Participatory Causal Modelling of Learning Systems Ben Hicks : Connected Intelligence Centre : University of Technology Sydney Learning Analytics (LA) aims to improve the learning process. This necessitates a *causal* interpretation of observational data. One way to model causal structure is by using causal Directed Acyclic Graphs (DAGs). The visual formalism of the model requires little technical knowledge to engage with, providing an opportunity for non-technical experts to remain engaged deep into the crafting of critical statistical assumptions about the learning system, including the importance of latent variables.



Constructing DAGs as a collaborative thinking

process

How can I collaborate with stakeholders?

A **DAG** is a graphical representation of how variables influence each other in a system. If we think changes in A result directly in changes in B we draw $A \to B$. This forms a **D**irected **G**raph (**DG**). If it is also **A**cyclic (no) loops) then we have a DAG, which has a precise mathematical translation to the joint probability distribution of the data which can be leveraged in a several ways.

How could these these models help LA?

Affordances of DAGs for Learning Analytics

Can I just have a go at drawing one with you now?

Abstract but accessible

Constructing a causal DAG requires minimal technical knowledge. This means it offers a way for non-technical Education experts to describe the important aspects of a learning system so that it can be leveraged by **Dat**a experts.



Interrogating the model, together

The graphical model can be interrogated at any point using prompts (such as the example below). This helps facilitate precision in thinking about what is important in the system.

From graph to causal claim

Any given causal DAG directly translates to how we can decompose a joint probability distribution, using do-Calculus (Pearl, 2009). For instance, the graph $X \leftarrow Y \rightarrow Z$ describes the factorization P(X,Y,Z) = P(X|Y)P(Y)P(Z|Y) and with it the conditional independence $X \perp Z | Y$.

What is important here is that each DAG describes how best to understand the causal relationships between variables. Typically, this is used to find causal relationships from observational data by understanding which variables to control for to minimise bias (Weidlich et al., 2022). This set of variables, for a given relationship, is known as the adjustment set.

Beginning with any two causally connected nodes in the graph: Does X influence Y directly, or is there some other variable (M) in be-

(M is called a *mediator* which we are inserting into the path $X \to Y$)

Are there other things (M) that X changes that in turn change Y? (Adds a new path with a mediator in it)



Abstraction allows comparison

Models built from a range of stakeholders can be compared. The formal requirements of constructing a causal DAG allow for easier comparison – the price to achieve this is paid through reducing the complexity of the system so it can be modelled.

Testable learning theories

Kitto et al. (2023) outline a way to use a causal DAG of a learning theory to generate a collection of implied conditional independence relationships. These can be used build tests to examine to what degree the theory holds in the observed data.





DAG informed dashboard design

An adjustment set of variables has implications beyond statistical models. If a dashboard shows a comparison between X and Y then an unbiased view of the data should compare these variables within levels of the adjustment set. This could be implemented with filters, slicing or aggregation.

PhD publication plan

This poster describes a slice of the work towards my PhD. Feedback welcome, and cases to take part in collaborative modelling very welcome!

A paper (not yet written) to help frame this kind of thinking.

ABSTRA Learning Ar effort to ensite to its devel that suppoor graphical cc providing a tially challe as they for its potentia to causal cl help us to r is illustrate modelling. CCSCOO • Computi Human-ce puting the KEYWO causal mod tion, diagra ACM Refer Ben Hicks, K Tinikking wit

Permission to classroom use for profit or c on the first pu must be hono to post on ser fee. Request j *LAK22, Marc* & 2022 Assoc ACM ISBN 97 https://doi.or

A paper and a commentary introducing using DAGs as a way to think about education system.

An idea of what to do when non-linear causation is required.

Ben Hicks

arning Analytics is full of situations where featur

GenAI in the classroom as a case in point.

CCS Concepts

Keywords

1 Introduction

2025 Copyright held by the owner/author(s). ACM ISBN 979-8-4007-0701-8/25/03 https://doi.org/10.1145/3706468.3706557

nent, their interactions with class er

cooperating or free-riding on the provisioning of feedback in a class activity is used as a case study. We show how our initially imple model can gradually be built up to help understand potential

• Applied computing \rightarrow E-learning: • General and reference \rightarrow Estimation; • Computing methodologies \rightarrow Modeling and simulation.

ame theory, learning analytics, missing data, learning theory

ACM Reference Format: Ben Hicks and Kitryk Kito. 2025. Game Theoretic Models of Intangible Learning Data. In LAK25: The 15th International Learning Analytics and Knowledge Conference (LAK 2025), March 63–07, 2025, Dublin, Ireland. ACM, New York, NY, USA, 7 pages. Histly:/doi.org/10.1145/3704668.3706557

Much of what is important to learning - thinking, social interaction, lecisions - is incredibly challenging to measure or even observe. This tension between what learning is and what artefacts of learning

cator responses as new situations arise, using the emergence of

Connected Intelligence Centre University of Technology Sydney Sydney, NSW, Australia ben.hicks@student.uts.edu.au

Collaborative Causal Modelling

with Learning Experts

A paper (in progress) building on the work

presented in this poster. I hope to find out:

RQ1: How well do these models represent

an experts understanding of a learning

RQ2: What are the affordances of

graphical causal models to develop

thinking and a shared understanding?

