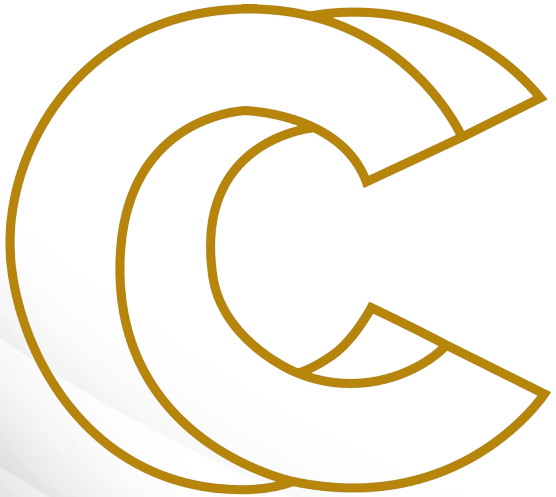
AI Meets Computational Social Science



EURO 4SEE

Marija Mitrović Dankulov
Institute of Physics Belgrade, NCC Serbia

About me: current roles

EURO 4SEE

INSTITUTE OF PHYSICS BELGRADE

Research Professor

Scientific Computing Laboratory

Head

Innovation Center Institute of Physics Belgrade

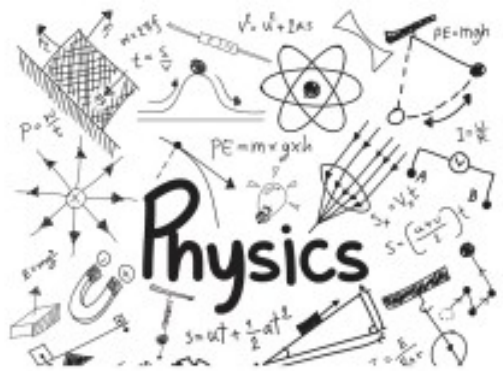
Lecturer at MSc:
Computing in Social Sciences,
Advanced Data Analytics

