



ORIGINAL ARTICLE

## Disciplines and the Categorization of Scientific Truth: The Case of Social Sciences in the Hebrew Wikipedia

Eyal Bar-Haim

University of Luxembourg

*For the general audience, Wikipedia is considered the source of “truth,” especially for scientific knowledge. While studies of Wikipedia usually focus on the accuracy of the knowledge within it, few studies have explored its hierarchy and categorization. This study aims to describe how scientific information is organized into disciplines in Wikipedia. I take as a case study the Hebrew Wikipedia and examine the representation and interrelations of five social sciences: sociology, anthropology, economics, political science, and psychology. I gather data from Wikipedia entries categorized under each of these disciplines and create a network that contains categories and subcategories derived from these entries. Using network analysis techniques, I estimate the strength of the relations between the disciplines. Results indicated that while sociology, anthropology, and political science are strongly linked to each other, psychology and economics are relatively isolated. As there is a hierarchical difference between these disciplines, the result is a hierarchical value of the scientific knowledge presented in these Wikipedia entries. An interesting case is the distance between economics and sociology, since under the subcategory “Inequality,” the entries are uniquely categorized under sociology or economics but rarely under both categories. I claim this is an example of a fractal walk in the distinction between the two disciplines.*

url:

### 1. Introduction.

Wikipedia is one of the most important knowledge preservation and accessibility projects of the information age (Wright 2011, Mesgari et al. 2015). Driven by the ideology of free information (Yam 2012), in less than two decades it has become the most important source of information for general audiences, especially regarding scientific knowledge (Jullien 2014).

To date, studies of Wikipedia have focused mainly on the accuracy and ideology behind its articles (Nov 2007). However, existing research has ignored the overall structure of Wikipedia and especially the way articles in Wikipedia are organized and categorized. As sociologists of knowledge demonstrate (Fabian 1975), the way in which knowledge is classified and categorized is (a) ideological, and (b) affects readers' perception of reality. In Wikipedia, writers and editors categorize articles into categories and sub-categories. This categorization helps a reader find other articles in the same subject or field and writers to focus on writing articles that belong only to a specific category. Nevertheless, the structure and hierarchy derived from such categorizations has rarely been studied.

In the present study, I explore the categorization of social science articles in the Hebrew Wikipedia. The Hebrew Wikipedia is an interesting case study of a very active Wikipedia that serves, almost uniquely, a community with specific geographic boundaries (the Hebrew-speaking community in Israel). Therefore, it is possible to distinguish similarities and differences between the academic community in Israel and the representation of science in the Hebrew Wikipedia.

Using network analysis tools, I was able to document the relations between different categories and sub-categories. Thus, it was possible to observe the connections and distance between different disciplines in social sciences as they are represented in Wikipedia.

Results indicated that as a category, sociology is strongly connected to other disciplines. It serves as a hub to various terms related to other disciplines in social science. In that sense, the depiction of sociology in the Hebrew Wikipedia is similar to an unspecified broad discipline, such as social science, than to a distinct discipline. The disciplines of economics and psychology are rather isolated, with only weak connections to other disciplines. Political science and anthropology were strongly connected to sociology, but with a very weak connection to other disciplines.

To interpret these results, I turn to the notion of scientific boundary-work (Gieryn 1983). I argue the distance between disciplines is a form of distinction that is practiced in order to acquire epistemic authority. The way in which social science knowledge in the Hebrew Wikipedia is organized reflects, and maybe even constructs, such boundary-work. This helps to establish hierarchy between the disciplines. I claim that by using measures of network centrality it is possible to observe such hierarchy.

