ALASI 2019 workshop planning

# Themes

Anything that re-occurs throughout as a running theme goes here.

## Question >> Answer

The **right** question is more important than an exact answer. What questions are important?

## Bridging the gap between the data and the question

We have the data, we have questions that we want answered (usually, and usually ill-defined), but there is a murky no-man’s land in between.

## Best way to share findings

Could include material from previous talk here.

## Obviously

Sometimes an obvious result is still useful. We wouldn’t want conflict between the analytics and our lived experience all the time. And a result such as more time means better performance is surely validating of learning design.

## Intuition

Human choices abound in the analysis process and are often smoothed over. They are necessarily based often on intuition / experience because performing a full analysis on every decision is unfeasible. But intuition / experience is built on getting to know the context and the data well, or forming relationships with those that do. Sometimes challenging to be aware of these decisions.

# Blending machine learning, graph theory and social network analysis to better understand student engagement and learning.

#### Following the white rabbit of trace data

##### ALASI 2019

## Intro

* Me
* Structure of the session; broad concepts, a specific story, and generalising / playing.
* Cues; dark slides, resources

## Part 0: Questions, Mapping

### The right question is more important than an exact answer.

* People are often seduced by accuracy. Comfort in definite truths, uncomfortable with uncertainty. If you can give sharp focus to a small part of the picture people feel like they understand (*the idiom of the blind man and the elephant*). What we chose to focus on is often dictated by the available data, …, and which questions lead naturally to definitive answers. Also, the right question is not just what is in front of you (the data).
* The analyst is often the blind man and what they can see is the data. Better to avoid getting a perfectly accurate picture of what is there at the expense of contributing to a more complete / useful picture as a whole.
* How people learn is not a field that lends itself well to exact answers. It involves people, the brain, and learning.
* Improving questions is often the job of a data scientist

### Mapping (embedding, feature engineering, wrangling, tidying)

* There is the question. There is the (raw) **trace data**. How we bridge between has different names, let’s call it mapping.
* Example of raw compared to what a machine needs. Maybe include some of the stats table as an example of decisions being made.

### Case study

* Data example from above
* How did students engage with the content?
* Did the forums help students learn?

## Part 1: Clustering

How did **students** engage with the content?

* Students. We can get precise answers to individual students. And we can average all students. But it is more useful to look at groups of students (clusters) but this is fuzzier. Machine learning is a tool that can do this, but it only understands certain things, so we must make decisions. These decisions *will* influence results, but hopefully in an informed way.
* Mapping: What’s important? My choices are… [Create handout of variables chosen with definitions]
* Discussion.
  + What did I miss?
  + What will be the most important in distinguishing groups of students?
  + In distinguishing academic performance?
  + Reveal what I would include (mean, median, sd of timestamp).
  + Optional resource – analysis of those features.
* Visualising the clusters, describing the clusters

## Part 2: Connecting to learning design

How did students engage with **the content**?

*Need sequence of content activity maps for these slides, including animations.*

* What does the content actually look like (network map, overlay to refine picture)
* How do the different clusters look on the network map?

## Part 3: Harmonics

How did students **engage** with the content?

* We have clicks; when, where, who. Imagine you’re the blind man again, and just listening to the clicks – how would you pick up engagement, or distraction?
* What does engagement look like when framed through a time-interval lens?
* Workshop: Hand out scale and classify activities
* Show spectrums… clusters and possibly content

## Part 4: Social Networks

Did the forums help students learn?

* Did students form cohesive networks? What kind of network evolved in the re-designed course compared to the previous course?

(Use reciprocity, transitivity and triadic census to try and articulate this and compare to random graphs of the same)

* Reciprocity
* Transitivity
* Traidic Census

## Part 5: Try this at home

## Handout 1 – Mapping (features selection)

Features I chose:

* **Accesses**: The number of times a student accessed the site (i.e. a single login session)
* **Clicks per access**: The average (mean) number of clicks the student made per access
* **Clicks standard deviation**: The standard deviation of the number of clicks per access
* **Time**: The total time the student spent on the site
* **Time median**: The average (median) time the student spent on the site per access
* **Time standard deviation**: The standard deviation of the time per access
* **Forum views**: The number of times the student viewed any forum related content

## Handout 1 – Mapping (features selection)

Features I chose:

* **Accesses**: The number of times a student accessed the site (i.e. a single login session)
* **Clicks per access**: The average (mean) number of clicks the student made per access
* **Clicks standard deviation**: The standard deviation of the number of clicks per access
* **Time**: The total time the student spent on the site
* **Time median**: The average (median) time the student spent on the site per access
* **Time standard deviation**: The standard deviation of the time per access
* **Forum views**: The number of times the student viewed any forum related content

## Handout 2 - Clusters

## D:\Users\bhicks\Documents\alasi.2019\clusters_A4_B3.png

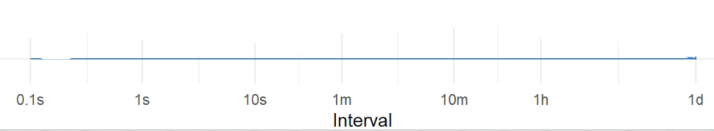
Subject A Cluster Summaries

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cluster | n | Grade | Forum | Accesses | Clicks pa | Clicks sd | Time | Time md | Time sd |
| 1 | 53 | 0.658 | 63 | 38 | 19 | 23 | ~14.49 hours | ~3.26 minutes | ~46.18 minutes |
| 2 | 62 | 0.760 | 147 | 69 | 24 | 30 | ~2.09 days | ~4.7 minutes | ~1.41 hours |
| 3 | 32 | 0.773 | 334 | 63 | 34 | 39 | ~3.38 days | ~17.24 minutes | ~2.04 hours |
| 4 | 14 | 0.818 | 712 | 168 | 23 | 26 | ~5.12 days | ~4.79 minutes | ~1.43 hours |

Subject B Summary

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cluster | n | Grade | Forum | Accesses | Clicks pa | Clicks sd | Time | Time md | Time sd |
| 1 | 78 | 0.687 | 87 | 44 | 17 | 20 | ~1.6 days | ~7.89 minutes | ~1.55 hours |
| 2 | 100 | 0.692 | 20 | 38 | 10 | 11 | ~17.65 hours | ~1.92 minutes | ~58.84 minutes |
| 3 | 9 | 0.840 | 703 | 126 | 17 | 16 | ~4.66 days | ~7.44 minutes | ~2.07 hours |

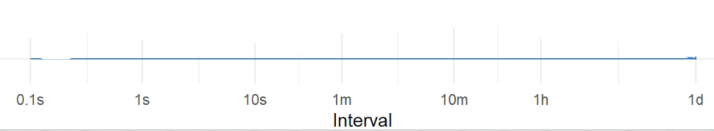
## Handout 3 – Spectrum Scale



Where would you place the following?

1. Navigating through a course site
2. Reading basic instructions
3. Forgetting to log out and coming back the next day
4. Looking at content but not deeply engaged, clicking around
5. Deeply engaged
6. Watching extended material or using an external tool

## Handout 3 – Spectrum Scale

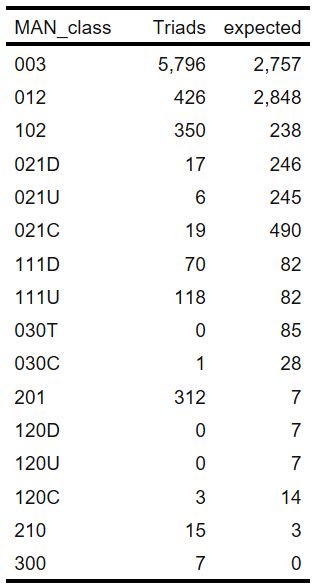
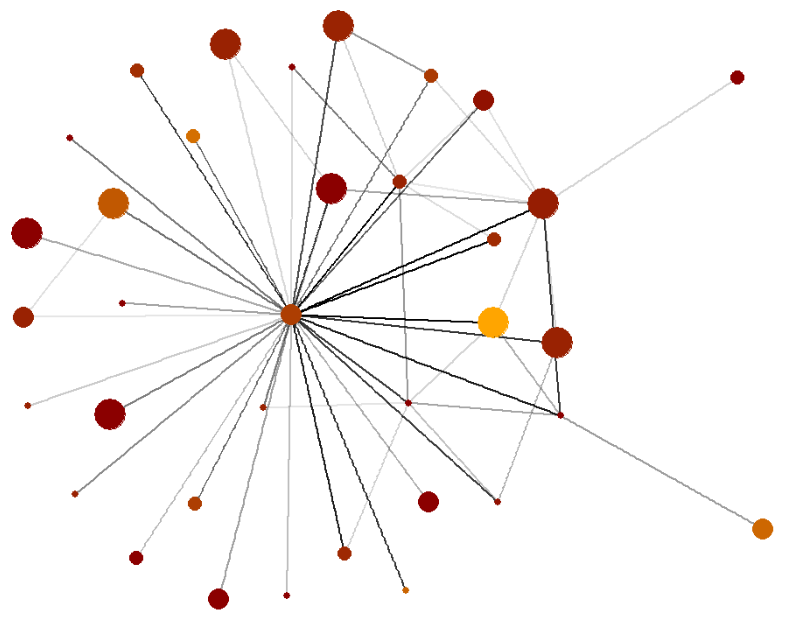


