JDHCS 2009
Volume 1 Number 1

The WikiPhil Portal: Visualizing Meaningful Philosophical Connections

Sofia J. Athenikos, Drexel University
Xia Lin, Drexel University

1. Introduction

Philosophy (and here we mean Western philosophy), dating back at least to approximately 600 BC, is one of the oldest of all academic disciplines and is, in particular, one of the core disciplines in the humanities. Indeed, it would not be a great overstatement to say that various ideas and ideologies borne out of philosophical reflections and discourses have exerted major formative influences on the Western civilization itself, directly or indirectly shaping its social, cultural, political, and economic underpinnings. Partly due to its long history, and partly due to the nature of the discipline itself, the domain of philosophy presents an extended lineage of philosophers and corresponding philosophical concepts, ideas, and doctrines. The rich fabric of semantic networks in philosophy can in turn be viewed from diverse angles, not only from the chronological perspective but also in terms of its subdisciplinary branches, domains of focus, and schools of thought.

Wikipedia, the free encyclopedia,[1] is an open-access, multilingual, collaborative Web encyclopedia project. Since its inception in 2001, Wikipedia has grown rapidly to become one of the most frequently sought resources on the Web. According to the statistics on the Wikipedia Web site, there are 75,000+ voluntary contributors working on 10 million+ articles in 250+ languages. As of the time of this writing, November 2008, the English version of Wikipedia contains more than 2.5 million articles. Corresponding to the rapid, exponential growth of the size of Wikipedia, recent years have witnessed a continuous growth in the number of computer science research articles concerned with Wikipedia. In particular, researchers working in the fields of natural language processing, text mining, information extraction, question answering, etc. have explored various ways to exploit the vast amount of lexical, semantic, and encyclopedic knowledge contained in Wikipedia. In addition, some Semantic Web researchers have turned to Wikipedia for clues to resolving the main roadblock on the way toward bringing the Semantic Web vision to reality, namely, the scarcity of structured data available on the Web.

How does the oldest discipline meet with the newest phenomenon on the Web? What interesting things can we say about philosophy based on the information contained in Wikipedia? The main objective of the ongoing project, entitled The WikiPhil Portal, is to extract, analyze, and visualize meaningful and interesting connections among philosophers and philosophical concepts via the automatic processing of the hyperlink structure and the textual and semantic content of Wikipedia articles. By doing so, we aim at creating a useful and user-friendly portal for researchers and students of philosophy as well as the general interested public, thereby contributing to the cause of digital humanities. In this paper, we report on the initial results we have obtained by extracting data concerning the hyperlink connections and influence relations involving 330 Western philosophers, with a focus on visualization of the results.

---

[1] http://www.wikipedia.org

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

2. Background and Related Work

The WikiPhil Portal project draws upon various research areas, topics, and trends.

First of all, the project employs Wikipedia as its primary source of data. As mentioned, Wikipedia has recently become a topic of intense interest among researchers who recognize its utility as a source of a vast amount of lexical, semantic, and encyclopedic knowledge that can be effectively exploited for various applications in natural language processing, text mining, information retrieval/ extraction, and knowledge engineering tasks. What makes Wikipedia a particularly valuable knowledge resource is the fact that it can be mined for knowledge based on its structural features, i.e., a dense hyperlink structure, a hierarchical category structure, and a set of structured templates, as well as based on its textual content.[2] Among the many research works on Wikipedia, particularly relevant to our project are those that involve network analysis using the hyperlink structure,[3] semantic information extraction using templates[4] and using the category structure.[5] As will be described later, our project involves extracting, analyzing, and visualizing semantic (conceptual and intellectual) network structures using both the hyperlink data and the infobox template data. In addition, we plan to incorporate the category data in the future.

The idea of extracting and visualizing network structures brings us to the second related area of research, namely, social network analysis. Social network analysis (SNA) has been used for some time not only in sociology but in such diverse disciplines as anthropology, biology, psychology, economics, information science, communication studies, and organizational studies. In particular, with the recent advent of Web 2.0,[6] characterized by the emergence of various collaborative authoring, blogging, bookmarking, tagging, networking, etc. sites that utilize combined social capital, Wikipedia being a prime example, SNA has become a key technique for capturing and exploiting information concerning various social connections and interactions for applications in search (e.g., Amer-Yahia et al. 2008), personalization, and recommender systems, as well as in intelligence and security informatics (e.g., Qin et al. 2005). Even though we have not attempted to compute various (normalized) centrality measures used in SNA,[7] we will show that the rankings of philosophers based on raw link connection and influence relation counts (approximating degree centrality measures) do in general correspond to their relative centrality in the domain of philosophy. We will also show that the network of philosophers emerging from the hyperlink and semantic data extracted from Wikipedia exhibits the characteristic of the *small world or six degrees of separation* phenomenon.[8]

---

[2] Zesch et al. 2007.

[3] Bellomi and Bonato 2005.

[4] Auer and Lehmann 2007.

[5] Ibid.

[6] Chernov et al. 2006.

[7] Wasserman and Faust 1994.

[8] Milgram 1967.

Finally, The WikiPhil Portal project is also, or even prominently, a project about information visualization as an intuitive and effective mode of knowledge representation. While the information processing mechanism operating inside the human brain is considered to be highly capable of pattern recognition, the abstract character and/or sheer scale and dimension of some data sets may pose obstacles to such cognitive processing. Information visualization, via the use of interactive, visual representations of abstract data, serves to amplify human cognition, thereby making it possible or easier to recognize the hidden patterns and structures that might not otherwise be apparent or comprehensible.[9] As we will illustrate with several examples, this is particularly the case with visualizing network structures. In this regard, we will introduce a simple yet effective graph simplification method, called *strongest link paths*, which is particularly helpful for the purpose of identifying the most dominant nodes and connections in network graphs.

3. Research Questions

Some of the initial research questions we have explored in The WikiPhil Portal project are as follows:

1. What kinds of network structures emerge from the information extracted from Wikipedia pages for philosophers (and philosophical concepts)?

2. How well do those structures represent the intellectual (and semantic) relationships among philosophers (and philosophical concepts)?

3. How can the information be effectively presented so as to facilitate discovery and exploration of the connections among philosophers (and philosophical concepts)?

Concerning Question 1, we shall see that the distribution of link connections and influence relations among philosophers exhibits power-law-like patterns, that the rankings of philosophers based on the simple link/relation count statistics generally reflect the relative centrality of the philosophers, and that the emergent network structure reveals small-world-like characteristics, with most of the philosophers being reachable within six degrees. Concerning Question 2, we shall see that the network structures emerging from the Wikipedia link data can give us fairly good approximations of the intellectual connections among philosophers, confirming the idea that the hyperlink structure, on the Web in general and within Wikipedia in particular, embodies a huge amount of latent human annotation.[10] Concerning Question 3, we will present a Web interface that incorporates both tabular and graphical formats of information representation.

4. Data Extraction

We implemented an initial prototype system in Java, using the Java servlet technology, in order to explore the above research questions. We used the English version of Wikipedia as the data source. Data extraction proceeded as follows:

---

9 Card et al. 1997; Tufte 1990.

10 Chakrabarti et al. 1999.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

1. Extract the names and page URLs of major Western philosophers from Wikipedia's 'Timeline of Western Philosophers' and 'Contemporary Philosophy' pages.

2. Extract the hyperlink data and data on the influence relationships among the philosophers from their respective Wikipedia pages, and store the data in a MySQL database.

3. Extract information for the graph and tree format visualizations of the data by querying the database, and store the results as XML files marked up with GraphML and TreeML.

