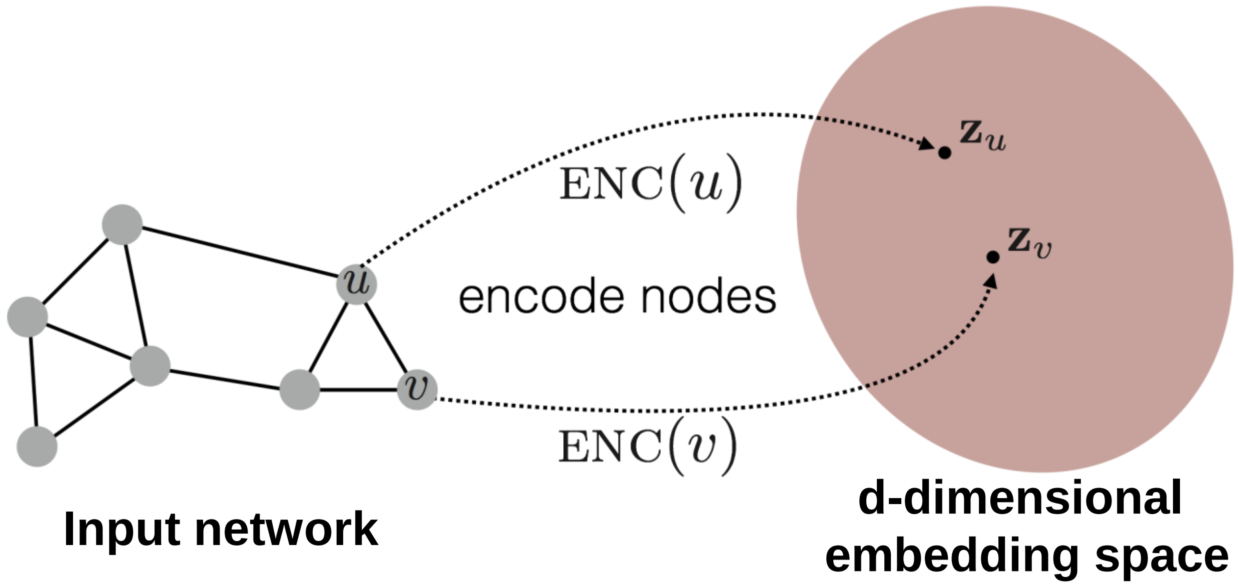
Embedding is not the most important!

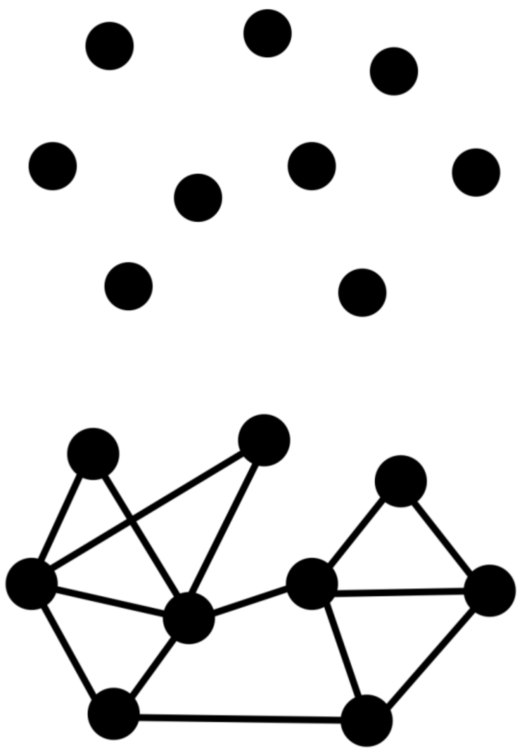


The embedding performs poorly when

- SNR is too low in dataset.
- the relationship is incorrect.
- the relationship is redundancy.

Dataset and relationship is much important!

How to build a graph? GNN?



Similarity:

PCC

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

An important topic!

Grale: Designing Networks for Graph Learning

Jonathan Halcrow[†], Alexandru Moșoi[‡], Sam Ruth[†], Bryan Perozzi[†]
†: Google Research
‡: YouTube
{halcrow,mosoi,samruth}@google.com,bperozzi@acm.org

- In real world applications, the choice of which edges to use for computation is the first step in any graph learning process.
- The choice of edges can drastically affect the performance of downstream semi-supervised learning systems.
- Grale: graph design for graphs with billions of nodes
- Grale operates by fusing together different measures of (potentially weak) similarity to create a graph which exhibits high task-specific homophily between its nodes.