AI Lead

HPC SERBIA

Industry Champion

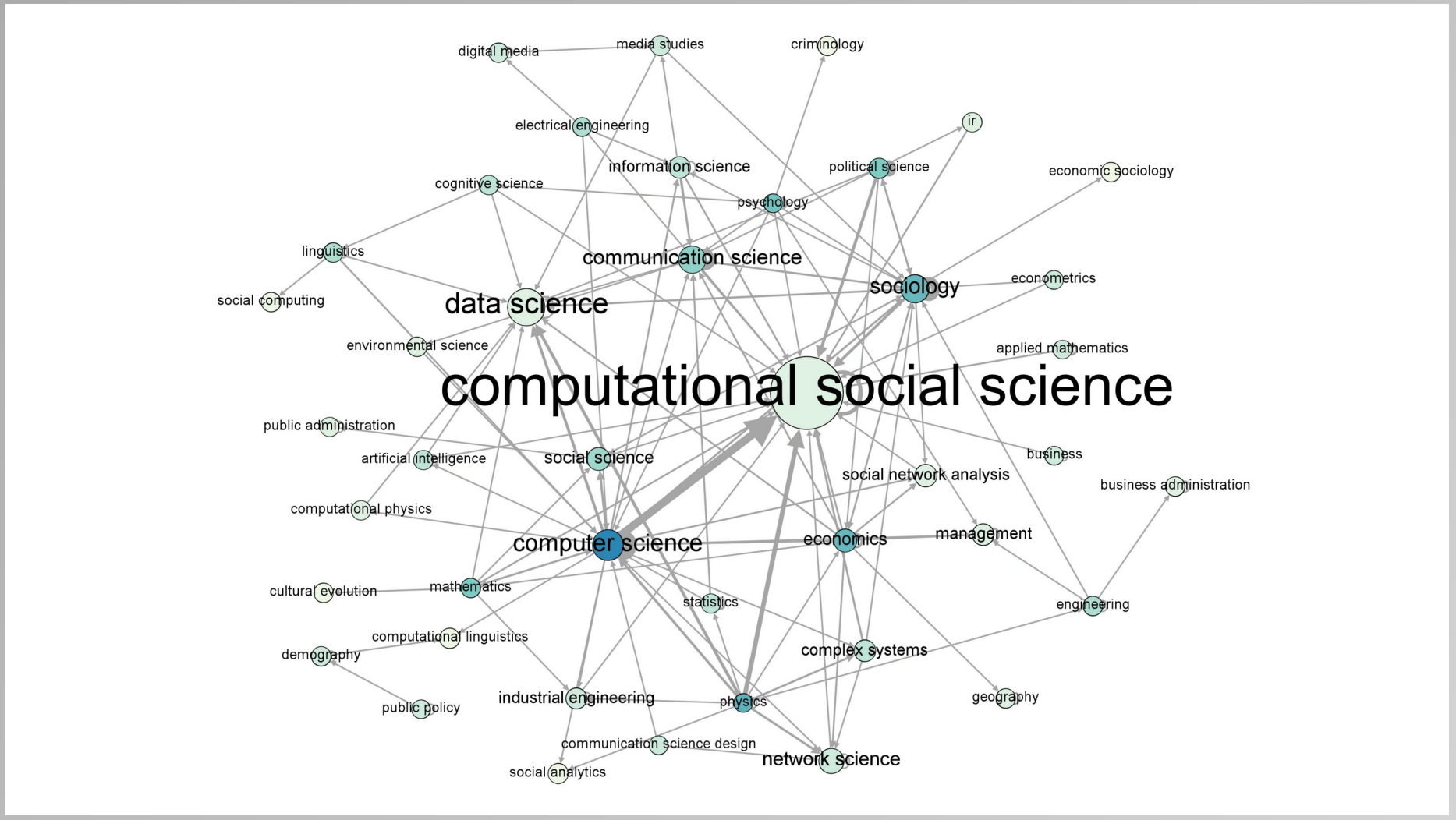
About me: my true passion



+

=

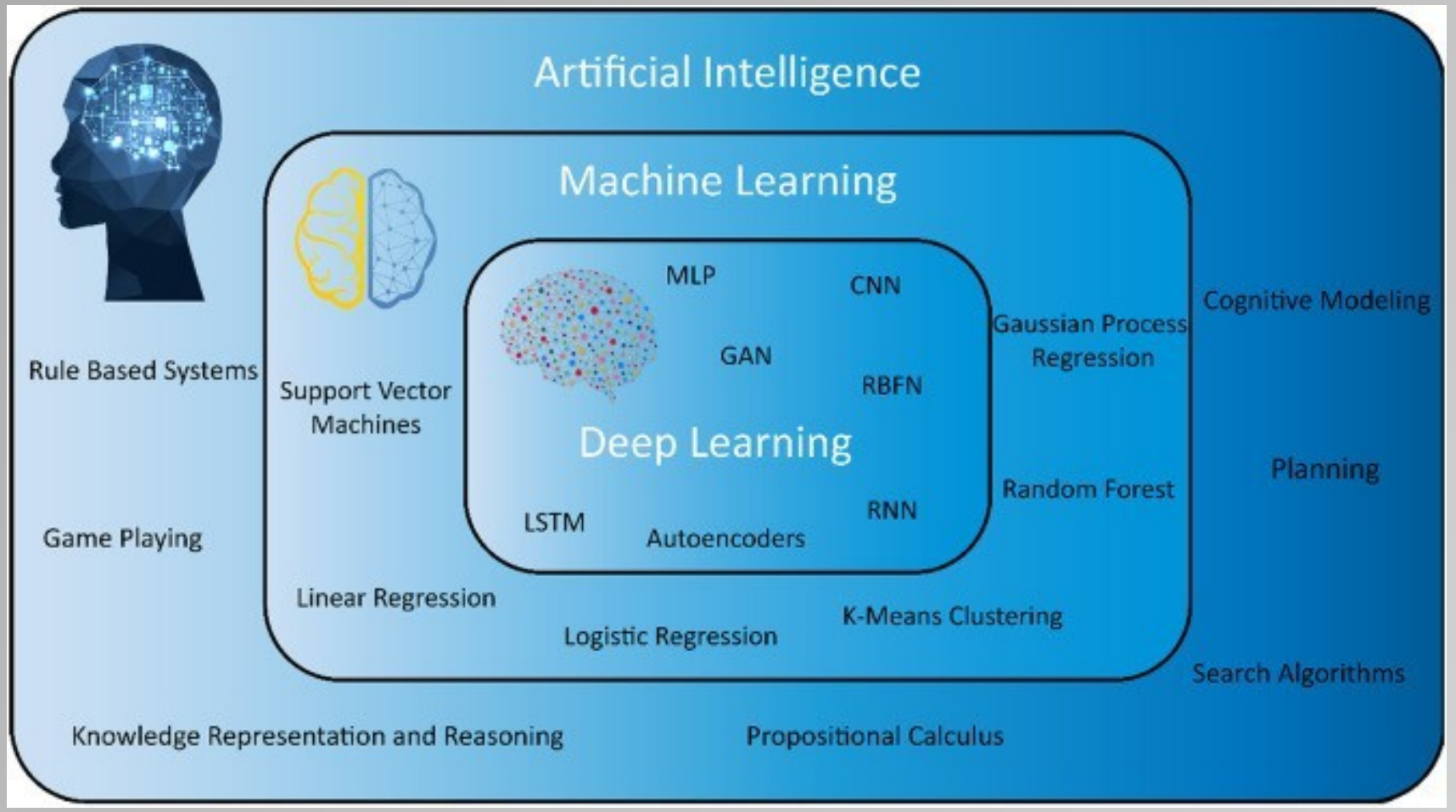
Computational Social Science - CSS



Source: https://communities.springernature.com/posts/computational-social-science-heralds-the-age-of-interdisciplinary-science

Artifical Intelligence - AI

EURO 4SEE



Source: https://www.linkedin.com/pulse/artificial-intelligence-automation-shaping-future-global-economy-6d1xe/

CSS beginnings

EURO 4SEE

- Physics and social sciences – more than 2 centuries long relationship
- Modeling social systems using physics and computer science models: Santa Fe Institute
-
- Early 21st century:
  - Web 2.0 – great data source
  - Computational resources
  - Physicists and computer scientists – more interested in social sciences
  - Social scientists became more interested in computational sciences

Our approach

EURO 4SEE

- **Empirical analysis -** Quantify collective behavior using methods of complex network theory and statistical physics to analyze empirical data
- **Agent-baed modeling -** data driven agent-based model of human interactions
- **Experiments -** in lab, controlled experiments with selected group of participants

Cyberemotions project

EURO 4SEE

- **Cyberemotions – Collective emtions in cyberspace –** project about the role of collective emotions in creating, forming and breaking-up e-communities
- **EU FP7 Funded project** from 2009-2013
- **Consortium:** 8 research organizations from Poland, Germany, Austria, United Kingdom, Slovenia
- **Datasets**: BBC Blog, Twitter, Digg, MySpace, IRC channels (with emotional content of messages)
- **Models:** Agent based models of emotional agents, **SentiStrength (** **https://mi-linux.wlv.ac.uk/~cm1993/sentistrength/**)
- SentiStrength classified London Olympics tweets with the results put up in lights on the EDF Energy London Eye

General structure of data

EURO 4SEE

Data:

- High-temporal resolution
- **User (U) :** unique ID; details about activity;
- **Posts/Comments (P/C):** posting time; ID's of connected users; texts;
- **Emotions (E):** Emotional content of text - 1 -positive, 0 - neutral, −1 negative, (-1,-5) - negative and (1,5) - positive

