

## All Roads Lead to Philosophy

Austin Barish

Georgetown University

**Abstract:** In this study, I analyze a phenomenon on Wikipedia in which repeatedly clicking “first link” of a webpage invariably takes a user to the Philosophy page. I examine the percent of pages on Wikipedia in which this idea holds true in an effort to understand how Wikipedia’s network is structured and what that means for its user navigability and understanding. Previous research indicates that users’ page navigation is heavily focused on the lead of a Wikipedia article, rarely venturing beyond the first paragraph<sup>1</sup>; therefore, I limit my analysis to the first several links in this section; further analysis with greater computing power could be done on the links within the entire article. Amongst these first several links, I seek to determine if there are any other link locations that reach a specific page with any abnormal frequencies, including the philosophy page. To conduct my analysis, I construct a network using Wikipedia pages as nodes and the links on the page as undirected links between nodes. Since I am focused on reaching the philosophy page, once I reach a page that has already been determined to reach the philosophy page, I move on to another root page. With the network, I examine average path lengths to the philosophy page, the neighbors of the philosophy page that most commonly direct to it, and the nature of the philosophy node itself that results in this phenomenon. My conclusions demonstrate the effectiveness of Wikipedia’s effort to make their introductory sentence and links broad as well as cementing of Philosophy as “the first science”.

**Keywords:** Wikipedia, Philosophy, Getting to Philosophy, First Page Philosophy, Navigability

## Introduction:

Wikipedia pages are built with the user's understanding in mind. To ensure consistency across pages and maintain reliability as a credible source, there are extensive guidelines on the structure of each page. As one of the most important components of a Wikipedia page, linked content and the content of the lead paragraph is tightly monitored. Links serve to “provide instant pathways to locations within and outside the project that can increase readers' understanding of the topic at hand.”<sup>1</sup> Users will click on links when a topic is unfamiliar to them, or they interested in learning more.

When arriving to a page, a user ought to have the topic explained to them as though they know little to nothing about it. The lead ought to frame the reader so as to “set the scene of the topic”.<sup>1</sup> Wikipedia explains the structure of the lead paragraph:

In Wikipedia, the lead section is an introduction to an article and a summary of its most important contents. It is located at the beginning of the article, before the table of contents and the first heading. It is not a news-style lead or "lede" paragraph.

The average Wikipedia visit is a few minutes long. The lead is the first thing most people will read upon arriving at an article, and may be the only portion of the article that they read. It gives the basics in a nutshell and cultivates interest in reading on—though not by teasing the reader or hinting at what follows. It should be written in a clear, accessible style with a neutral point of view.<sup>1</sup>

Wikipedia goes on to outline how the opening paragraph and sentence ought to be structured. They explain that the “[Opening paragraph] It should establish the context in which the topic is being considered by supplying the set of circumstances or facts that surround it. If appropriate, it should give the location and time.”<sup>1</sup> For example, a building's first link will most likely be its

location. Within that paragraph, its opening sentence is critical for my study as it will contain the first link. Editors are instructed that “the first sentence should tell the nonspecialist reader what or who the subject is, and often when or where.”<sup>1</sup> They go on to provide explicit instructions on what the first linked topic ought to be in an article:

The first sentence should provide links to the broader or more elementary topics that are important to the article's topic or place it into the context where it is notable.

For example, an article about a building or location should include a link to the broader geographical area of which it is a part.

**Arugam Bay** is a [bay](#) on the [Indian Ocean](#) in the dry zone of [Sri Lanka's](#) southeast coast.

In an article about a technical or jargon term, the first sentence or paragraph should normally contain a link to the field of study that the term comes from.

In [heraldry](#), **tinctures** are the colours used to [emblazon](#) a [coat of arms](#).

The first sentence of an article about a person should link to the page or pages about the topic where the person achieved prominence.

**Harvey Lavan "Van" Cliburn Jr.** (July 12, 1934 – February 27, 2013) was an American [pianist](#) who achieved worldwide recognition in 1958 at age 23, when he won the first quadrennial [International Tchaikovsky Piano Competition](#) in Moscow, at the height of the [Cold War](#).

Exactly what provides the context needed to understand a given topic varies greatly from topic to topic.<sup>1</sup>

As you can see, the first link of a page will be increasingly broad as you continue to click the first link. These instructions create a picture of how a topic like philosophy can be at the center of Wikipedia's first link network. Conversely, it is doubtful that such a center exists for another link placement. Even just the second link in an article can be increasingly specific, moving laterally or even backwards in specificity rather than towards larger hubs such as philosophy. Take one of Wikipedia's examples in Harvey Lavan "Van" Cliburn Jr; his first link path begins with pianist then continues as follows: piano, keyboard instrument, musical instrument, music, art, creativity, psychology, mind, thought, consciousness, awareness, **philosophy**. With each passing link you can sense that your destiny on the philosophy page grows closer; the topics are broader and the connection from it to philosophy feels increasingly obvious. However, if we were to follow the second link, International Tchaikovsky Piano Competition, we find ourselves on the following path: Saint Petersburg, Russia, Eastern Europe, Ural Mountains, Eurasia, Europe, peninsulas, mainland, continent, regions, Earth's surface, hemispheres, etc. Unlike with the first link, the second link gets stuck in geographic limbo without ever getting closer to a central topic like Philosophy. I will explore what a second link network looks like further in my analysis.