Notation

consider a general multiclass learning setting

a partially labeled set of points $\mathcal{X} = \{x_1, x_2, \ldots, x_V\}$

the first L points have class labels $\mathrm{Y} = \{\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_L\}$

y_k a one-hot vector of dimension C

$x_i = \{x_{i,1}, x_{i,2}, \ldots x_{i,d}\}$

Each sub-representation $x_{i,d}$ has its own natural distance measure κ_d

Loss function

a classification problem: $y_k = \hat{y}(\mathrm{x_k})$

$$\mathcal{L} = -\sum_{i \in \mathcal{Y}} \sum_{c \in C} y_{i,c} \log \hat{y}_{i,c}$$

assume a graph G = (V , E):

$$\hat{y}_{i,c}^{(n+1)} = \alpha y_{i,c} + \beta \frac{\sum_{j \in \mathcal{N}_i} w_{i,j} \hat{y}_{j,c}^{(n)}}{\sum_{j \in \mathcal{N}_i} w_{i,j}}$$

$$\mathcal{L} = \sum_{i,j \in E} w_{i,j} \sum_{c \in C} |\hat{y}_{i,c} - \hat{y}_{j,c}| + \sum_{i \in V, c \in C} |\hat{y}_{i,c} - y_{i,c}|$$

Potts model type loss function

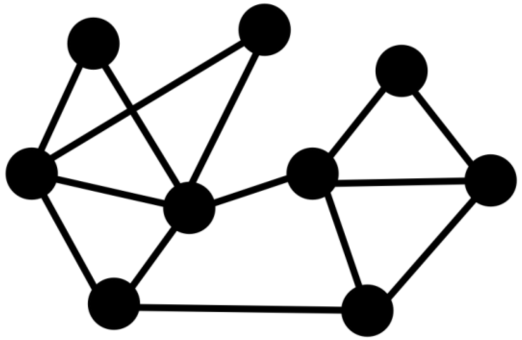
GRALE

Design **task specific graphs** by reducing the original task to training a classifier on pairs of points

$$w_{ij} := \log G(x_i, x_j) \quad G : \mathrm{R}^d x \mathrm{R}^d \to \mathrm{R} \quad 0 < G < 1 \quad w_{ij} < 0$$

$$G(x_i, x_j) = f(\kappa_1(x_i, x_j), \kappa_2(x_i, x_j), \ldots, \kappa_d(x_i, x_j))$$

$$\mathcal{L} = -\sum_{c \in C} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{X}} y_{i,c}, y_{j,c} \log G(x_i, x_j)$$



