Digital Twins

graphical formulations for managing complexity & uncertainty

Multicore World 2025

Professor Karen E. Willcox

Associate Vice President for Research

Director, Oden Institute for Computational Engineering & Sciences

Professor of Aerospace Engineering & Engineering Mechanics

University of Texas at Austin

ODEN INSTITUTE

FOR COMPUTATIONAL ENGINEERING & SCIENCES

ODEN

50+ Years of Leadership

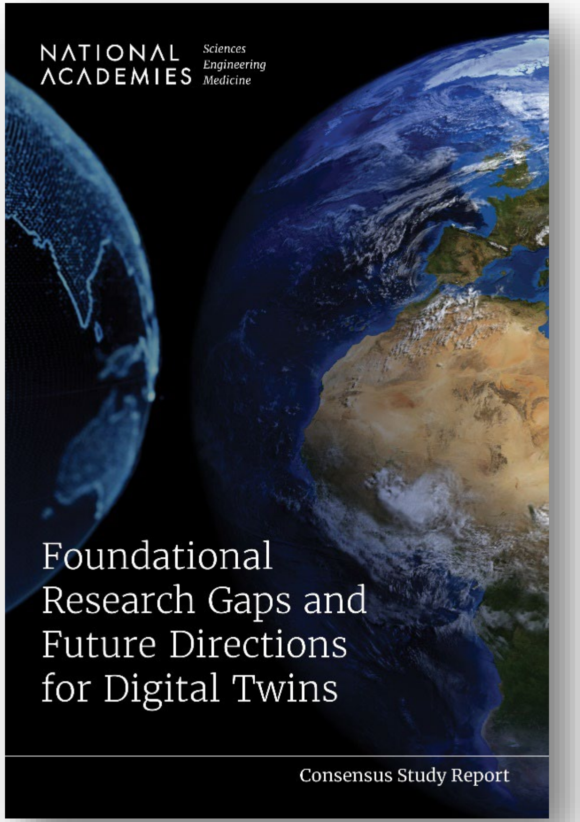
in Interdisciplinary Research & Education in Computational Engineering & Sciences

1973

2025

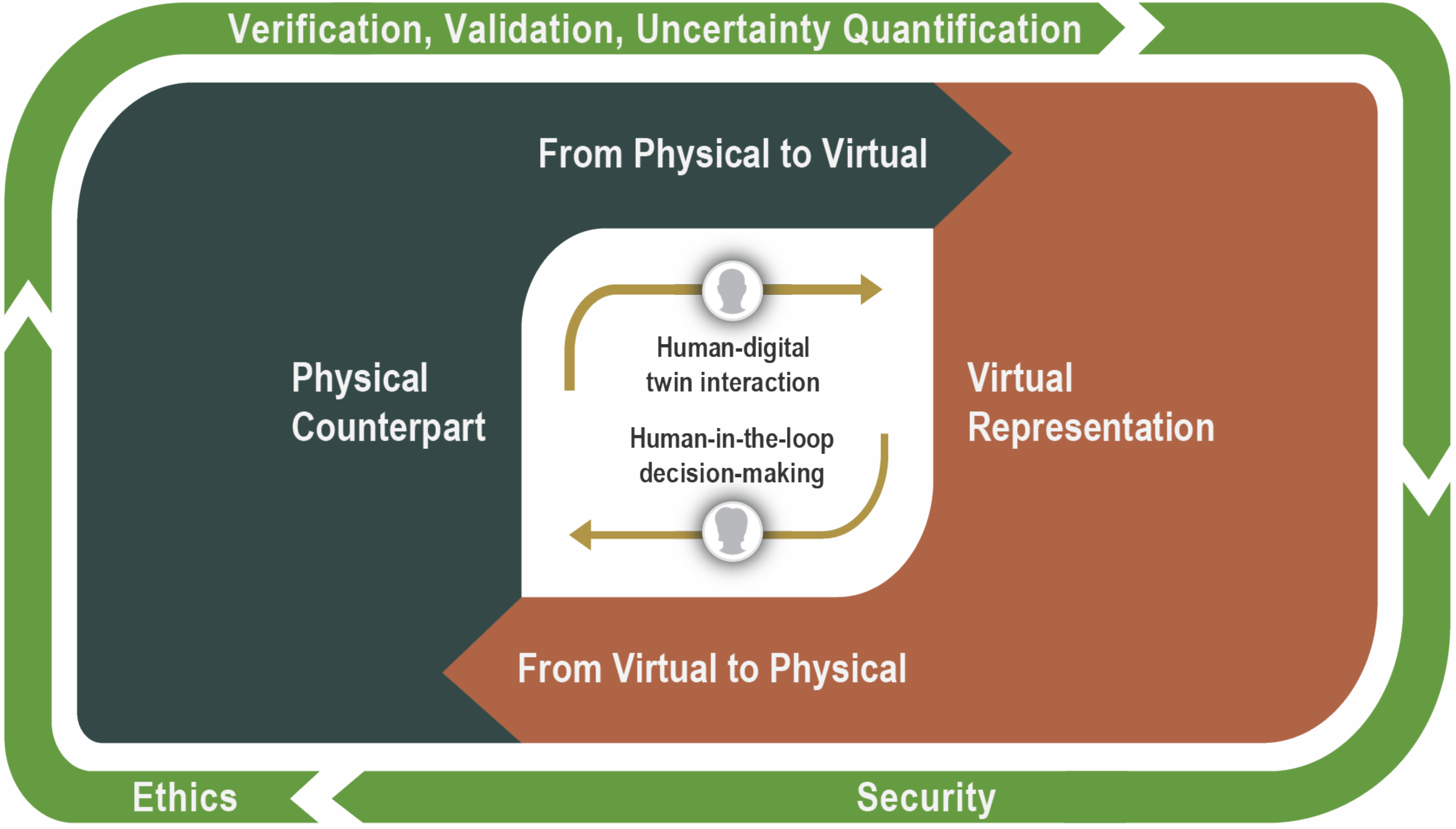
25	Research Centers/Groups
23	Principal Faculty
119	Affiliated & Core Faculty
100	Research Staff/Postdocs
29	Administrative & IT Staff
108	CSEM Students
188	Graduate Research Asst
30	Visiting Faculty Fellows
$128M	Active Research

A digital twin is a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems), is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value. The bidirectional interaction between the virtual and the physical is central to the digital twin.



National Academies Study on Foundational Research Gaps and Future Opportunities for Digital Twins (2023)

A Digital Twin provides a new mathematical paradigm for integrating data, models & decisions



A Digital Twin is more than just simulation and modeling

National Academies Study on Foundational Research Gaps and Future Opportunities for Digital Twins (2023)

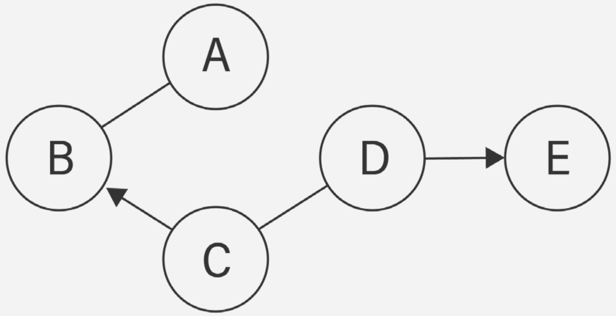
Outline

1 **GRAPHICAL FORMULATION**
of a digital twin

2 **EDUCATIONAL DIGITAL TWIN**
for Texas postsecondary pathways

This work was supported in part by Texas Higher Education Coordinating Board (THECB) and the U.S. Department of Energy (grant DE-SC0021239). The opinions and conclusions expressed are those of the authors and do not necessarily represent the opinions or policy of the THECB.

1 GRAPHICAL REPRESENTATION of a digital twin



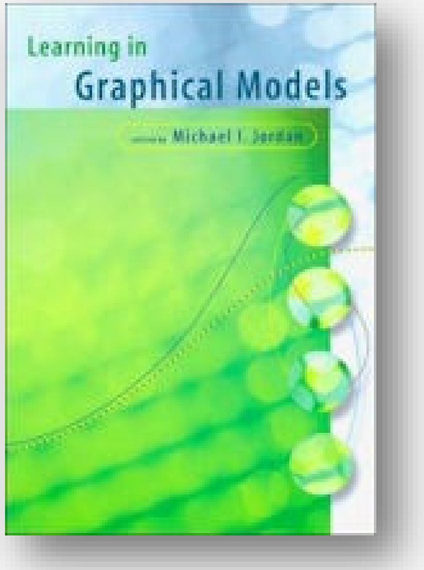
$$G = (V, E)$$

$$V = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}\}$$

$$E = \{\mathrm{A} \leftrightarrow \mathrm{B}, \mathrm{C} \rightarrow \mathrm{B}, \\ \mathrm{C} \leftrightarrow \mathrm{D}, \mathrm{D} \rightarrow \mathrm{E}\}$$

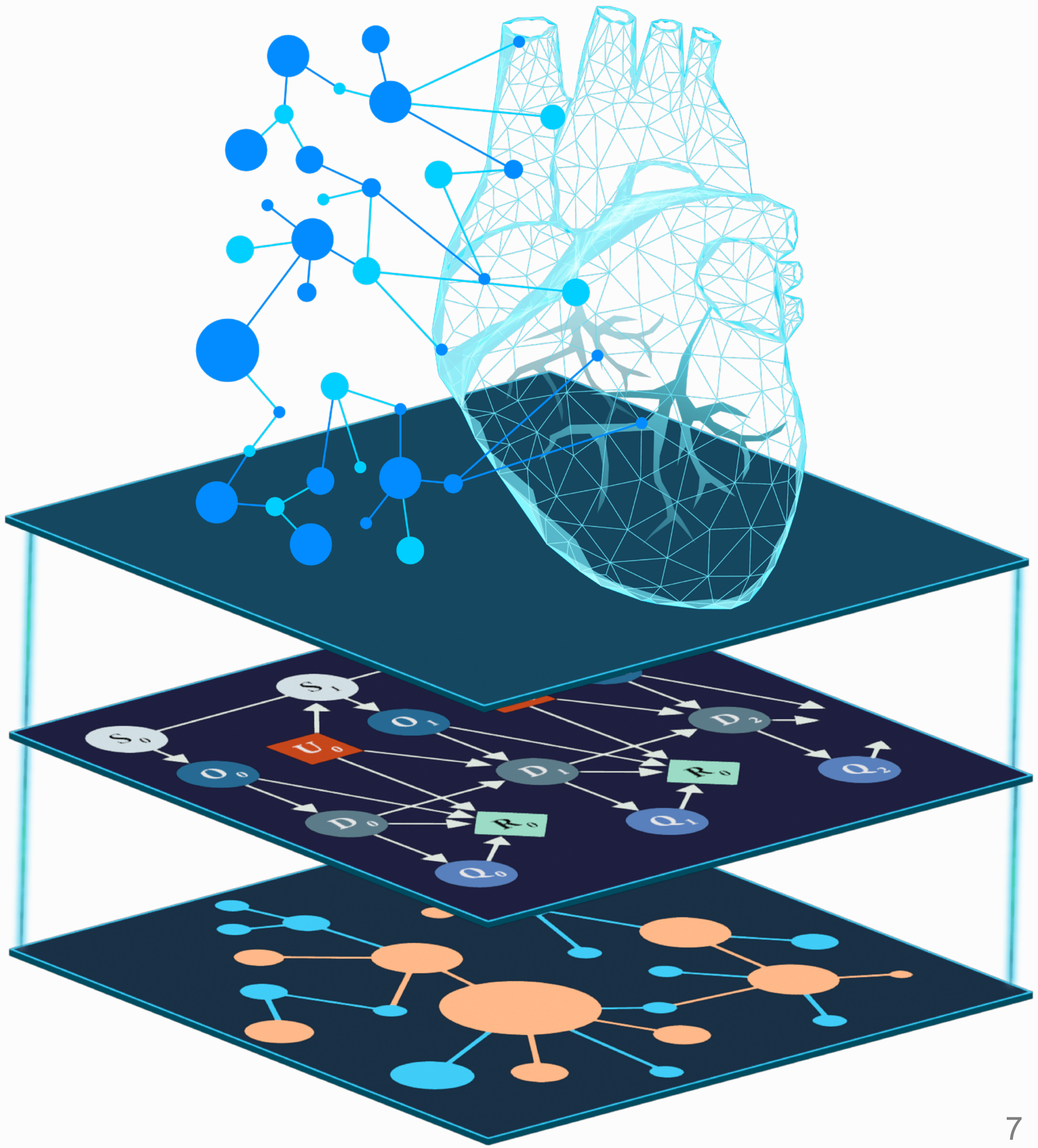
"Graphical models, a marriage between probability theory and graph theory, provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering — uncertainty and complexity."