As a result of Step 1, we obtained a chronological list of 330 philosophers. (It must be noted that the list, obtained from the aforementioned Wikipedia pages, includes influential thinkers, writers, theologians, scientists, etc., some of whom are of Arabic/mid-Eastern origin.) In Step 2, we extracted the data on the influence relations by using the 'Influenced' and 'Influenced by' attributes in the infoboxes contained in the Wikipedia pages for some of the philosophers. We augmented the extracted data by adding inferred data, in cases where the influence relationship information involving two philosophers appears in only one philosopher's Wikipedia page. Both the MySQL database tables created in Step 2 and the XML files created in Step 3 are used to present the query results via the Web interface that we describe later.

5. Analysis of Results

Through the procedures described above, we extracted a total of 3706 links (counting only those links that involve the 330 philosophers) and a total of 723 influence relationships among the philosophers from the collection of 330 philosopher pages downloaded from Wikipedia on 29 May 2008. Table 1 summarizes the basic statistics on the extraction results.

Total number of links	Total number of unique link connections	Average number of unique link connections per philosopher	Total number of unique influence relations
3706	2456	7 (7.44)	723

Table 1. Basic statistics on the extracted data

It must be mentioned that Wikipedia article pages for 15 out of 330 philosophers do not have any out-links (i.e., outgoing links directing to the other philosopher pages) and that those for 44 philosophers do not have any in-links (i.e., incoming links from the other philosopher pages). Also, only 229 philosopher pages contain infoboxes from (some of) which we could extract the data on the influence relations.

We computed philosopher rankings based on the statistics. The philosopher who is linked to the largest number of philosophers turned out to be Hegel with 46 unique out-links, whereas the one who is linked from the most philosophers turned out to be Aristotle with 91 unique in-links. Both Kant and Hegel rank as number one in terms of the number of bi-directional links, each having 29 unique bi-link connections. In terms of the direct influences, Kant, Aristotle, and Plato rank as number one to three, having 38, 31, and 28 philosophers influenced by each, respectively. The philosopher who is influenced by the largest number of philosophers turned out to be Heidegger with 18 influenced-by relations.

Table 2 shows top 5 philosophers in terms of the number of unique out-link/in-link/bi-link connections and the number of unique influenced/influenced-by relations. As shown, the high-ranked philosophers in general include major figures in Western philosophy, reflecting their centrality. In particular, Plato, who is arguably the most central figure, appears in the top-5 lists for all three forms of link connections as well as the influenced relation.

Rank	# of out-links	# of in-links	# of bi-links	# of phil influenced	# of phil influenced-by
1	Hegel (46)	Aristotle (91)	Kant, Hegel (29)	Kant (38)	Heidegger (18)
2	Kant (40)	Plato (83)	Plato (25)	Aristotle (31)	Popper (14)
3	Heidegger (35)	Kant (78)	Heidegger (24)	Plato (28)	Descartes, Hegel (13)
4	Plato (31)	Russell (52)	Descartes (21)	Hegel (25)	Nietzsche, Habermas (12)
5	Hume (30)	Descartes, Hegel (51)	Russell (19)	Marx, Nietzsche (18)	Aquinas, Spinoza, Sartre, Wittgenstein (11)

Table 2. Rankings of philosophers based on the statistics

We also examined the distribution of hyperlink connections and influence relations across different ranks of philosophers. Figs. 1 and 2 show the distribution of out-link connections and in-link connections, respectively. The distribution exhibits a power-law-like pattern, showing that a small number of higher-ranked philosophers are linked to/from a large number of philosophers while a large number of lower-ranked philosophers are connected to/from only a small number of other philosophers. Similar patterns were observed with respect to the distribution of bi-links and influenced/influenced-by relations.

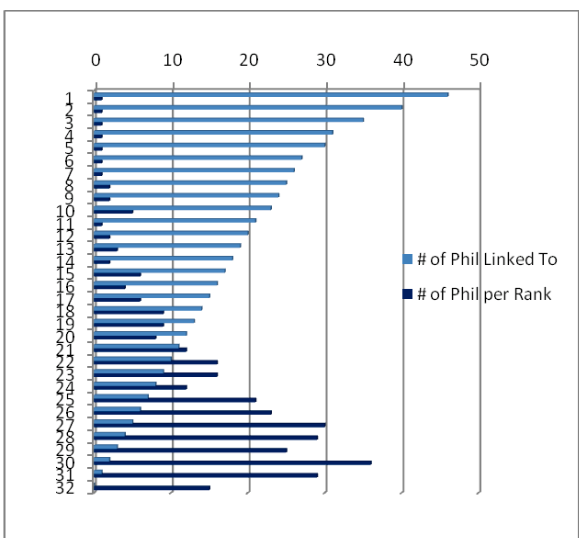

Fig. 1. Distribution of out-link connections

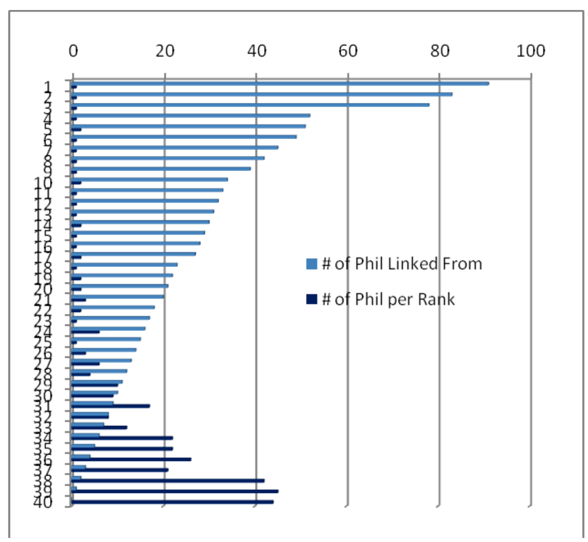

Fig. 2. Distribution of in-link connections

Finally, we checked to see if the network of philosophers derived from the extraction results show smallworld-like characteristics. Specifically, we counted the number of philosophers that are connected from Thales, the first philosopher on our chronological list of 330 philosophers, via direct (1st-degree) out-links (resp. influenced relations), 2nd-degree out-links (resp. influenced relations), and so forth. The results showed that 279 out of 285 philosophers that have in-links are connected from Thales within 5 degrees. Similarly, 181 out of 190 philosophers who have influenced-by relations can be reached from Thales within 6 degrees. (Those isolated philosophers who could not be reached within 5 (resp. 6) degrees via out-links (resp. influenced relations) remained unreachable even if we increased the number of degrees.) Table 3 shows the cumulative counts of philosophers connected from Thales per degree of separation and in total.

	Degree 1	Degree 2	Degree 3	Degree 4	Degree 5	Degree 6	Total
Out-link connections	12	101	233	275	279	279	279
Influenced relations	3	8	49	142	177	181	181

Table 3. Six degrees of separation involving Thales

6. Visualization and Exploration

6.1 Tools and Modalities

We used Prefuse information visualization toolkit[11] to enable interactive visualization of the data concerning the hyperlink connections and influence relationships among 330 philosophers. In particular, we implemented the following four modalities of data presentation using the Prefuse API: graph view, colored graph view, radial graph view, and fisheye tree view.

Fig. 3 shows a colored graph view representing 46 unique out-link connections originating from Hegel. The colored graph view makes it easy to identify stronger/weaker connections by differentiating the node colors according to the edge weights (corresponding to the link counts). The philosopher nodes shown in blue (representing the strongest link connection with a link count ≥ 5) include those of Kant and Marx, correctly reflecting the strength of their intellectual connections with Hegel (shown as the central node colored in red).