GRALE

一种计算相似性的神经网络

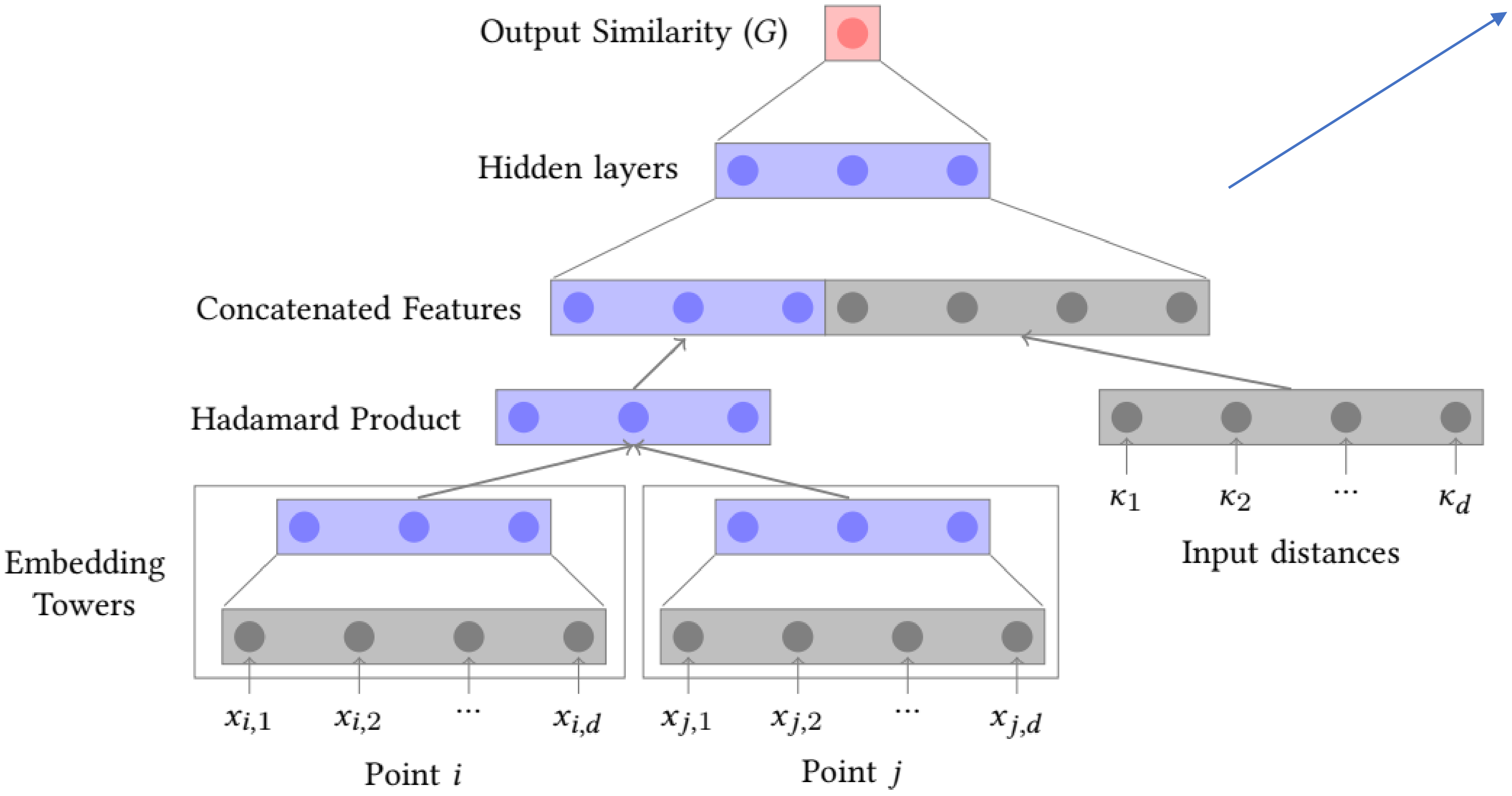
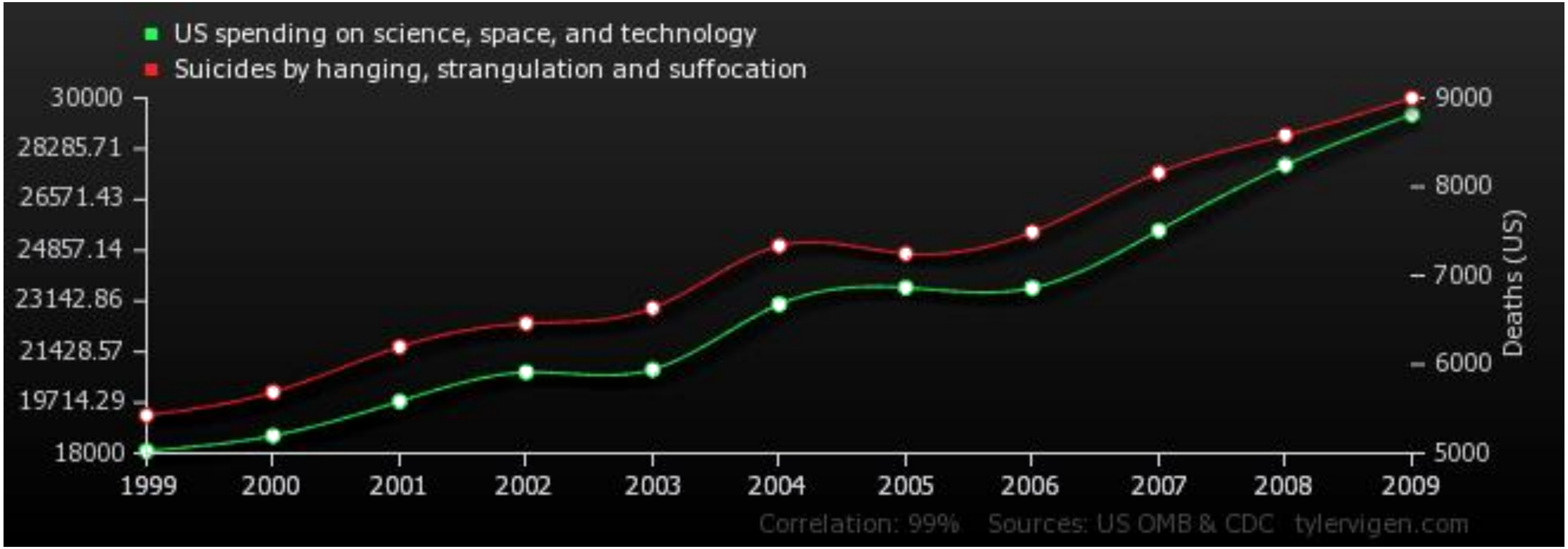


Figure 2: The Grale Neural Network model. The gray nodes are inputs, the blue are hidden layers, and red is the output. The network architecture combines a standard two-tower model with natural distances in the input feature spaces. Weights are shared between the two towers. The Hadamard product (pointwise multiplication) of the two towers is used to give us a pairwise embedding. We treat this as an additional set of distance features to augment the input distance features $\kappa_1(x_i, x_j), \ldots, \kappa_d(x_i, x_j)$. This combined set acts as an input to the second part of the model which computes a single similarity score.