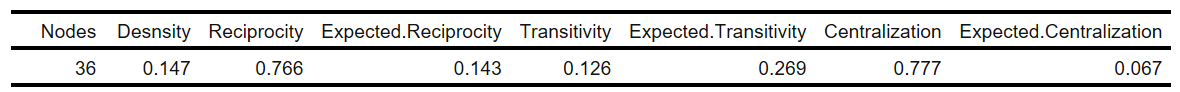
Where would you place the following?

1. Navigating through a course site
2. Reading basic instructions
3. Forgetting to log out and coming back the next day
4. Looking at content but not deeply engaged, clicking around
5. Deeply engaged
6. Watching extended material or using an external tool

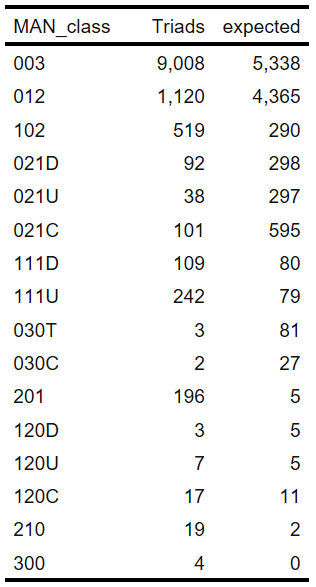
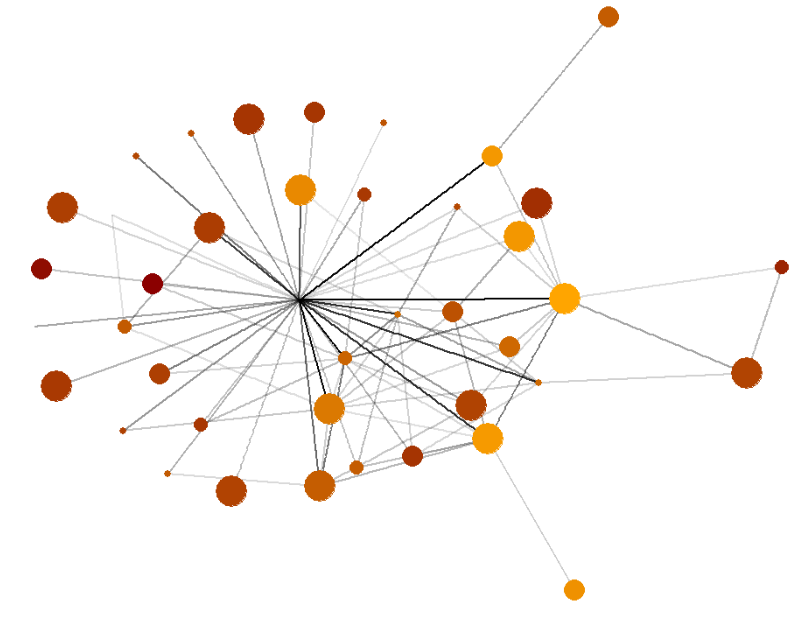
## Handout 4 – Forum Comparison

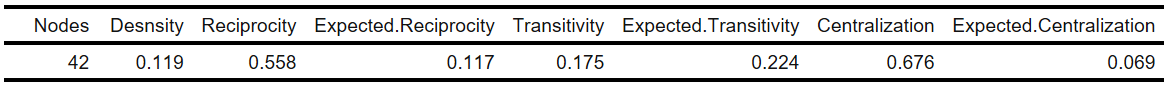
### 2017





### 2018





### **Mutual-Asymmetric-Null Triadic Classes**

The Mutual-Asymmetric-Null encoding system classifies triad connections in a directed graph. The three numbers represent, in order, the number of (M)utual edges, the number of (A)symmetric edges, and the number of (N)ull or no edges between nodes. The possible connections (and their names) are outlined below:

