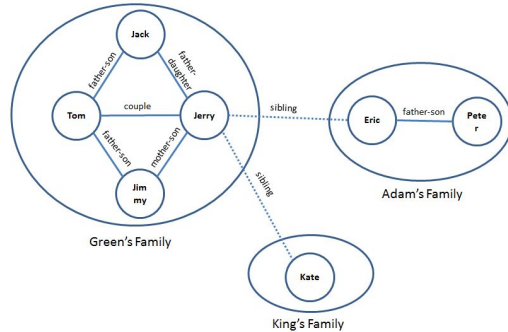
Téma: Domácnosti, technologie a civilizace

“Další zaměření přináší problémy spojené se šířením misinformací, desinformací, malinformací a propagandy v informačním prostředí internetu a sociálních sítí, neboť jsou často zprostředkované algoritmy umělé inteligence a posilované s nimi spojenými mechanismy – jako jsou komory ozvěn a filtrační bubliny. To může ohrozit nejen soudržnost domácností, ale celé společnosti, a výrazně ohrozit schopnost společnosti jako celku čelit globálním výzvám současnosti.”

Výzkumné cíle:

- technologie a biotechnologie chytré domácnosti
- inteligentní sídla
- **umělá inteligence v civilizačním procesu**
- genderová dimenze technologií v domácnosti
- **modelování s cílenými hypotézami mechanismů vývoje**

Model rodiny



Sít'ový model

- Společenství lidí je reprezentováno grafem
- Vrcholy (uzly) jsou lidé, hrany určují vztahy/komunikaci

Rodina

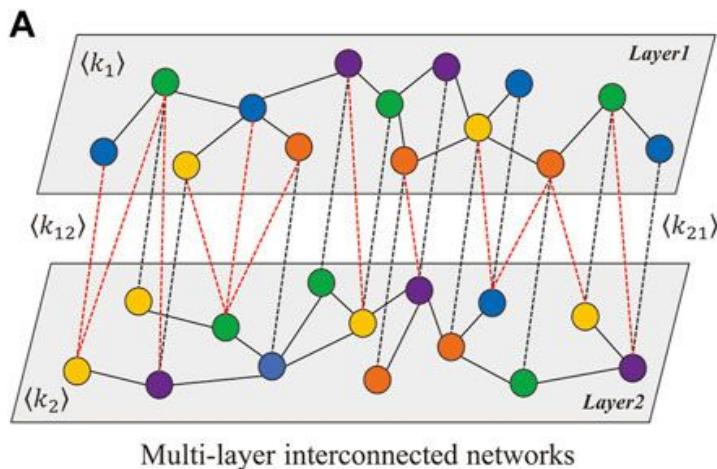
- Specifický typ grafu - malý počet vrcholů, velký počet hran uvnitř rodiny
- Relativně řídké propojení rodin mezi sebou

Rodinné interakce

Graf komunikací může mít více vrstev:

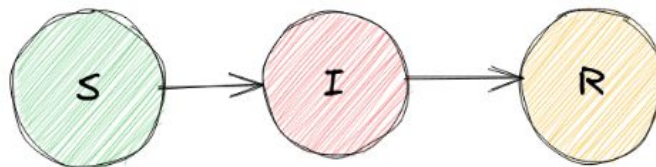
- Přímá komunikace, média, sociální sítě

Tradiční model šíření informace je inspirován epidemiologickými **kompartmentovými modely**



Jedinec má různé stavy vzhledem k informacím:

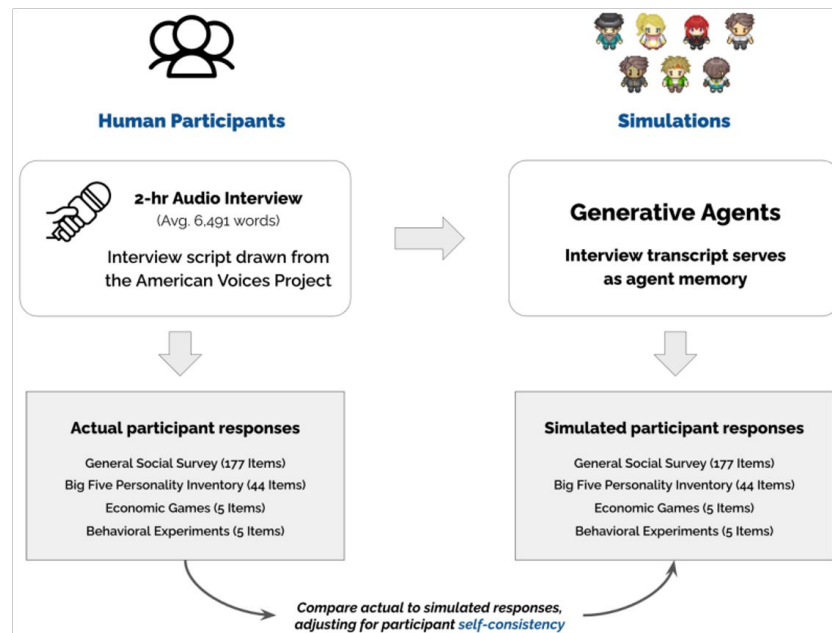
- **S**usceptible - může být ovlivněn informací
- **I**nfected - aktivně šíří dané informace
- **H**esitate - pochybující
- **R**emoved - není ovlivněn informací



Generativní agenti

Místo jednoduchých stochastických modelů využijme **velký jazykový model (LLM)** pro modelování jednotlivých lidí.

- Nutno správně vytvořit **personu** modelující konkrétní vlastnosti člověka
- Nutno doplnit LLM o další mechanismy, které samotný chat typicky nemá (**paměť**, plánování, zdůvodňování)
- Nutno experimenty založit a validovat na skutečných **datech**.



J. S. Park, et al. Generative Agents: Interactive Simulacra of Human Behavior. (2023)

J. S. Park, et al. Generative Agent Simulations of 1,000 People. (2024)

Příklad: Syntetický volební panel

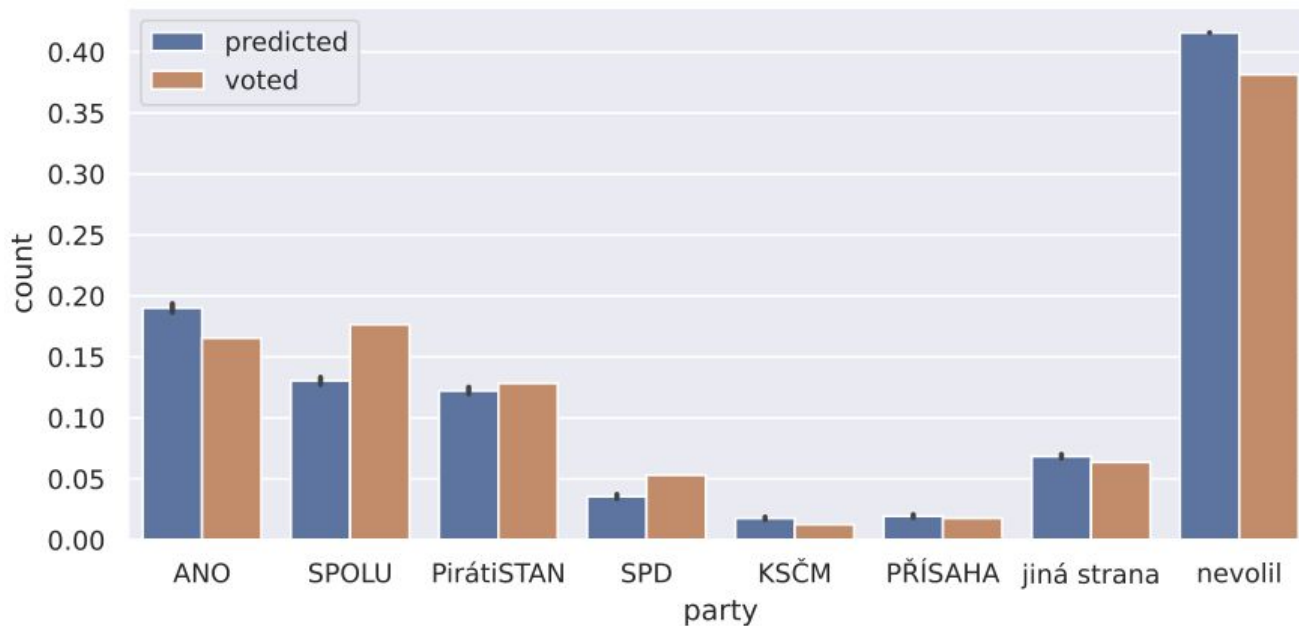
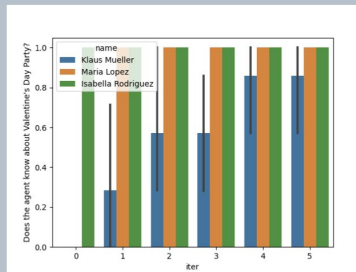


Figure 1. Predicted votes by Command r+ [2] vs respondent votes from Society of Distrust [5].