[11] http://prefuse.org

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

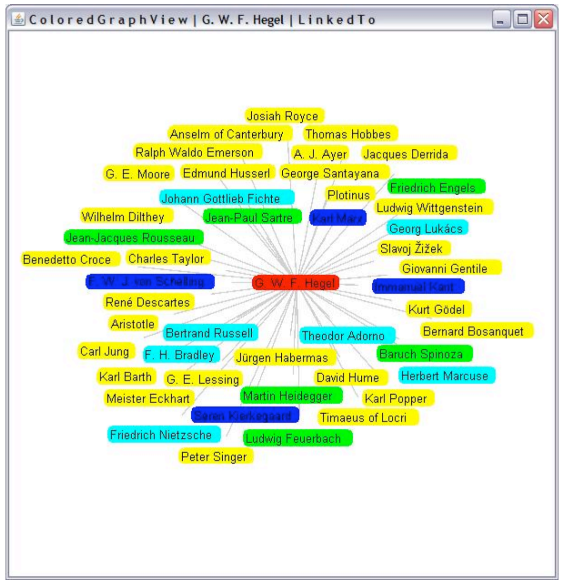

Fig. 3. Out-links from Hegel

Fig. 4 presents a radial graph view showing 91 unique in-link connections directed to Aristotle. As shown, some philosopher names are overlapped with other names, due to the large number of nodes uniformly distributed along the perimeter of the circle. However, when one moves the mouse over a node, the corresponding philosopher name and the link count appear both as tool tips and near the bottom of the window, as shown in Fig. 4 with the example of the node representing Plato. One can also search for a particular philosopher by entering the name in the search box.

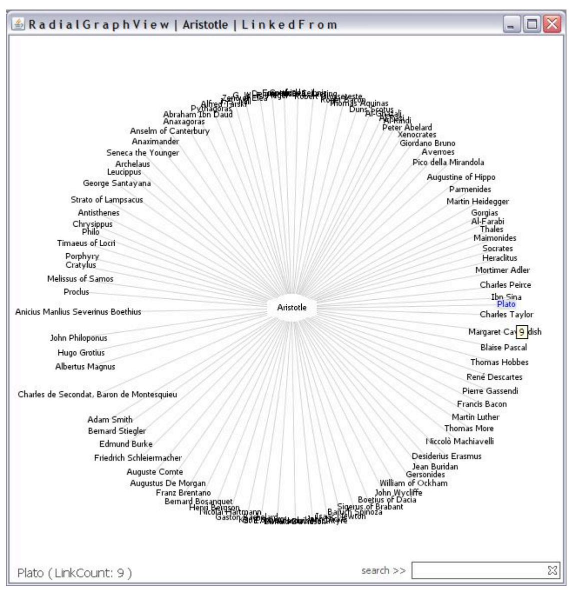

Fig. 4. In-links to Aristotle

Fig. 5 presents a radial graph view showing 29 unique bi-link connections involving Kant and 29 unique bi-link connections involving Hegel. One can easily identify not only the philosophers who

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

are bi-linked with either Kant or Hegel, but also those philosophers who are bi-linked with both Kant and Hegel, one of whom is Heidegger shown at the center in Fig. 5. One can also select a different node as the central node by clicking on it and observe the instant transformation of the graph into a slightly different shape while retaining the same nodes and links.

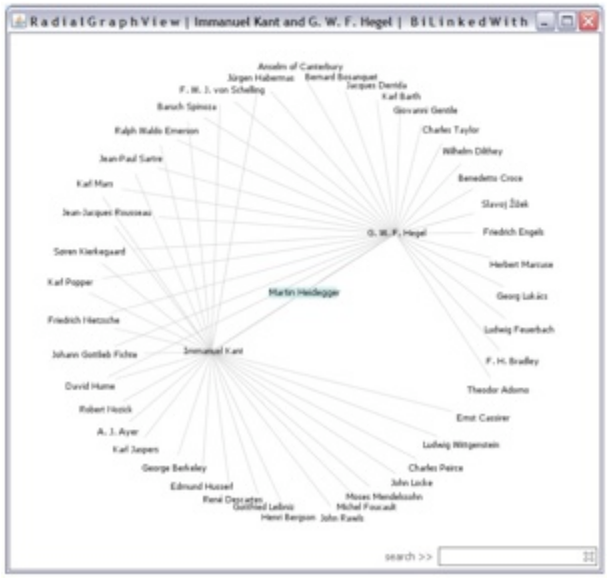

Fig. 5. Bi-links with Kant and Hegel

Fig. 6 shows a (partially expanded) fisheye tree view representing 5 degrees of influences from Plato. The interactive tree view expands/contracts nodes according to user selection. When the window first opens, one initially sees the links to 28 philosophers directly influenced by Plato, as shown in Fig. 7. But when one clicks on one of the 28 philosopher nodes, the selected node expands to reveal links to the philosophers who are in turn influenced by the selected philosopher. And this goes on up to 5 degrees of influence connections. The tree view is thus particularly useful for exploring different connection/relation paths. One can easily explore many different influence paths originating from Plato, other than the one shown in Fig. 6 (i.e. Plato → Aristotle → Descartes → Kant → Hegel → Heidegger), by selecting different nodes in the tree. Fig. 8 shows a partially expanded fisheye tree view representing 5 degrees of influences to Heidegger, tracing the path in Fig. 6 in the reverse direction.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