"Fundamental to the idea of a graphical model is the notion of modularity: a complex system is built by combining simpler parts. Probability theory serves as the glue whereby the parts are combined, ensuring that the system as a whole is consistent and providing ways to interface models to data."



Graphical models emphasize relationships
→ key to a digital twin as "more than just simulation & modeling"

GRAPHICAL REPRESENTATION
of a digital twin

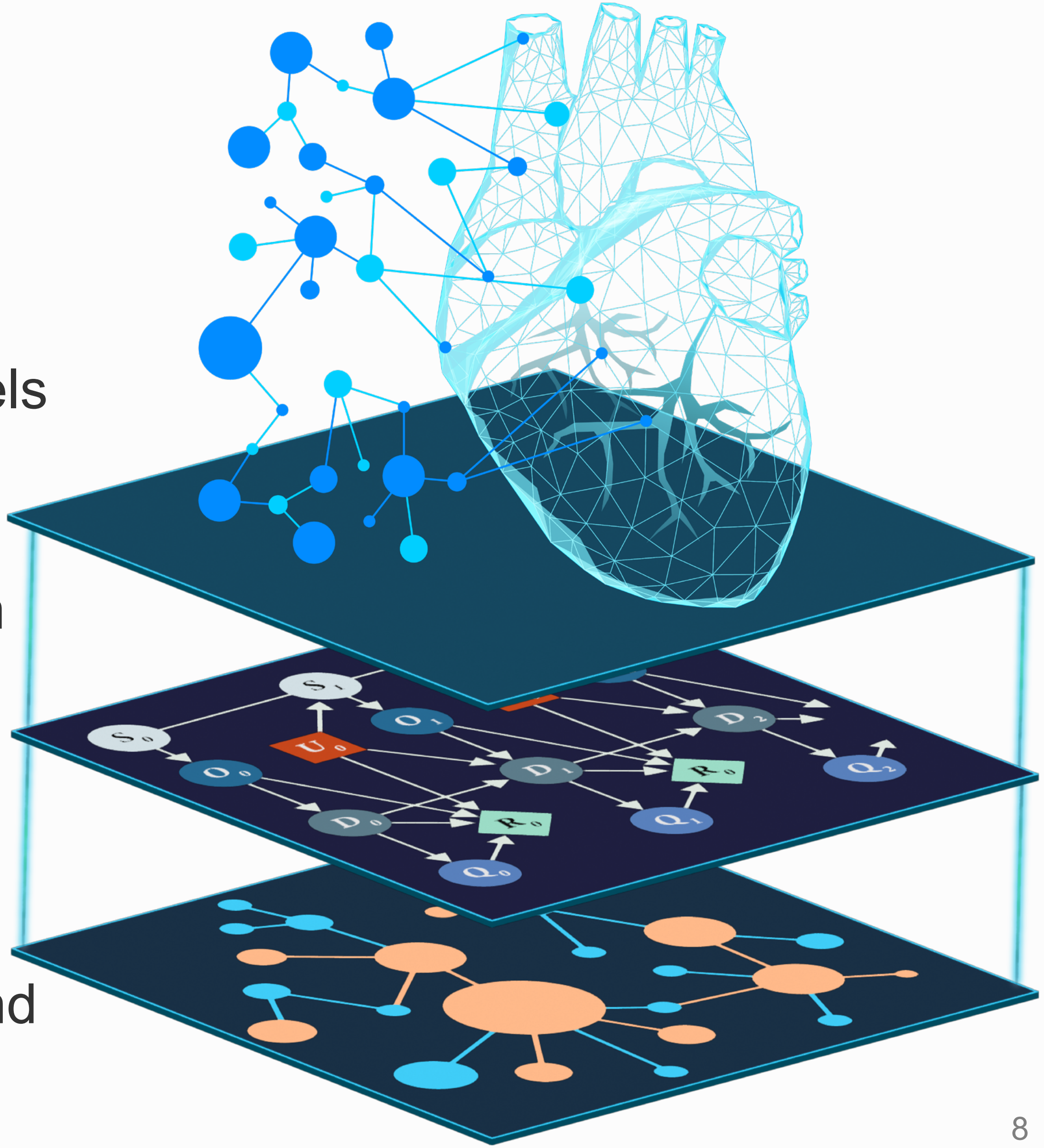


We use a multilayered graphical formulation to manage complexity and embed uncertainty quantification

DIGITAL TWIN: FOUNDATIONAL LAYER

Challenges and needs

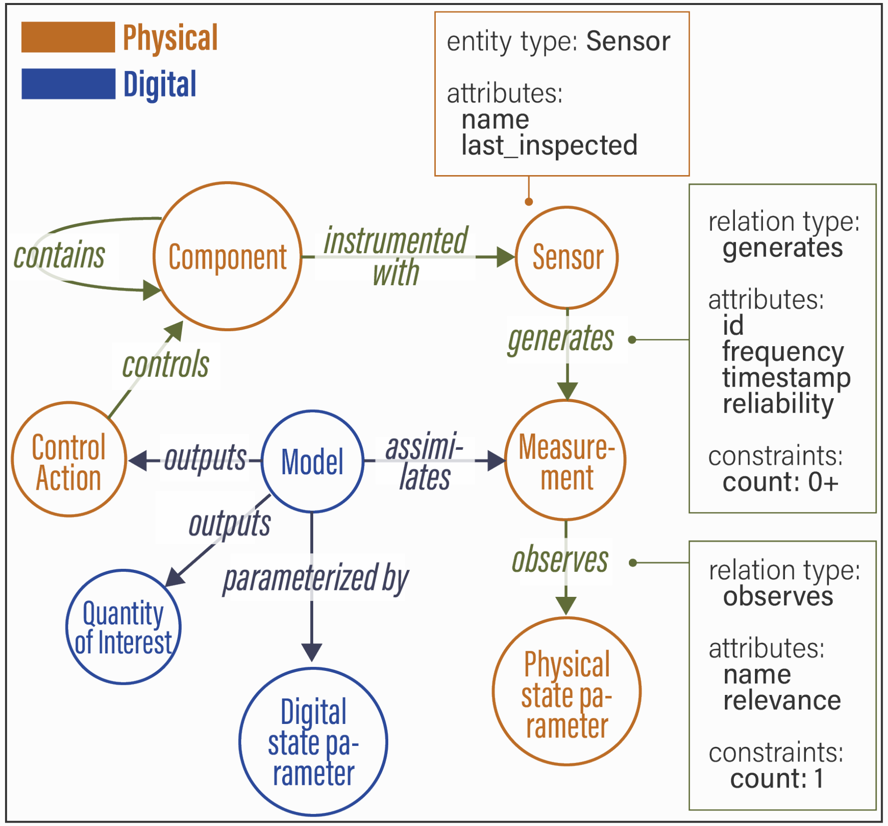
- Knowledge diversity: Data vary widely in source, format, and subject
- Dense interconnectivity: Mappings from digital representations to the physical world with high levels of complexity
- Knowledge integrity: Bidirectional data flows can introduce errors that propagate through the system



A knowledge graph as the foundational layer for digital twins

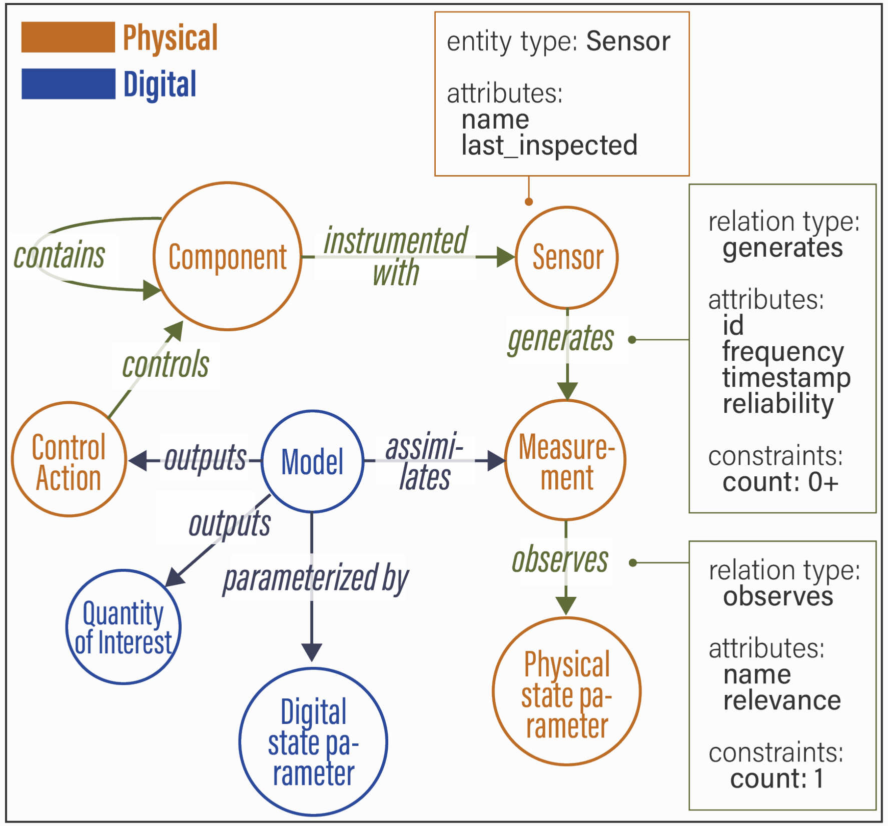
- A mathematical structure to support semantic knowledge organization, scalable computations and bidirectional data flow as a basis for digital twins

DIGITAL TWIN ONTOLOGY AND KNOWLEDGE GRAPH

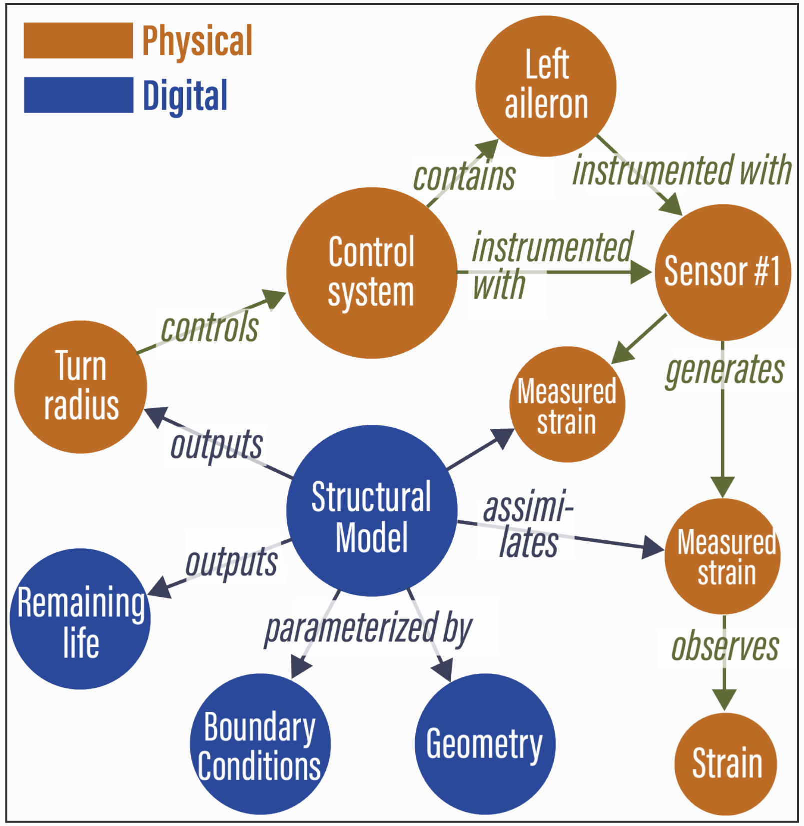


Ontology: for an unmanned aerial vehicle specify the entities, relationships between entities, and rules for entities and relationships