Emotion classifier

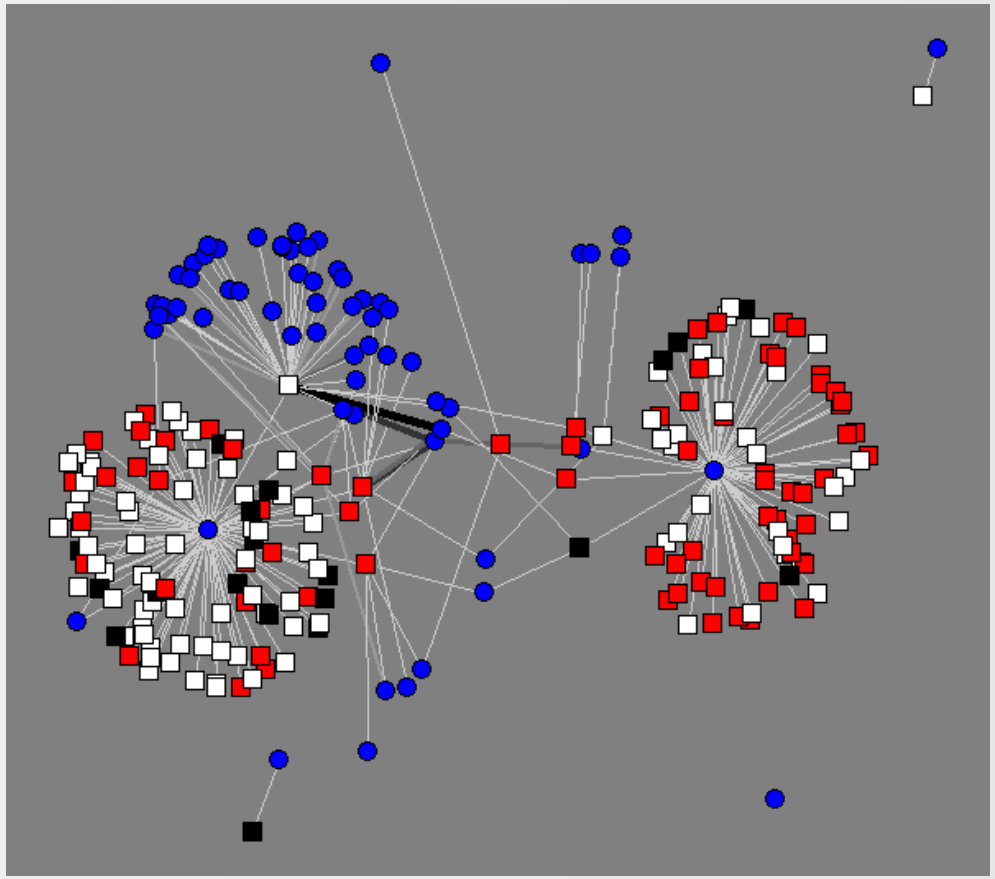
EURO 4SEE

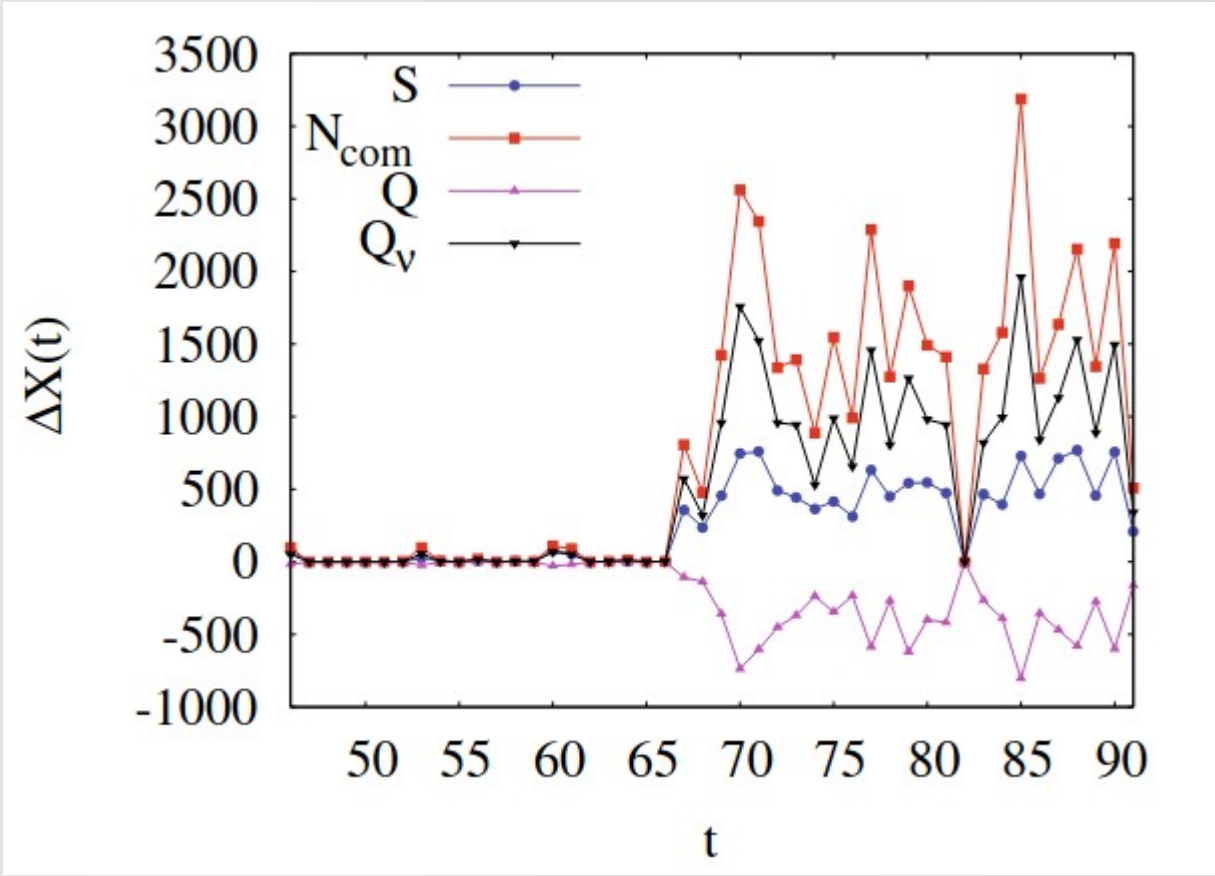
- **Extraction of emotions** – a classification problem