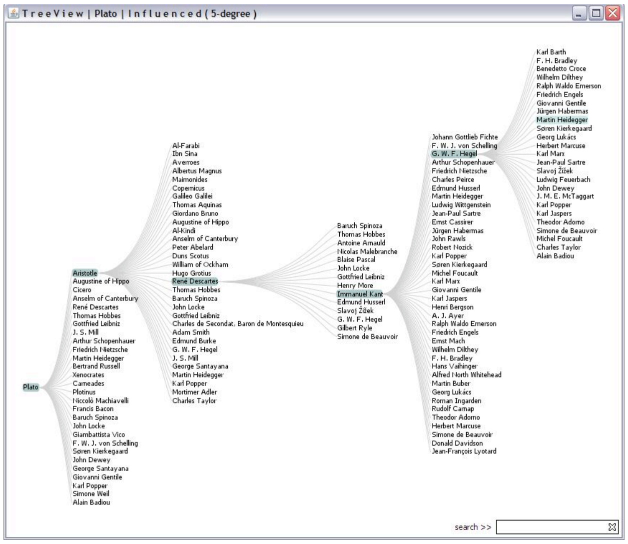

Fig. 6. Five degrees of Influences from Plato

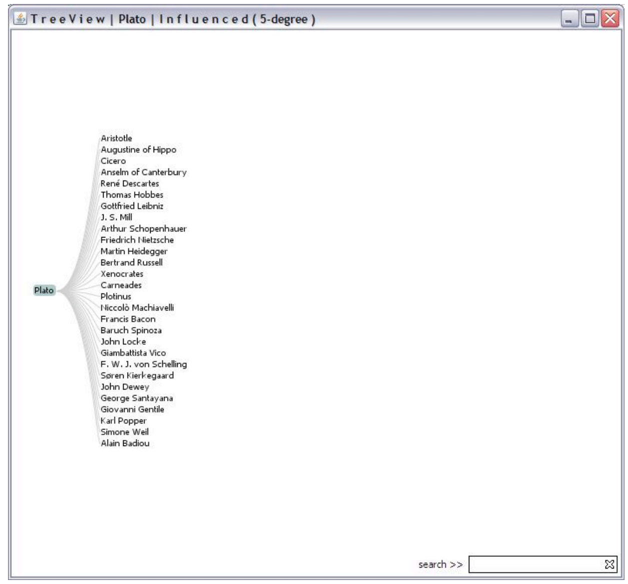

Fig. 7. Direct influences from Plato

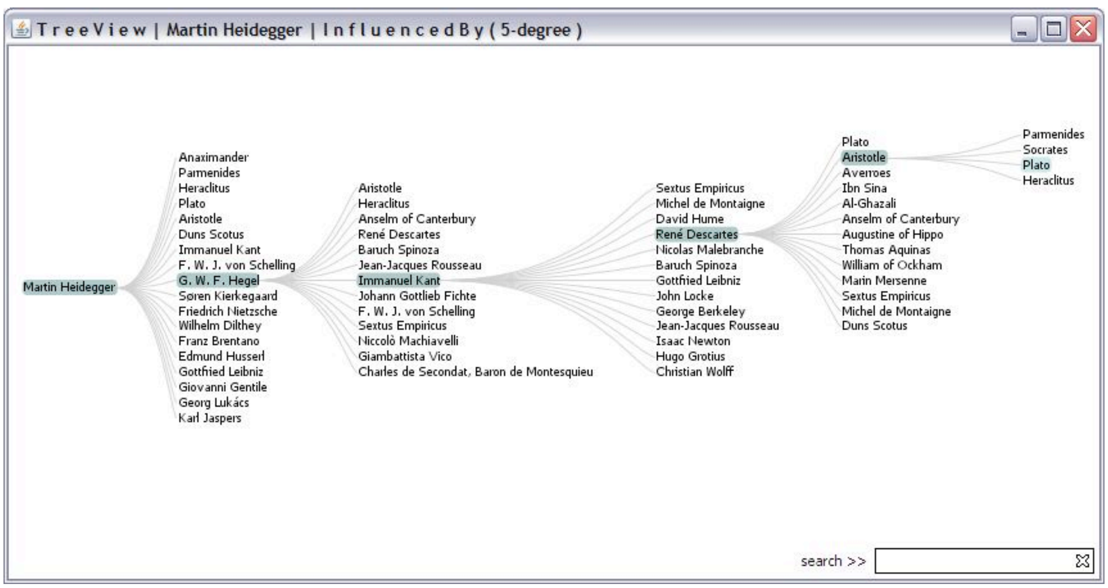

Fig. 8. Five degrees of Influences to Heidegger

Fig. 9 shows a graph view representing all 2456 unique hyperlink connections among 323 philosophers (excluding 7 philosophers who have no out-links/in-links/bi-links). The program that generates the graph view is based on the spring-embedder algorithm[12]. Spring embedder is a force-directed graph layout algorithm, which simulates a mechanical system in which the edges correspond to springs and the nodes correspond to the rings in the springs. The nodes are drawn toward or repelled away from one another depending on the forces acting on them through the edges, until the

[12] Eades 1984.

entire graph settles at some point of minimum force. Fig. 9 represents a relatively settled graph, which shows the majority of the nodes, except several outlier nodes, being densely concentrated toward the center, reflecting the existence of multiple link connections among most of the nodes. Beyond the overall structure of the graph, however, it is hard to recognize any significant substructures.

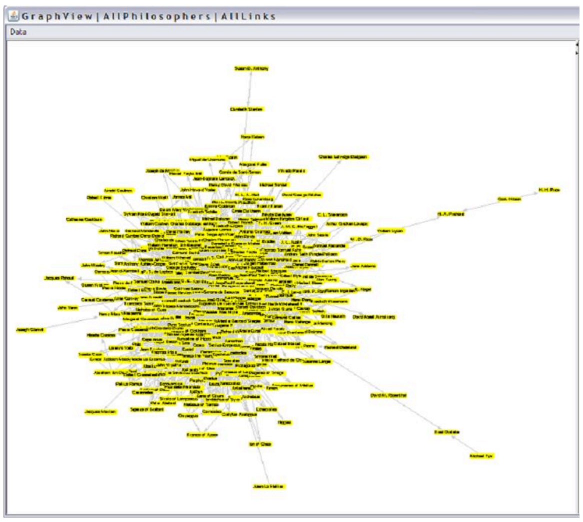

Fig. 9. All links among philosophers

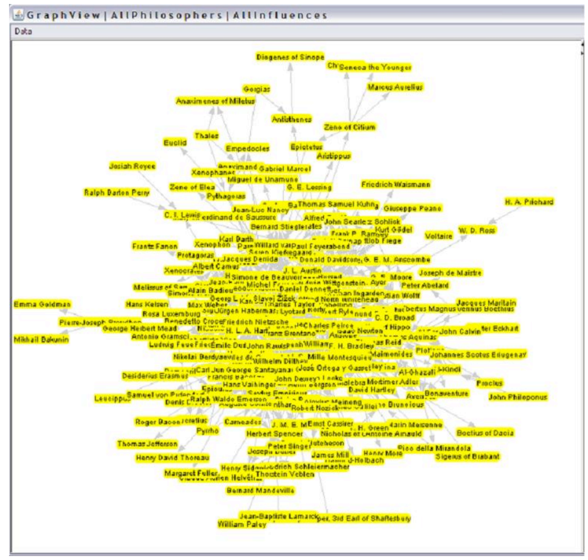

Fig. 10. All influences among philosophers

Fig. 10 shows a graph view representing all 723 unique influence relations among 223 philosophers (excluding 107 philosophers who have no influenced/influenced-by relations). The overall structure of the graph is similar to the one shown in Fig. 6, with most of the nodes densely packed together toward the center. Similarly as in Fig. 6, it is hard to recognize particularly dominant nodes or particularly conspicuous connections.

6.2 Strongest Link Paths (SLP) Method for Graph Simplification

As we have seen in the previous subsection, the graph representing all link connections or all influence relations among philosophers turned out to be quite complex, with no easily recognizable substructures. In order to identify significant and interesting relations and connections from the dense web of links, therefore, we devised a simple yet effective graph simplification method, called the *strongest link paths*, which reduces the complexity of the graph topology by selecting only the strongest link connections.

The basic idea to select, for each node in a given network graph, only the strongest out-link/in-link/ bilink (in terms of the edge weight corresponding to the link count or other connection strength measure), taking a greedy algorithmic approach. The resultant graph contains only a single link per source node, thereby substantially simplifying the graph topology. (By 'source node' here we mean a given node for which we identify the strongest out-link originating from it, the strongest in-link pointing to it, or the strongest bi-link connecting it with another node, depending on which link connection is at issue.) Even in the worst case (i.e., where each node in a network has bi-directional

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

link connections with all the other nodes), the upper bound of the number of residual links is at most equal to the total number of nodes in a given network.

How much data simplification do we achieve by using the strongest paths method on our data involving 2456 hyperlink connections and 723 influence relations among 330 philosophers? Table 4 shows the ratio of the number of residual links to the total number of links and the percentage of link reduction.

	Out-links	In-links	Bi-links	Influenced	Influenced-by
Ratio of residual links	315/2456	286/2456	243/2456	160/723	190/723
Percentage of link reduction	87.17%	88.36%	90.11%	77.87%	73.72%

Table 4. Statistics on link reduction via the strongest link paths method

While the strongest link paths method can effect substantial data simplification, as shown in Table 4, such simplification no doubt can incur some information loss in certain cases. For example, suppose that philosopher A has 10 (repeated) links to philosopher C and 9 links to philosopher D and that philosopher B has 1 link to philosopher C and 5 links to philosopher D. In such a case, the method will pick only links A→C and B→D, for philosopher A and philosopher B, respectively, supposing that link counts are used as edge weights. This is so, even though the strength of link A→D is only slightly less than that of A→C, and it is actually greater than that of B→D. In that sense, we lose the higher-strength link connection A→D in favor of the relatively lower-strength connection B→D.

Nevertheless, by isolating the strongest link *per each source node* and by observing the patterns of convergence and divergence of those links, we can easily detect the most dominant nodes that are strongly connected to/from/with large numbers of other nodes, the groups of nodes that are connected to/from/with a common dominant node, and the connections among such groups/ dominant nodes. As such, the method is quite useful for identifying dominant figures and intra-and inter-group connections within a social network, which may not otherwise be apparent in the original, unsimplified graph.

Note that, the fact that a given philosopher is connected with a large number of other philosophers via out-links/in-links/bi-links, by itself, does not guarantee that the node corresponding to the philosopher will show up as a dominant node when the strongest link paths method is applied. For, we can imagine a case where each hyperlink connection with a given target (philosopher) node is not the strongest one *with respect to the source node*. The fact that the nodes representing major philosophers turn out to be dominant nodes to which large numbers of strongest link connections converge, as will be shown, thus does not simply (re-)confirm the fact that those nodes are central nodes in terms of degree centrality.