Příklad: Simulace šíření informací

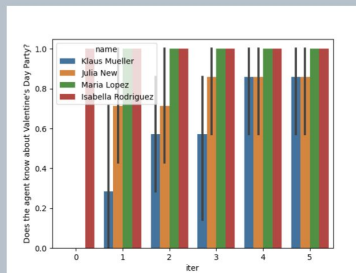
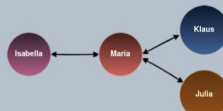
Experiment I

- 3 agents, personalities taken from [1]
- the target piece of information: *"Isabella organises a party"*
- at the start other agents do not know it



Experiment II

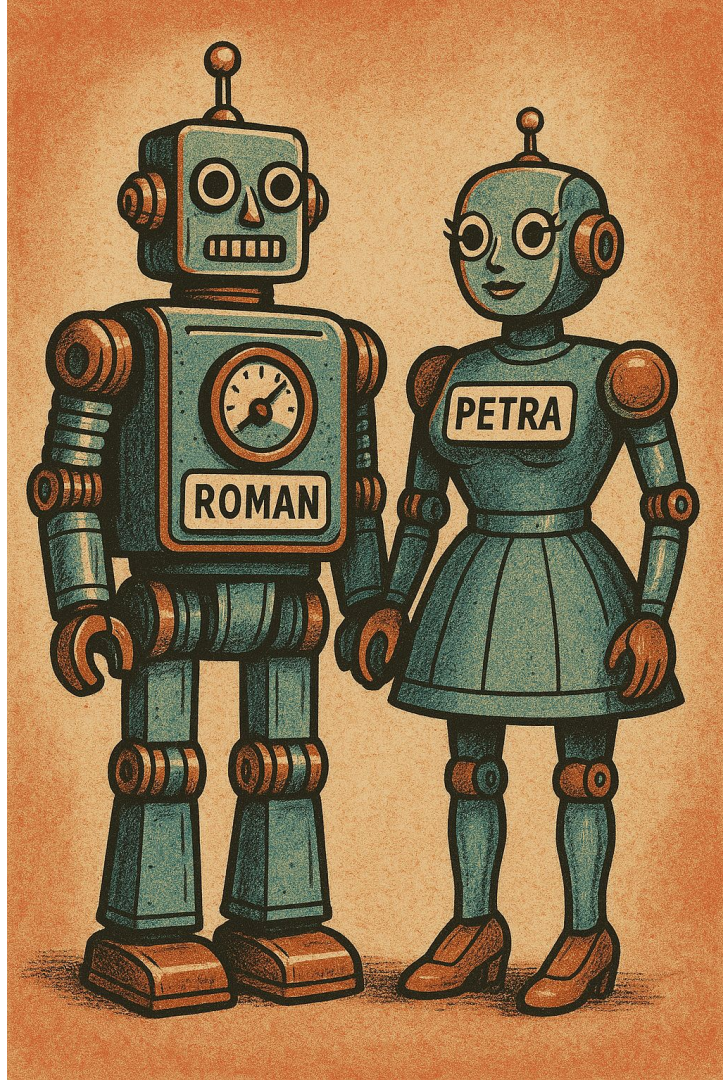
- an additional agent Julia, who loves partying



Mixtral-8x7B-Instruct-v0.1[2] used in all experiments.

2024-06-20 10:55:28,068 - INFO - Julia New meets Klaus Mueller

2024-06-20 10:55:28,068 - INFO - JULIA NEW: Hi Klaus! It's great to see you again. I heard from Maria that you're working on a research paper about the effects of gentrification in low-income communities. I'm really interested in that topic, especially since I've noticed some changes in my neighborhood that could be related to gentrification.



Závěrem



Máme metody AI pro simulaci interakcí, šíření informací a dalších komunikačních fenoménů.

Dokážeme modelovat rodiny a částečně i vztahy mezi nimi (simulace jsou časově náročné).

Hledáme zajímavé výzkumné otázky a data od partnerů v projektu.

Těšíme se na spolupráci