GraphNEx

Graph Neural Networks for Explainable Artificial Intelligence

graphnex.eecs.qmul.ac.uk

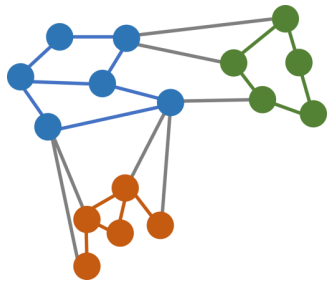
GraphNEx

Queen Mary University of London

EPFL

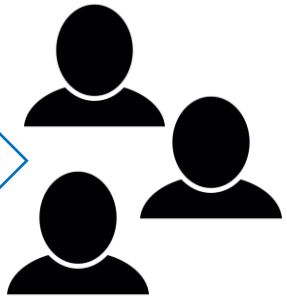
Université de Lyon

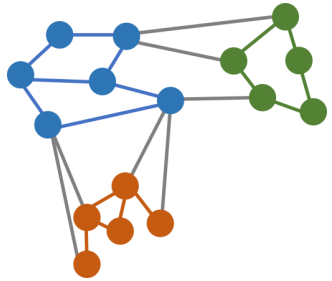
ENS de Lyon



Graph construction & refinement
Concepts & rules discovery
Attention-driven concept selection

Personalisation, gamification
& user feedback



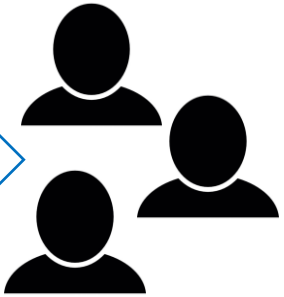


Graph construction & refinement
Concepts & rules discovery
Attention-driven concept selection

Systems genetics

Privacy protection

Personalisation, gamification & user feedback



Systems Genetics

Privacy Protection

The problem

Classic systems genetics approaches

- focus on pairwise comparisons between individual biological data layers
- fail to capture **complex patterns of interactions** across many biological data layers

AI systems capture secondary information that

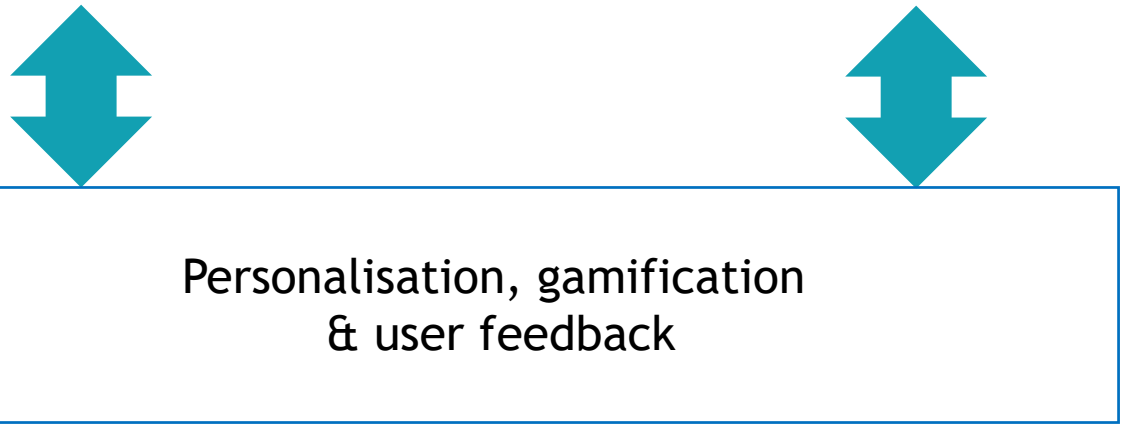
- is **irrelevant** for the intended purpose
- but may **bias** the prediction and
- may **infringe** upon the privacy of users