By applying the strongest link paths method, which is simpler than other graph scaling methods such as pathfinder network[13] or main path analysis,[14] therefore, we can not only achieve substantial

[13] Schvaneveldt, Durso, and Dearholt 1989.

[14] Hummon and Doreian 1989.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

data reduction but also obtain a fairly meaningful representation of the dominant figures and intellectual connections within the network of philosophers even from the simple hyperlink data.

We use three variations of the strongest link paths method to the network of philosophers as follows:

A. For out-links/in-links/bi-links among all philosophers:

1. Retrieve a set of philosophers (P) who have out-links (resp. in-links, bi-links).
2. For each philosopher p in set P, retrieve a set of philosophers (Q) with whom p has out-link (resp. in-link, bi-link) connections.
3. Select philosopher q from set Q such that the count of out-link from p to q (resp. in-link from q to p, bi-links between p and q) is the highest.
4. Add edge p→q (resp. p←q, p→q) for node p.

B. For influenced/influenced-by relations among all philosophers:

1. Retrieve a set of philosophers (P) who have influenced (resp. influenced-by) relations.
2. For each philosopher p in set P, retrieve a set of philosophers (Q) with whom p has influenced (resp. influenced-by) relations.
3. Select philosopher q from set Q such that q influenced the greatest number of philosophers.
4. Add edge p→q (resp. p←q) for node p.

C. For extended out-links/influences from Thales:

1. Retrieve a set of philosophers (Q) to whom Thales has direct out-links (resp. influenced relations).
2. Add edges Thales→q for each philosopher q in set Q.
3. Select philosopher q from set Q such that q has the highest out-link count (resp. q is the chronologically earliest philosopher).
4. Retrieve a set of philosophers (R) with whom q has out-link connections (resp. influenced relations) (i.e., with whom Thales has 2nd-degree out-link connections (resp. influenced relations)).
5. Add edges q→r for each philosopher r in set R.
6. Repeat steps 4-5 for the remaining philosophers in set Q, in the decreasing order of the out-link count (resp. in the chronological order), but without adding edges to the nodes already connected previously.
7. Repeat steps 3-6 for from 3rd-degree up to 5th-degree out-links (resp. 6th-degree influenced relations).

Below we illustrate the application of the method using visualization.

6.3 Application of SLP

6.3.1 Strongest out-link/in-link/bi-link connections among philosophers

We first apply the strongest link paths method to out-links, that is, we select, for each philosopher node (for 315 philosophers who have out-links), the link that goes out to a philosopher with whom the given philosopher is connected via the greatest number of outgoing links. Fig. 11 shows the resultant graph, which consists of distinct clusters clearly separated from one another, despite the fact that no explicit clustering method was used in the process.

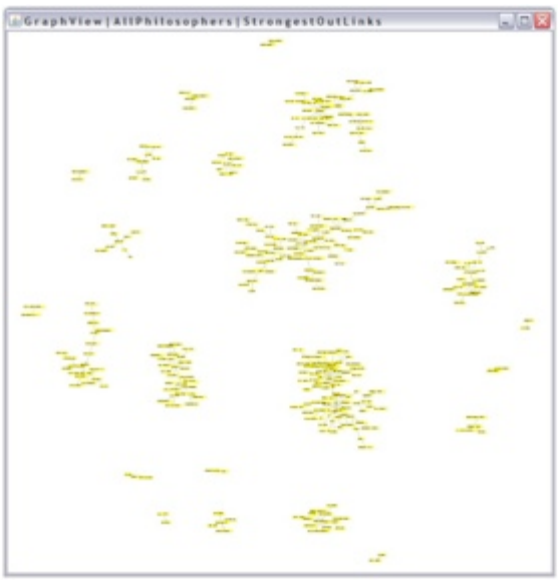

Fig. 11. Strongest out-link connections among philosophers

Fig. 12. Strongest out-link cluster centering on Plato and Aristotle

Fig. 12 shows the close-up (via the zoom-in function) of the largest, most concentrated cluster in Fig. 11. The cluster centers on Plato and Aristotle, reflecting the fact that a large number of philosophers have the strongest out-links pointing to these two central figures in Western philosophy. (In the graph view, the red-colored node represents the node that is currently focused on by the user, and the blue-colored nodes represent the nodes that are directly connected with the red-colored focus node.)

Figs. 13 and 14 show the next two dominant clusters in Fig. 11, centering on Descartes and Leibniz, and on Kant, respectively.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

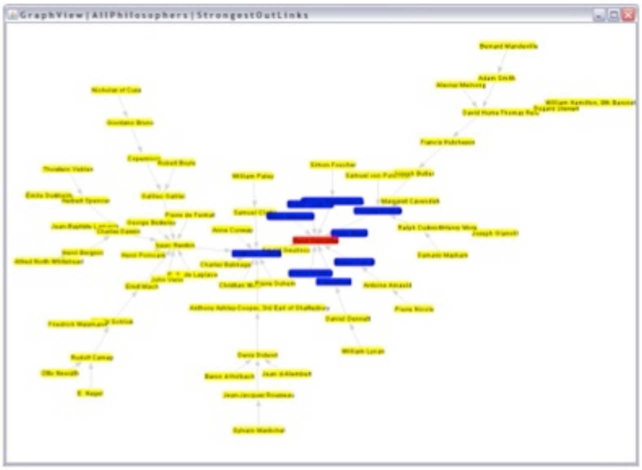

Fig. 13. Strongest out-link cluster centering on Descartes and Leibniz

Fig. 14. Strongest out-link cluster centering on Kant

Next, we apply the strongest link paths method to in-links, that is, we select, for each philosopher node (for 286 philosophers who have in-links), the link that comes from a philosopher with whom the given philosopher is connected via the greatest number of incoming links. The resultant graph consists of distinct clusters, as in the case of the strongest out-links. Figs. 15-17 show the three most dominant clusters in the graph, centering on Hegel, Kant, and Plato, respectively.

Fig. 15. Strongest in-link cluster centering on Hegel

Fig. 16. Strongest in-link cluster centering on Kant

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

Fig. 17. Strongest in-link cluster centering on Plato

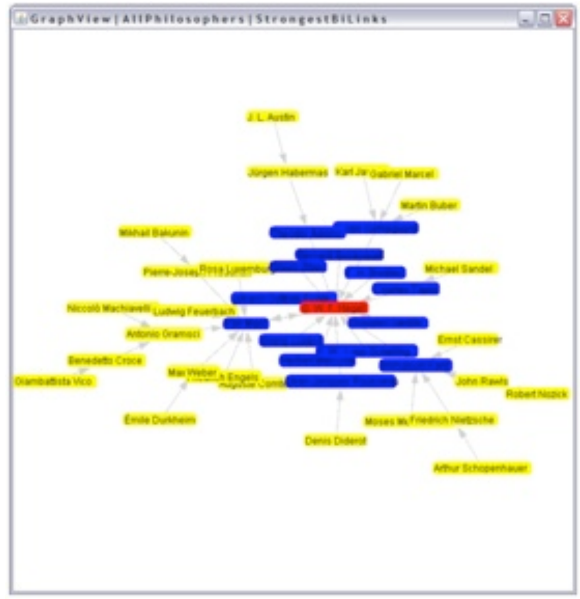

Fig. 18. Strongest bi-link cluster centering on Hegel (with Kant and Marx)

Finally, we apply the strongest link paths method to bi-links, that is, we select, for each philosopher node (for 243 philosophers who have bi-link connections), the link that connects to a philosopher with whom the given philosopher is connected via the greatest number of bi-directional links. The shape of the resultant graph is again similar to that of the graph containing the strongest out-links. Figs. 18-20 show the three most dominant clusters in the resulting graph, centering on Hegel (with Kant and Marx), Plato and Aristotle, and Heidegger (with Husserl and Sartre), respectively.