### **Science, disciplines and hierarchy**

Although the sociological discussion regarding the classification of scientific disciplines can be traced back to Comte (1830), the study of the relationship between the structure of disciplines within science and the hierarchy of knowledge can be attributed to the seminal work of Ben-David and Collins (1966). They suggested that social factors, rather than purely scientific progress, determine whether a scientific idea will become a distinct discipline. Bourdieu (1975), who claimed that distinction in science enables scientists to gain scientific authority, emphasized the social importance of a distinction between different scientific fields and disciplines. He also noted that there was always a social hierarchy between disciplines. Gieryn (1983, 1999) referred to this process as “boundary-work.” This term is used to describe the way in which actors distinguish themselves (or eliminate distinction) from other actors who practice similar activity (Small 1999). While at first these practices served to distinguish “science” from “non-science” in order to gain epistemic authority, later it also occurred when distinguishing between different scientific disciplines (Lammont & Monlar 2002).

According to this perception of science, the ongoing process of boundary-work enforces disciplines to distinguish themselves in establishing authority and prestige. Moreover, even after gaining recognition as a distinct discipline, the process continues. The discipline increases its presence in more fields to acquire more authority, usually through struggle with other disciplines looking to gain the same authority over the same fields. However, the struggle to access more fields might be a double-edged sword, as too much presence can affect the ability to distinguish between competing disciplines. The cases of Weber and Simmel provide us a good example. Weber, according to Swedberg (1999), expanded the field of economics to create “economic sociology,” but ended with a distinguished field that became part of the sociological discipline (and with no epistemic authority in economics). In contrast, according to Deflem (2003), Simmel’s sociology of money could not be

accepted in its time due to the epistemic authority of the economists. However, as the discipline expanded, the work of Simmel gained a larger audience within sociology.

This process is usually regarded as top to bottom. Scientists and other important stakeholders in the scientific world seek to gain more visibility in broader fields. An important example is the “Gulbenkian commission” (1996), where important scholars, headed by Immanuel Wallerstein, suggested to eliminate boundaries between the social sciences and replace them with thematic interdisciplinary subjects. They made clear that the distinction between social science disciplines (as well as between natural and social science) was at least partly the result of a struggle for prestige and claim that creating a more holistic (“open”) social science would raise its prestige.

### **Categorization and boundary-work**

The practice of boundary-work in the process of categorization dates to the eighteenth century. Darnton (1984) studied the categorization of science in Diderot’s *Encyclopedia*. He found that the process of categorization of subjects and disciplines in the *Encyclopedia* followed the logic of boundary-work in the sense that the aim of the categorization was to establish distinct fields with a clear hierarchy between the fields. The result, which Diderot called “the tree of knowledge,” represented an innovative approach for its time: the hierarchal categorization of scientific fields. The tree of knowledge was a powerful tool in establishing the separation between analytical, empirical, and metaphysical disciplines as presented in the *Encyclopedia* and for establishing hierarchy between different disciplines. Moreover, in some cases where there was not enough separation, the competition between disciplines ended with subordination, whereby several disciplines were perceived to be included in the other, making the latter more prestigious.

Other examples pointed to the effect of boundary-work in establishing disciplines in a specific context. Gaziano (1996) thoroughly described the boundary-work needed to create change in the hierarchy of disciplines. For example, within a given hierarchy present at the University of Chicago in the 1920s, sociology needed to differentiate itself from social work, while also associating itself with biology, but without being subordinate to it. Thus, it could become more prestigious within the hierarchy of disciplines of the interwar period and could gain epistemic authority with specific sociologists (such as Park and the Chicago School).

However, the boundary-work does not end when the discipline is treated as a distinct one. Disciplines continue to compete with other close disciplines over subjects, fields, and authority. Abbot (2001) further describes the notion of “association without subordination.” He suggests that the categorization of disciplines (in a similar manner to the “tree of knowledge”) leads to a fractal scale, in which the disciplines that are close to each other (for example, disciplines within the broad social science discipline) demonstrate similar relations to broader disciplines, such as the relations between social science and natural science.

Abbot and Moody (2004) suggested the relations between close disciplines might create a situation in which ideas that already exist in one discipline would be reinvented in a close discipline. He referred to this situation as a “fractal walk.” This enables close disciplines to maintain their distinction without being subordinate to each other, while also enabling some form of cross-discipline collaboration. However, too much collaboration and association might result in the subordination of one of the disciplines into another (Gal & Irvine 1995), or a complete merging of several disciplines into one (Reigersberg & Elsbeth 2010, Burack, Iaroci, & Motttron 2002).