The objective

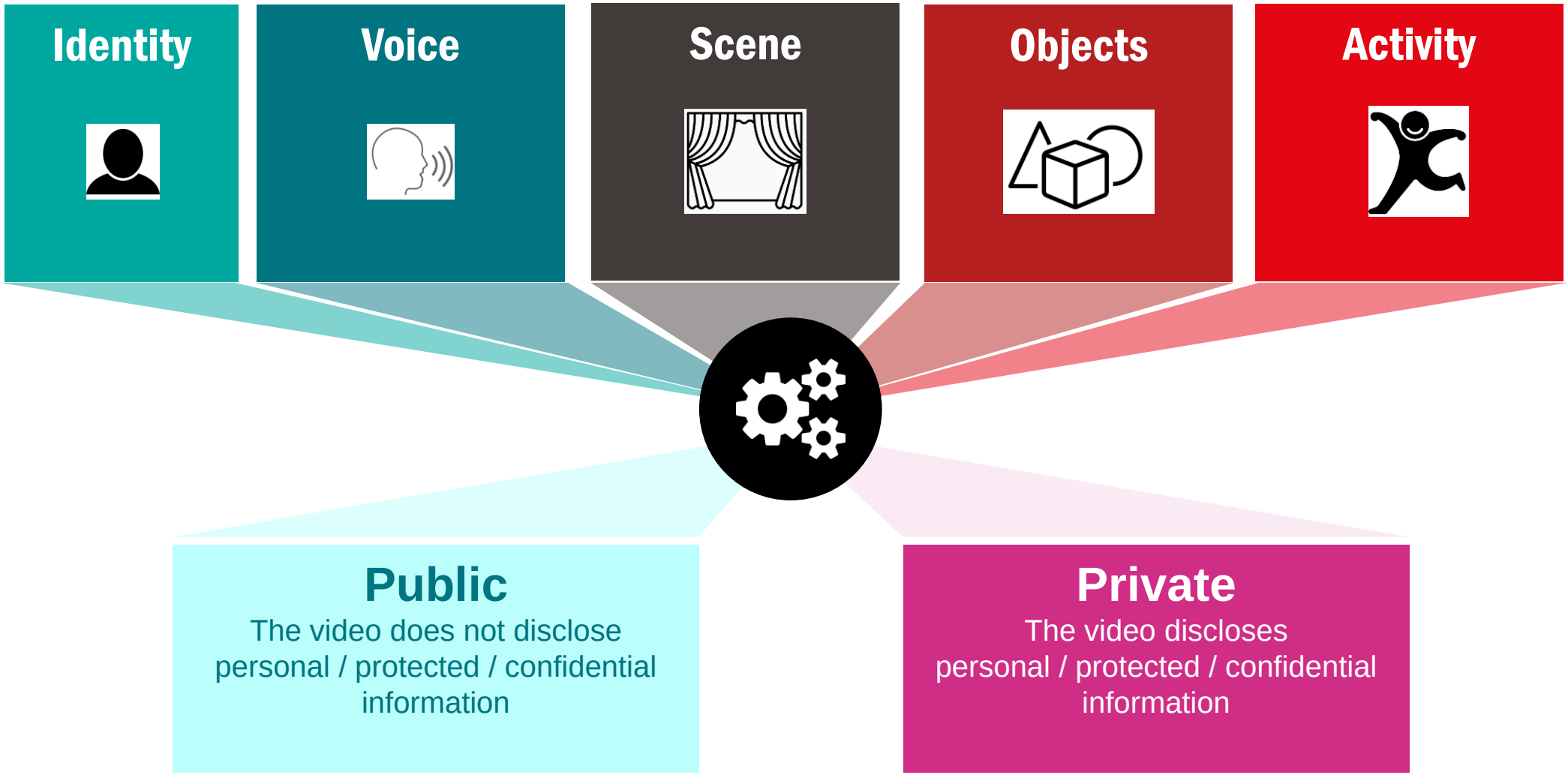
To capture complex patterns across multiple biological **data layers**