Datasets: BBC Blogs (June 2007-February 2009) posts – 3792, comments 80873, users – 21462; **Digg** (February 2009 – April 2009) posts – 3984, comments – 917708, users - 82201

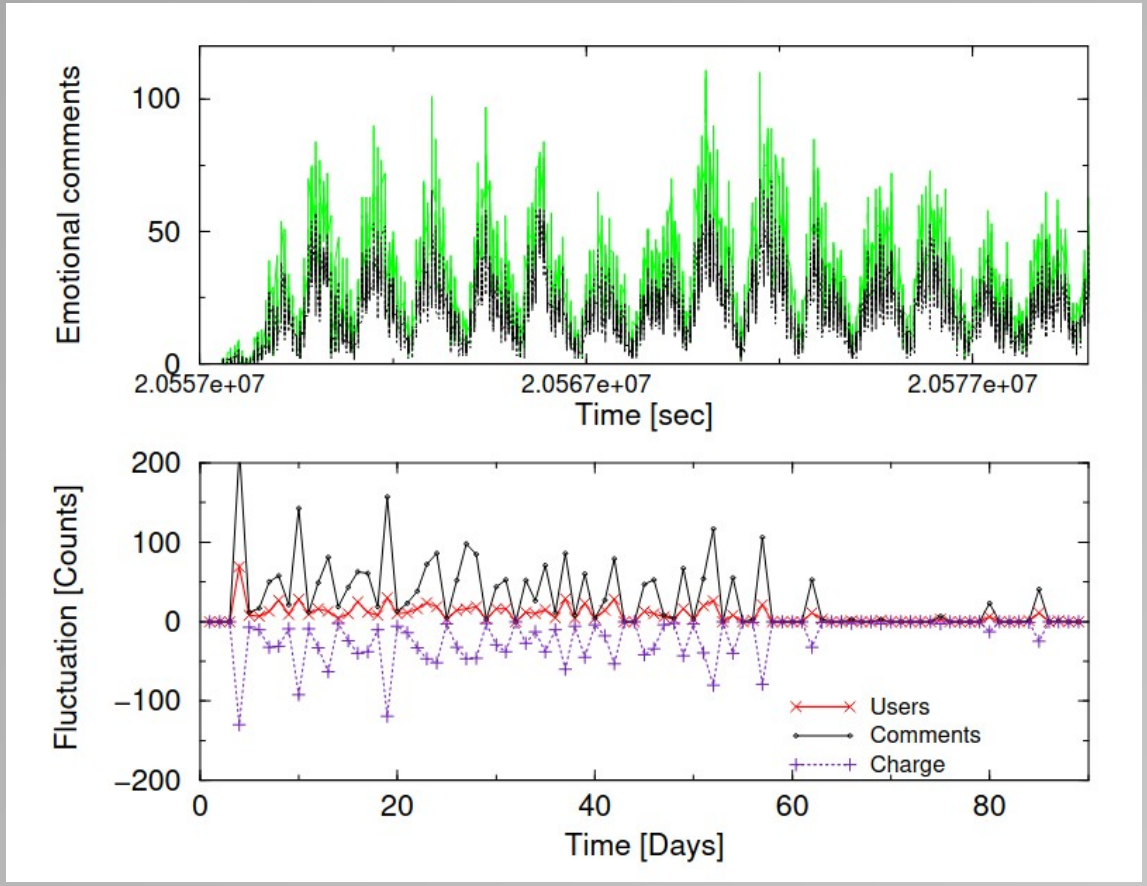
- **Supervised:**
  - document D – objective or subjective, positive or negative
  - language model classifier
  - D represented by tokens, n-gram approximation
  - BLOGS06 dataset – 16 481 – objective, 7930 – negative documents, 9 968 – positive
  - 70% accuracy

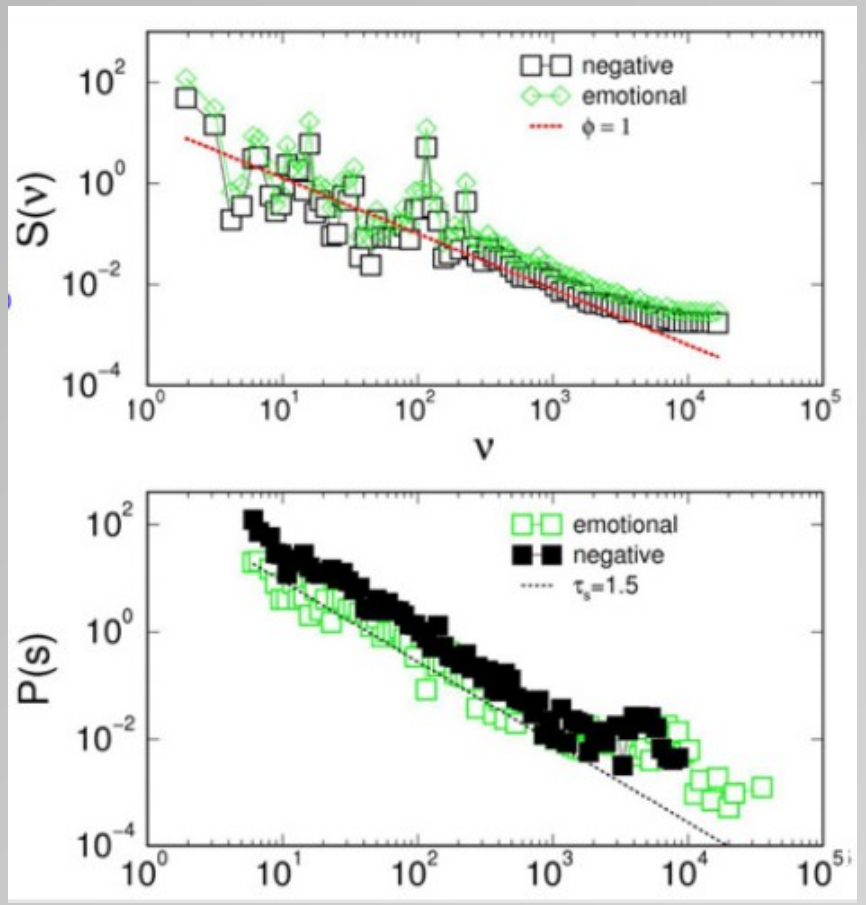
Popular BBC Blogs





Discussion driven Diggs - SOC





CTRUST project

EURO 4SEE

CTRUST - Topology-derived methods for the analysis of collective trust dynamics – project about understanding rise and fall of collective trust in online communities