Another implication of categorization as boundary-work is the possible emergence of a new discipline. Rodgers (2012) showed that the emergence of social psychology could be attributed to the

boundary-work that had been done by sociologists and psychologists to distinguish these disciplines. For instance, social psychology began almost simultaneously as a field of study in both sociology and psychology. This arose as scholars from both disciplines who studied social psychology introduced topics that could not be categorized into sociology or psychology (such as insect behavior). This enabled the emergence of social psychology as a distinct discipline.

To summarize so far, the categorization of scientific discipline is a form of boundary-work (Gieryn 1983) that aims to establish a distinction between disciplines but also to form a hierarchy among them. Disciplines that enable one to distinguish them from others while grasping an abundance of topics and fields of study, might be considered more prestigious than others might, up to a point that some disciplines would become regarded as sub-disciplines within it (Gaziano 1996). However, the distinction between disciplines also creates an opportunity for similar innovations to be formed separately in each discipline (Abbot 2002, Moody 2004). In time, these cross-discipline innovations, which Moody (2004) refers to as “fractal walk,” might be the start of a new discipline or a new interdisciplinary field of study (Small 1999, Rodgers 2012). Next, I describe how scientific knowledge is constructed in Wikipedia.

### **Wikipedia and the hierarchy of knowledge**

The establishment of Wikipedia is regarded as a challenge to traditional hierarchies of knowledge (Yam 2012). Unlike traditional encyclopedias, initially anyone, regardless of credentials, could contribute knowledge to it by writing and editing articles (later, this policy was slightly changed, restricting certain users and applying some sort of hierarchy, see for example Staub & Hodel 2016). This led to the creation of the largest and most popular encyclopedia that ever existed but it also led to concerns in respect to the accuracy and importance of the knowledge available in Wikipedia.

Garfunkel (2008) criticized the lack of hierarchy between truth and fiction in Wikipedia. Articles that describe fictional characters, events, or places appear in Wikipedia in a very similar manner to actual people. While this is also the case with traditional encyclopedias, which contain articles regarding characters and places from various mythologies, in Wikipedia there are much more of these articles and they are often heavily biased toward contemporary cultural products (movies, television shows, etc.). Their abundance and the amount of work that people invest in writing and improving them undermines the ability of Wikipedia to distinguish reality from fiction (Garfunkel 2008).

Others have questioned the scientific accuracy of Wikipedia. In a prominent paper, Giles (2005) claimed the accuracy of Wikipedia is on par with the prestigious Encyclopedia Britannica. However, later studies found mixed results in regards to specific scientific fields. For example, Clauson and colleagues (2008) found that in the case of pharmacology, Wikipedia performed poorer than a traditional pharmaceutical database. Similar results were found by Zhang and colleagues (2014) regarding scientific knowledge related to terrorism. On the other hand, Teplitskiy, Lu, and Duede (2016) found that Wikipedia articles followed the scientific consensus in the field, citing high impact articles and prestigious scholars, with a small bias toward open access papers.

However, the above scholars did not study the scientific boundary-work in Wikipedia. Specifically, they did not question the pattern in which the Wikipedia articles are categorized within scientific fields and ignored the implicit hierarchy based on this categorization. For example, Zhang and colleagues (2014), who studied the scientific representation of terror in Wikipedia, used an external source of taxonomy to identify articles dealing with terror to avoid the potential bias based on the taxonomy provided by Wikipedia itself. Hypothetically, Wikipedia might not have considered

articles examined by Sun, Chung, and Lim as being terror-related and in turn, they might not have examined articles that were categorized by Wikipedia as being terror-related.