DIGITAL TWIN ONTOLOGY AND KNOWLEDGE GRAPH

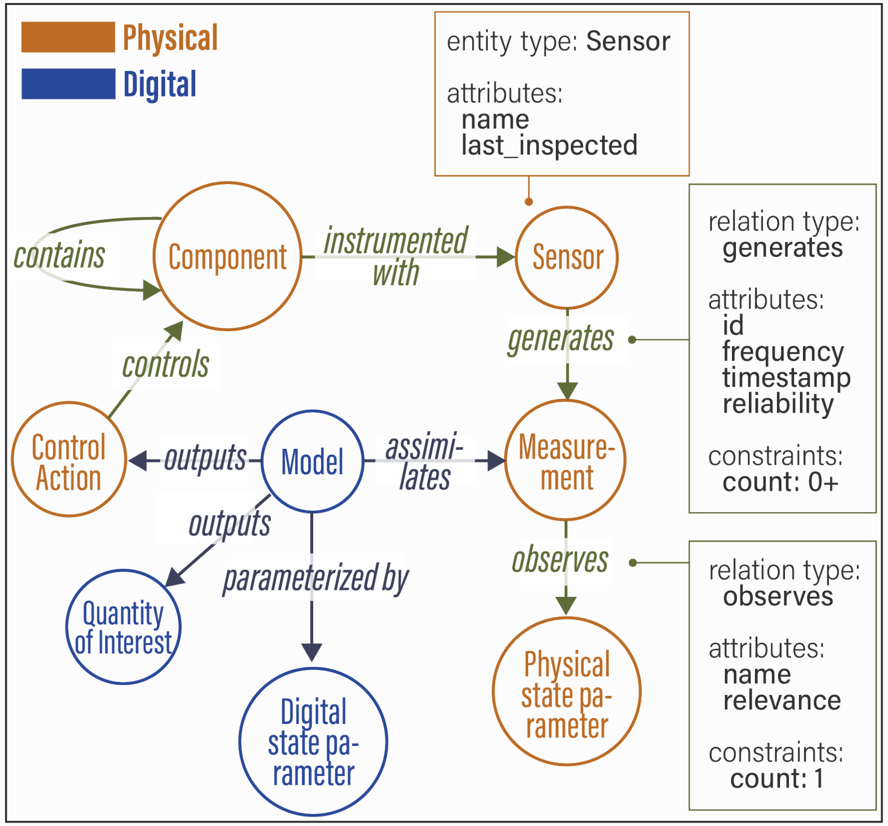


Ontology: for an unmanned aerial vehicle specify the entities, relationships between entities, and rules for entities and relationships

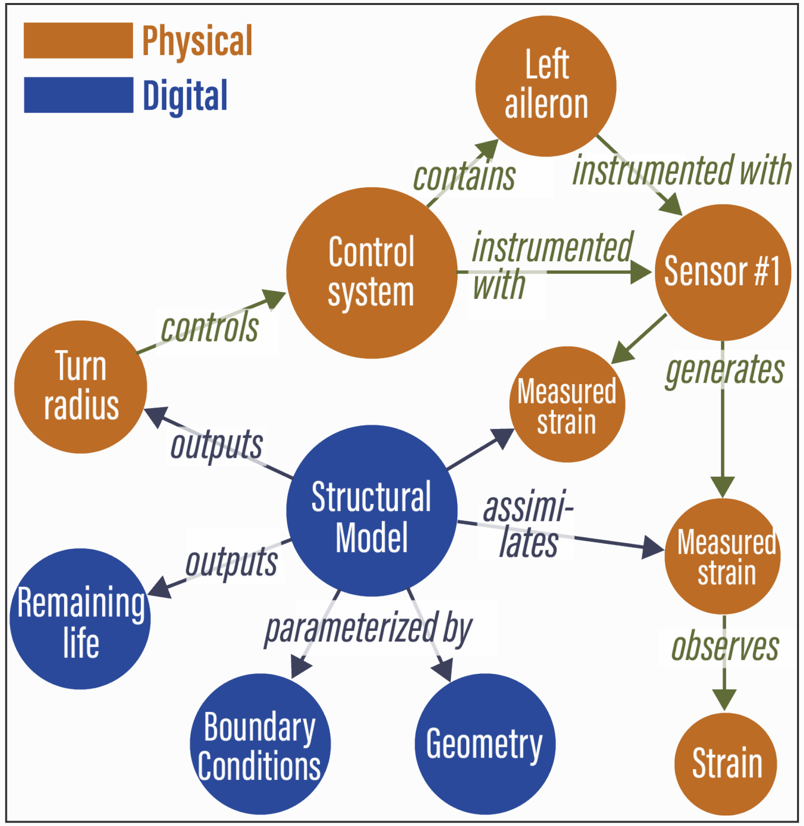


Knowledge Graph: Notional example instantiated by applying the ontology

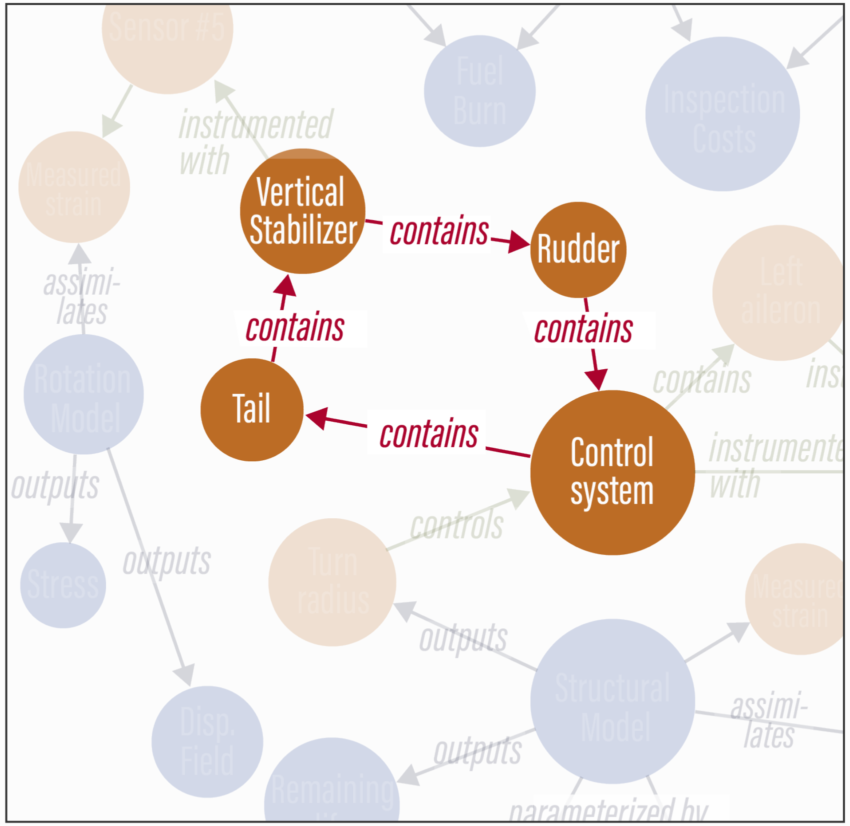
DIGITAL TWIN ONTOLOGY AND KNOWLEDGE GRAPH



Ontology: for an unmanned aerial vehicle specify the entities, relationships between entities, and rules for entities and relationships



Knowledge Graph: Notional example instantiated by applying the ontology



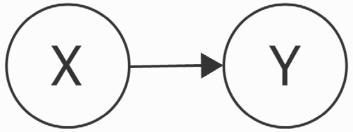
Illegal cycle: formalized as a mathematical statement annotated with type semantics, discovered by depth-first-search traversal through the graph

DIGITAL TWIN: PREDICTIVE LAYER

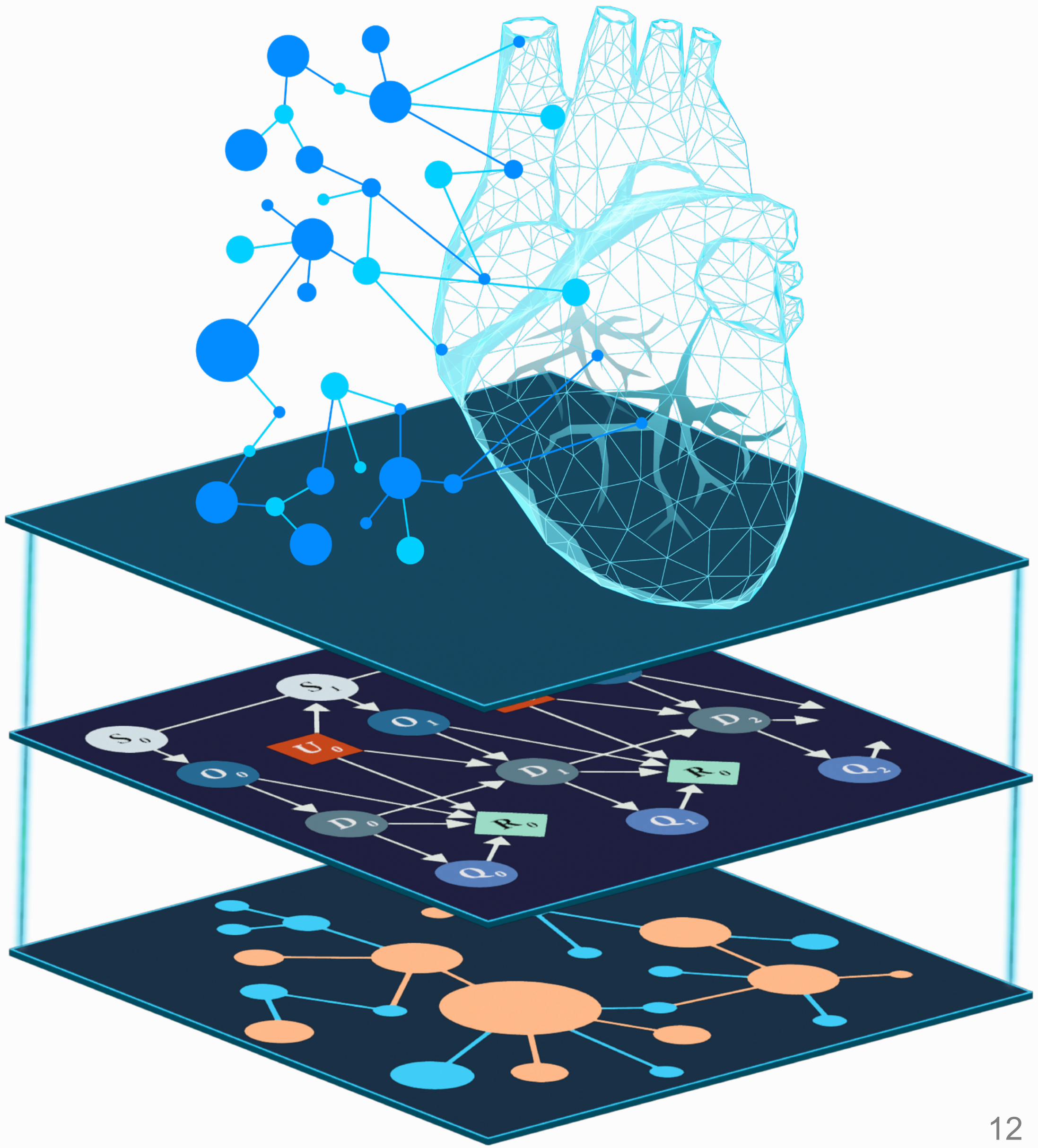
Challenges and needs

- Uncertainty in the physical world
  e.g., material properties, loading conditions, etc.
- Uncertainty in the digital world
  e.g., model parameters, initial conditions, forcing, boundary conditions

