

Hands-on Training on the SoBigData Platform

Case Study: Identifying Epistemic Enclaves and Understanding Polarisation

Giulio Rossetti, Senior Researcher at CNR-ISTI

Andrea Failla, PhD students in the Italian National PhD Program in Artificial Intelligence

This short, hands-on tutorial focuses on how to detect **epistemic enclaves**—also known as echo chambers, using **CDlib**, and how to study **polarisation dynamics** through opinion-evolution models implemented in **NDlib**. Rather than offering purely theoretical insights, the course provides practical workflows for uncovering fragmented information ecosystems, analysing structural drivers of echo chambers, and simulating how opinions evolve and diverge on both real-world and synthetic social graphs.

What You'll Learn

The tutorial provides a complete, end-to-end introduction to network and opinion analysis in Python. You'll learn to:

- Load, manipulate, and explore complex networks using standard formats and curated datasets.
- Identify epistemic enclaves using **CDlib** by applying community detection algorithms (Louvain, Leiden, Infomap, label propagation), and assessing community cohesion, isolation, and fragmentation.
- Evaluate and compare community partitions to reveal structural patterns consistent with selective exposure and echo-chamber effects.
- Simulate opinion dynamics with **NDlib**, focusing on two key models for studying polarisation:
 - **Deffuant–Weisbuch (DW)**: A bounded-confidence model that reveals when consensus emerges, when moderate disagreement persists, and when polarisation crystallises.
 - **Algorithmic Bias Model**: A framework for simulating how recommendation biases or feed-ranking algorithms amplify selective exposure, intensify clustering, and foster polarised belief distributions.
- Interpret opinion trajectories, quantify polarisation, detect consensus or fragmentation patterns, and compare intervention scenarios or structural modifications.
- Use synthetic social-network data generated through **YSocial**, an agent-based environment that simulates interactions, discussions, and misinformation dynamics, providing a controlled sandbox for studying echo chambers and polarisation mechanisms.

Key Tools and Topics

The tutorial follows a project-based approach, ensuring proficiency with modern libraries for network and diffusion analysis:

- **CDlib**: A comprehensive framework for community detection and evaluation. You'll run multiple algorithms, compute community quality metrics (modularity, conductance, internal density), compare partitions, and visualise the structural signatures of epistemic enclaves.
- **NDlib**: A flexible library for simulating opinion dynamics. You'll explore how **Deffuant–Weisbuch** and **Algorithmic Bias** models shape belief trajectories, drive or mitigate polarisation, and interact with network structure.
- **YSocial**: Used to generate synthetic interaction networks and narrative flows. These datasets allow you to test echo-chamber detection pipelines, analyse the temporal evolution of communities, and simulate polarisation under realistic social and behavioural assumptions.
- **NetworkX** (support library): Provides the underlying tools for network creation, manipulation, and visualisation, forming the computational backbone for CDlib and NDlib workflows.

By the end of the day, participants will be able to load social data, mathematically identify isolated 'epistemic enclaves,' and simulate how algorithmic bias can drive these enclaves toward polarization, using both real and synthetic social data. This equips students with practical skills for research in computational social science, digital behaviour, and the analysis of online information ecosystems. All hands-on activities throughout the training sessions are performed on the SoBigData platform.

Program

Morning Session: The Anatomy of Segregation (3 Hours)

09:30 – 11:00 | Module 1: Foundations – Mapping the Terrain (Tool: NetworkX)

Focus: Loading, manipulating, and exploring complex networks.

- **15 min:** Theory: Graph data structures.
- 45 min: Hands-on: Data Wrangling & Exploration.
 - Loading standard formats (GraphML, GEXF, Edge Lists).
 - **Metric Analysis:** Degree distribution, clustering and path lengths to establish the "baseline" connectivity of the network.
 - **Filtering:** Subgraph extraction - isolating the "core" participants from the periphery.
- 30 min: Visual Exploration:
 - Force-directed layouts to inspect potential clustering.

11:00 – 11:15 | Coffee Break

11:15 – 12:45 | Module 2: Detecting Epistemic Enclaves (Tool: CDlib)

Focus: Identifying cohesive subgroups and assessing their isolation.

- **15 min:** Theory: From "Community" to "Enclave."

- 45 min: Hands-on: Comparative Partitioning.
 - Louvain (Modularity), Leiden (Refined Modularity), Infomap (Flow-based), and Label Propagation (Diffusion-based).
 - Comparison: Using Adjusted Rand Index (ARI) or Normalized Mutual Information (NMI) to see if these algorithms find the *same* enclaves or different ones.
- 30 min: Epistemic Enclaves:
 - Global Scale (network assortativity)
 - Meso-scale evaluation (community purity and separation)
 - Local evaluation (node assortativity/conformity)

12H45—13H45 lunch

Afternoon Session: The Mechanics of Division (3 Hours)

13:45 – 15:15 | Module 3: Simulating Polarisation Dynamics (Tool: NDlib)

Focus: Simulating how opinions fracture and crystallize.

- **15 min:** Theory: Bounded Confidence vs. Algorithmic Bias.
- **45 min:** Hands-on: The Deffuant–Weisbuch (DW) Model.
 - Setting the "confidence threshold"
 - **Experiment:** Varying the "confidence threshold" to observe the phase transition:
 - Phase 1: Emergence of Consensus (High tolerance).
 - Phase 2: Persistent Disagreement.
 - Phase 3: Crystallisation of Polarisation (Low tolerance).
- **30 min:** Hands-on: The Algorithmic Bias Model.
 - Simulating a "feed-ranking" system: Bias the neighbor selection probability based on opinion similarity.
 - **Analysis:** Interpreting opinion trajectories (Time vs. Opinion plots) to prove that recommendation algorithms accelerate selective exposure.

15:15 – 15:30 | Coffee Break

15:30 – 17:00 | Module 4: The Controlled Sandbox (Synthesis) (Tool: YSocial)

Focus: Synthetic agents to study enclaves in a controlled environment.

- **20 min:** Theory: The Social Digital Twin. Using generative ABM to test "what-if" scenarios.
- 40 min: Hands-on: Defining a Simulation.
 - Base concepts: Agent, Slot, Action, Algorithmic curation
- 30 min: Student project
 - Synthetic data analysis: Leverage CDlib and NDlib to characterize YSocial simulation data

Required Competencies and Materials

Technical Prerequisites

- **Python Proficiency:** Intermediate-level knowledge of Python is essential. Participants should be comfortable with basic data structures (lists, dictionaries), control flow, and defining functions.
- **Statistical Literacy:** Familiarity with basic statistical concepts (e.g., mean, distribution) is helpful for understanding metric analysis.
- **Basic Network Concepts:** A rudimentary understanding of what a graph/network is (nodes and edges) will be beneficial, though core concepts will be covered.

Software and Setup

Python libraries:

A preconfigured JupyterLab environment will be made available on the SoBigData RI.

Participants that would like to replicate the hands-on sessions locally on their machine will require the following minimal configuration.

- Python Environment: Python 3.10+
(We recommend using Anaconda or Miniconda for environment management).
- Core Libraries:
 - Jupyter Notebook or JupyterLab (for running the hands-on sessions).
 - [networkx](#) (`pip install networkx`)
 - [cdlib](#) (`pip install cdlib`)
 - [ndlib](#) (`pip install ndlib`)

YSocial:

[ysocial](https://y-not.social/)(<https://y-not.social/>)

While installation instructions will be provided, it is recommended that the YSocial software package is already downloaded and installed on the participants' machines.