The scarcity of studies examining Wikipedia categorization as boundary work is problematic because studies of Wikipedia categories (Lemer & Lomi 2018) strongly relate the association of articles into categories to the attention they get from Wikipedia editors as well as their perceived quality. The categorization of a Wikipedia article into a certain category affects who will edit it, how it would be evaluated and, as a result, how both the editors and the audience (Lemer & Lomi 2018) would accept it. Most studies examining Wikipedia categories have found that the categories contain accurate knowledge (Nastase & Strube 2008, Ponzetto & Navigli 2009). They claim that Wikipedia's categories successfully represent the taxonomy of scientific knowledge. However, these studies were mostly focused on the actual categorization and ignored any alternative categorizations. Vivaldi and colleagues (2012) addressed these issues and compared the categories in several scientific fields (health education, math, and ecology) in the Spanish Wikipedia in terms of their correspondence with text-books in these fields. They found a considerable variation, for instance, where health education in Wikipedia was much more defined in line with the relevant text book in comparison with math and ecology, and especially with ecology. They partly explained these differences as due to the interdisciplinary nature of ecology but they did not suggest concrete conclusions regarding the role of categorization in defining scientific hierarchy.

In this study, I suggest treating Wikipedia as a site for scientific boundary-work. I claim that the act of categorization, regardless of the question of its accuracy, is a part of the boundary-work for categorizing disciplines to gain scientific authority for the practitioners of these disciplines. The perception of the structure of science, as it is molded by consistent boundary-work between and within disciplines, is mirrored and established in the categorization process in Wikipedia. The result is a network of categories connected to disciplines, where disciplines with more authority are connected to more categories, which represent different domains. In the next part, I review the specific site for this analysis, i.e., the Hebrew Wikipedia.

### **The Hebrew Wikipedia**

Founded in 2003, the Hebrew Wikipedia is a medium-sized Wikipedia (according to the number of articles) with 217,623 articles, ranking it 39 out of 298 official Wikipedia sites (Wikipedia 2018). However, regarding the update frequency, which usually serves as measurement of quality, it is ranked 12. This is reflected in the repeated controversies among Wikipedia users and the public regarding the deletion of articles due to their political bias or lack of scientific interest (Leshnick & Livio 2016). The debates regarding edits and deletions in the Hebrew Wikipedia are usually grounded in a modernist conception of knowledge, whereby articles might be deleted if they do not follow traditional hierarchies of encyclopedic information even if they are considered legitimate in Wikipedia in other languages (Shachaf & Hara 2010, Leshnick & Livio 2016).

Another important aspect of the Hebrew Wikipedia is its representativeness in terms of community. The majority of Hebrew speakers using the Hebrew Wikipedia live in Israel (Hoover 2016). Therefore, it would be safe to assume the main source of interrelation with the Hebrew Wikipedia is the Hebrew-speaking Israeli society.

Categorization in the Hebrew Wikipedia follows a different guideline ("Wikipedia: Category [Hebrew]" n.d.) than the guideline provided by the English Wikipedia ("Wikipedia: Category" n.d.) In the Hebrew Wikipedia, it is possible to categorize articles under different sub-categories within the same broad category (parallel categorization). It is also possible to leave an article without a category. Moreover, rules for creating categories and sorting under categories are less strict and

more ambiguous than in the English guidelines. The only requisition for creating a category is that at the time of its creation there must be at least five articles that can be categorized under it. However, according to the discussion page, this rule is rarely enforced and categories can exist with only two articles that relate to them. Reviewing the discussion page of the guidelines suggests that Hebrew Wikipedia editors see much importance in the existence of categories in Wikipedia, both for editors and readers, but feel that there is confusion regarding the categorization process itself (“Wikipedia talk: Category [Hebrew]” n.d.). In that sense, the Hebrew Wikipedia makes an interesting case study as categorization rules are less enforced than other Wikipedias, suggesting that the categorization reflects more of the intuition of the Wikipedia contributors.

Overall, the Hebrew Wikipedia provides an interesting case study to explore Wikipedia categorization as a scientific boundary-work. To focus our study, I decided to explore a case study of social science. I chose this rather broad field due to its relevance to public debates in Israel. Unlike other contexts (for example, climate science or evolutionary biology in the U.S.), most political debates that involve science are related to social science. Therefore, I expect it to be the most relevant field for debate in the Hebrew Wikipedia.