Science Fund Republic of Serbia – PRIZMA program

Consortium: Institute of Physics Belgrade, Faculty of Philosophy Novi Sad, Institute of Nuclear Science VINČA

Datasets: MADOC: Multi-Platform Aggregated Dataset of Online Communities
https://zenodo.org/records/14637314

Empirical analysis: scripts and resutls available on Github

Agent-based modeling: models and scripts available on Github

Science Fund of the Republic of Serbia

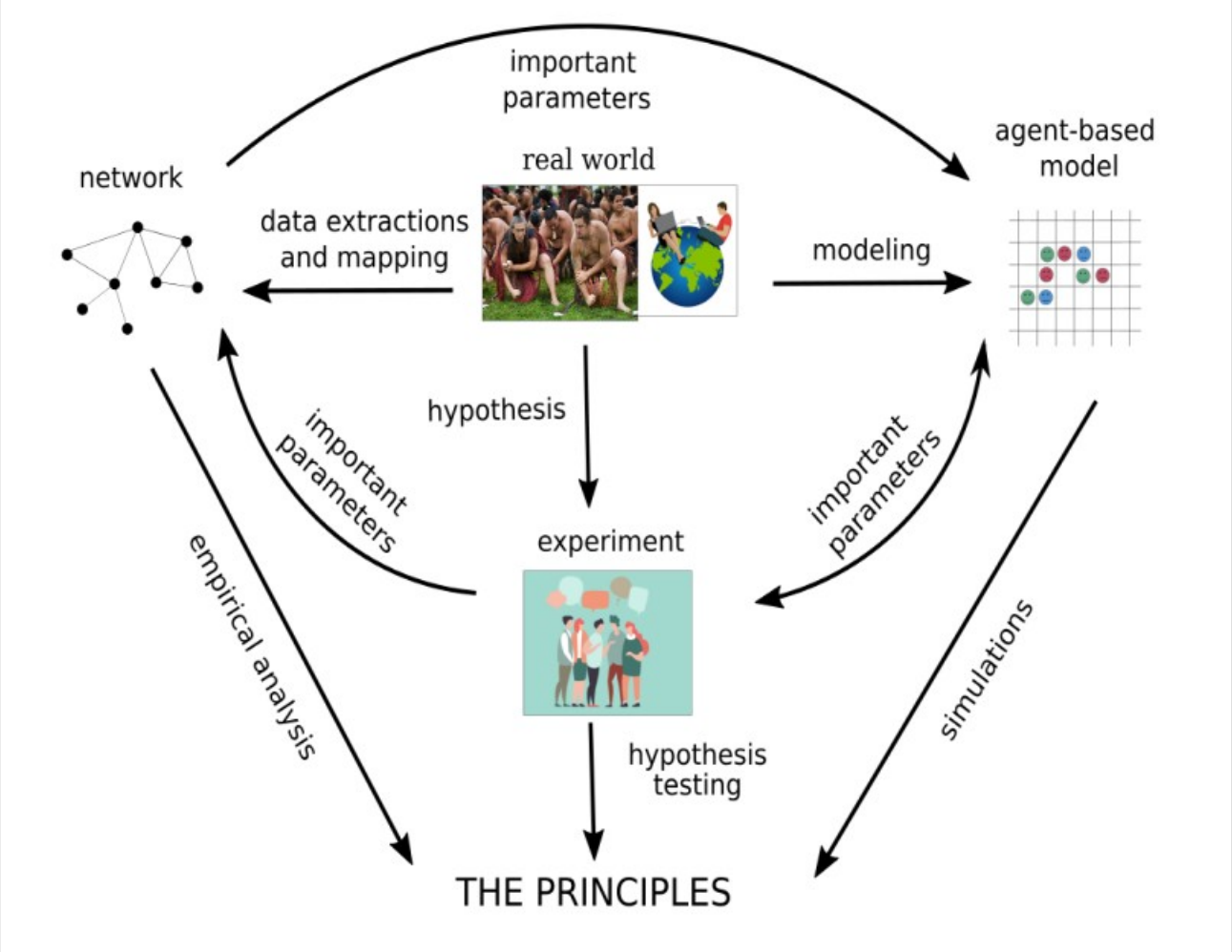
CTRUST

INSTITUTE OF PHYSICS BELGRADE

ФИЛОЗОФСКИ ФАКУЛТЕТ · НОВИ САД

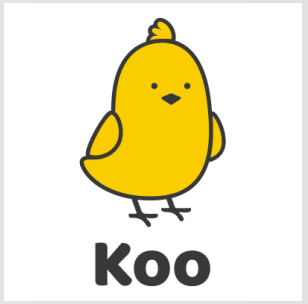
www.ctrust.ac.rs

CTRUST approach



MADOC dataset

EURO 4SEE



MADOC dataset: 18.9M posts, 236M comments, 23.1M users

AI usage: topic alignment, sentiment analysis

Application: collective trust, community dynamics, migration of users between platforms, toxic behavior

Paper: **https://arxiv.org/abs/2501.12886**, accepted at ICWSM‘25
https://www.icwsm.org/2025/index.html

Source: **https://zenodo.org/records/14637314**

MADOC: topic alignment

EURO 4SEE

- **Groups from Reddit and VOAT**: funny, gaming, gifs, pics, videos, technology, fatpeoplehate, GrateAwakening, MillionDollarExtreme, CringeAnarchy, KokatuInAction, MensRights