Interested in modelling a system for this

work? e: ben.hicks@student.uts.edu.au

system?

SeLAR SRELTY by MESEANNER

Causation and the Interplay Between Learning

Ben Hicks, Joshua Weidlich, Kirsty Kitto, Simon Buckingham Shum and Hendrik Drachsle

Motz et al. (2023) make the point that learning analytics should be more frequently measuring learning outcomes

and making interventions in order to better align with its stated goals. These two aspects of their critique are

symptomatic of an underlying need for a more formal modelling of causality. We comment on how this might be expected in an emerging field and offer a potential way forward.

Distinguishing between learning and learning outcomes is key for understanding learning interventions

Motz et al. (2023) lament the misalignment between the stated aim of learning analytics (LA) and the current state of resea

questioning how we can "understand and improve learning" if we rarely measure learning outcomes or intervene in learning systems. We are in broad agreement with this stance but wish to interrogate to what degree this might be expected in an

As an emerging field LA does not yet have well established theoretical frameworks (Kitto et al., 2023). LA is still exploring

testable theories of learning, and as such we would not yet expect widespread measurement of a complex phenomenon such as learning, despite the abundance of data. Something is missing between the deluge of data recorded in learning environments and our understanding of the learning processes that leave digital traces. A breadth of research may still be required in order to establish an empirical basis upon which more formal theoretical models and measures of success can be built. The lack of

to establish an empirical basis upon which more formal necerical models and measures of success can be outil. The lack of measurement of learning outcomes may be a symptom of the challenges present for an immature field, or indeed any field struggling to theorize and operationalize the complex constructs that we commonly designate as "learning." Motz et al. (2023) acknowledge this immaturity of the field in reference to the deficiency in active interventions. We think these two challenges for LA in meeting its stated goals — measuring learning outcomes and making interventions — are linked and stem from the informality in how we think about the causal structure of LA systems. Being mindful and precise about how we obstrate its original to mediline learning (Group 2010). Immediate learning incoming learning the prime concernent

we abstract is critical to modelling learning (Essa, 2019). Improving *learning* is one thing, but improving *learning outcomes* is something entirely different. How we think about and model the causal relationships around this distinction impacts framatically on how we might use LA to guide interventions. To be clear, here we see learning as the (unseen) cognitive process of understanding more about the world, and the learning outcome as some, possibly measurable, proxy of this process (Wilsevy et al., 2019). We take a slightly narrow view of learning outcomes here as data intensive fields such as LA gravitate towards *measurable* outcomes. For instance, learning is what is happening in the student's mind, leading to improved

erformance in an assessment --- the learning outcome. Optimizing purely for a measurable learning outcome could lead to

mptore a statistic sector and the statistic possible provide statistic possible statistic

the consequential validity (Messick, 1989). In this view, interventions can correctly target the variable of interest (thus valid in the usual sense) but still produce downstream educational consequences that may be unintended or undesired.

we might want to start "privileging the measurement of learning outcomes... to approximate a shared trajectory toward

optimizing and improving them" (Motz et al., 2023, p. 9). One way of thinking about consequences of LA interventions is

ion skills, possibly improving learning outcomes without improving learning, a phenomen

arcical interventions such as simplifying learning content or making assessments easier. Less obvious could be an intervention

Graphical causal models provide a way to transparently model causal structure

emerging, complex field, and how thinking about causal structure offers a possible direction forwar

Outcomes and Learning Interventions

Notes for Practice

1. Commentary

Causal models, theory, learning outcome

References

Ferguson, R., Khosravi, H., Kovanović, V., Viberg, O., Aggarwal, A., Brinkhuis, M., ... & Yan, V. X. (2023). Aligning the goals of learning analytics with its research scholarship: An open peer commentary approach. Journal of Learning Analytics, 10(2), 14-50. – "Causation and the Interplay Between Learning Analytics and Learning Interventions" is part of this.

Game Theoretic Models of Intangible Learning Data Hicks, B., Kitto, K., Payne, L., & Buckingham Shum, S. Kirsty Kitto

Connected Intelligence Centr

University of Technology, Sydney Sydney, NSW, Australia kirsty.kitto@uts.edu.au

may always be so [18]. However, the difficulty of observing the

to think carefully and critically about what data might be missing from LA and how this might impact our models. This paper will

entifying and supporting students' at risk of failure. Howe

accessful and others are not [17]. In short, learning is a cor

ese intangible data are key to the learning experience, wh

RQ: How can game theoretic models connect intangible data t

the data that we have available in LA'

components of learning [9, 12]. This

gible, such as a student's disposition, engager

be critical contextual features of an educational system matate apossible to measure. This leaves an empty space where essential at as insising from our analysis. This paper proposes the of Game Theoretic models as a way to explore that empty space and potentially even to generate synthetic data for our models. This paper win propose one new method for progressing in our modelling over intangible learning data. Cooperating or free-riding on the provisioning of feedback to the provisioning of feedback to the provisioning of feedback to processes, interactions between learners and their environment

Much of what is important to learning. Hunking, social interaction, decisions - is incredibly challenging to measure or even observed through decisions - is incredibly challenging to measure or even observed through enter of the order of t

(2022, March). Thinking with causal models: A visual formalism for collaboratively crafting assumptions.