## 2. Data

### Five disciplines in social science

To examine the representation and categorization of social science in the Hebrew Wikipedia, I chose five distinct disciplines within the broad subject of social science: psychology, sociology, political science, anthropology, and economics. The five disciplines represent the core of social sciences in the Hebrew-speaking community. Each of these disciplines is represented by social science faculty in all Hebrew-speaking universities. These are the only disciplines, aside from communication and media studies, that are uniquely represented as social science at all of these universities. I did not include communication and media studies due to the ambiguity of the Hebrew translation of the term communication (תקשורת).<sup>1</sup>

### Gathering data from Wikipedia

Data were gathered during the period September 4–5, 2014. To obtain the data, I used a modified version of the WikipediR2 package (<https://cran.r-project.org/web/packages/WikipediR/index.html>), which enables obtaining category trees from Wiki-type pages. The package is a wrapper for the MediaWiki API and thus it is useful for obtaining data from Wikipedia.

By using this package, it was possible to obtain the pages listed in a Wikipedia category and similarly, to obtain the categories that are listed in each Wikipedia page. I started with the Hebrew names of five social sciences disciplines as categories: psychology (פסיכולוגיה), sociology (סוציולוגיה), political science (מדע המדינה), anthropology (אנתרופולוגיה), and economics (כלכלה). I obtained the titles of each article that was listed under each of these categories. Then, for each article, I listed all the categories under which it was listed.

<sup>1</sup> In an unreported analysis, I created a network based on communication and media studies, and found that categories that were related to it were too wide, spanning from computer network protocols to speech therapy.

<sup>2</sup> I modified it to be able to analyze, among other things, text in Hebrew. The modified version can be obtained from the author of the present paper.

The result was a list of categories connected to each other by the articles. For example, if an article like “deviation,” was listed under two categories, like “Sociology” and “Psychology,” this meant that there was a (first-degree) connection between those two categories (i.e., sociology and psychology). However, the article could also be listed under the category “Crime,” under which is also listed the article “Terror,” which is also listed under the category “Political Science.” In that case, there would be a second-degree connection between “Political Science” and “Sociology” and between “Political Science” and “Psychology.” The result can thus be interpreted as representing a network, in which the categories are the vertices.

To simplify the network, I omitted the articles themselves and left only the categories. I also omitted irrelevant categories like “Articles in need of editing” and “Scrubs” (a term used to identify very short articles). The final list of categories included 635 categories.

### **Translation of Hebrew terms**

The original list obtained was in Hebrew. To translate it, I first automatically translated the list using the Google Translate interface available on the Google Docs spreadsheet. I then produced parallel networks from the original list and the translated list to assess for differences (e.g., due to terms with multiple meanings in each language). Fortunately, the networks were identical. In the final stage, I gave the translations to two independent referees to assess their accuracy. The referees suggested only minor changes.

### **3. Method**

Based on the definitions above, I created an actor-by-actor matrix. I then calculated, using the igraph package for R (Csardi & Nepusz 2015), three different centrality measurements: betweenness, eigenvalue centrality and Google’s PageRank.

#### **Betweenness**

According to Freeman (1978), the betweenness index measures the number of shortest paths that start from a vertex. In the current context, disciplines that serve as a category for many articles would have ranked higher in the measure of betweenness centrality than disciplines with a smaller number of categories, even if those disciplines were more connected to the other disciplines. In that sense, the betweenness centrality measurement serves as an indication of the “strength” of a discipline (or a category within a discipline) in terms of the number of categories that are related to it. According to the idea of boundary-work, disciplines with a high level of betweenness are more distinguished than others, while maintaining a relatively large domain. Therefore, they can be considered more prestigious.