Fig. 19. Strongest bi-link cluster centering on Plato and Aristotle

Fig. 20. Strongest bi-link cluster centering on Heidegger (with Husserl and Sartre)

6.3.2 Strongest influenced/influenced-by relations

Now we apply the strongest link paths method to influenced relations, by selecting, for each philosopher node (for 160 philosophers who have influenced some other philosophers), the link that connects to a philosopher who, among all philosophers influenced by the given philosopher, (in turn) influenced the greatest number of philosophers. The resultant graph consists of distinct clusters, as in the case of the strongest out-link/in-link/bi-link connections. The two most dominant clusters that appear in the graph, however, are relatively more concentrated, which reflects the variation in the link selection criterion used for influenced/influenced-by relations as versus the one used for ordinary hyperlink connections. Fig. 21 shows the most dominant cluster centering on Kant and Hegel.

Published by: The Division of the Humanities at the University of Chicago

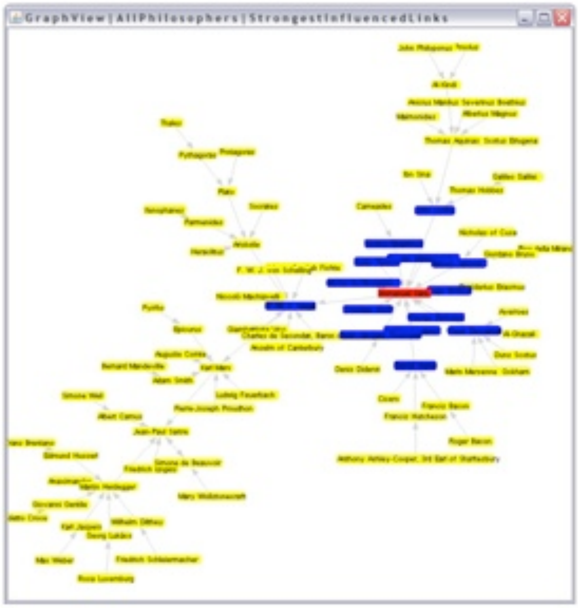

Fig. 21. Strongest influenced relation cluster centering on Kant and Hegel

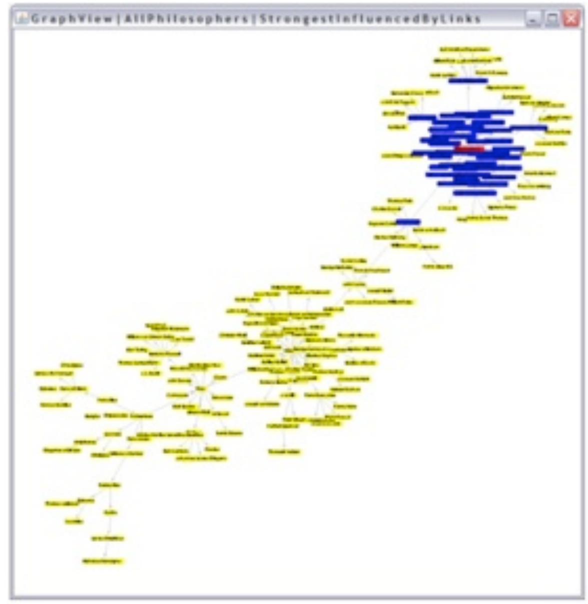

Fig. 22. Strongest influenced-by relation cluster centering on Kant et al.

Next, we apply the strongest link paths method to influenced-by relations, by selecting, for each philosopher node (for 190 philosophers who are influenced by some other philosophers), the link that originates from a philosopher who, among all the philosophers that influenced the given philosopher, influenced the greatest number of philosophers (including the given philosopher). The resultant graph consists mainly of one giant cluster at the center, with a few, very small clusters distributed near the periphery. Fig. 22 shows the dominant cluster, which contains subclusters centering on Kant (with Hegel, Nietzsche, and Wittgenstein), Hume, Locke, Aristotle, Plato, and Parmenides (from right to left).

6.3.3 Extended out-links/influences from Thales

Now we apply the strongest link paths method to extended out-links originating from Thales. That is, we first add 12 direct out-links from Thales and then, for the 12 target philosopher nodes found in the previous step, we in turn add out-links originating from them, starting from the node representing the philosopher who has the greatest number of out-links and continuing on to the other nodes in the decreasing order of out-link count, but without adding links to nodes that have already been connected. The process continues until we have added all non-overlapping out-links to 279 philosophers who can be reached within 5 degrees of separation from Thales. Fig. 23 shows the resultant graph, in which the node representing Thales is shown in red and the 12 philosophers directly out-linked from Thales are shown in blue. The large cluster on the right centers on Aristotle, with smaller clusters centering on Kant (farthest right) and Heidegger (bottom right). The three clusters below Thales center on Plato, Anaximander, and Nietzsche (from right to left). Finally, the large cluster on the left centers on Russell, with small subclusters near the top that center on Hume, Mill, and Leibniz (from right to left).

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

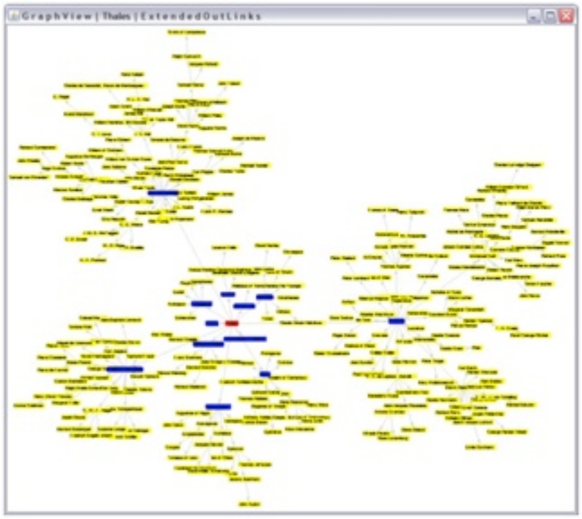

Fig. 23. Graph view of the extended out-links originating from Thales

Fig. 24. Radial graph view of the extended out-links originating from Thales

Fig. 24 shows the identical, extended out-links originating from Thales, rendered in a radial graph view. Even though it is hard to read the labels of the many nodes without zooming in (or without selecting the nodes individually), one can still recognize the dominant philosopher nodes identified in the graph view, near the Thales node at the center.

Fig. 25 on the following page shows a fully expanded tree view of the extended out-links originating from Thales.

Fig. 25. Tree view of the extended out-links originating from Thales

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

Next, we apply the strongest link paths method to extended influences originating from Thales. That is, we first add 3 links representing direct influence connections from Thales and then, for the 3 target nodes corresponding to the 3 philosophers who were directly influenced by Thales, we in turn add links representing influence connections originating from them, starting from the node representing the philosopher who is chronologically earliest and continuing on to the other nodes in the chronological order, but without adding links to nodes that have already been connected. The process continues until we have added all non-overlapping influence links to 181 philosophers reachable within 6 degrees of separation from Thales. Fig. 26 shows the resultant graph. The largest cluster on the right centers on Plato, who is within 2 degrees of influences from Thales (shown in red) via Pythagoras (shown in blue around the center). The large cluster on the left, on the other hand, centers on Aristotle, who is within 3 degrees of influences from Thales via Pythagoras and Parmenides (to the left of Pythagoras). The cluster toward the bottom centers on Heidegger, who is within 2 degrees of influences from Thales via Anaximander (shown in blue). The Plato cluster contains the rightmost subcluster centering on Russell. The Parmenides cluster (between Pythagoras at the center and Aristotle on the left) mainly consists of two small subclusters centering on Nietzsche (top) and Spinoza (bottom).

