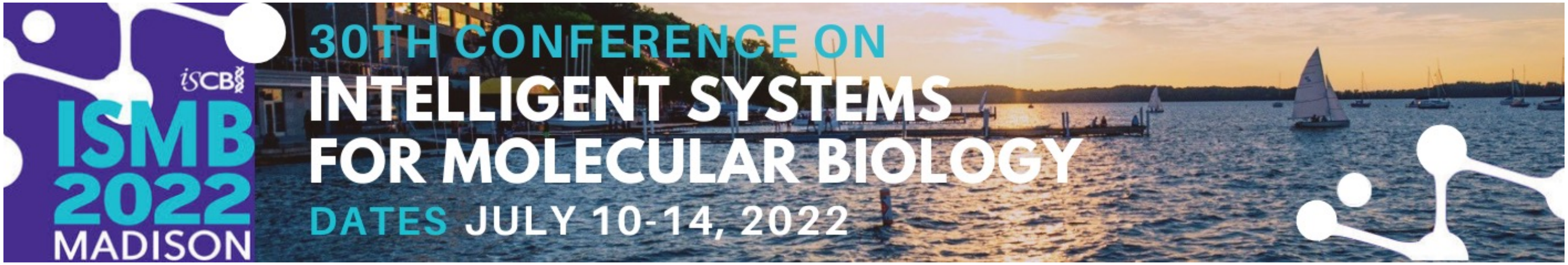
Towards Precision Medicine with Graph Representation Learning

Michelle M. Li & Marinka Zitnik

Department of Biomedical Informatics
Broad Institute of Harvard and MIT
Harvard Data Science