Spurious Correlations

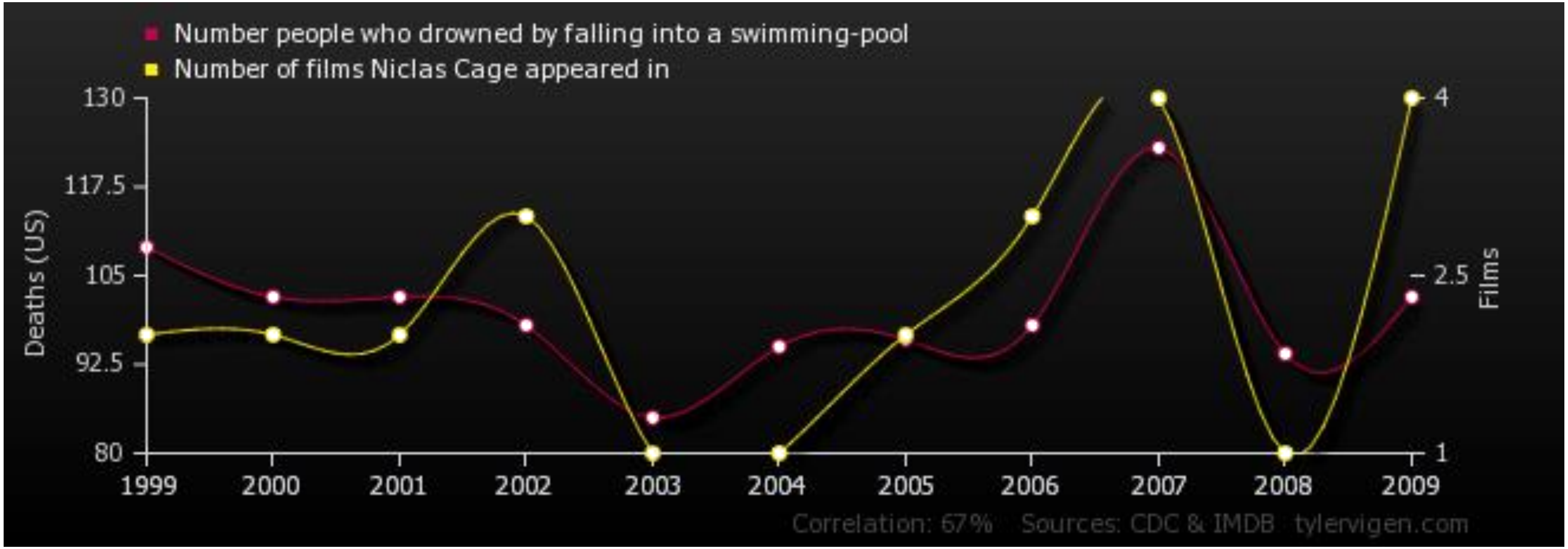
- Statistical significance: Standard threshold is 5%, meaning would be observed by chance 1 out of 20 times
- But if you look at enough relationships you'll find "significant" ones by chance
- Spurious means there's no real connection

Spurious Correlations



US Spending on Science, Space and Technology Correlates with Suicides by Hanging, Strangulation and Suffocation

Spurious Correlations



Number of People Who Drowned by Falling into a Swimming Pool
Correlates with Number of Nicholas Cage Films

Big Data Analysis Requires Causality

- Big data is not randomized
- We have to compare treatments when patients and providers choose what they think is best (AB test)
- We want to choose treatments and design interventions
- Causal inference is crucial

If We Don’t Think About Causality . . .

- Our predictions can be fragile
- Our interventions can be misguided

Causal Inference

Relating Graph Neural Networks to Structural Causal Models

Matej Zečević MATEJ.ZECEVIC@CS.TU-DARMSTADT.DE
Computer Science Department, TU Darmstadt

Devendra Singh Dhami DEVENDRA.DHAMI@CS.TU-DARMSTADT.DE
Computer Science Department, TU Darmstadt

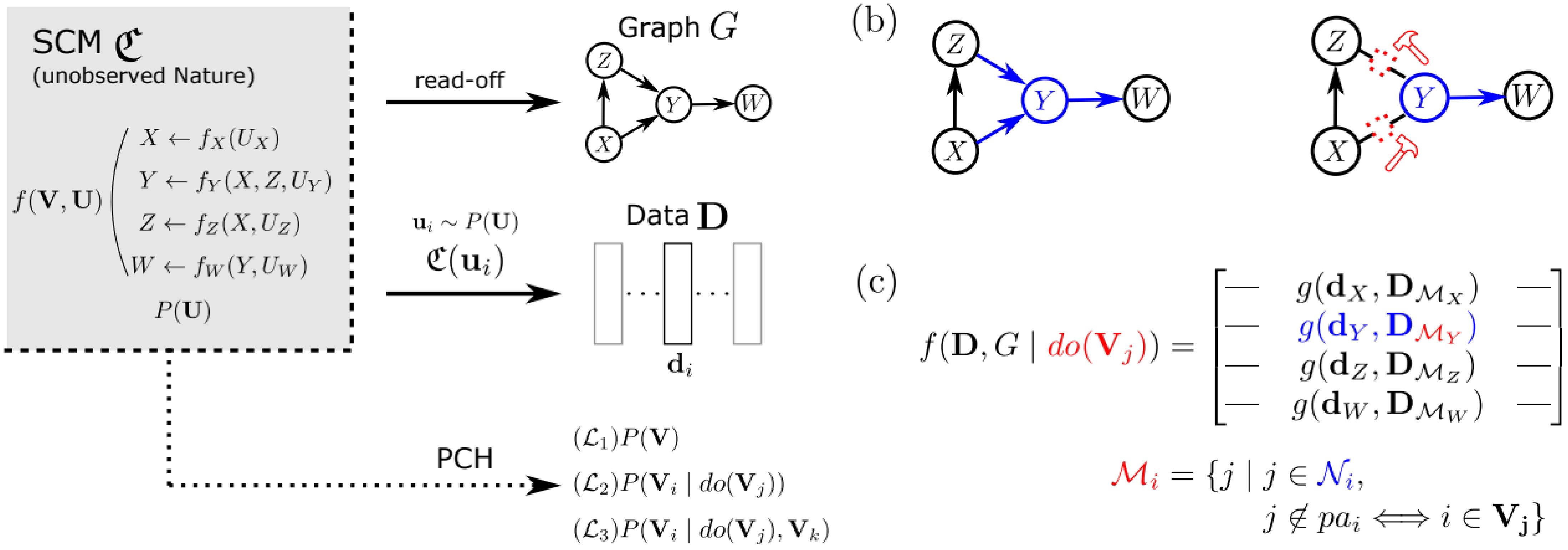
Petar Veličković PETARV@GOOGLE.COM
DeepMind, London

