What's So Special About Philosophy?

Unraveling Wikipedia's First Link Network

Abstract

rewrite abstract

I. Results

I.1 Traversals

For each article, we the first link from one page to another up to an invalid link or cycle. By tracking the articles along each path, we can assign each article a rank measuring the number of paths containing a particular article. Through our traversal algorithm, we measure a total of 232.4 million traversals through the network of first links.

Remarkably, the distribution of traversals per article exihibits a power-law distribution ((citation + justification)) for justification look at definition of power-law; consider log-log with line. Of the 11 million articles, 78% have 0 traversals and 99% have fewer than 100 traversals. Of the highest ranking 100 articles, fewer than 30 articles have an overwhelming majority, ((7 million, confirm)) of the total number of traversals.

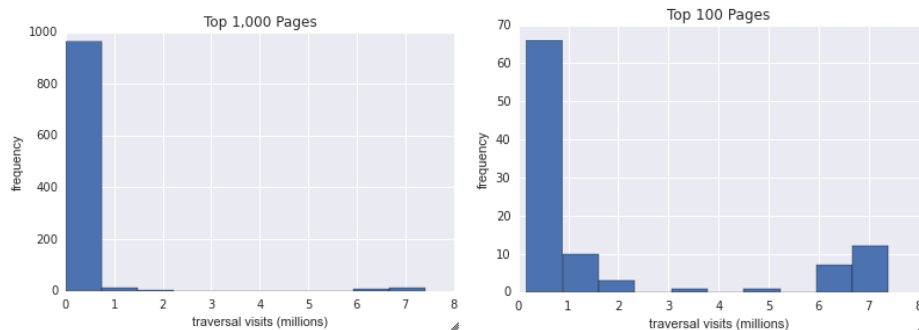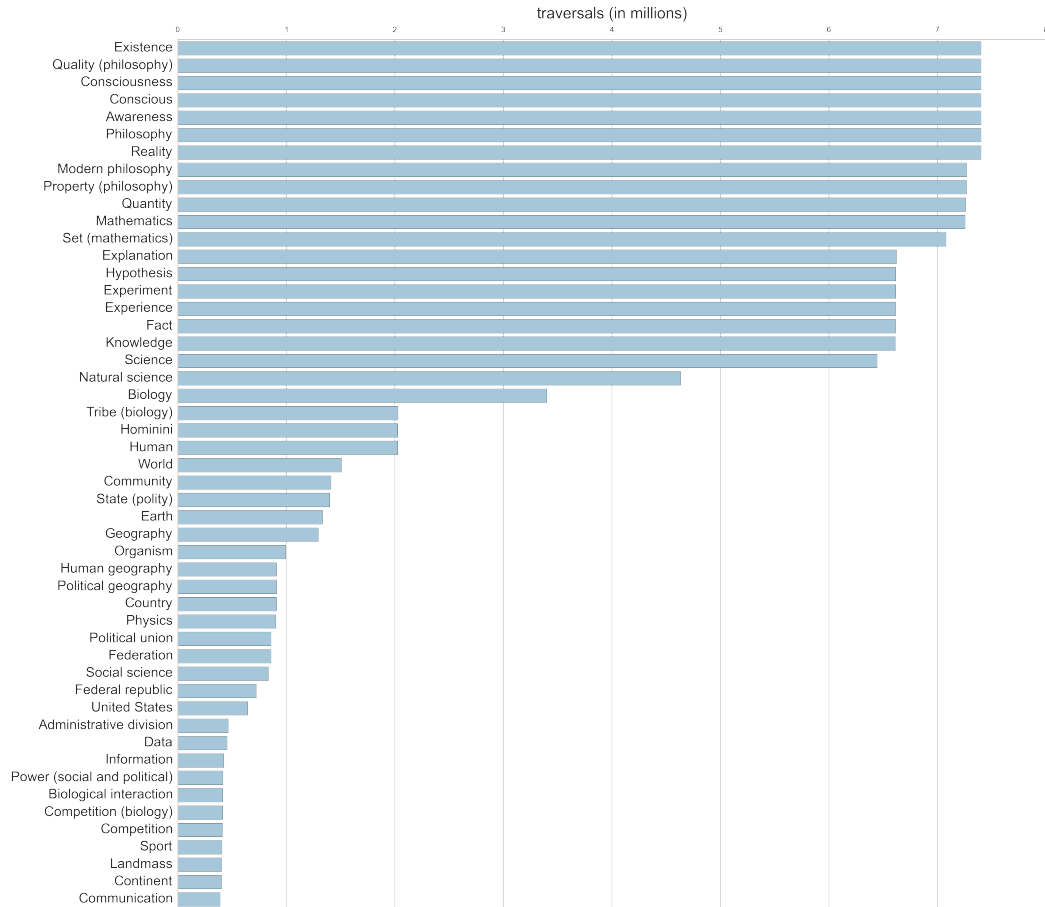*Figure 1: distribution of traversals*