zitniklab.hms.harvard.edu/biomedgraphml



Tutorial VT4

July 7, 2022 at 9am – 1pm CDT

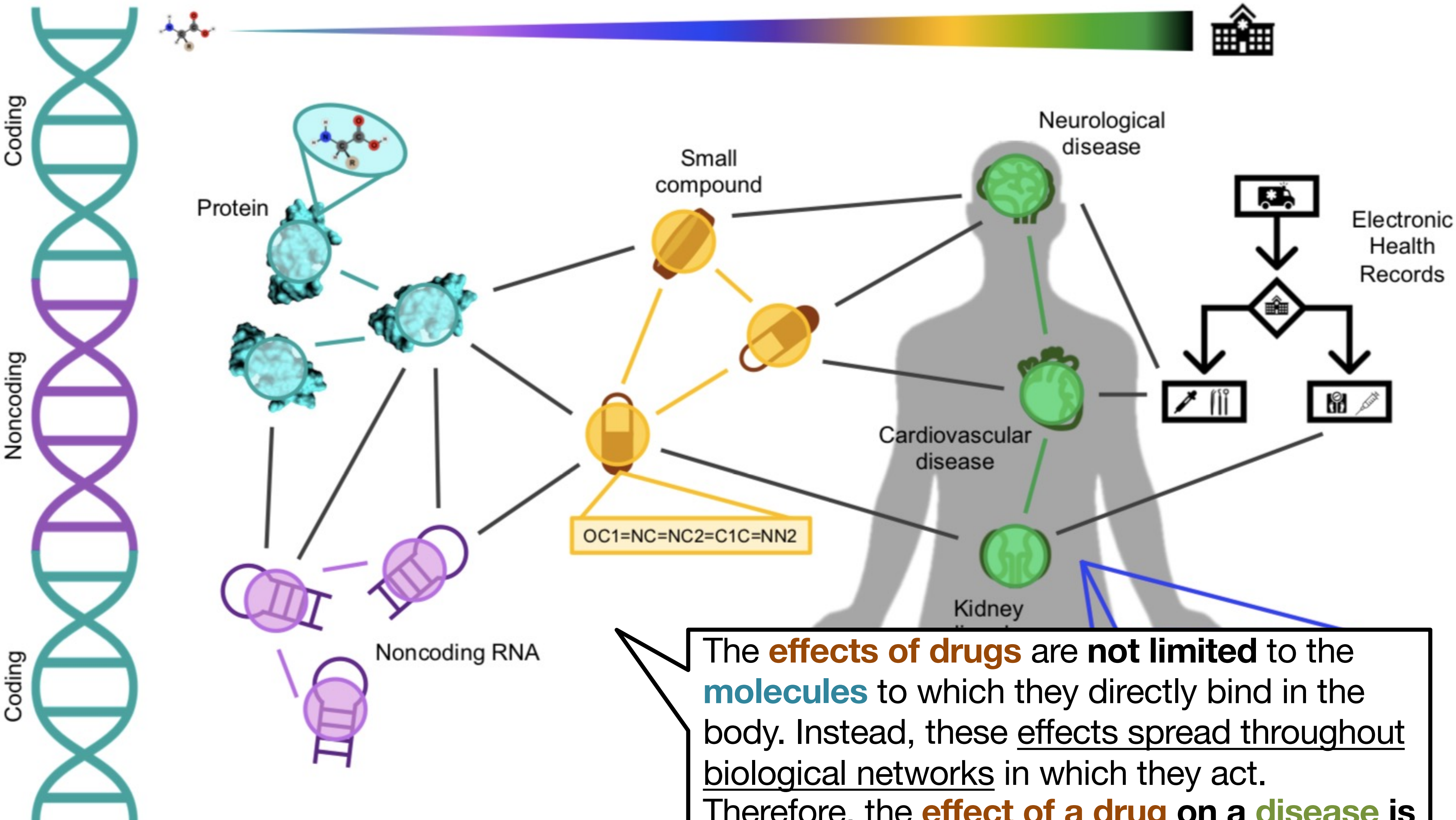
All tutorial materials are available at

zitniklab.hms.harvard.edu/biomedgraphml

JUNO Live Logistics

- Chat & emoji buttons: Interact with us and others in the tutorial
- Q&A button: Ask us questions & upvote your favorite questions
- Poll button: Participate in our mini polls throughout the tutorial
  - Where are you from (e.g., geographically, institution)? *Fill in the blank.*
  - What is your position (e.g., PhD student, data scientist, postdoc, clinician)? *Fill in the blank.*
  - How would you rate your familiarity with graph representation learning (1 = novice, 5 = expert)? *Rating.*
  - How would you rate your familiarity with biology/medicine (1 = novice, 5 = expert)? *Rating.*
  - What do you hope to get out of this tutorial? *Fill in the blank.*

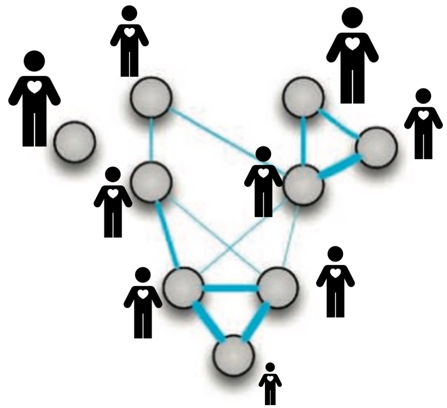
Biology is interconnected



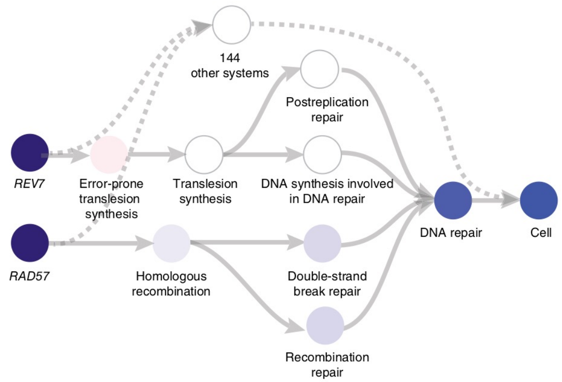
The **effects of drugs** are **not limited** to the **molecules** to which they directly bind in the body. Instead, these effects spread throughout biological networks in which they act. Therefore, the **effect of a drug on a disease is inherently a network phenomenon**