#### **Eigenvalue centrality**

The eigenvalue centrality index (Bonacich 1987) is based on the notion that a vertex is more central if it is connected to other central vertices (Ruhnau 2000). In that sense, it is a measurement of the influence or of the importance of a vertex in a given network. In the context of the present study, I expected the five disciplines to be central (due to the way the network was constructed). Therefore, when a discipline or a category is strongly connected to other disciplines, it is expected to be ranked high in this index. For categories that were not one of the five disciplines, a high level of eigenvalue centrality might imply a broader discipline (such as social science or humanities) or an emerging discipline, as described by Moody (2004).

## Google's PageRank

Google's PageRank is considered to be an improvement of the eigenvalue centrality index (Brin & Page 1998). Like the eigenvalue centrality index, it provides a measurement of the importance of each vertex. However, it can account for the weights in the network and therefore provides a more refined measurement.

## 4. Results

### The complete network

I begin with presenting the complete network in Figure 1. The orange nodes represent Wikipedia categories. The widths of the edges represent the number of repetitions of the connection. Nodes that are connected to more than three edges are labeled and nodes with more than 20 edges are labeled with a bigger font. The latter represents the five disciplines. Sociology serves as a hub and is highly connected to anthropology and political science but has fewer direct connections with psychology and economics. These two disciplines are more isolated than the others, and are only loosely connected to each other, a fact that might be surprising considering the popularity of "behavioral economics."

Social science, to which the five disciplines are usually regarded as sub-disciplines, is indeed quite connected to some disciplines and sub-categories. However, it is not connected to political science and economics. Beside social science, the other well-connected categories are culture, stratification and inequality. Culture appears to be strongly connected to sociology and anthropology. Stratification and inequality are connected to sociology, economics and political science.

Table 1 presents the results of the investigations into the connectivity of the five disciplines. Sociology has more direct connections to other disciplines than any other discipline and is indeed connected to all the other disciplines. Sociology has the highest connectivity to anthropology. This seems to be in line with the common perception in Israel regarding the connection between sociology and anthropology. All universities, except for the University of Haifa, have joint departments for sociology and anthropology. Political science has a direct connection only to sociology, whereby even pages listed under the category "Political Economy" are connected to political science but not to economy. The relatively high connectivity of psychology and anthropology cannot be attributed to common core terms, such as taboo, which are categorized uniquely only under anthropology and psychology. However, the fact that some prominent Israeli anthropologists have a background in psychology (for example, Yehuda Goodman, Yoram Bilu, Zali Gurevitch) might play some part in explaining the connection.