- **Topic alignement for Koo and Bluesky:**
  - - combined communities from Reddit and VOAT for each group
  - - removed: stop words, common words (occuring in more than 99%), removed documents with less than 20 and more than 2000 words
  - - Latent Dirichlet Allocation – determined topics within each group and corresponding 20 keywords
  - **-** unique set of words per group - filtered to retain only the least frequen English words
  - - **basic filter:** documents containing at least one community keyword
  - - **strict filter**: documents containing at least two different community keywords

MADOC: sentiment

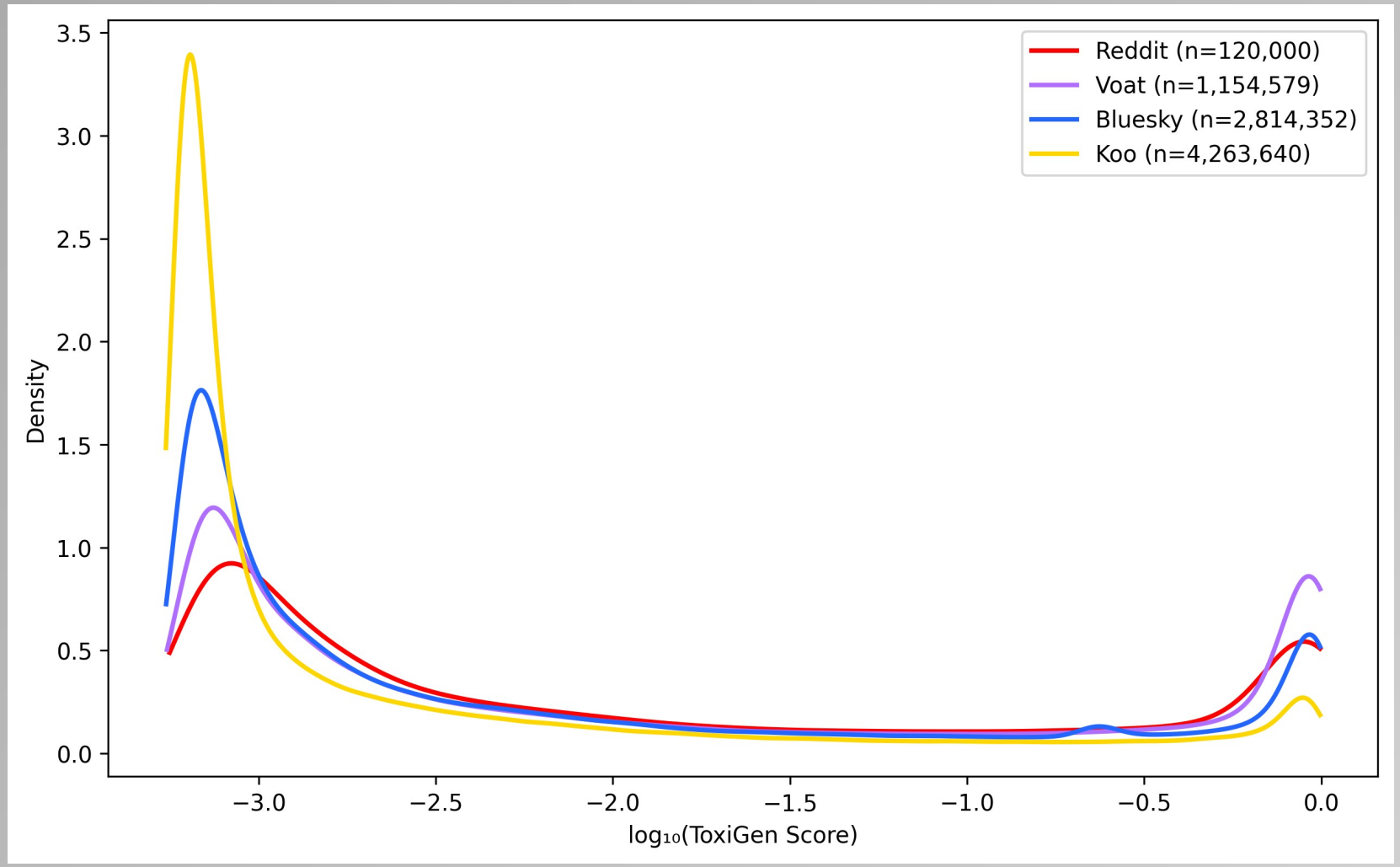
EURO 4SEE

- **VADER:** sentiment of posts and comments
  - VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media
  - https://github.com/cjhutto/vaderSentiment
  - for each post/comment we provided compound value between -1.0 and 1.0

- **TextBlob:**
  - TextBlob: Simplified Text Processing – python library for text processing
  - https://textblob.readthedocs.io/en/dev/
  - for each post/comment we provide polarity between -1.0 and 1.0, and objectivity between 0.0 (objective) and 1.0 (subjective)