Graph Representation Learning for Biomedicine, *Nature Biomedical Engineering* (in press), 2022, arXiv:2104.04883

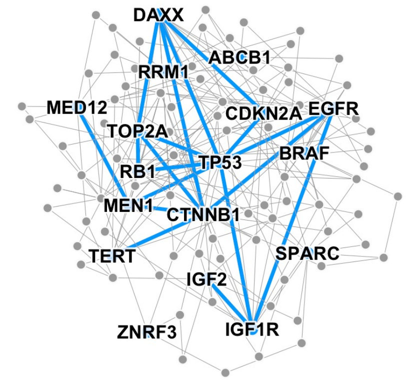
Networks are a general language for describing and modeling complex systems



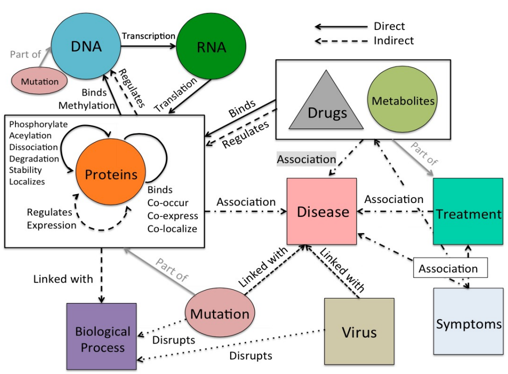
Patient networks



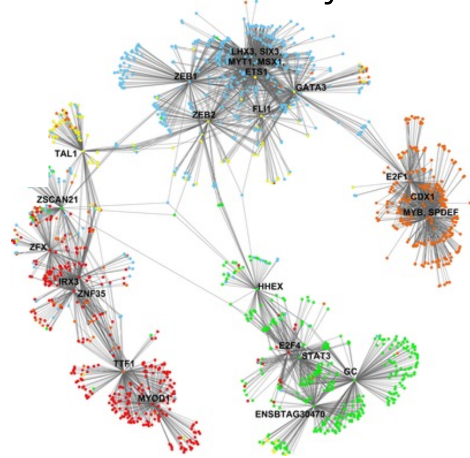
Hierarchies of cell systems



Disease pathways



Biomedical knowledge graphs



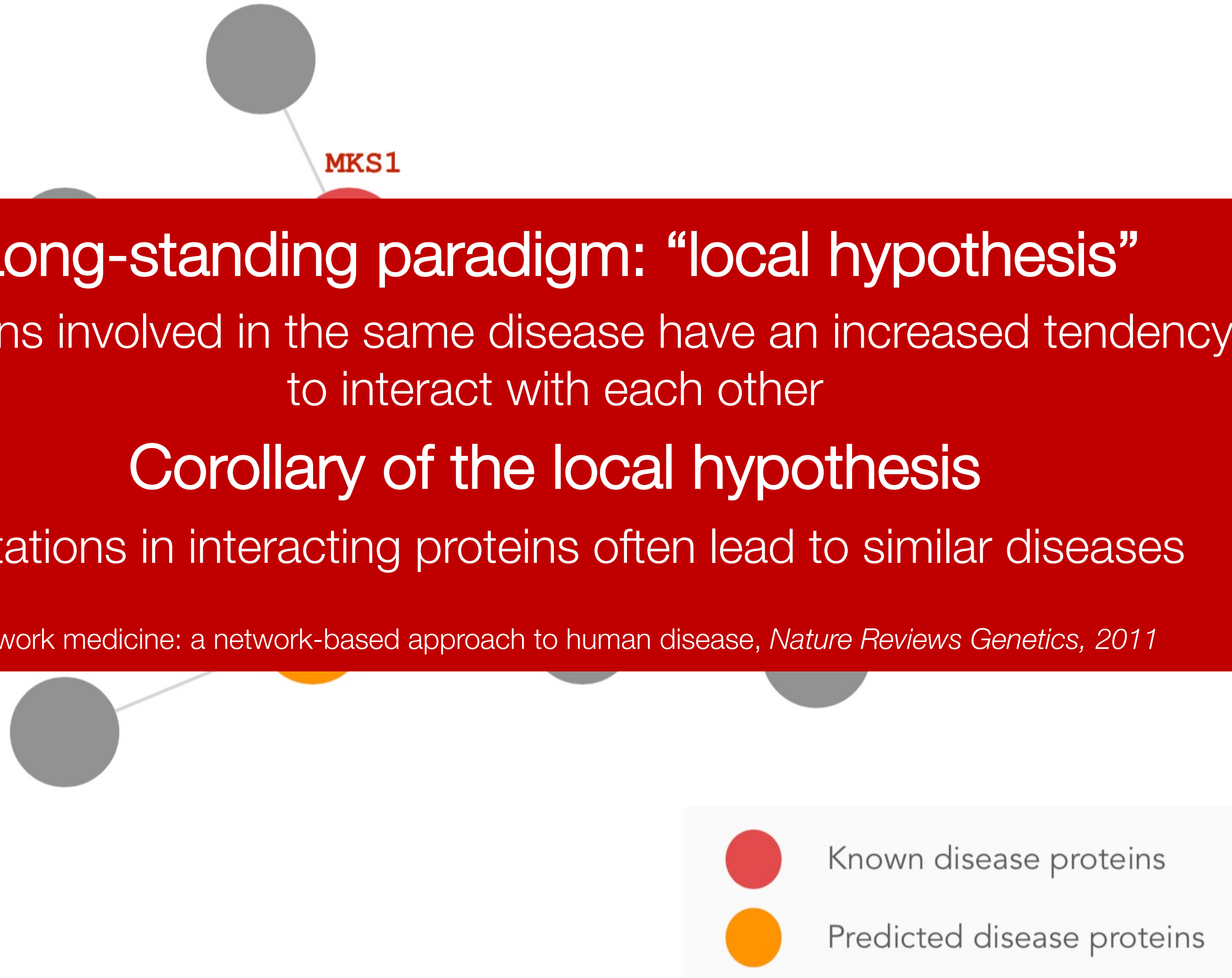
Gene interaction networks

Cell-cell similarity networks

Graph Representation Learning for Biomedicine, *Nature Biomedical Engineering* (in press), 2022, arXiv:2104.04883

Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities, *Information Fusion* 2019

Why networks in biology?



Long-standing paradigm: “local hypothesis”

Proteins involved in the same disease have an increased tendency to interact with each other

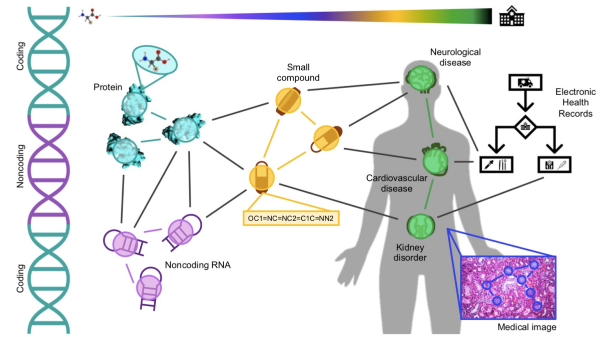
Corollary of the local hypothesis

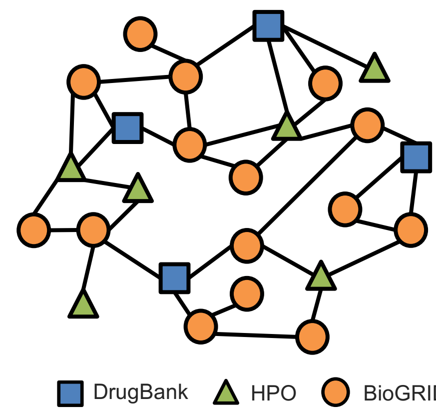
Mutations in interacting proteins often lead to similar diseases

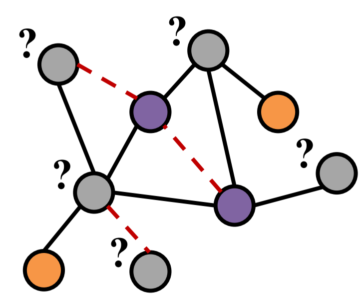
Network medicine: a network-based approach to human disease, *Nature Reviews Genetics, 2011*

Why are biological networks challenging?