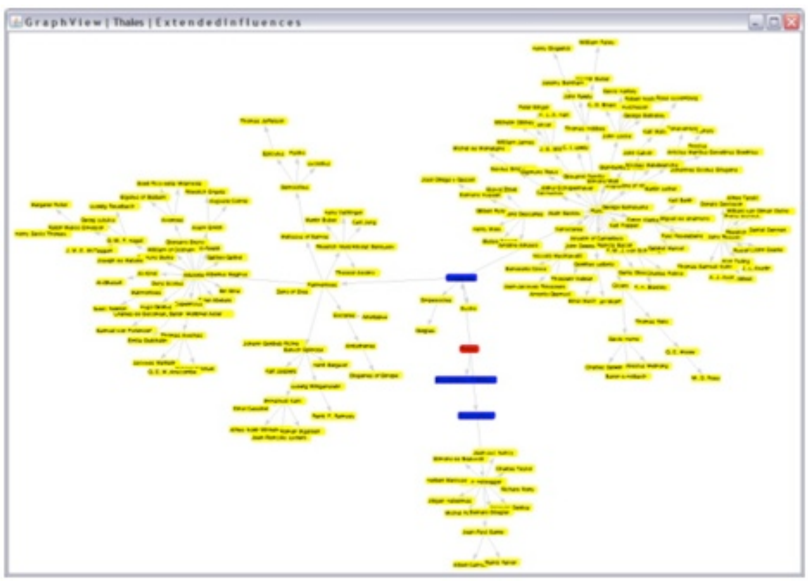

Fig. 26. Graph view of the extended influences originating from Thales

Fig. 27 shows the identical, extended influence connections from Thales, rendered in a radial graph view. Fig. 28 shows a fully expanded tree view of the extended influences originating from Thales.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

Fig. 27. Radial graph view of the extended influences originating from Thales

Fig. 28. Tree view of the extended influences originating from Thales

7. Web Interface

We created a Web interface via which the user can issue queries on the hyperlink connections and influence relationships involving the 330 philosophers and explore the results displayed using diverse modalities of information visualization illustrated in the previous section as well as in the tabular format. Fig. 29 shows the home page of The WikiPhil Portal project. (As of the time of this writing, we have not yet made the project Web site publicly accessible. But we plan to do so in the future.)

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

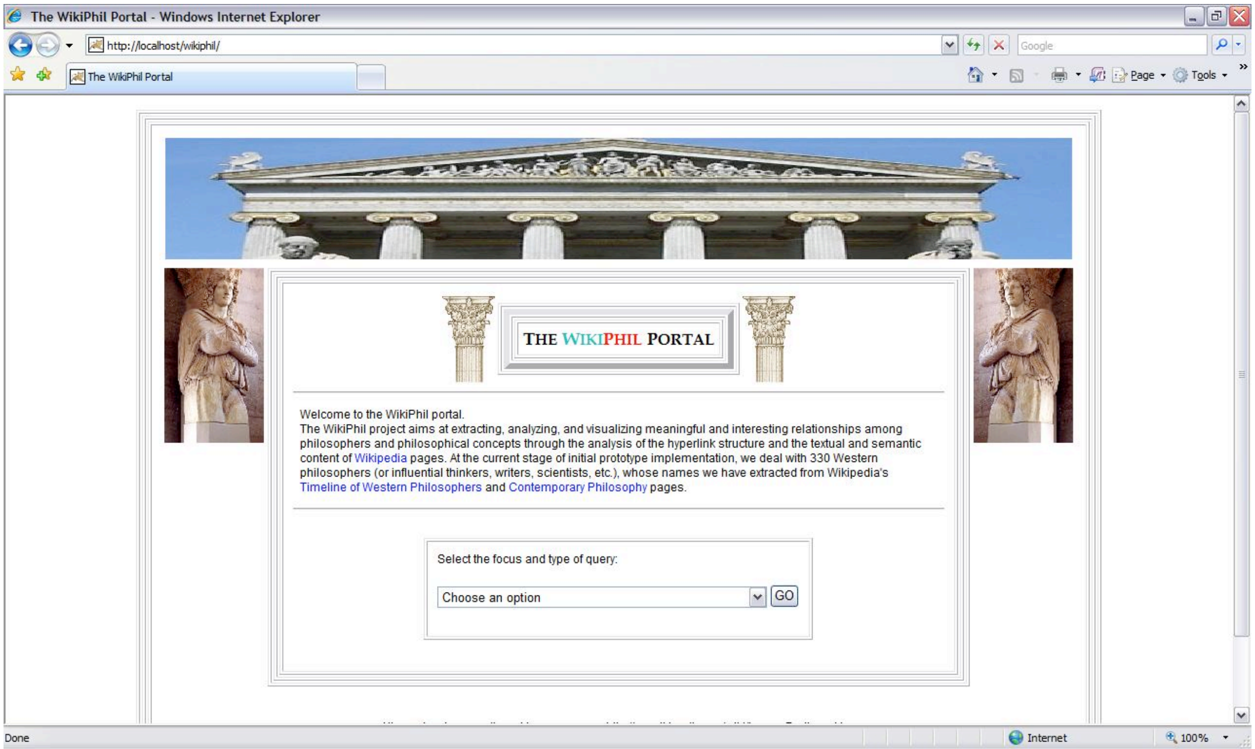

Fig. 29. Home page of The WikiPhil Portal project

Table 5 summarizes the options for the type and focus of query that can be selected by the user. In case a query type involving 1 philosopher or 2 philosophers is selected, the user can then select the particular philosopher(s) to be the focus of the query, from (an) alphabetical list(s) of 330 philosophers.

	1 Philosopher	2 Philosophers	All 330 Philosophers
Link connections	X	X	X
Influence relations	X	X	X
Rankings w.r.t. link connections			X
Rankings w.r.t. influence relations			X

Table 5. Options for the type and focus of query

Fig. 30 presents a sample query result page, showing the result of the query concerning link connections involving Plato. As shown, the user can explore the results using tree views, colored graph views, and radial graph views, by clicking on the appropriate links. The user can also see the results in the table inside a scrollable frame, as shown in Fig. 31. If the user clicks on a philosopher name, the Wikipedia page for the philosopher pops up in a new window.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