In LAK22: 12th international learning analytics and knowledge conference (pp. 250-259).

Hicks, B., Kitto, K. (2025, March). Game Theoretic Models of Intangible Learning Data. In LAK25: The 15th International Learning Analytics and Knowledge Conference.

Kitto, K., Hicks, B., & Buckingham Shum, S. (2023). Using causal models to bridge the divide between big data and educational theory. British Journal of Educational *Technology*, *54*(5), 1095-1124.

> Pearl, J. (2009). Causality (2nd ed.). Cambridge: Cambridge University Press.

Weidlich, J., Gašević, D., & Drachsler, H. (2022). Causal inference and bias in learning analytics: A primer on pitfalls using directed acyclic graphs. *Journal of Learning* Analytics, 9(3), 183-199.

Describing bounds for locally linear causation

When we model a system using a DAG we are describing a system with linear causation – effects are additive. In reality most learning systems will exhibit nonlinear causation if we look closely enough.

This (more philosophical) paper will argue that the key is to describe the boundaries within which the system is sufficiently describe with linear causation (such as a DAG).

Ben Hicks	Kirsty Kitto
University of Technology Sydney	University of Technology Sydney
Sydney, New South Wales, Australia	Sydney, New South Wales, Australia
Charles Sturt University	Kirsty.Kitto@uts.edu.au
Albury, New South Wales, Australia	
bhicks@csu.edu.au	
Leonie Payne	Simon Buckingham Shum
University of Technology Sydney	University of Technology Sydney
Sydney, New South Wales, Australia	Sydney, New South Wales, Australia
Leonie.E.Payne@student.uts.edu.au	Simon.BuckinghamShum@uts.edu.au
Т	1 INTRODUCTION
lytics (LA) is a bricolage field that requires a concerted	To think at all is to forget. To know at all is to abstract.
e that all stakeholders it affects are able to contribute	When we reason scientifically, we go further and deeper
ment in a meaningful manner. We need mechanisms	with thought. We become deliberate with what to forget and mindful of how to abstract Essa [17, p34]
sal models can help us to span the disciplinary divide.	I consider A colotica (I A) is constituted as formed to an a briteland
ew apparatus to help educators understand, and poten-	field [19] that requires people to work in a 'middle space' between
e, the technical models developed by LA practitioners	the learning and analytical sciences [30, 58] to bridge epistemic
We briefly introduce causal modelling, highlighting	boundaries [20]. As such, it is a field that can suffer from problems
enefits in helping the field to move from associations	of communication, where people from very different cultural back-
son about complex statistical models. The approach	grounds talk past one another. Often, we see educational experts
by applying it to the well known problem of at-risk	excluded from the conversation, unable to judge of evaluate nighty
	learning, and other methods that quickly come to resemble a black
	box [40]. This difficulty in traversing disciplinary boundaries leads
CEPTS	to a number of critical challenges for the field around who gets to
g methodologies \rightarrow Modeling and simulation; •	Participate in defining the questions that the field explores [64], and
tered computing → Collaborative and social com-	how educational <i>Theory</i> can be used to inform results. Furthermore,
ory, concepts and paradigms.	can make it difficult to Intervene in the system even with strong
DS	statistical results.
directed acyclic graphs, transdisciplinary collabora-	In this paper we will argue that many of these problems can
matic reasoning	be addressed, at least partially, by thinking with causal models. We
an Formati	Directed Acyclic Graphs (DAGs) can be used with little mathemat-
ty Kitto, Leonie Payne, and Simon Buckingham Shum. 2022.	ical expertise to provide a highly interpretable artefact that sup-
ausal models: A visual formalism for collaboratively crafting	ports genuine transdisciplinary collaboration among technical and
LAK22: 12th International Learning Analytics and Knowledge	non-technical stakeholders. After elaborating upon the problems
K22), March 21–25, 2022, Online, USA. ACM, New York, NY, https://doi.org/10.1145/3506860.3506899	introduced above (Section 2), and then providing a brief introduc-
	tion to the framework first introduced by Pearl [41] (Section 3), we will demonstrate its utility in thinking about complex advectional
ke digital or hard copies of all or part of this work for personal or	processes by reasoning through a real world scenario encountered
rearted without fee provided that copies are not made or distributed sercial advantage and that copies bear this notice and the full citation Copyrights for components of this work owned by others than ACM	by one of the authors in his position as an LA professional (in Section 4).
Abstracting with credit is permitted. To copy otherwise, or republish, or to redistribute to lists, requires prior specific permission and/or a	1.1 Challenges facing the field of Learning
hissions from permissions@acm.org	Analytice
on for Computing Machinery.	Principal the technical shallowers of softwaring useful to form
4503-9573-1/22/03\$15.00	beyond the technical challenges of gathering useful data from mul-