Kristian Kersting KERSTING@CS.TU-DARMSTADT.DE
Computer Science Department, Centre for Cognitive Science, TU Darmstadt, and Hessian Center for AI (hessian.AI)

https://www.youtube.com/watch?v=XC-Bfg3dO0I&ab_channel=ValenceDiscovery
Pearl Causal Hierarchy (PCH) - 馒头and花卷 - 博客园 (cnblogs.com)
https://www.nature.com/articles/s41598-021-87411-8
https://proceedings.mlr.press/v139/lin21d.html
CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting | Proceedings of the AAAI Conference on Artificial Intelligence

Relating Graph Neural Networks to Structural Causal Models

AB实验的形式化



Judea Pearl提出的“因果阶梯”论（Pearl Causal Hierarchy，PCH）

CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting

构建差分方程
来考虑因果性

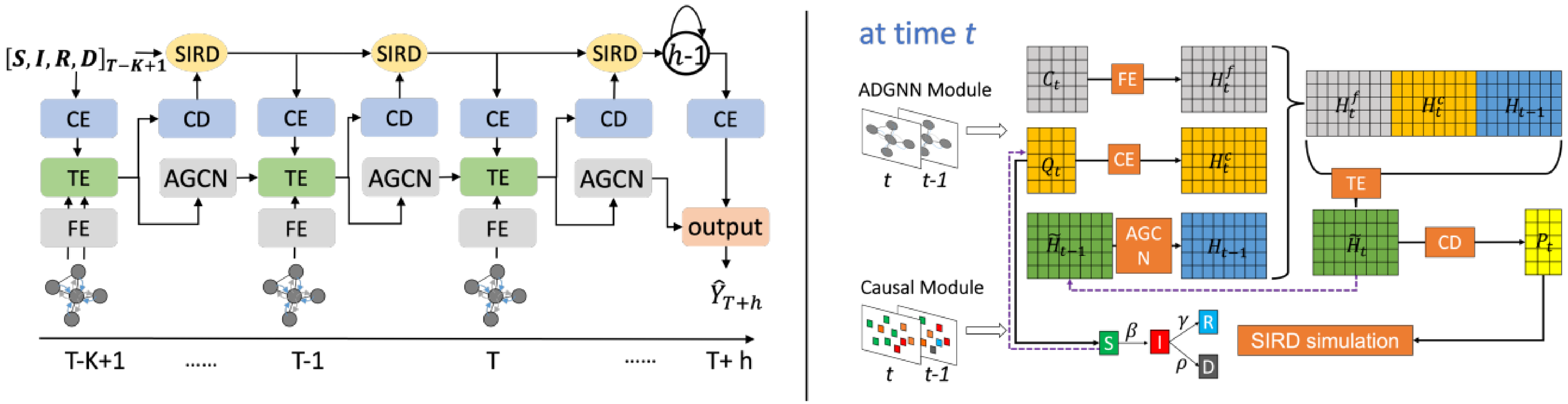


Figure 1: Framework of CausalGNN which consists of a causal module and an attention-based dynamic GNN module.

FE: feature encoding

CE: causal encoding

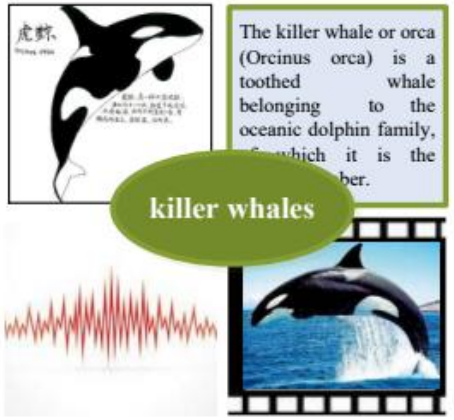
AGCN: dynamic graph encoding

TE: temporal encoding

CD: causal decoding

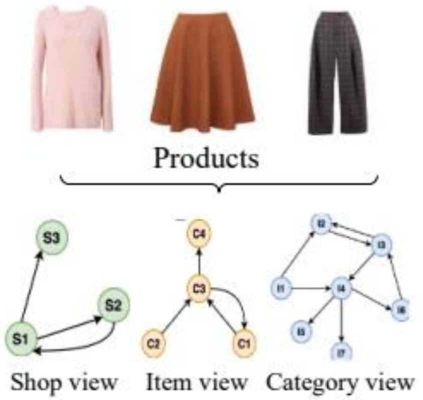
SIRD: susceptible(S)-infected(I)- recovered(R)-deceased(D)