Probabilistic graphical models



- Use a graphical structure to encode uncertainty and interdependence between random variables
- Nodes are random variables
- Edges are conditional probabilities representing relationships between random variables

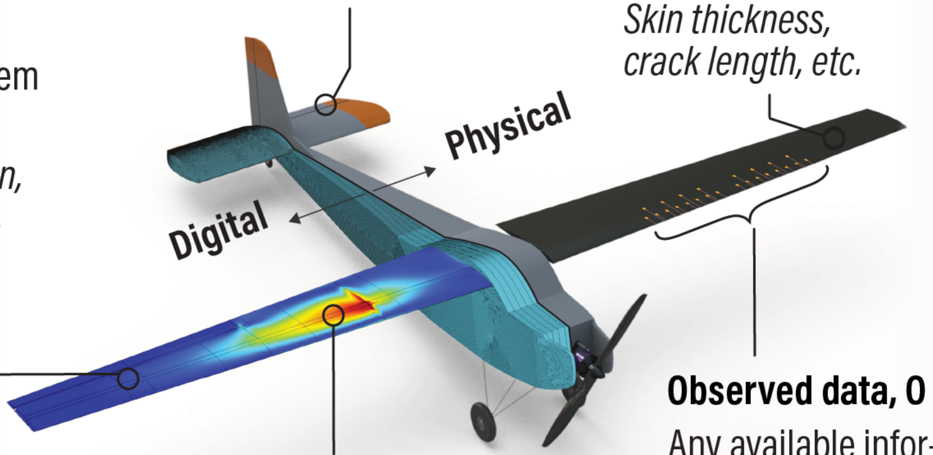


DIGITAL TWIN PROBABILISTIC GRAPHICAL MODEL

Control
Any available action or decision that influences the physical asset
In-flight maneuvers, maintenance inspection decisions,

Reward, R
Quantifies overall performance of the twinned system
Mission success, fuel burn, maintenance, inspection costs, etc.

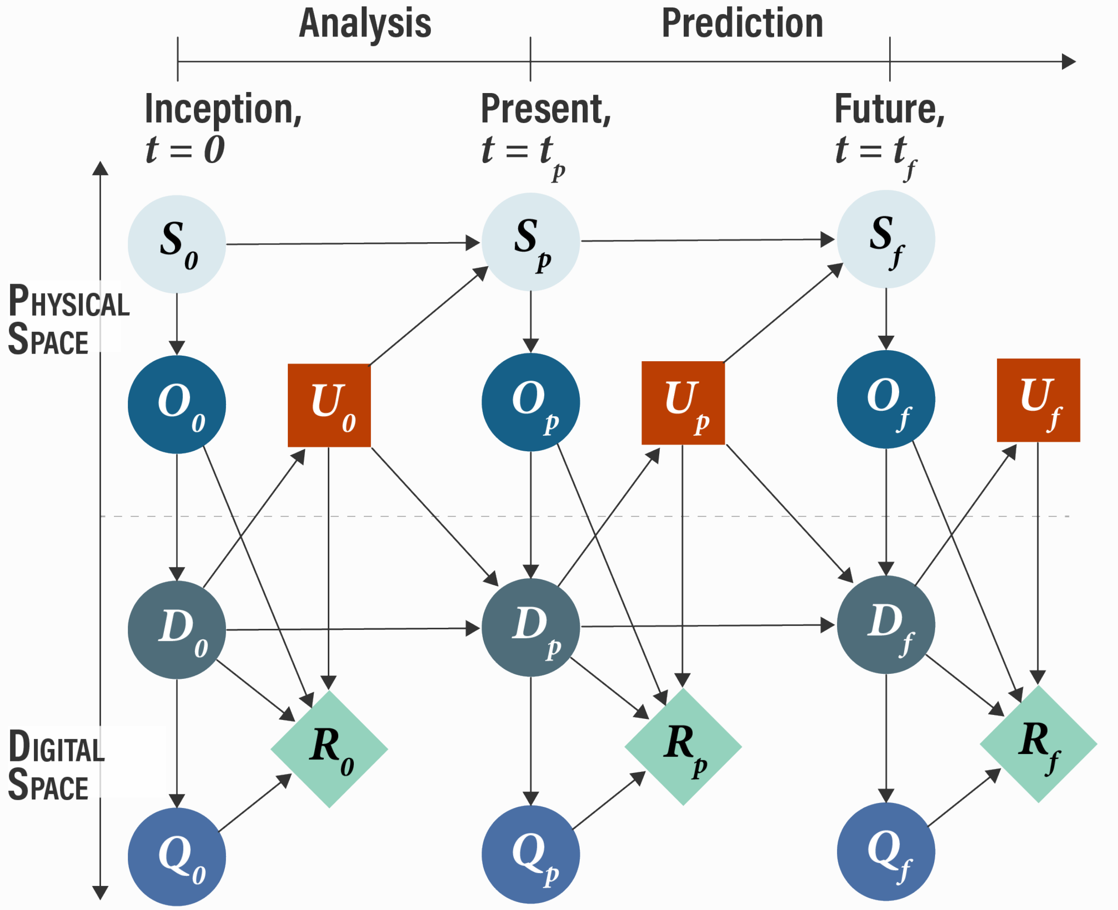
Physical State, S
Parametrized state of the physical asset
Skin thickness, crack length, etc.



Digital State, D
Parameters (model inputs) that define the computational models comprising the digital twin
Geometry, structural parameters, etc.

Quantities of Interest, Q
Quantities describing the asset and estimated via model outputs
Stress, strain, displacement fields, failure stress,

Observed data, O
Any available information describing the state of the physical system
Strain, accelerometer, inspection



Kapteyn, Pretorius, Willcox
Nature Computational Science, 2021

DIGITAL TWIN PROBABILISTIC GRAPHICAL MODEL

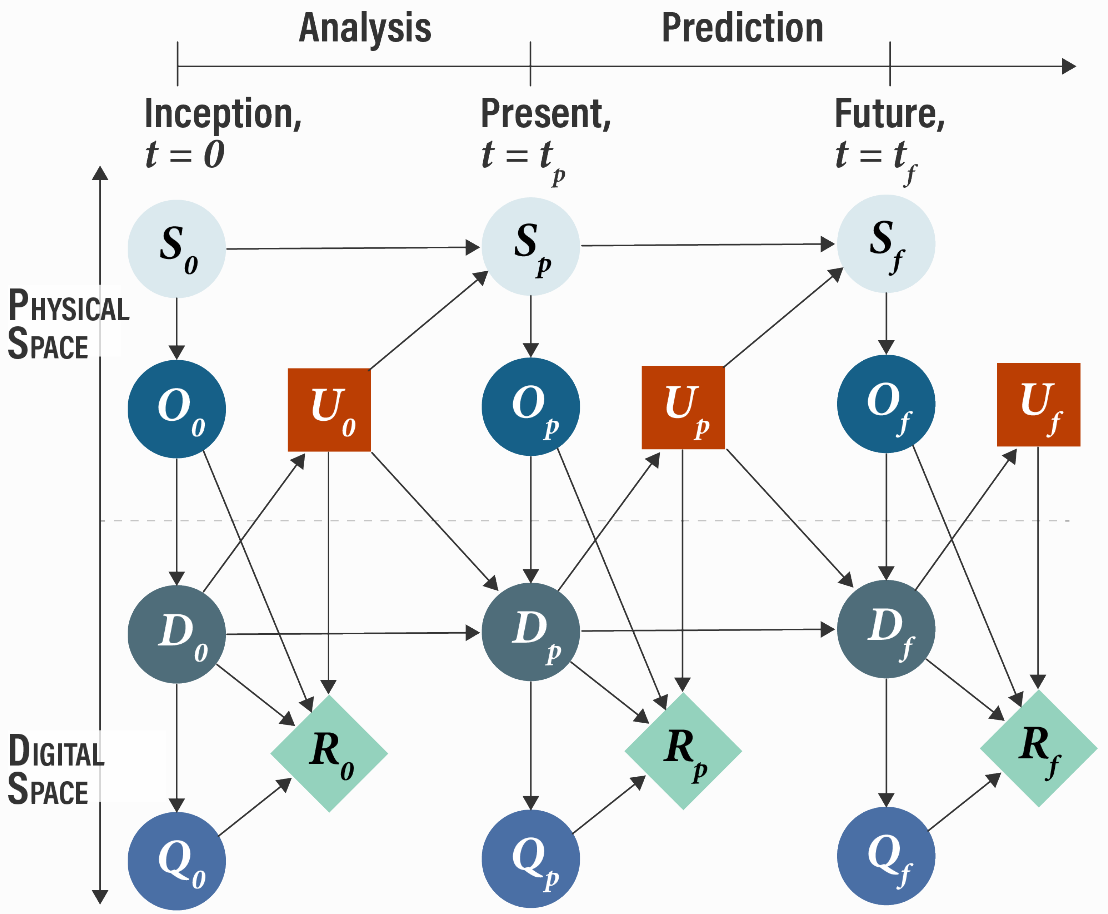
Let $\mathbf{X}$ denote the set of random variables in the PGM representation of the digital twin:

$$\mathbf{X} = \{S_0, O_0, D_0, Q_0, U_0, R_0, \ldots \\ S_f, O_f, D_F, Q_f, U_f, R_f\}$$

The PGM is a directed acyclic graph that encodes the joint probability distribution $P(\mathbf{X})$

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i))$$

where $\mathrm{Pa}(X_i)$ denotes the parents of node X_i



Kapteyn, Pretorius, Willcox
Nature Computational Science, 2021

2 EDUCATIONAL DIGITAL TWIN

Luwen Huang
CS PhD student
UT Austin

Huang, L. and Willcox, K. Educational Digital Twin: Tackling complexity in educational big data. IEEE International Conference on Big Data, December 2024.

Huang, L., Kapteyn, M. and Willcox, K. Digital twin: Graph formulations for managing complexity and uncertainty. IEEE DigitalTwin, December 2024.

Huang, L., Bicol, K. and Willcox, K. Modeling COVID-19 disruptions via network mapping of the Common Core Mathematics Standards. *Computers in Education Journal*, Vol. 13, No. 2, 2023.

Huang, L. and Willcox, K., Network models and sensor layers to design adaptive learning using educational mapping. *Design Science*, Vol. 7, 2021.

Willcox K. and Huang, L., Network models for mapping educational data, *Design Science*, Vol. 3, 2017.

