# Topological / Categorical Framework for Analysis and the Information Loss through Space Transitions and something about Learning Analytics

Research Proposal for PhD – Ben Hicks

## Context

Learning Analytics, and the process of analytics in general, can be viewed as a sequence of mappings / transformations from reality, to data, to some kind of artefact (text, graphic, conversation, email, number) **[and a narrative??]** and finally then interpreted by some observer. Through each step various necessary simplifications are made, be they assumptions, sampling, modelling, visualising, dimension reduction or even the creation of metaphor for clearer explanation. The end goal is to try and convey some *meaning* about the beginning reality to some person, but the choices and processes made along the way can distort or *over* simplify the original truth. All analysis and storytelling is a shadow of the world. The question is can we be sure enough about the complexity of the original object of interest to know that the silhouette captures essential (and possibly generalizable) information or is it just particular to this particular set of data.

I intend to look at the transitions between these spaces and attempt to quantify the loss of structure, and hopefully develop tools to help people understand the limitations of analysing complex systems and the boundaries of the conclusions they can draw.

## Aims and Objectives

* Developing a mathematical framework for viewing the process of moving from reality to an analysis product (or further towards an end users interpretation of the result…).
* Testing this framework and process on data analysis performed on a complex system, ideally in the field of Learning Analytics and understanding student’s interaction with learning design.
* Developing tools to help others make a sanity check on their analysis by analysing the complexity of the spaces they are moving between and highlighting the (possibly) hidden assumptions and simplifications used.

## Literature Review

*The complex*

Stuff on complexity theory (), topologies (distinguishing between metrizable and non-metrizable topological spaces – see <http://homepage.math.uiowa.edu/~jsimon/COURSES/M132Fall07/MetrizationTheorem_v5.pdf>)

*The analysis product*

Grammar of graphics, scedastics

Narrative? Sense making?

*Moving between the two*

Categories, information flow reference

## Research Methodology

Although there is little in the application to data science or analysis, the tools available in Category Theory and Topology could provide a flexible enough lens to both define the various spaces as well as descript the morphisms between them, even the basic notion of equivalence relations could be sufficient in parts.

*I’m not really sure what to write here. Mathematics??*

## Research plan and timetable

In rough 6-month section:

* Research into previous work on complexity, applications of category theory, information theory, topology (in particular complexity-analogous measures such as metrizability). Sketching ideas for objects and morphisms in the analysis process and conversations with people with experience in the field at the different stages of analysis.
* Formalising the mathematical structure to be used, and linking this with the research. Identifying holes in the research that need to be filled in order to apply to process of learning analytics. Exploring options for a data set to apply this too; this could be leveraged with my current work with CSU analysing student engagement data (which is certainly complex enough). Applying for ethics once a data set has been decided upon. Beginning literature review formal writing on research conducted in last checkpoint.
* Polishing the framework and first round of applying the framework to a data set. Finalising literature review. Possible paper on framework.
* Refining application of model to data analysis process; exploring options for quantifying space-complexity metric and whether it is possible to retroactively look at an analysis and quantify information loss.
* Further exploration and refinement of information loss metric that can highlight hidden assumptions and simplifications; and how broadly applicable it could be. Drafting thesis.
* Finalising thesis. Fixing numerous errors.

## References

## Notes

Complex Systems – a bit all over the shop, but a good lens. See Bruce Edmonds thesis – ‘Syntactic Measures of Complexity’

Epistemic Cut

Grammar of Graphics – see Willconson book. Might be a way to specify the codomain of the mapping form ‘learning space’ to artefact / analysis.

Human Interaction conference

IEEE transactions on visualisation and computer graphics – Quality Metrics in High Dimension data visualisation (has a good table)

**Graph theoretic Scagnostics – Wilkonson; see who has cited this.**

John Barwise and Seligman; Information formed framework to model information flow “Semantic Information” Stanford Encyclopedia

Data storytelling; trying to make explicit;

* Assumptions
* Info loss (gain?)
* Where?
* Tools to mitigate loss

Other: Can you predict a replication problem based on some measure of info loss used in the analysis?