Multi-view data

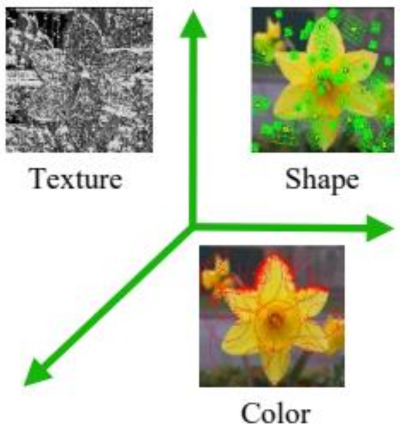


(a) A object is described by text, video, audio

(b) A news is reported by different languages



(c) A product can be represented by multi-view graphs



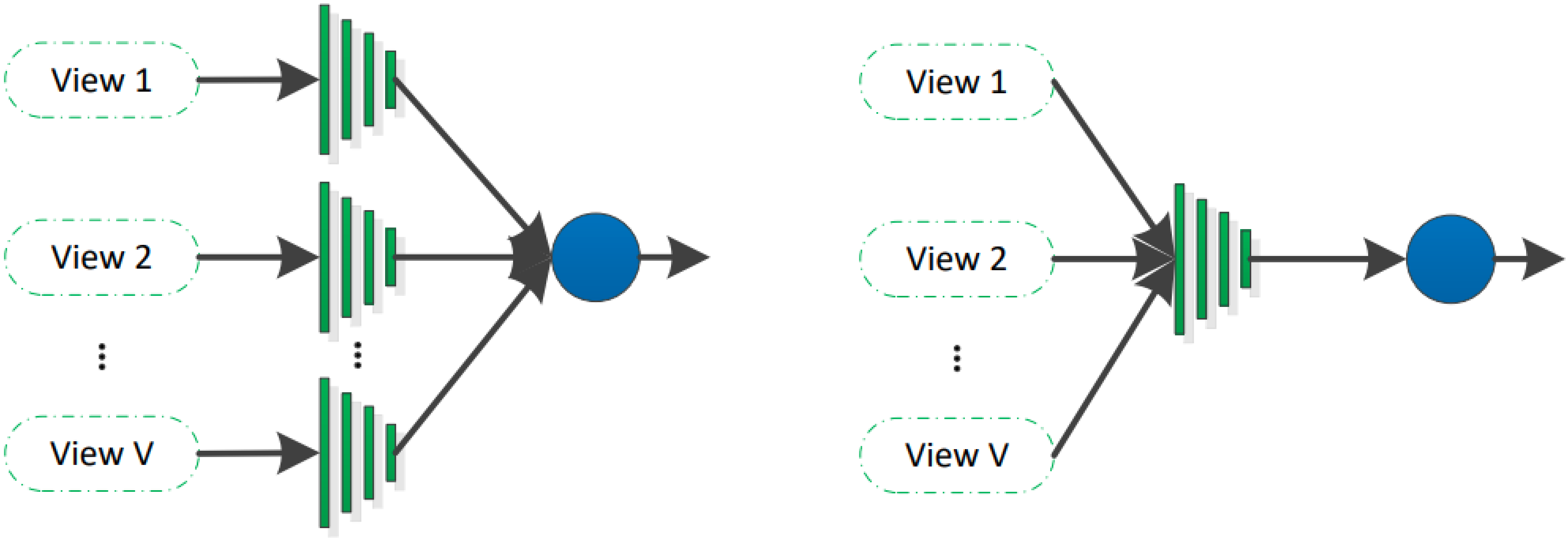
(d) An image is depicted by heterogonous features

(e) A social image along with user tags

(f) One action can be captured from different camera viewpoints

Multi-view learning

- Multi-modal?