To highlight the **key elements** of the patterns

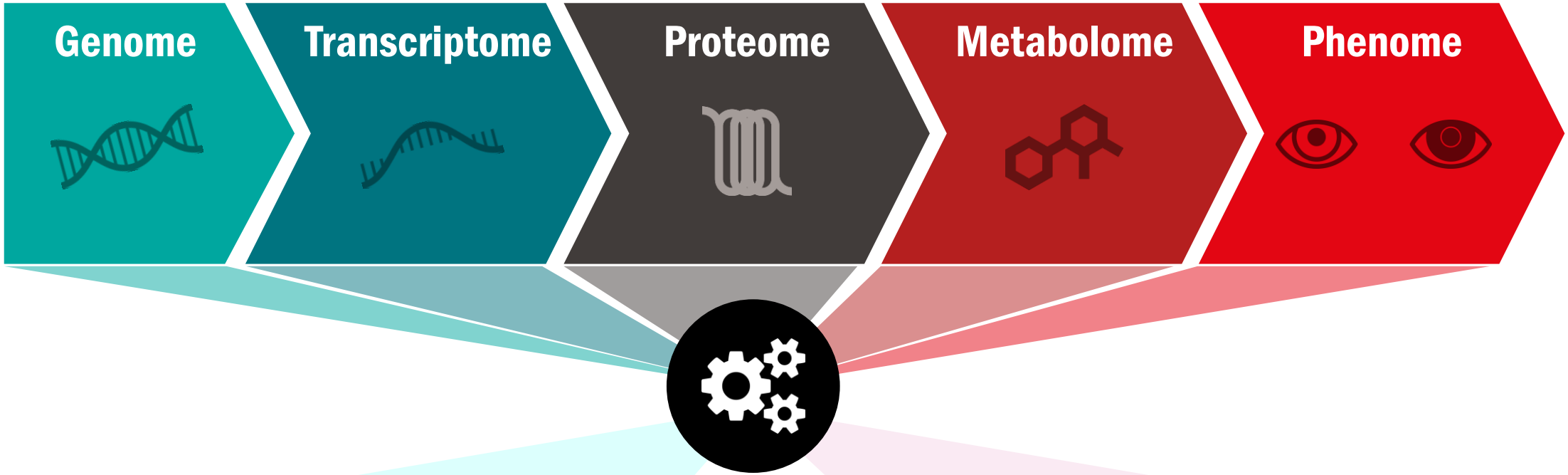
To discover complex and/or hidden **relationships** between similar, hierarchical or correlated (private) concepts in audio and video data



GraphNEx
Graph Neural Networks
for Explainable Artificial Intelligence



GraphNEx
Graph Neural Networks
for Explainable Artificial Intelligence



Disease

GeneY associated with disease
AND High expression in Liver
AND High blood pressure

No disease

GeneX associated with no disease
AND Low expression in Liver
AND Normal blood pressure

Adapted from:
- Li, H. et *al.* An Integrated Systems Genetics and Omics Toolkit to Probe Gene Function. *Cell Syst* 6**,** 90–120.e4 (2018).
- Zitnik, M. et *al.* Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Inf Fusion* 50, 71-92 (2019).

90 strains

Phenotyping tests

Activity
Cognition
Fitness
Vision
Metabolism
Body composition
Glucose tolerance

Birth

20 months

Bio-molecular assays

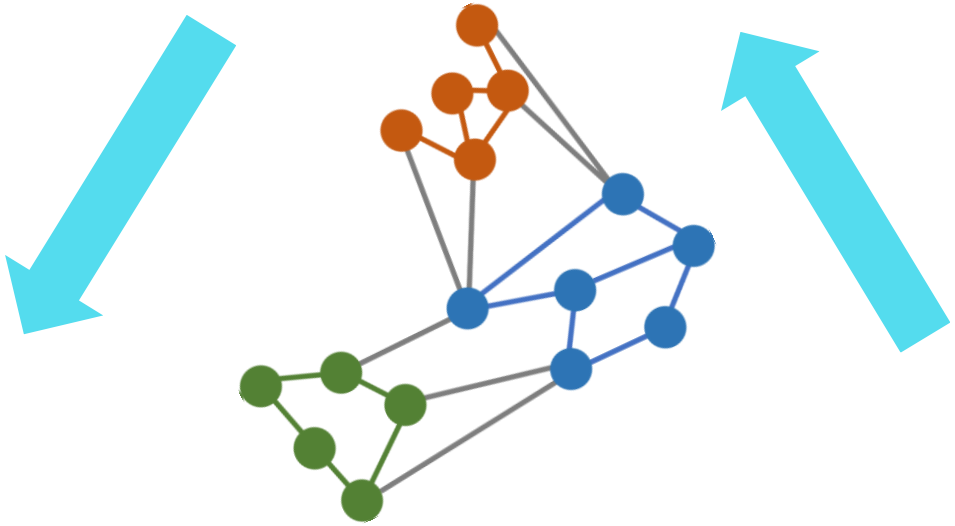
1-35 tissues
25,000 genes transcripts
2,622 proteins
979 metabolites
~200 lipids

GraphNEx
Graph Neural Networks
for Explainable Artificial Intelligence

Multiple sources of information

Networks

Biological network
Entities in images/audio/videos



Learn

Graph Machine Learning

Explain

On which inputs do the predictions depend?