1

Figure 2: highest ranking articles by number of traversals

The ranking by traversals reveals a cyclic structure marked by identical traversal numbers for the highest ranking 7 articles. By analyzing the traversal feeds, we can identify which articles in particular are the funnels or highest ranking within a particular cluster.
TODO
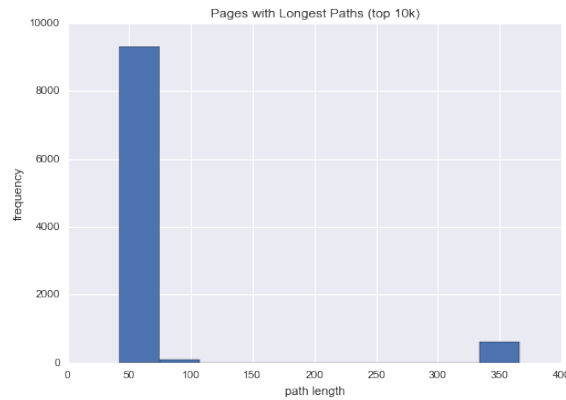((look at other descriptives about a distribution or even descriptors of power-law))

I.2   Traversal Path Length

By tracking the number of traversals to a repeated article or an invalid link, we can measure the path length starting at any given article. We discovered the longest path length is 365, following 365 calendar days of Orthodox Liturgics, where the first link of today's Liturgic article links to the next day's and so on. Curiously, at on the last calendar day the article links back to the first day. Similarly lengthy traversal paths included articles on the history of a nation on a particular year, for example 1953 in Scotland or 1560s in architecture.

Of the 11 million articles in our sample, 5.5 million had a path lenght of zero, meaning the first link is an invalid link or the article links to itslef (see section of cycles). The most frequent path length is 29, with an interquartile range of (26, 30). The distribution of

path lengths is similarly power-law like with few pages at the extreme of 355 traversals in their path length, while the majority is between 26-30.

Figure 3: Longest Path Lengths

I.3   Basins of Attraction and Traversal Feeds

A basin of attraction in our network is collection of path-connected articles. These articles do not necessarily form a loop, but may end at an invalid link. These basins are one way of measuring the center the of a network. If we rank the basins of attraction based on the accumulated number of ttraversals we find: many basins around philosophy:

Figure 4: Longest Path Lengths

	rank	loop
visits		
6613248	10	[Fact, Experience, Experiment, Hypothesis, Exp...
6444926	11	[Knowledge, Fact, Experience, Experiment, Hypo...
4633960	12	[Science, Knowledge, Fact, Experience, Experim...
3400690	13	[Natural science, Science, Knowledge, Fact, Ex...
2028827	14	[Biology, Natural science, Science, Knowledge,...
2027606	15	[Tribe (biology), Biology, Natural science, Sc...
2027495	16	[Hominini, Tribe (biology), Biology, Natural s...
1509110	17	[Human, Hominini, Tribe (biology), Biology, Na...
1412646	18	[]

*rank 18 are invalid links

i.e., pages linking to files, external pages, another part of the same article, wiktionary etc.

Excluding any basins containing Philosophy, we find basins of attractions around Community, Landmass, Federal Government, Belief System, Presentation, and similarly foundational human concepts. This remarkable results confirms the foundational quality these ideas have in our english-based society. Whereby these concepts naturally arose out fo the links structure formed by teh owrk of many individuals working independtly on many topics often far aflied from one another. Curiously, gas and India both appear right after the foundational ideas, perhaps indicating the importance of these perhaps geographically or by population (second most populous country in the world) and gas which could refer to one fothe four fundamental states of matter or one of the most prized resources in the world.

((maybe add graphic))

Although identifying basins of attraction identifies a group of fundamental ideas, using the accumulated traversals, we can't identify whether one or more of the pages is a funnel. Among Philosophy, Existence, Knowledge, one page may be a funnel while the remaining simply happened to be connected to an article with many traversals. By measuring, where the traversal feeds, that is the number of traversals an article feeds into a loop, we can identify special funnel pages, which are part of a larger basin of attraction. Remarkable, within the Philosophy basin, Philosophy seems to be the singular funnel, accumulating an overwhelming majority of the traversals in the philosophy basin, with reality much farther behind (contributing a mere .2% of the traversals in the loop):

Figure 5: Top Funnels

	page	feeds
7948850	Philosophy	7374892
224026	Presentation	30799
9030902	Tree of life (biology)	29274
1344349	Southeast Europe	25745
11029885	Feudalism	19276
632584	Census-Designated Place	17483
7652704	United States Constitution	13952
7974918	Reality	13416
8629119	Health care	10762
7739754	BBC	8945
7580925	Hip Hop Music	7166
4495967	Consciousness	6587
5516532	Balkans	6547
3381363	Quality (philosophy)	5712
5866358	Biological system	5568