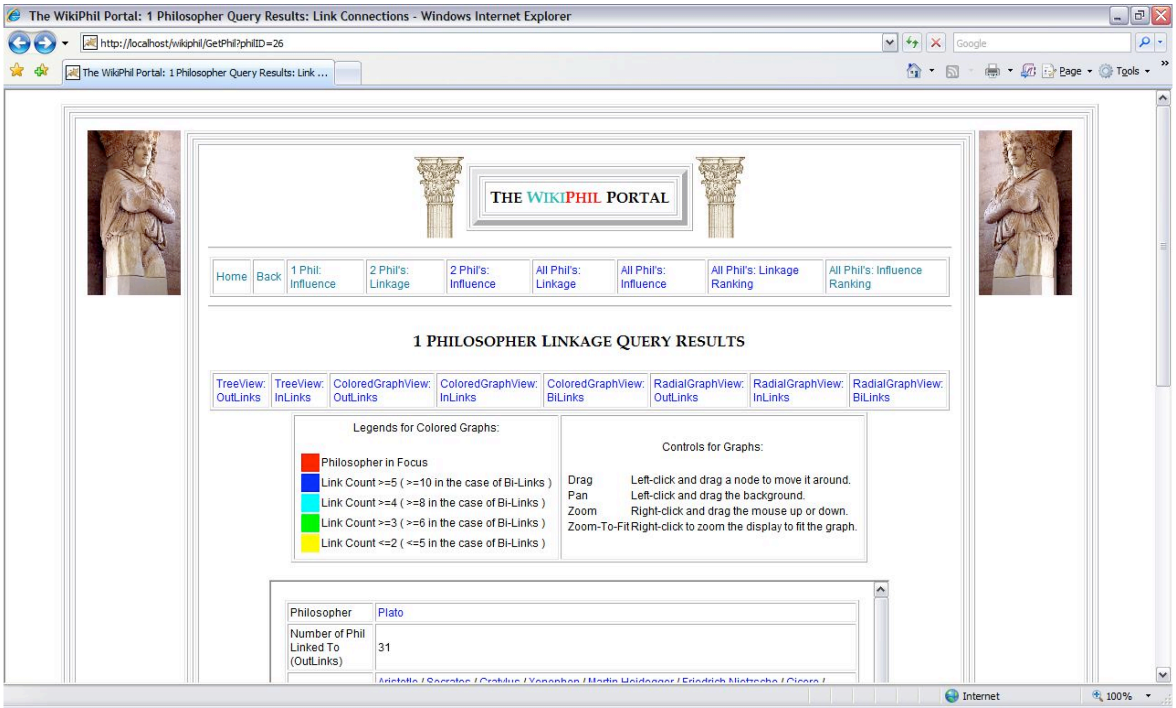

Fig. 30. Query result page for the link connections involving Plato

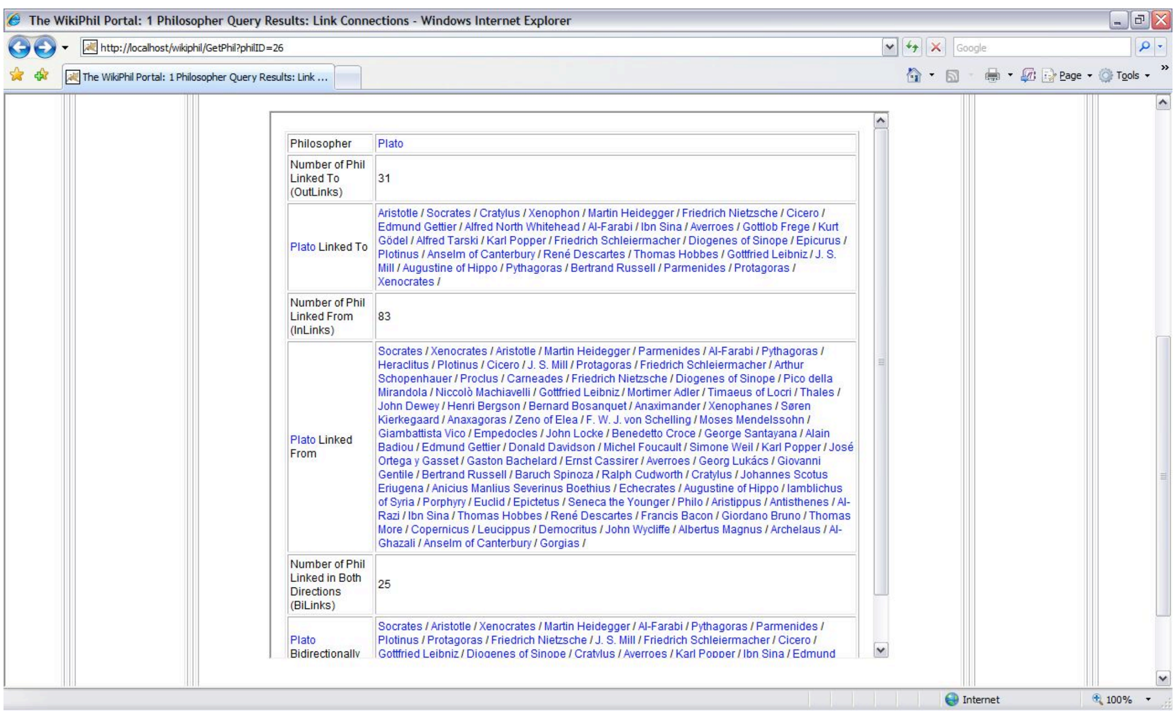

Fig. 31. Query result table summarizing the link connections involving Plato

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago

8. Conclusion and Future Work

The WikiPhil Portal project is conceived as a venue of the convergence of art, (computer) science, and philosophy. It aims to facilitate the discovery and exploration of the hidden and known connections among philosophers and philosophical concepts via an aesthetically appealing interface based on the data extracted from Wikipedia. From our initial results involving 330 major philosophers reported in this paper, we have found that even the simple hyperlink connections can reveal interesting relations among the philosophers, especially when they are presented through various forms of interactive visualization. We are currently working to compare the results obtained from the link data extracted from Wikipedia against the results obtained by using author co-citation patterns. Specifically, we plan to compare the extracted Wikipedia data set against a subset of Thomson Reuters Arts & Humanities Citation Index[15] data containing 1.26 million records covering the period of 1988-1997. Besides the work currently under way, the future work will include extending the approach presented in this paper in order to extract, analyze, and visualize the relations among philosophical concepts and to include more philosopher pages from Wikipedia. We are also considering further research directions including: (1) construction/visualization of semantic networks on philosophers and philosophical concepts, utilizing more semantic content from Wikipedia, (2) semantic search/query incorporating reasoning, and (3) temporal/geospatial mapping of the evolution of concepts. It would be also interesting to compare the results obtained from Wikipedia with those from using prominent philosophical resources on the Web such as the Stanford Encyclopedia of Philosophy or The Internet Encyclopedia of Philosophy.

Note: Please refer to the *Supplementary* files linked from the Abstract page for this article on the JDHCS website for larger and higher quality versions of the tables and figures displayed above. The latest version of the project, renamed *WikiPhiloSofia* is accessible at: http://research.cis.drexel.edu:8080/sofia/WPS/

References

Amer-Yahia, S., M. Benedikt, L.V.S. Lakshmanan, and J. Stoyanovichy. 2008. Efficient network-aware search in collaborative tagging sites. Paper presented in *Proceedings of the 34th international conference on very large data bases (VLDB 2008).*

Auer, S., and J. Lehmann. 2007. What have Innsbruck and Leipzig in common?: Extracting semantics from wiki content. Paper presented in Proceedings of 4th *European semantic Web conference (ESWC 2007).*

Bellomi, F., and R. Bonato. 2005. Network analysis for Wikipedia. Paper presented in P*roceedings of the First international Wikimedia Conference* (Wikimania 2005).

Card, S.K., J.D. Mackinlay, and B. Shneiderman, ed. 1997. *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufman Publishers.

Chakrabarti, S., et al. 1999. Mining the Web's link structure. *Computer* 32 (8): 60-67.

[15] http://www.thomsonreuters.com/products_services/scientific/Arts_Humanities_Citation_Index

Chernov, S., T. Iofciu, W. Nejdl, and X. Zhou. 2006. Extracting semantic relationships between Wikipedia categories. Paper presented in *Proceedings of the first workshop on semantics wikis (SemWiki 2006)* at the third European semantic Web conference.

Eades, P. 1984. A heuristic for graph drawing. *Congressus Numerantium* 42: 149-160.

Hummon, N.P., and P. Doreian. 1989. Connectivity in a citation network: The development of DNA theory. *Social Networks* 11:39-63.

Milgram, S. 1967. The small world problem. *Psychology Today* 1(1): 60–67.

O'Reilly, T. 2005. "What is Web 2.0": Design patterns and business models for the next generation of software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

Qin, J., J.J. Xu, D. Hu, M. Sageman, and H. Chen. 2005. Analyzing terrorist networks: A case study of the global Salafi Jihad network. Paper presented in *Proceedings of the third IEEE international conference on intelligence and security informatics (IEEE ISI 2005).*

Schvaneveldt, R.W., F.T. Durso, and D.W. Dearholt. 1989. Network structures in proximity data. *The psychology of learning and motivation: Advances in research and theory*, vol. 24, ed. G. Bower, 249-284. New York: Academic Press.

Tufte, E.R. 1990. *Envisioning information*. Cheshire: Graphics Press.

Wasserman, S., and K. Faust. 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Zesch, T., I. Gurevych, and M. Mühlhäuser. 2007. Analyzing and accessing Wikipedia as a lexical semantic resource. Paper presented in *Proceedings of the biannual conference of the society for computational linguistics and language technology*.

URL: http://jdhcs.uchicago.edu/
Published by: The Division of the Humanities at the University of Chicago