



What's popular?

## Description

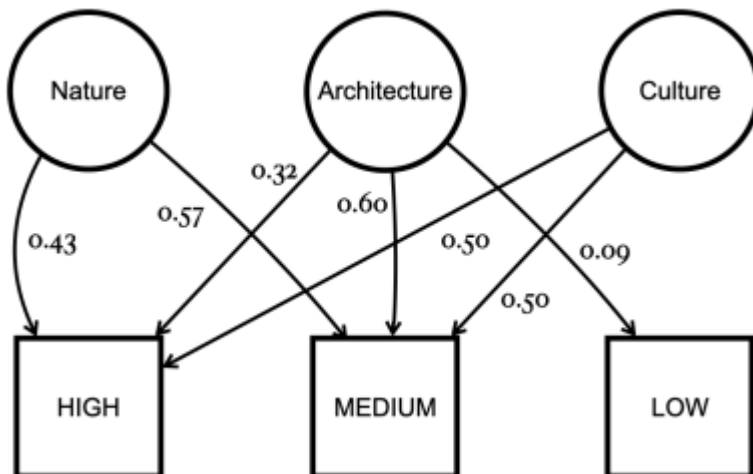
What's it like to be liked? The WordPress blogging platform delivers analytics for each blog post providing the author with stats on visits, likes and comments. You would expect older posts to have more visits than recent visits. So it's hard to compare the popularity of posts. But you can compare these stats if you just note the number of visits in the busiest month for each post.

I tabulated those figures for each post category for my last 100 posts. That enabled me to establish a crude system for classifying the popularity of each post as High, MEDIUM or LOW, according to how many visits the post received in its best month.

I then reviewed what category each post belonged to (e.g. Architecture, Nature, Culture) and worked out the proportion of HIGH, MEDIUM and LOW for each category. These proportions then indicate a probability. If the post is about Architecture then there's a 0.32 probability that it rates HIGH in popularity. If it's about Nature there's a 0.5 probability that it rates HIGH. I show this in the following table. For simplicity, I'll only deal with 3 of these categories, shown here highlighted in grey.

	HIGH	MEDIUM	LOW
Architecture	0.32	0.60	0.09
Artificial Intelligence	0.67	0.33	0.00
Body	0.00	1.00	0.00
Culture	0.50	0.50	0.00
Economics	0.00	0.83	0.17
Ethics	0.33	0.67	0.00
Film	0.75	0.00	0.25
Media	0.00	0.50	0.50
Metaphor	1.00	0.00	0.00
Nature	0.43	0.57	0.00
Play	0.50	0.50	0.00
Research	0.00	1.00	0.00
Society	0.00	0.67	0.33
Voice	0.43	0.57	0.00

This is how these probabilities look as a network graph.



According to these figures, if I know that a blog post is about Nature it is likely to be more popular than posts about Architecture (0.43 versus 0.32). That could be a useful finding. If a blogger wants to be popular they should write more posts about Nature than about Architecture.

### When all you have is an observation

The rectangular nodes in the network are results, effects, outcomes, or conclusions of a straightforward deductive process. In Markov process terminology they are referred to as *observations*.

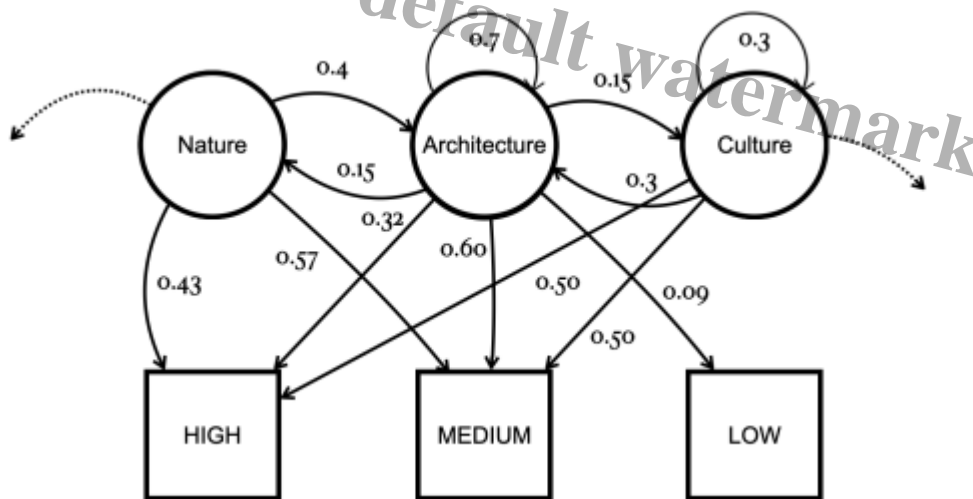
Let's assume that popularity is something that can be *observed*, thanks to the ubiquity of published statistics about likes, numbers of hits, star ratings, etc.

If a post has high popularity, what is the probability that it is about Nature? That's harder to determine. As it happens, in my sample of 100 consecutive blog posts, nearly half are about Architecture, less than a tenth are about Culture or Nature. So the calculation about whether or not the post is in the Nature category would have to take account of the relative frequency of that category of posts.

## Incorporating sequencing information

As I outlined [previously](#), blog posts are published in a time sequence. So we might want to use this information to predict what will happen next: e.g. if I've just looked at a highly popular post, what is the probability that the blogger will publish the next post in the time sequence that is also highly popular.

For this calculation, further information is required. The Markov network shown in my previous post captures information about the relationship between blog post categories in temporal sequence. The combination produces something like this.



That combines probabilities about the temporal relationships between the blog post categories and the probabilities of the popularity of each category. I'm not going to do any calculations with that here. In Markov process terminology, the circular nodes and their relationships are *hidden*; the rectangular nodes are what is *observed*.

The sequential model (circles and arrows) doesn't look very hidden here. It is hidden in the sense that it lies behind the observations, which are visible. You could say the relationships between the observations are difficult to discern. The complicated hidden sequential model mediates the relationships between the observations. The hidden components may also be unknown, and have to be derived by inference, or [abduction](#), and drawing on methods from [AI](#).

There's a very helpful [video](#) by Luis Serrano that explains this method of Hidden Markov modelling and how to perform calculations with it. His simple example involves inferring what the weather is like from someone's mood. We don't know what the weather is like as we are speaking to this person online. He won't talk about the weather, but we know his mood. Usually a good mood means it's sunny, but not always. The Markov method allows you to calculate what a succession of daily moods

tells you about a succession of hidden weather events.

## Applying Markov models

Serrano explains some more serious applications of the Markov approach. The most usual is in parsing, translating and interpreting strings of text. Text is after all a time sequenced phenomenon. Robot or drone localization is another application area. The observations are what the robot's image recognition system detects in its environment through sensors and the hidden states are the most likely coordinates of the robot's location. In genetics, Markov procedures are also used in making inferences about DNA sequences.

The spaces inspected by planners and designers can be modelled as Markov domains. Any spatial domain can be turned into a problem about sequencing. Think of real estate: the changes in property values along a series of houses in a street. A plan of a city is produced by sequential operations such as drawing, tracing and scanning.



## References

- Rabiner, L.R., and B.H. Juang. 1986. An introduction to hidden Markov models. *IEEE Acoustics, Speech and Signal Processing Magazine*, (3)4-16.
- Serrano, Luis. 2018. A friendly introduction to Bayes Theorem and Hidden Markov Models. *YouTube*, 27 March. Available online: <https://www.youtube.com/watch?v=kqSzLo9fenk> (accessed 27 September 2020).

## Category

1. Architecture

## Tags

1. abduction
2. Markov process

**Date Created**

October 3, 2020

**Author**

rcoyne99

*default watermark*