2 EDUCATIONAL DIGITAL TWIN

S_t **<u>Physical State</u>**

Student enrollments, engagement, outcomes, demographics;
course enrollments, success rates;
institutional resources, faculty capacity

U_t **<u>Control inputs</u>**

Curriculum changes, student advising, tutoring interventions, added resources

O_t **<u>Observational data</u>**

Grades, attendance records, assessment results, institutional records, engagement analytics

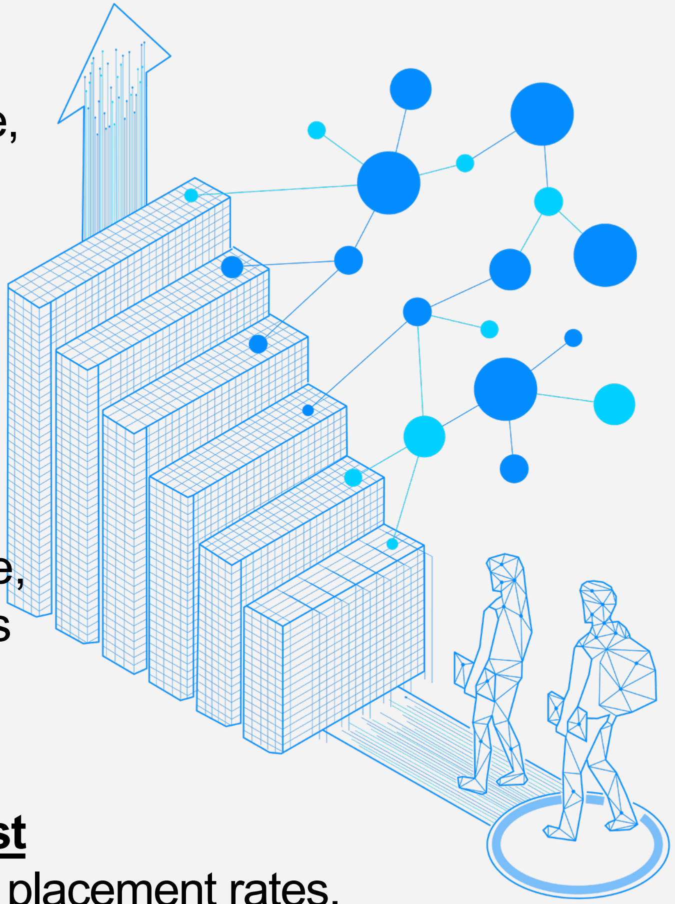
2 EDUCATIONAL DIGITAL TWIN

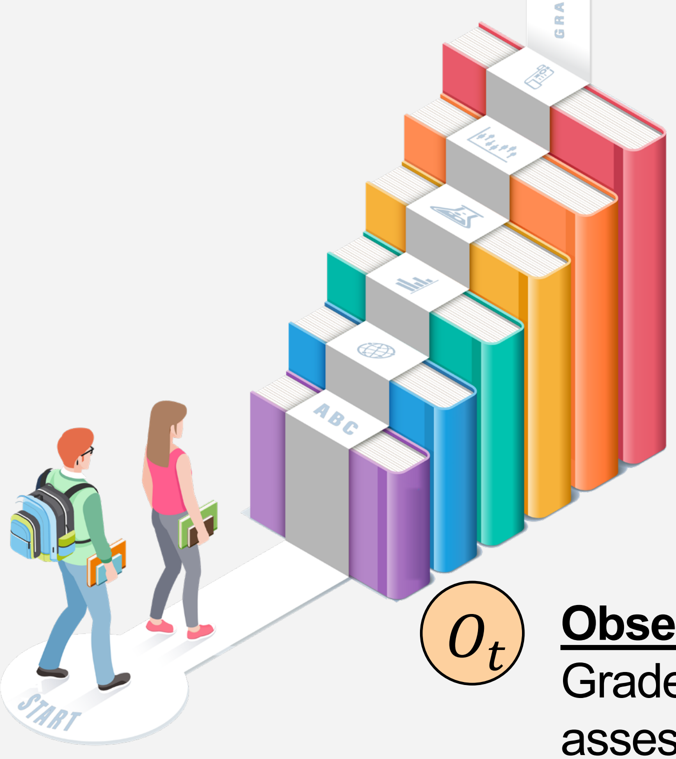
S_t **Physical State**
Student enrollments, engagement, outcomes, demographics;
course enrollments, success rates;
institutional resources, faculty capacity

D_t **Digital Twin State**
Virtual representation at the student, course, program, institution, and state levels

R_t **Reward**
Student outcomes, student performance, drop-out rates, costs

Q_t **Quantities of Interest**
Graduation rates, job placement rates, teacher-student ratios, advising time





U_t **Control inputs**
Curriculum changes, student advising, tutoring interventions, added resources

O_t **Observational data**
Grades, attendance records, assessment results, institutional records, engagement analytics

Educational Pathways

Academic trajectories are nonlinear, multi-faceted and time-dependent

Fall 2015 | Spring 2015 | Fall 2016 | Spring 2016 | Fall 2017

Jennifer enrolls in community college for the first time.

She meets the Texas requirements in Reading and Writing, but not in math.

She enrolls in College Algebra I, Drama I, Biology I and English I.

She fails College Algebra I at the end of term.

She does not enroll the following term.

She re-enrolls.

She takes Pre-Algebra II, Govt I, and Engl II.

Part of recommended sequence of courses for the Biology pathway.

She passes all classes.

She earns a Core Completer certificate.

She transfers to a four-year institution.

All data shown are notional.

The Scale & Complexity of Educational Pathways in Texas

Volume: millions of students, hundreds of public postsecondary institutions

Variety: Grade reports, HTML, excel spreadsheets, etc.

Velocity: Semester-based grade reports, multiyear policy changes

Veracity: Accurate course records, self-reported student intentions

Value: Insights support both student interventions and policy making across Texas

Variability: Multiple use cases across multiple scales e.g. advising tool for counselors vs. student nightly registration status vs. institutional resource decisions

Degree Plans

ASSOCIATE DEGREE

Biology

TSI Liable

RECOMMENDED COURSE SEQUENC

First Year

Fall

ENGL 1301	Composition I
Creative Arts Elective - Core Curriculum	
Mathematics Elective - Core Curriculum	
Select one of the following:	
MATH 1414	College Algebra
MATH 2412	Pre-Calculus Math

All data shown are notional

(a) Selected columns from Report CMB001 ("Student")

student	fice	gender	ethnic	ecodis	...
15	1	1	2	1	
16	2	0	7	1	
17	2	0	7	1	

(b) Selected columns from Report CMB00S ("Student Schedule")

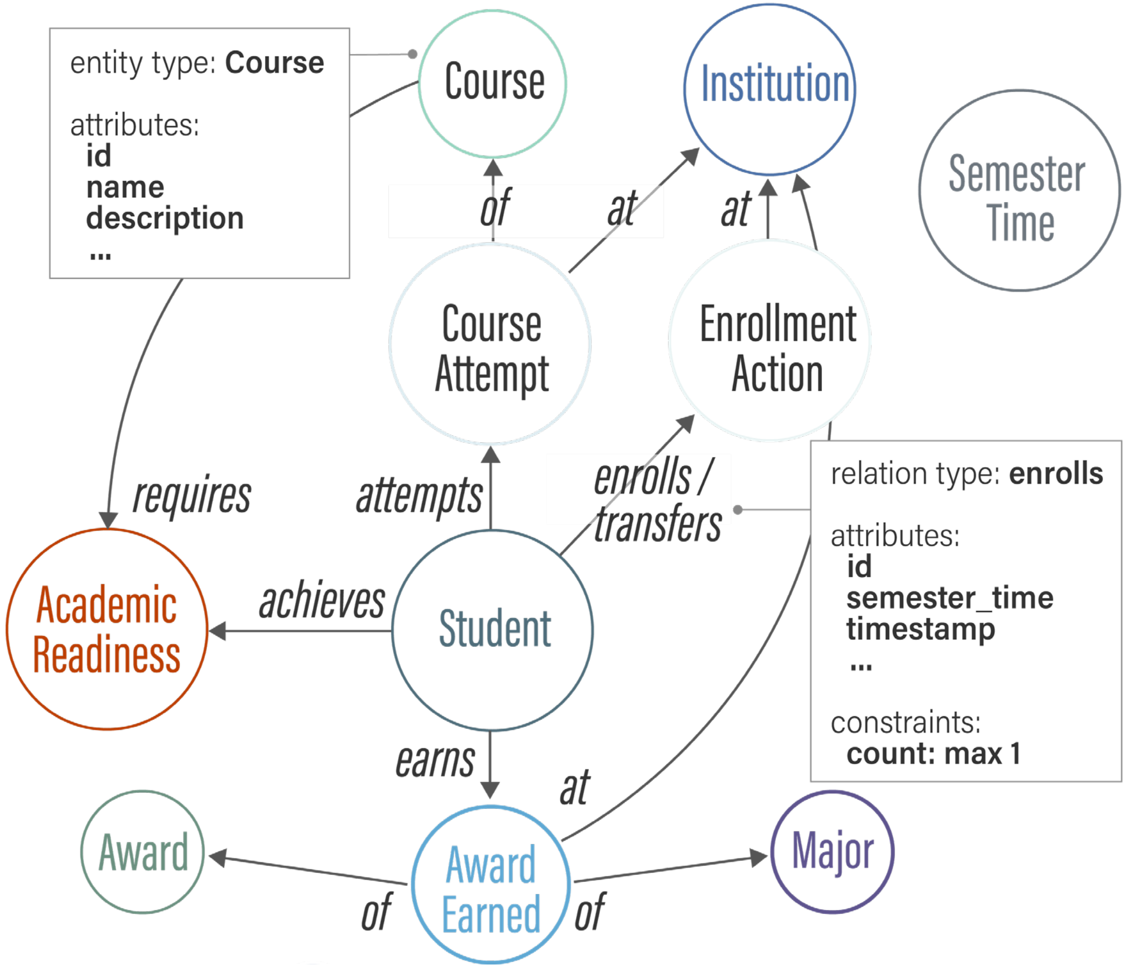
student	course	grade	credit	fcl	...
15	Math 1	A	TRUE	1	
15	Eng 1	B	TRUE	0	
16	Math 1	C	TRUE	3	
17	Eng 1	A	FALSE	4	
17	Econ 2	A	FALSE	4	

(c) Selected columns from Report CMB009 ("Graduation"):

student	fice	degree	level	major	type	...
9	1	AA	1	012	Academic	
13	1	ATC	2	156	Technical	
25	2	CCC	5	213	Tech-Prep	

Educational Digital Twin Knowledge Graph

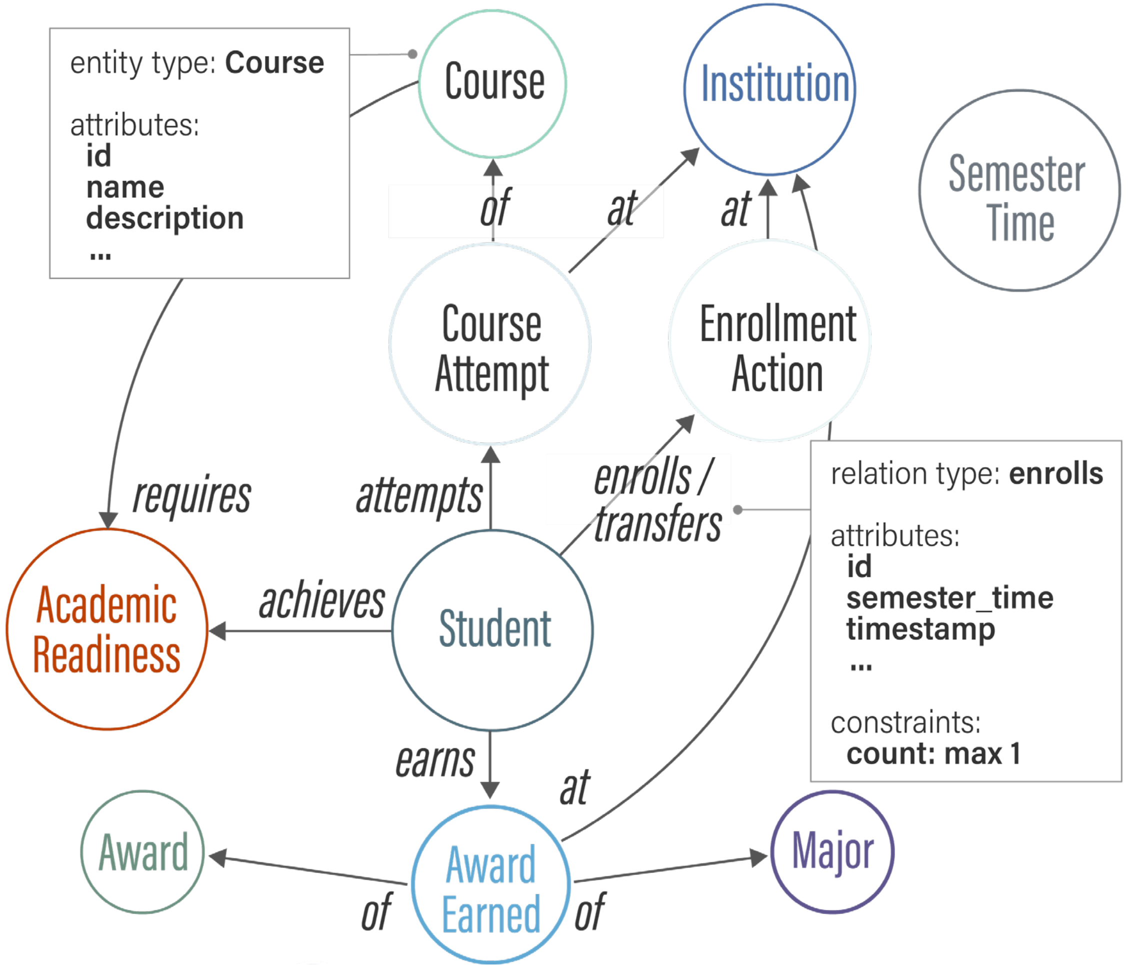
EDT-KG ontology



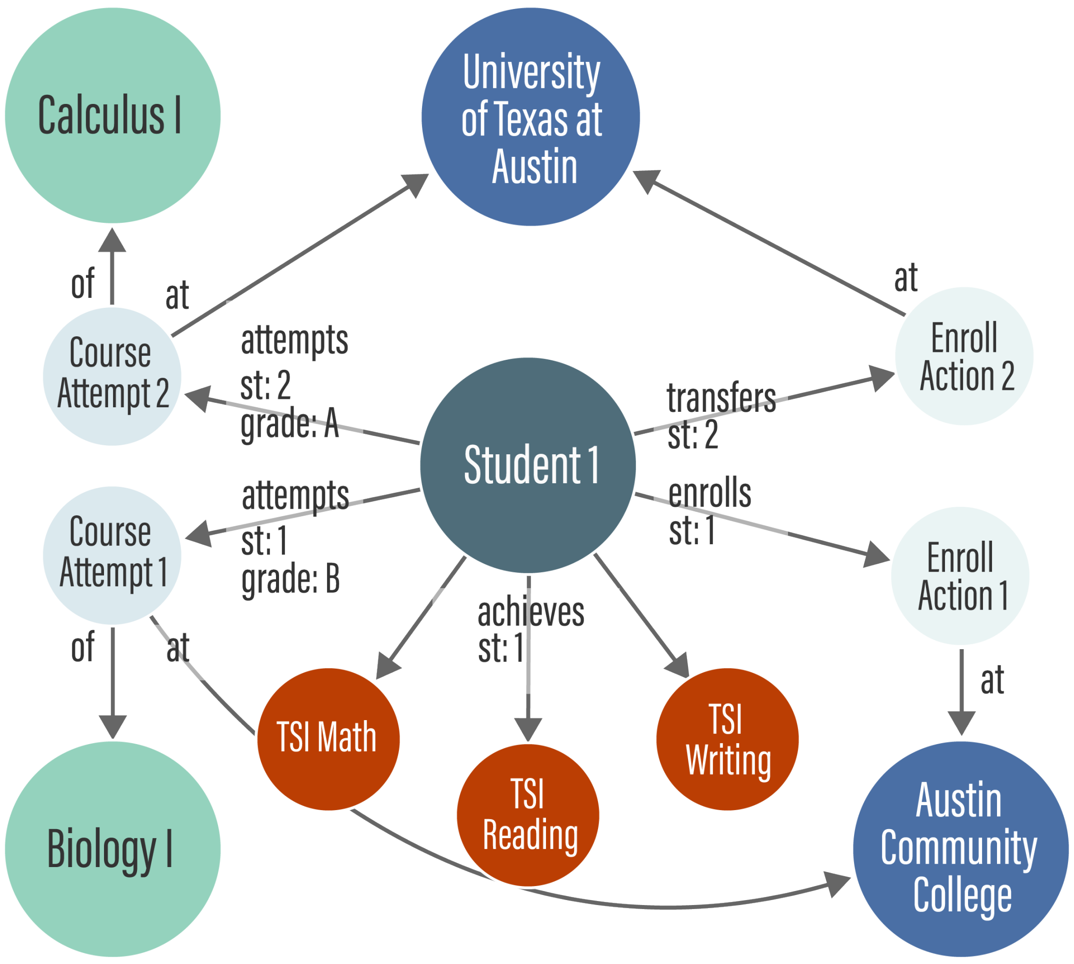
Educational Digital Twin Knowledge Graph

All data shown are notional

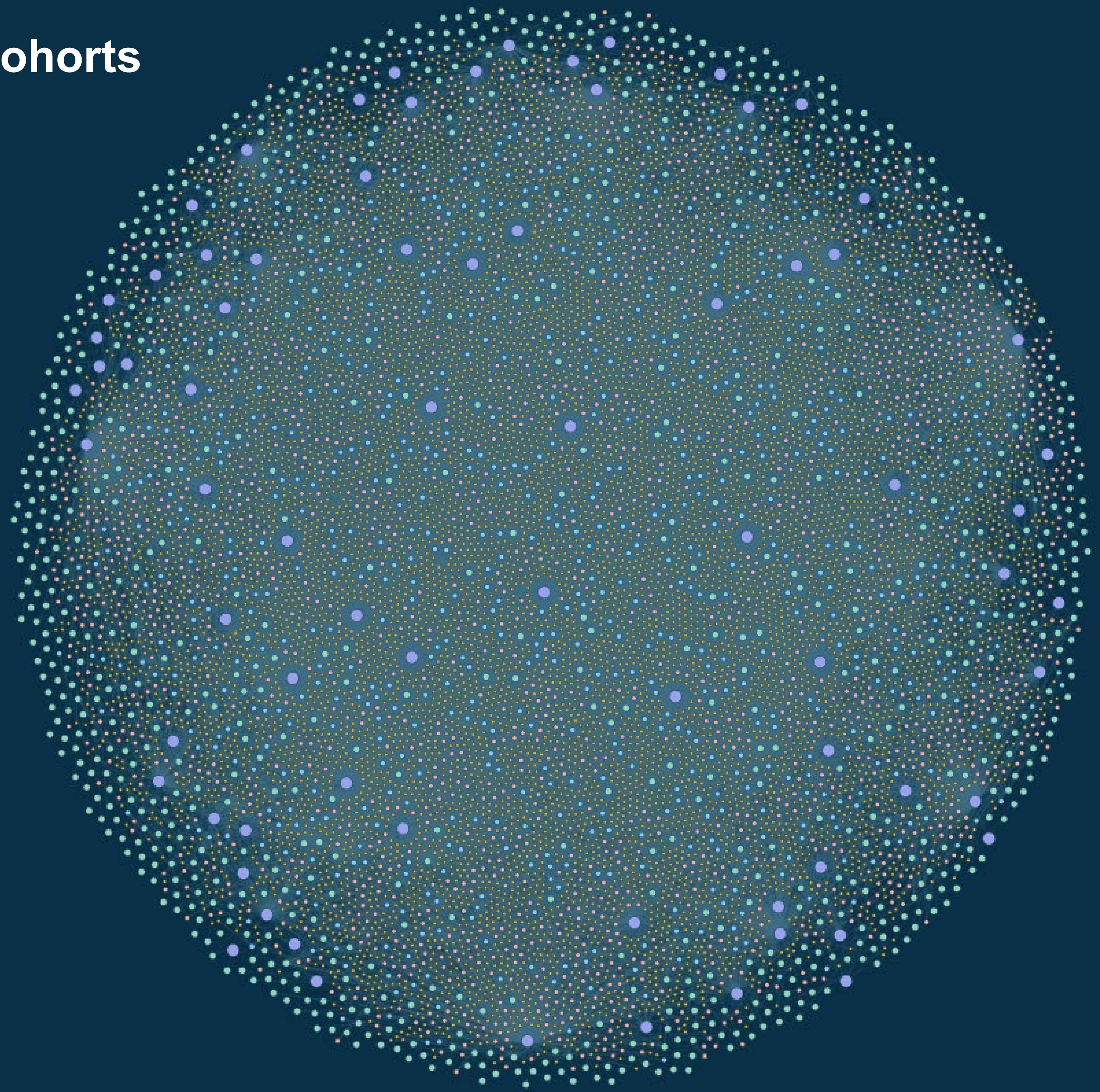
EDT-KG ontology



EDT-KG instantiation (illustrative)



7 majors spanning 6 cohorts
from 2012 – 2020
15,901,092 nodes
39,130,256 edges

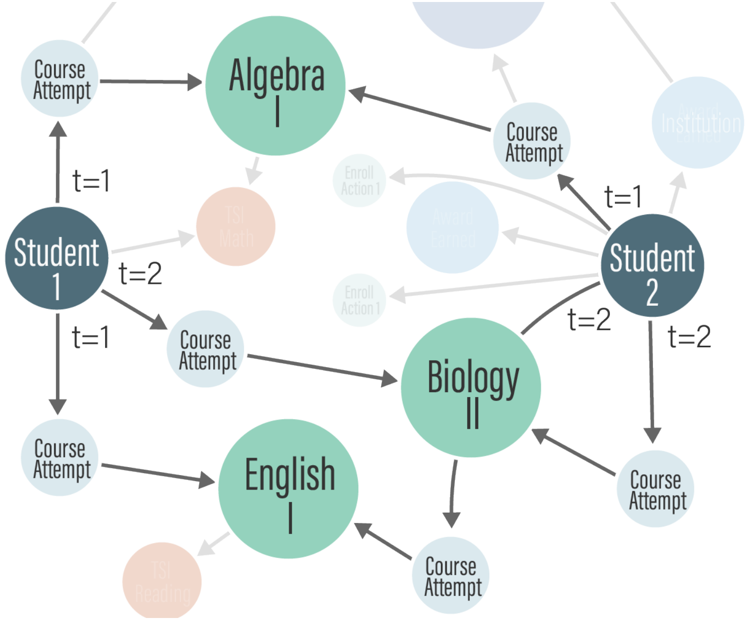


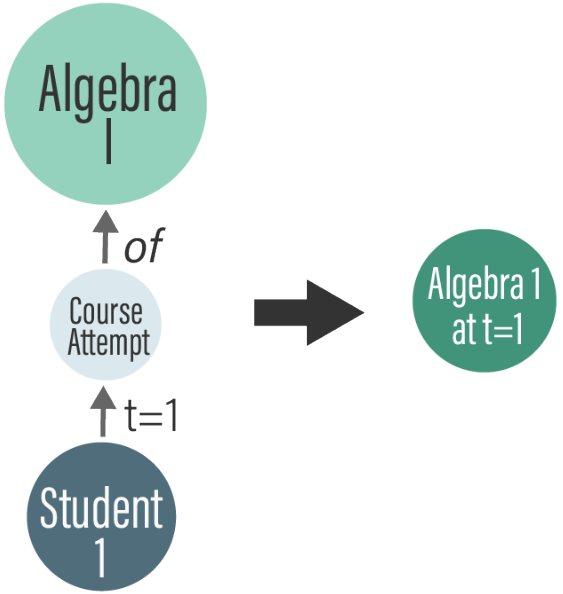
Authors: Luwen Huang and Karen E. Willcox