There is special focus on the very beginning of a Wikipedia page because that is where users devote most of their attention. Dimitrov et. al. utilize click data from Wikipedia's navigation logs to construct a heat map of where users are clicking the most on Wikipedia pages. The heat map illustrates two clear dark red, high density, lines at the beginning of the page directly where the lead is located, demonstrating that users highest click rate is on links within the first few lines of the opening paragraph. The rest of the page is sparse beyond a preference for links on the left side of pages, a phenomenon the authors themselves do not fully understand.<sup>3</sup>

However, the high click rate within the lead indicates to us that understanding the nature of the network of the first few links in an article is indicative of the nature of the network that users are typically interacting with.

Research has already been done into the size of the Giant Connected Component (GCC) of nodes that connect to the philosophy node. In a study of Wikipedia's navigability by language, as of 2017, 97.0% of pages in English will reach the philosophy page<sup>2</sup>, a slight increase of around 2.5% since 2011.<sup>4</sup> These numbers fluctuate across languages, with some languages have a center on pages such as Psychology in Spanish or Person in Japanese each with varying sizes but still having the majority of nodes reach these pages<sup>2</sup>; my study will only be focused on the English network of Wikipedia. In the future, it would be interesting to study this phenomenon in other languages as I have done with English. In particular, previous studies indicate that Dutch has the smallest GCC with just 67.0% of nodes in its GCC.<sup>2</sup> I would like to compare its network to English to understand this discrepancy.

If you would like to see how this network is formed beyond clicking through Wikipedia webpages on your own, the online page <https://www.xefer.com/wikipedia> will quickly build out a network of pages and their first links until you reach the philosophy page. This is a helpful tool that is good to visualize what this can look like in practice. However, it was designed to *always* reach the philosophy page even for those pages that manage to avoid the philosophy page. It does this by skipping to the second link on a page when it realizes it will not be able to reach the philosophy page through the first link.<sup>5</sup> Therefore, we need to construct our own network if we want to understand these disconnected nodes.

To understand how a node can be disconnected, we ought to look at what makes philosophy the center of the network. If you click on the first link on the philosophy page, you

will find yourself back on the philosophy page in 6 clicks. This self-loop forms a bottom of sorts to the network as nothing beyond the 5 pages you reach from the philosophy page can be found from there. Amongst those 5 pages, as I will later show, philosophy is by far the largest node by density, making it the logical choice for the center. For another node to avoid the philosophy node, it would require a similar cycle. Therefore, it is going to be a broad topic as it has to be something that could similarly be in the first sentence of a Wikipedia page. This eliminates super specific pages from consideration despite them being the intuitive guess for what might manage to avoid philosophy.

A page's neighbors will remain within semantically related to that page amongst links in the lead. In a study that constructed Wikipedia's network using the first ten links in an article as a node's edges, it was determined that the nodes will form into communities of semantically related terms.<sup>6</sup> For example, the mathematics page will be in a community of other topics related to math such as physics. For our sakes, this is an important result as it helps to paint a picture of what the branches stemming from philosophy's neighbors will look like. For example, we can now expect all scientific terms to be connected in communities allowing them all to pass through the science page on their way to the philosophy page.

Beyond some of the quicker results such as the size of the GCC, the average path length to philosophy, the number of disconnected components, and the nature of networks from other link locations, I will also look extensively at the neighbors of the philosophy node. If the philosophy node is removed from the network, how large is the remaining GCC and what is its largest node? I hypothesize that the awareness node and its connecting parts will form the basis of the GCC and that the network will not shrink by more than 10%. However, that if awareness were to be removed as well, the GCC would shrink dramatically as the awareness node serves as

a bridge between all scientific topics and all locations-based topics (buildings, monuments, historical figures). Finally, I will test whether distance from the philosophy node and node density can be demonstrated to be statistically significant in relation as I would expect due to the generality of the topics.

## References

1. “Wikipedia:Manual of Style.” *Wikipedia*, 2023,  
[https://en.wikipedia.org/wiki/Wikipedia:Manual\\_of\\_Style](https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style). Accessed 8 July 2023.
2. Evaluating and Improving Navigability of Wikipedia: A Comparative Study of Eight Language Editions, 2016, Daniel Lamprecht, Markus Strohmaier, OpenSym '16: Proceedings of the 12th International Symposium on Open Collaboration
3. Visual Positions of Links and Clicks on Wikipedia, 2, 2016, Dimitar Dimitrov, Markus Stromaier, WWW '16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web
4. “Wikipedia:Getting to Philosophy.” *Wikipedia*,  
[https://en.wikipedia.org/wiki/Wikipedia:Getting\\_to\\_Philosophy](https://en.wikipedia.org/wiki/Wikipedia:Getting_to_Philosophy). Accessed 8 July 2023.
5. “All Roads Lead to “Philosophy.”” *xef*, 2011, <https://www.xef.com/2011/05/wikipedia>. Accessed 20 July 2023.
6. Extracting Domain Knowledge by Complex Networks Analysis of Wikipedia Entries, 3, 2015, Neven Matas, Ana Meštrović, 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)