- Heterogeneous interactions that span from molecules to whole populations
  - Challenge: Computationally operationalize these data & make them amenable to ML
- Requires data from diverse sources, including experimental readouts, curated annotations, metadata
  - Challenge: Capture all factors necessary to understand a phenomenon (e.g. disease)
- Noisy due to inherent natural variations & limitations of measurement platforms
  - Challenge: Handle missing data, repeated measurements, and contradictory observations



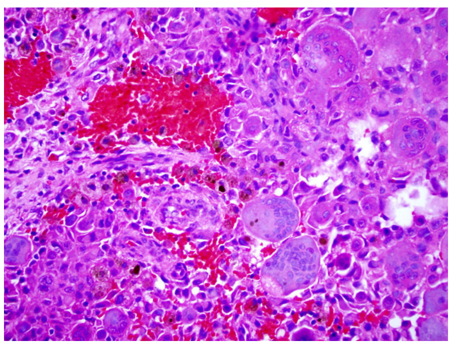


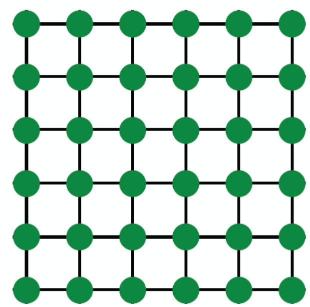


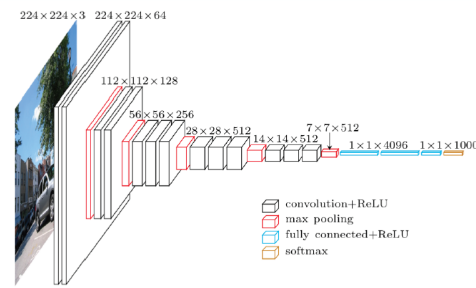
Graph Representation Learning for Biomedicine, *Nature Biomedical Engineering* (in press), 2022, arXiv:2104.04883

Classic deep learning

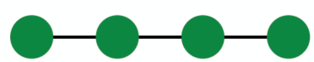
- Primarily designed for grids or simple sequences:
  - CNNs for fixed-size images/grids…