By measuring the number of traversals an article feeds into a loop, rather than measuring the accumulation, we measure how remarkable the traversals are around Philosophy. The second ranking page on teh list, Presentation, is dwarfed at a mere .4% of the traversal feeds Philosophy holds. Nonetheless, the list reveals a curious list of top pages including biology as the foundation of life, presentation a fundamental human act, as well as ideas of geographical and political import. Naturally, the page for Census-Designated Place makes an appearance on the top 10, as an accumulation place for the geographical regions in the US. Other topics of cultural and political importance today are Health Care (around which much political debate, science, and economics has been discussed), engaging a large number of our population ((cite something)). As well as Hip Hop or BBC, which are of cultural import. Other high ranking pages include historical events such as the Cold War, Fossil Fuel.

I.4   Network Cycles

By tracking traversals we can also measure true network cycles (loops) rather than a simple chain of connected articles with a dead end. While many of the lengthy paths we discussed in an earlier section terminate at an invalid link, the Eastern Orthodox Liturgics do form a true cycle in the network with a lenght of 365, one article per calendar day looping back to the January first. Other, less lengthy cycles also exist with lengths between 60-75 traversals. These cycles include Historical pages, Judicial Pages such as the Legislative Assembly of Ontario and various Japanese eras dating back to the 8th or 9th centuries.

Since articles iwth a path-length of two yielded curious results, we also measured and ranked the top 2-cycles in the network by accumulated traversals:

Figure 6: highest ranking 2-Cycles

	page	path length	visits
5918	Lecture	2	31000
58203	Presentation	2	31000
44072	Report	2	30771
23737	Examples of feudalism	2	19280
52724	Feudalism	2	19280
69026	Dynasty	2	15833
67007	Health care	2	10777
56597	Medical case management	2	10777
47675	Broadcasting House	2	8952
19184	BBC	2	8952
43228	Secondary education	2	7563
36797	Secondary school	2	7563
13436	Dam	2	7043
28743	Reservoir	2	7043
53676	Biological network	2	5574
79245	Biological system	2	5574
24215	Circulatory system	2	5266
80703	Photography	2	4594
11189	Photograph	2	4594

Of the 11 million artilces, only 83.6k are true 2-cycles. The average cycle traversals is 8, fewer than the typical article. However, Presentation does hold 31k traversals.

The top articles tended to be mere synoms rather than meaningful connections: Lec-

ture and Presentation; Secondary eduaction adn seconday school; biological network and biological system; photogay and photograph. Outside the higest ranking articles, we does see a more meaningful connection among articles in 2-cycles: inventor to product: Voere -> VEC-91; book to author: Anatomy of Britain -> Anthony Sampson; event to organizer: Poetry Bus Tour -> Wave Books.

We also measured 3-cycles, cycles with length fewer than 100 etc. 3-Cycles were still expressing similar ideas or capturing synomous articles for similar ideas: Tree of life (biology) -> Tree of life (disambiguation) and Tree of life. Similarly with Cinema of India, Indian Cinemay and Telugu cinema There were 74K 3-cycles. Once we looked beyond cycles of lenght 10 however, philosophy seemed to dominate along with the remaining list of higest ranking articles–this confirms the typical path length for high ranking pages is typically less than 10.

Figure 7: highest ranking 3-Cycles

	page	path length	visits
2558	Balkans	3	32333
71197	Balkan peninsula	3	32333
41373	Southeast Europe	3	32333
13023	Narrative	3	30716
33929	Tree of life (biology)	3	29302
20874	Tree of life (disambiguation)	3	29302
29882	Tree of life	3	29302
3043	Tree of life (science)	3	29266
11563	United States Constitution	3	13968
10199	Supremacy Clause	3	13968
69587	Article Six of the United States Constitution	3	13968
41849	Bulgaria	3	9719
65328	Order (honour)	3	8663
66125	Article Three of the United States Constitution	3	7591
8642	Southeastern Europe	3	6107
73868	Albania	3	5773
36667	List of Presidents of the United States	3	4485
51710	Cinema of India	3	4228

Style Guide

- Web Terms

    - **article** instead of page
    - **invalid link** (instead of dead or bad link)

- Measurements

    - **traversals** = node weight
    - **path length** = number of articles traversed until an invalid link or a repeated article
    - **traversal feeds** = traversals before a cycle is reached.

- Network Properties

    - **cycle** (permutation from abstract algebra) instead of loop
    - **path-connected articles** for both cycles and articles along the same path leading to an invalid link
    - **basin of attraction** a path-connected group of articles (not necessarily at loop) with a high ranking number of traversals.