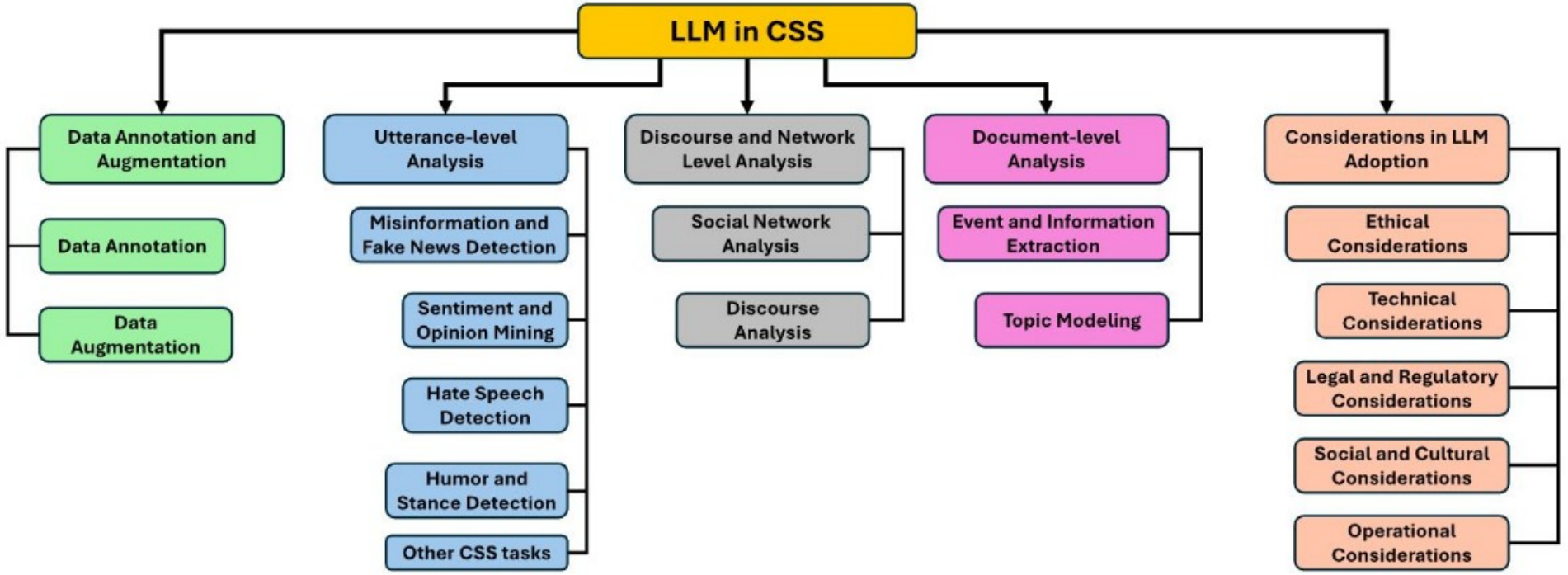
- **ToxiGen RoBERTa:**
  - used to detect implicit hate speech
  - https://huggingface.co/tomh/toxigen_roberta
  - toxicity between 0.0 (non-toxic) to 1.0

MADOC: statictics



CSS and LLMs

EURO 4SEE



Thapa, S., Shiwakoti, S., Shah, S. B., Adhikari, S., Veeramani, H., Nasim, M., & Naseem, U. (2025). Large language models (LLM) in computational social science: prospects, current state, and challenges. *Social Network Analysis and Mining*, *15*(1), 1–30. https://doi.org/10.1007/s13278-025-01428-9

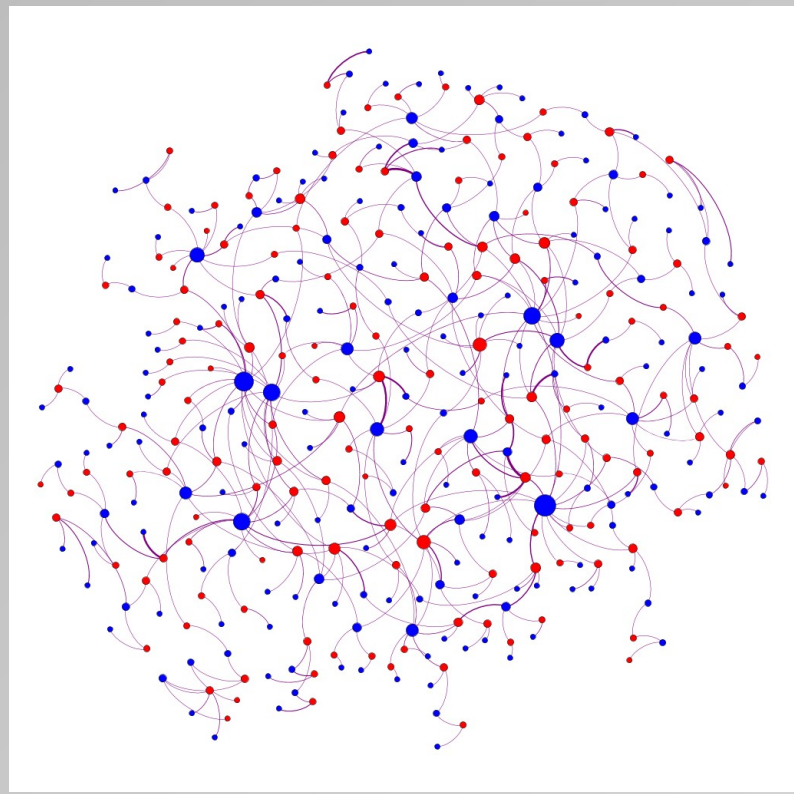
Agent-based modeling before LLMs

EURO 4SEE

Agent‘s state defined by equation and interact through evolving networks network of posts and users

$$a_i(t+1) = \begin{cases} (1-\gamma_a)a_i(t) + [h_i^a(t) + qh_{mf}^a(t)](d_1 + d_2(a_i(t) - a_i(t)^2))(1-a_i(t)) & \text{if } \Delta t_i < 1 \\ (1-\gamma_a)a_i(t) & \text{otherwise} \end{cases}$$

$$v_i(t+1) = \begin{cases} (1-\gamma_v)v_i(t) + [h_i^v(t) + qh_{mf}^v(t)](t)(c_1 + c_2(v_i(t) - v_i(t)^3))(1-|v_i|) & \text{if } \Delta t_i < 1 \\ (1-\gamma_v)v_i(t) & \text{otherwise} \end{cases}$$



Toy-like models

Pro: simple enough to understand rough picture

Cons: very far from realistic situations and hard to directly compare with empirical results

Agent-based modeling with LLMs

EURO 4SEE

LLMs revolution in agent-besed modeling:

- LLMs as agents – several agent types

- platforms that recreate specific setting: X, Reddit, other platforms

- other parameters to make the systems more realistic

- simulation of real systems, scenario testing, synthetic data creation

- platforms: Ysocial: https://ysocialtwin.github.io/, MachineSoM: https://github.com/zjunlp/MachineSoM, WelfareDiplomacy: https://github.com/mukobi/welfare-diplomacy