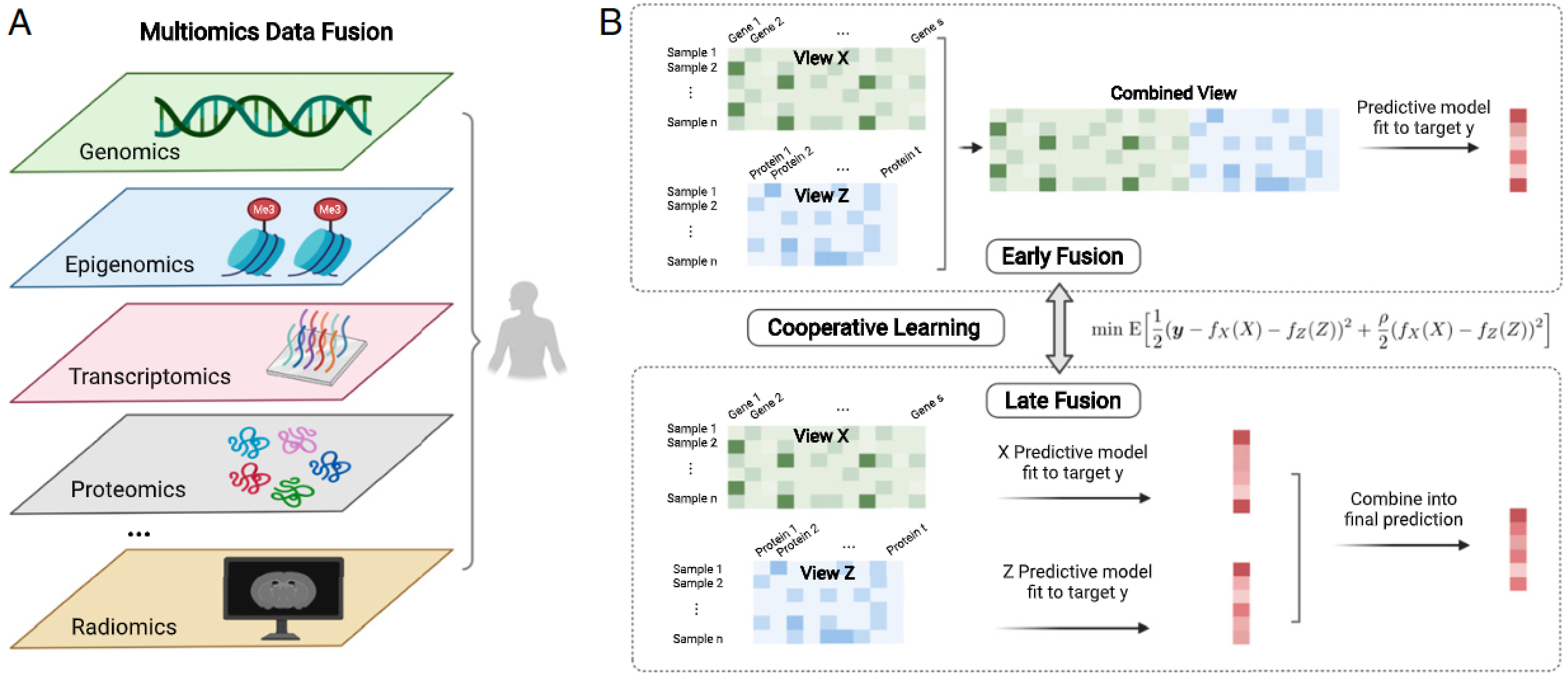


(a) One-view-one-network strategy

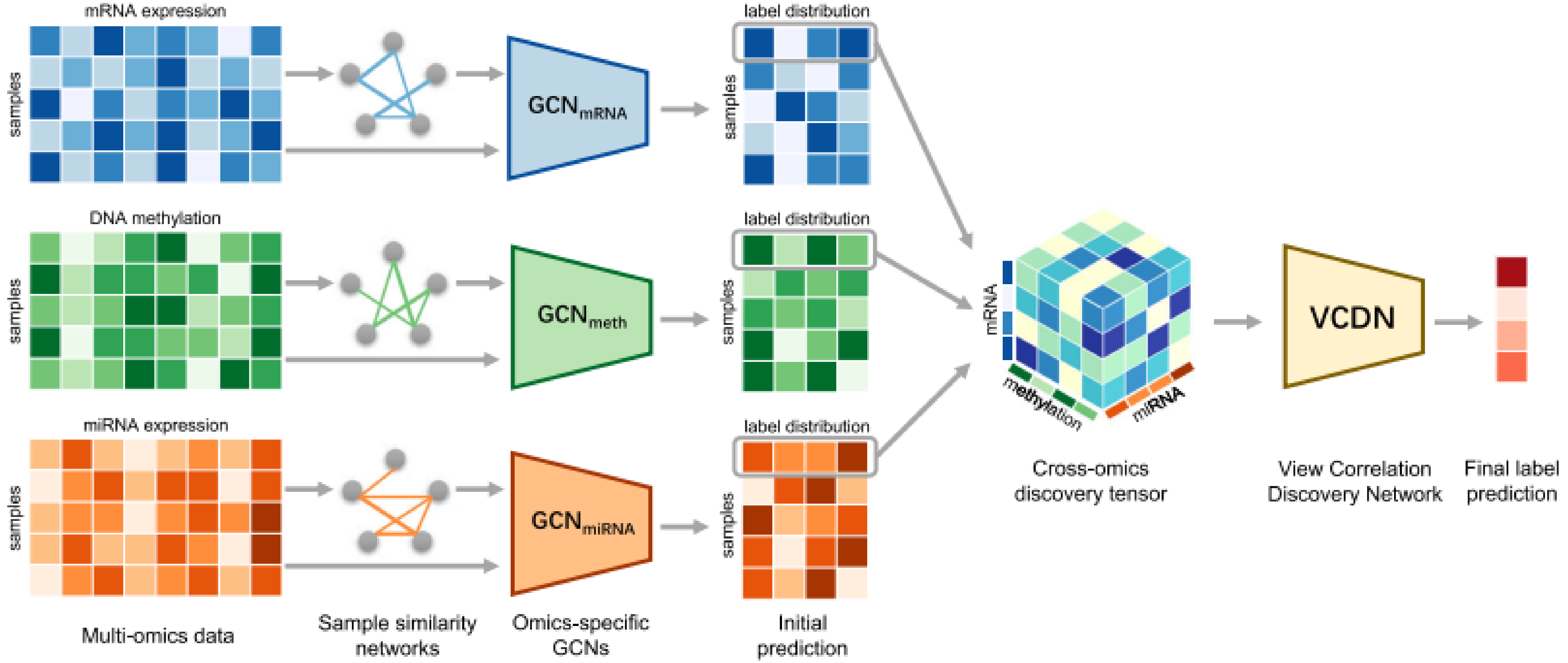
(b) Multi-view-one-network strategy

[Deep multi-view learning methods: A review - ScienceDirect]()