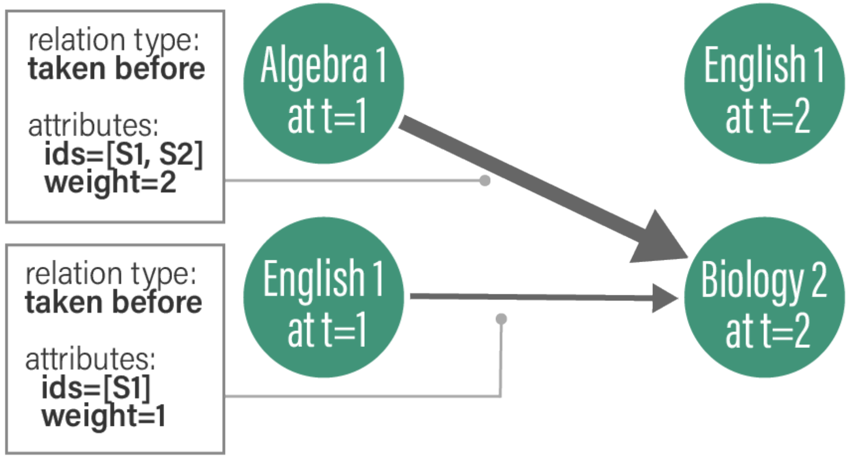
Multiscale Queries: Graph Transformations

All data shown are notional

Enable **scalable** bidirectional flow between physical and digital







Selection

Defines selection of vertices and edges with logical predicates

Projection

Create new vertices and edges based on selection

Aggregation

Traverse over selection and projection to attach attributes

The Educational Digital Twin Probabilistic Graphical Model

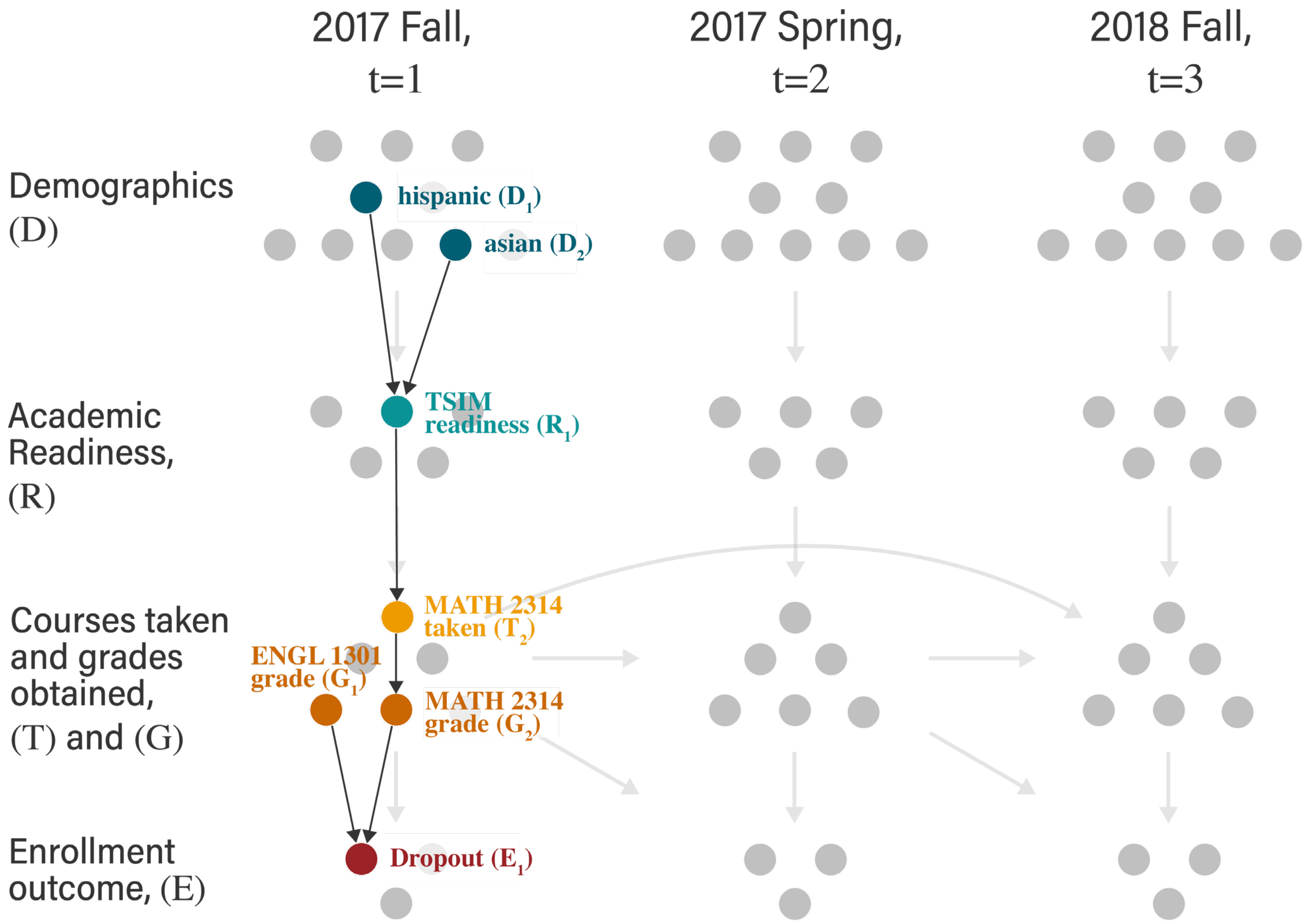
Let $\mathrm{P}(\mathbf{X})$ denote the multivariate joint distribution of our variables, where $\mathbf{X} = \{\mathrm{D}_1, \mathrm{D}_{2,} \mathrm{R}_1 \ldots, \mathrm{E}_\mathrm{N}\}$

We learn the posterior $\mathrm{P}(\boldsymbol{\theta} \mid \mathbf{X^*})$

$$P(\theta|X^*) \propto P(\theta)P(X^*|\theta)$$

We construct the EDT PGM to make learning $\mathrm{P}(\boldsymbol{\theta} \mid \mathbf{X^*})$ tractable

- D_1 and D_2 are independent of each other, with independent posteriors $\mathrm{P}(\boldsymbol{\theta}_2 \mid \mathrm{D}_2)$ and $\mathrm{P}(\boldsymbol{\theta}_1 \mid \mathrm{D}_1)$
- T_2 is dependent on R_1 with posterior, $\mathrm{P}(\boldsymbol{\theta}_{\mathrm{R}1} \mid \mathrm{R}_1)$.
  - …

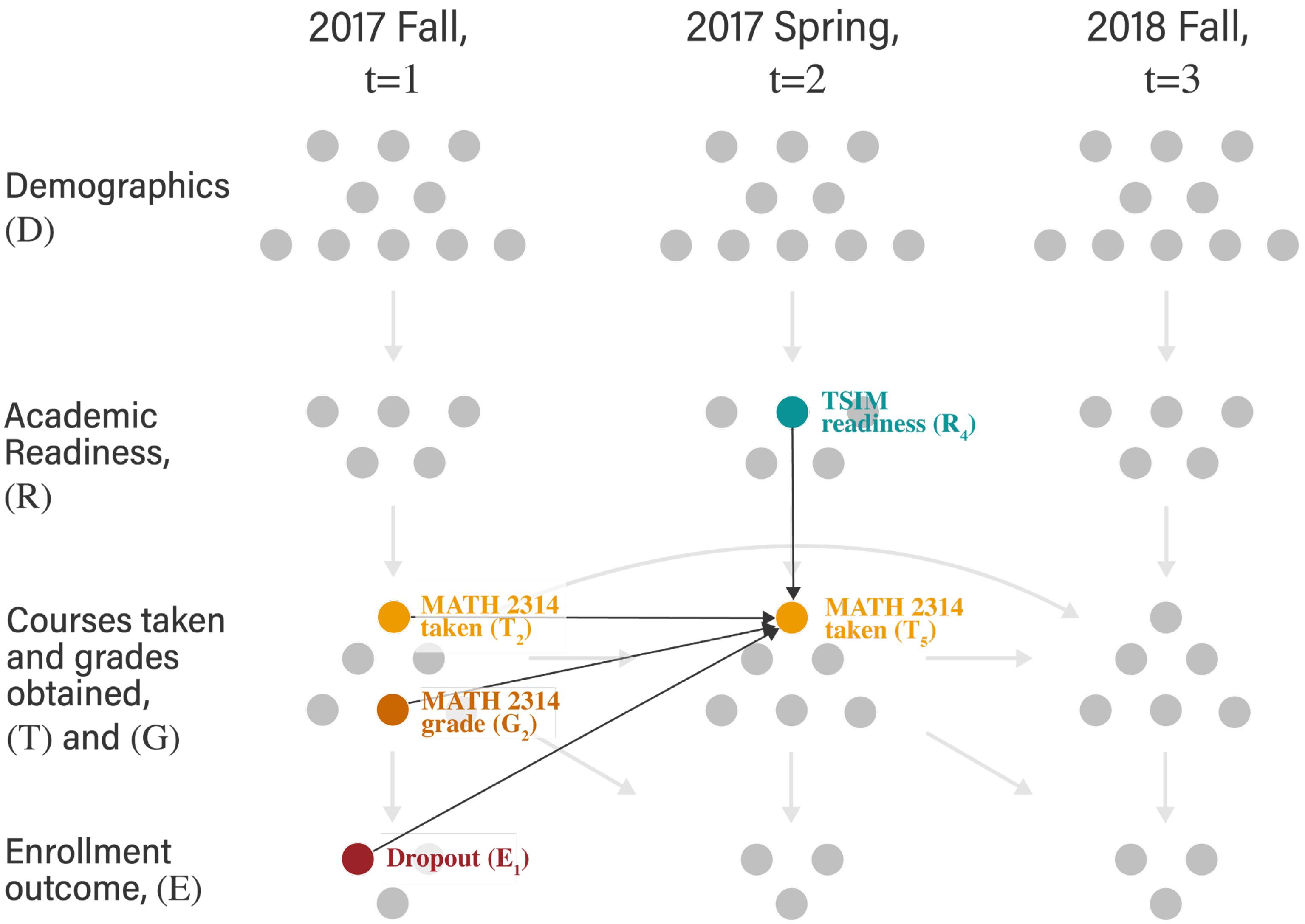


The Educational Digital Twin Probabilistic Graphical Model

Unroll PGM to continue forward in time

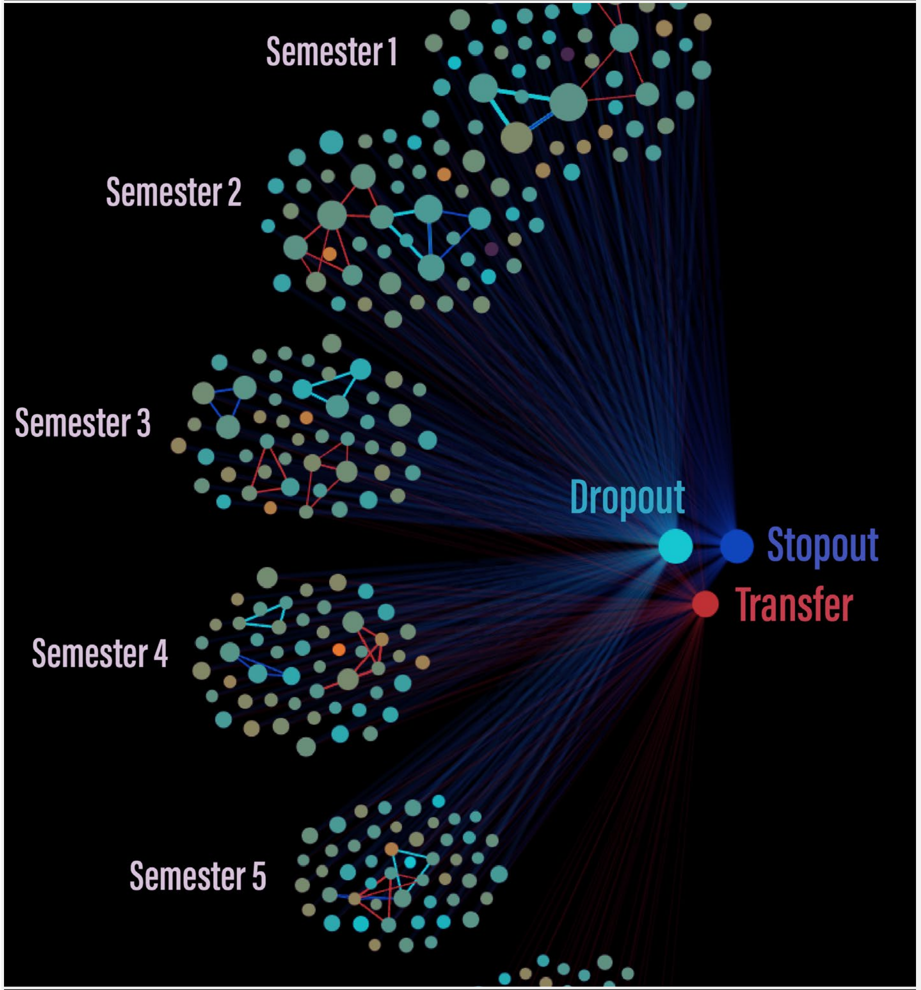
As we unroll more timesteps, use feature engineering to create new summary statistic for courses taken e.g. “max difficulty passed”