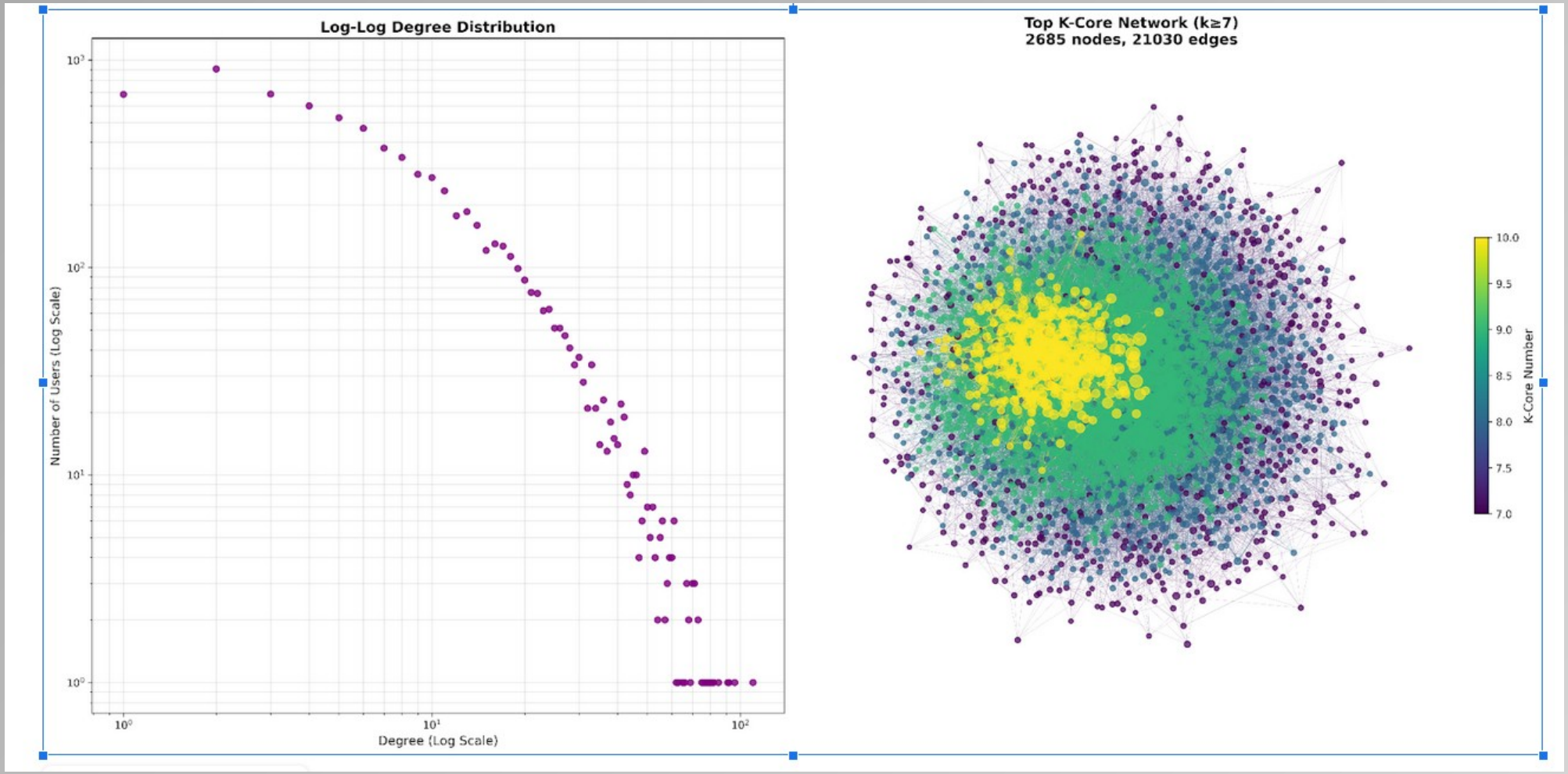
CTRUST: Agent-based model

EURO 4SEE

- Adapted version of YSocial to simulate Reddit dynamics: https://github.com/atomashevic/YClient-Reddit
- **LLM -** Dolphin3 LLM
- **Agents‘ detailed psychological profiles:** demographic characteristics, political affiliations, Big Five personality traits, education levels, and critically, configurable toxicity propensities ranging from "absolutely no" to "extremely" toxic behavior
- **Community:** technology

CTRUST: Agent-based model

EURO 4SEE



CTRUST: TEAM

EURO 4SEE

Aleksandar Tomašević
Faculty of Philosophy
Novi Sad

Ana Vranić
Institute of Physics
Belgrade

Darja Cvetković
Institute of Physics
Belgrade

Boris Stupovski
Institute of Physics
Belgrade

CTRUST: TEAM

EURO 4SEE

Slobodan Maletić
Institute of Nuclear Sciences VINČA

Miroslav Anđelković
Institute of Nuclear Sciences VINČA

Sara Major
Faculty of Philosophy Novi Sad

Aleksandra Alorić
Microsoft

Conclusions

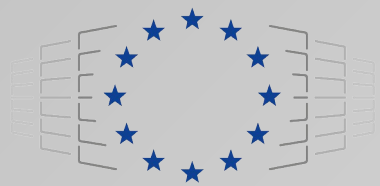
EURO 4SEE

- AI is a part of computational social science – information extraction, data curation, prediction, simulation
- Explainability is a must have
- Computational social science needs to part of AI

EURO 4SEE

Thanks!

Co-funded by
the European Union



EuroHPC
Joint Undertaking

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101191697. The JU receives support from the Digital Europe Programme and Germany, Türkiye, Republic of North Macedonia, Montenegro, Serbia, Bosnia and Herzegovina.

Science Fund
of the Republic of Serbia

This research was supported by the Science Fund of the Republic of Serbia, 7416, Topology-derived methods for the analysis of collective trust dynamics – CTRUST.