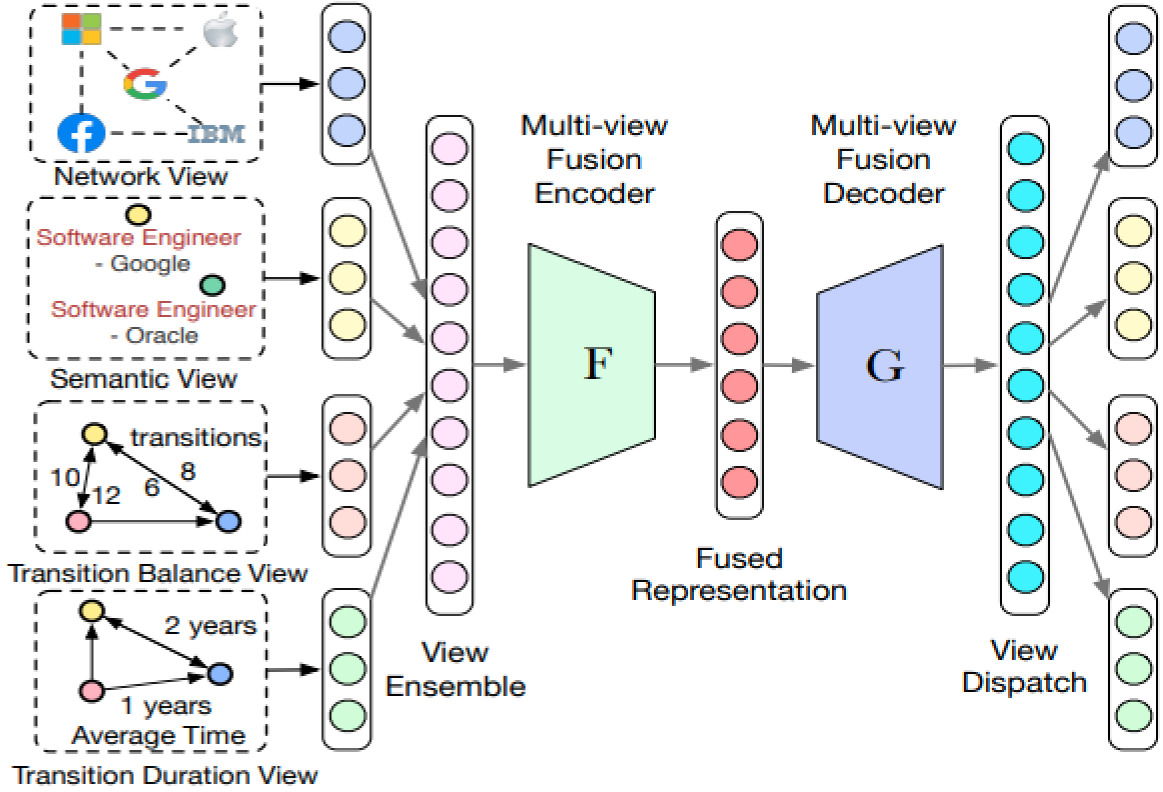
Multi-view learning



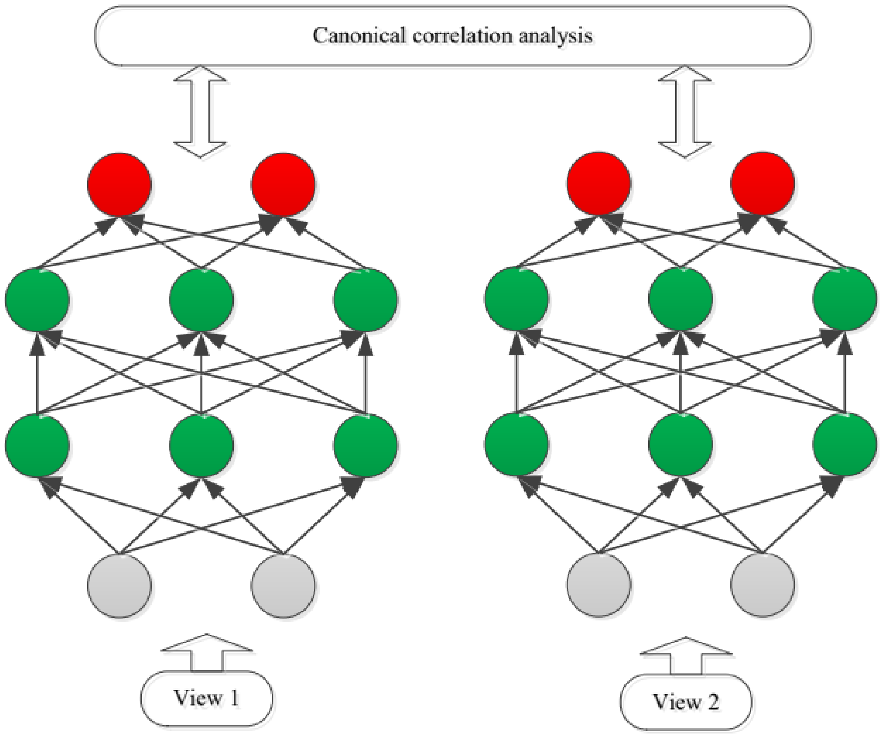
Multi-view learning



How to fuse multi-view data



Auto-encoder



canonical correlation analysis

[Deep multi-view learning methods: A review - ScienceDirect]