Interact

GraphNEx
Graph Neural Networks
for Explainable Artificial Intelligence

Graph Machine Learning

Learning from several sources of information on one or multiple graphs

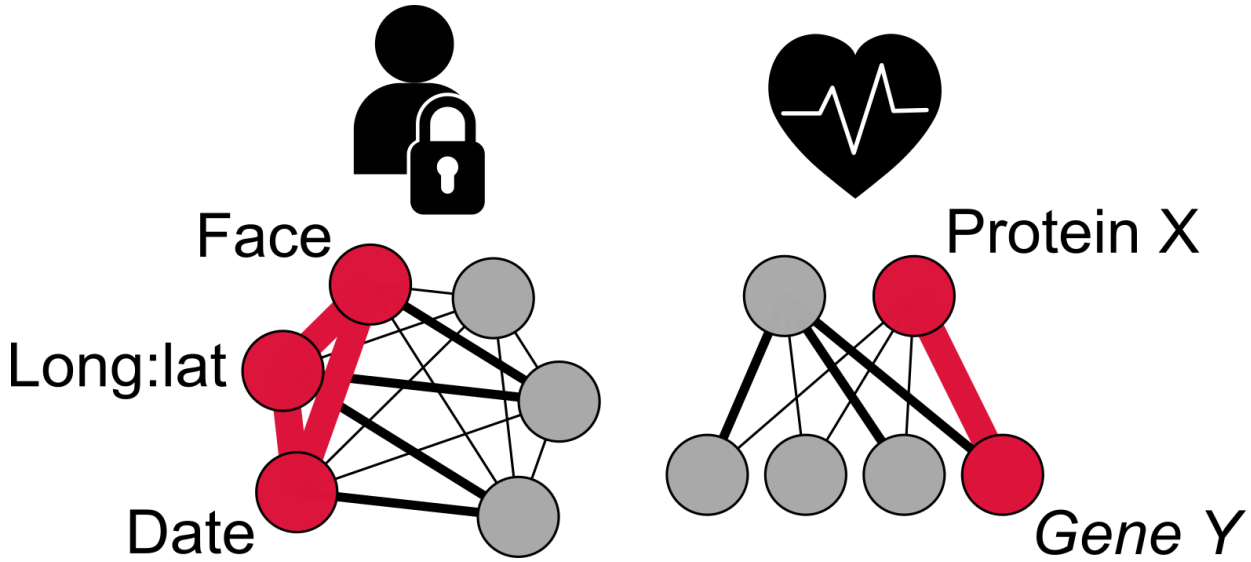
- Prior knowledge: curated databases, biological ontologies, privacy laws and regulations
- Experimentally derived information (co-expression graph, protein-protein interactions, entity co-occurrence)
- Graph coarsening for explanations to emerge
  - private concepts: hierarchy of general concepts
  - list of genes: larger groups of genes with high co-expression or with similar functions

Challenges

- Heterogeneous viewpoints on genes / on people captured in videos
- Strong interdependencies only sparsely known
- Missing data or explicit concepts

Explanations: visualize, evaluate, interact

- User-guided explainability in GraphNEx:
  - Investigating evaluation strategies for different stakeholders
  - Gamification interfaces for users with different levels of expertise
  - Stakeholder-based adaptation of the graph structure
- Post-hoc explainability in GraphNEx:
  - Diagnostic graph for probing non-explainable AI systems
  - Quantifying how responses correlate with the concepts in the (explainable) concept graph
  - Assessing the reliability and robustness of explanations



GraphNEx
Graph Neural Networks
for Explainable Artificial Intelligence

GraphNEx

Graph Neural Networks for Explainable Artificial Intelligence

graphnex.eecs.qmul.ac.uk

Queen Mary University of London

EPFL

Université de Lyon

ENS de Lyon

GraphNEx
Graph Neural Networks
for Explainable Artificial Intelligence