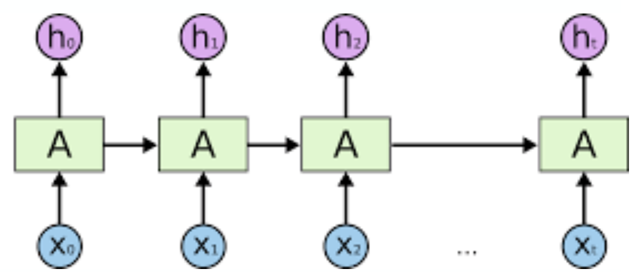






  - RNNs for text and sequences…

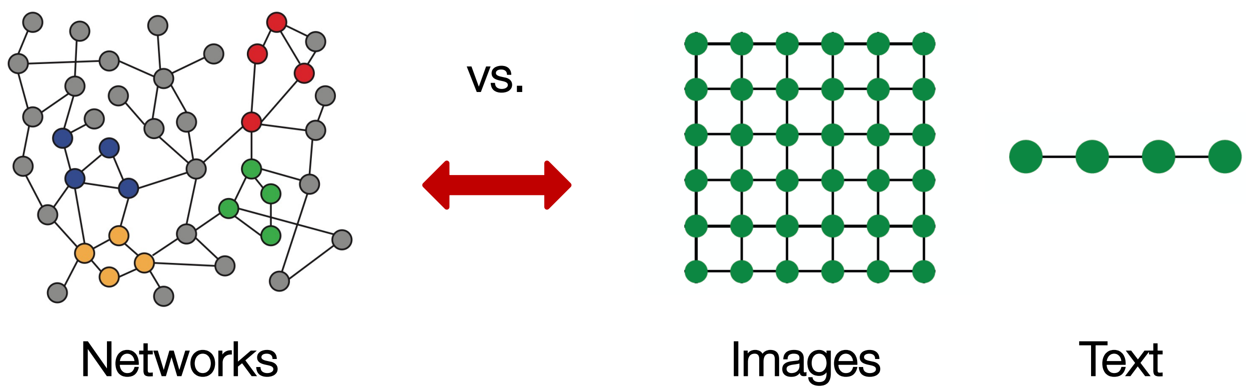




Graph Representation Learning for Biomedicine, *Nature Biomedical Engineering* (in press), 2022, arXiv:2104.04883

Classic deep learning

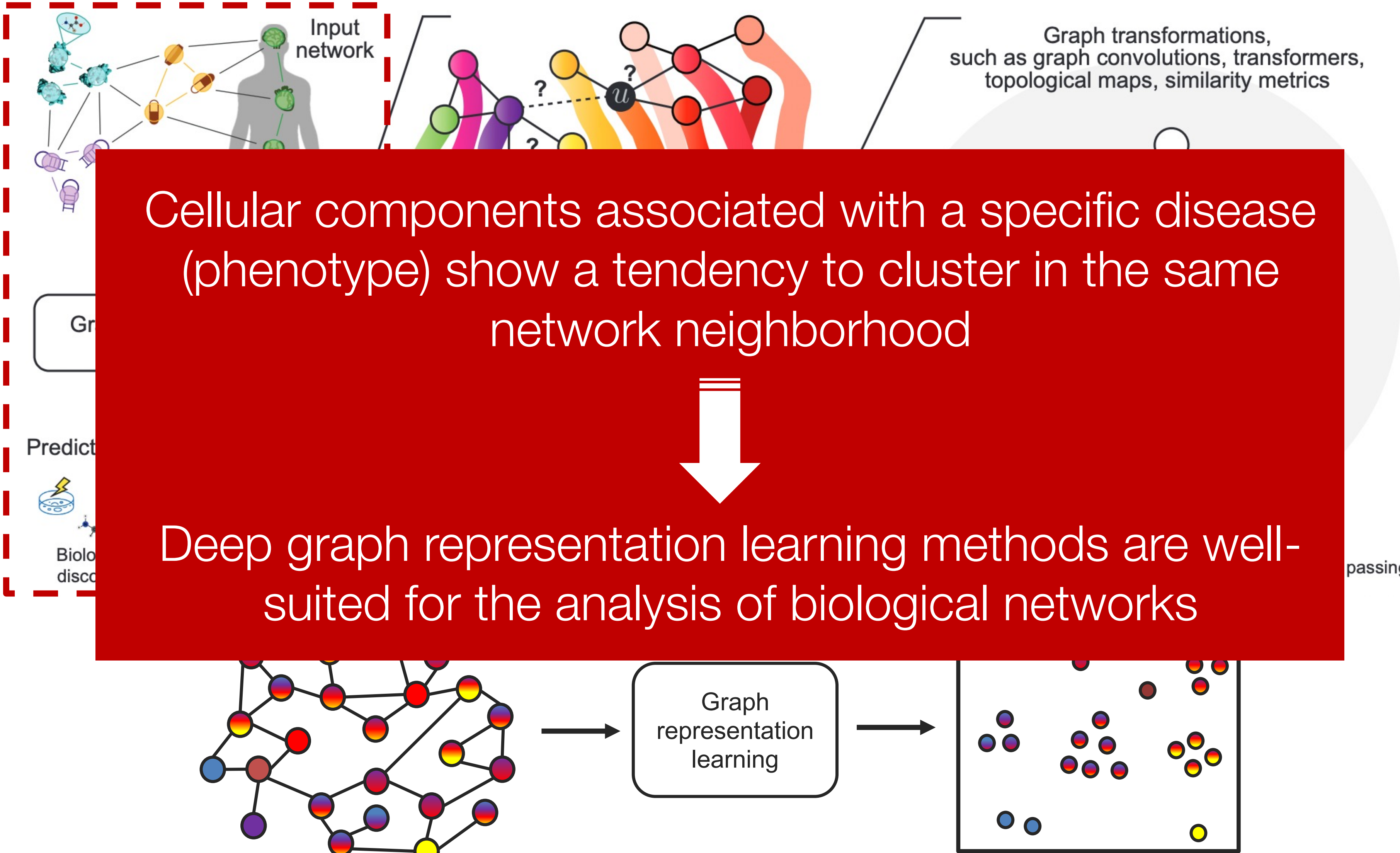
- Networks are far more complex!
  - Arbitrary size and complex topological structure (i.e., no spatial locality like grids)