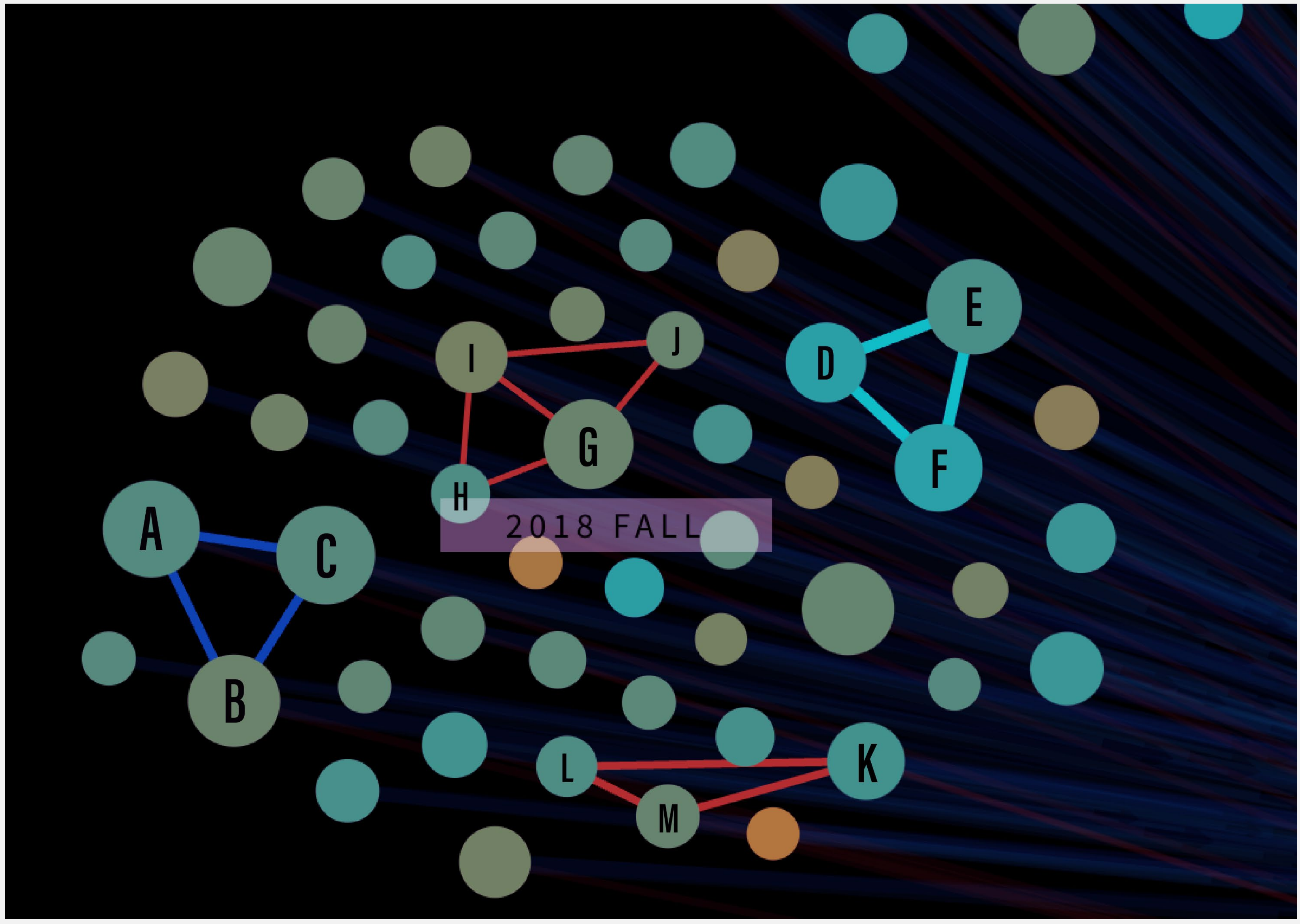
PGM structure can be determined manually using expert knowledge (e.g., `MATH 2314` *`requires`* `MATH 2311`) or learned from data



The University of Texas at Austin
Oden Institute for Computational Engineering and Sciences

TRANSFORMATION: COURSE OUTCOMES



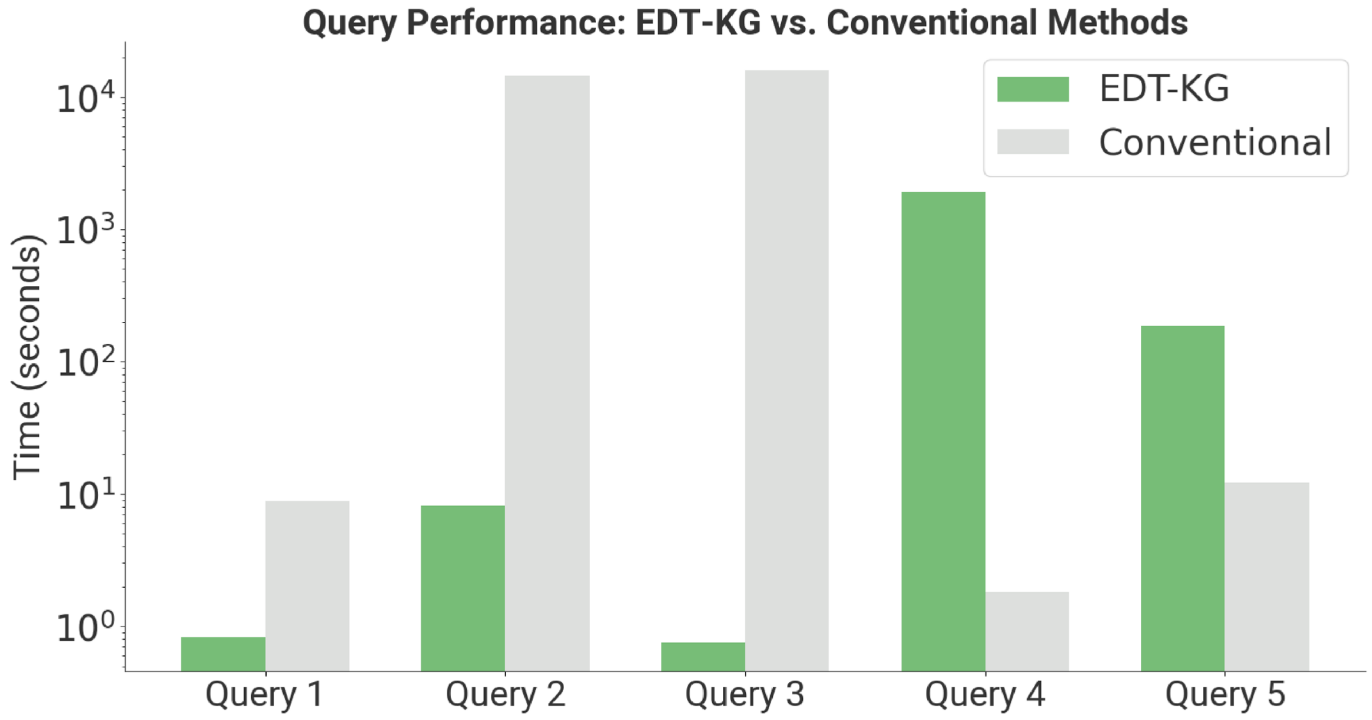


Scalability

Comparison against table-based methods used in educational analytics

Read queries
- Orders of magnitude faster, especially with pathway-based operations
- Semantically clearer to write and debug

Write queries
- More time-consuming, but is a time investment into more efficient read queries in the future
- Maintains model alignment with real world



Query 1: Who are the students who stop-out?

Query 2: Given identical TSI readiness profiles, what are the shortest and longest pathways to first certification?

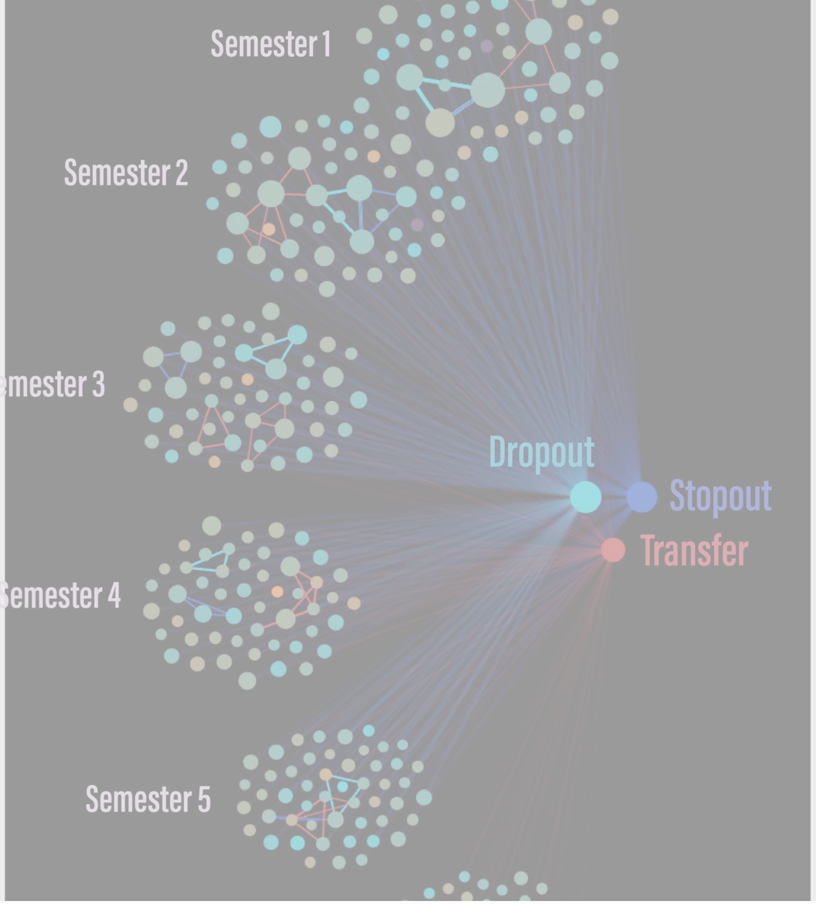
Query 3: What are the top sequences of courses taken prior to dropout?

Query 4: Update student data with new records from latest semester.

Query 5: Restructure graph such that `ecodis` attribute is a vertex and STUDENT vertices point to new vertices via edges typed *`of status`*.

Summary

- Digital twin paradigm for educational big data
  - Improved decision-making and student outcomes
  - Go beyond isolated snapshots of data
  - Large-scale and dynamic with millions of student records across Texas postsecondary domain



- Limitations
  - Limited by privacy restrictions
  - Limited sharing to only authorized stakeholders
  - Culture change, adoption (Thanks to Dan and the TACC team!)

- Future directions
  - Security and privacy for digital twins: Information-theoretic approaches to digital twin security that leverage graphical model
  - Extending our construct to larger slices of the student population