  - No fixed node ordering or reference point
  - Often dynamic and have multimodal features

Graph Representation Learning for Biomedicine, *Nature Biomedical Engineering* (in press), 2022, arXiv:2104.04883

Graph representation learning realizes key network principles for data-rich biomedicine



Cellular components associated with a specific disease (phenotype) show a tendency to cluster in the same network neighborhood

↓

Deep graph representation learning methods are well-suited for the analysis of biological networks

Graph Representation Learning for Biomedicine, *Nature Biomedical Engineering* (in press), 2022, arXiv:2104.04883

Tutorial Topics & Objectives

1. Methods: Summarize and contrast the major paradigms of graph representation learning

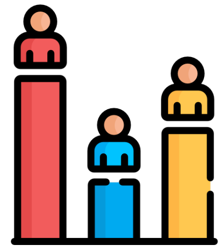
   Network diffusion, shallow network embeddings, graph neural networks, equivariant neural networks

2. Applications: Determine a graph representation learning method's utility for the biomedical learning task and network of interest

   Fundamental biological discoveries and precision medicine enabled by graph representation learning

3. Hands-on exercises: Identify new opportunities in biomedicine to leverage graph representation learning methods

   Demos, implementation details, tools, and tips



Time for a poll question about…

NETWORKS FOR BIOMEDICINE

1. Which of the following is a long-standing paradigm that empowers the use of networks for biology and medicine? *Multiple choice*
2. Why are classic deep learning methods unsuited to handle biomedical networks? *Select many*

Resources

- Books & survey papers
  - William Hamilton, *Graph Representation Learning* (morganclaypool.com/doi/abs/10.2200/S01045ED1V01Y202009AIM046)
  - Li et al., Graph Representation Learning for Biomedicine (arxiv.org/abs/2104.04883)

- Keynotes & seminars
  - Michael Bronstein, "Geometric Deep Learning: The Erlangen Programme of ML" (ICLR 2021 keynote) (youtube.com/watch?v=w6Pw4MOzMuo)
  - Broad Institute Models, Inference & Algorithms: Actionable machine learning for drug discovery; Primer on graph representation learning (youtube.com/watch?v=9YpTYdru0Rg)
  - Stanford University (CS224W Lecture): Graph neural networks in computational biology (youtube.com/watch?v=_hy9AgZXhbQ)
  - AI Cures Drug Discovery Conference (youtube.com/watch?v=wNXSkISMTw8)

- Conferences & summer schools
  - London Geometry and Machine Learning Summer School (logml.ai)
  - Learning on Graphs Conference (logconference.github.io)

Resources

- Software & packages
  - PyTorch Geometric
  - NetworkX
  - Stanford Network Analysis Platform (SNAP)
- Tutorials & code bases
  - Pytorch Geometric Colab Notebooks (pytorch-geometric.readthedocs.io/en/latest/notes/colabs.html)
  - Zitnik Lab Graph ML Tutorials (github.com/mims-harvard/graphml-tutorials)
  - Stanford University's CS224 (web.stanford.edu/class/cs224w)
- Datasets
  - Precision Medicine Oriented Knowledge Graph (PrimeKG) (zitniklab.hms.harvard.edu/projects/PrimeKG)
  - Therapeutic Data Commons (TDC) (tdcommons.ai)
  - BioSNAP (snap.stanford.edu/biodata/)
  - Open Graph Benchmark (OGB) (ogb.stanford.edu)