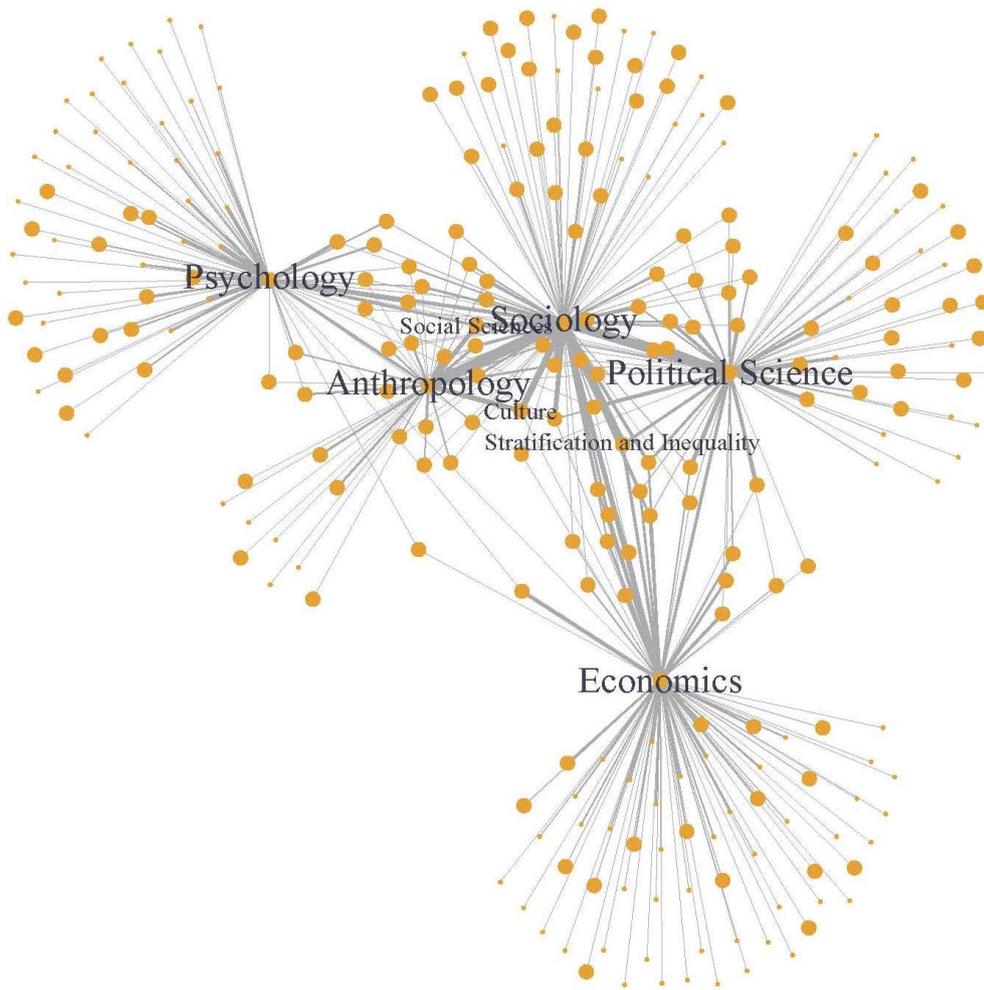


Figure 1: The Complete Network of Social Science Categories in Wikipedia

	Anthropology	Economics	Political Science	Psychology	Sociology
Anthropology	-	0	0	7	33
Economics	-	-	0	2	13
Political Science	-	-	-	0	28
Psychology	-	-	-	-	10
Sociology	-	-	-	-	-

Table 1: Number of Direct Connections between Disciplines

### Centrality

Table 2 presents the ranking of the top 10 categories based on three major centrality measurements: betweenness, eigenvalue, and Google’s PageRank. As noted in the methods section, while these are all centrality measurements, their interpretation is very different in each case and is largely dependent on the context. In the context of this research, the betweenness centrality measurement describes the strength of a discipline as a separate discipline as presented in the Hebrew Wikipedia.

To gain a high level of betweenness, a node must serve as a bottleneck, i.e., separating a part of the network from other parts. A discipline with a high betweenness level has a unique and large domain in the form of sub-categories that are minimally shared with other disciplines or categories. This is in line with the notion of boundary-work. If a discipline has gained control of a wide variety of subjects with no other discipline acquiring control of these same subjects, it would be considered the authority regarding these subjects and would gain more prestige.

According to the betweenness centrality measure, the discipline of economics has the highest level of centrality. Psychology is next, followed by sociology and political science. However, anthropology is ranked only sixth, below organizations, which is not even considered a discipline in this research (nor according to Wikipedia itself). However, it gains its centrality due to its closeness to economics. The relatively low rank of anthropology is due to the connections of this discipline to other disciplines (directly and indirectly) together with the small number of unique sub-categories. I claim that this represents a belittling perception of anthropology among the Hebrew Wikipedia writers and editors.

Among the other categories listed with the highest ranked in terms of the betweenness centrality measure are social psychology and social networks. Social psychology is usually defined as a separate discipline (Rodgers 2012). Social networks is not a distinct discipline but rather an interdisciplinary field strongly connected to disciplines outside social science like computer science and mathematics. Other terms are more connected to social action, such as anarchism and methods of protest, and provide a connection between social science and political activity.

Rank	Betweenness	Google's PageRank
1	Economics	15251.68
2	Psychology	13892.92
3	Sociology	13626.88
4	Political Science	11682.03
5	Organizations	5849.30
6	Anthropology	3340.53
7	Social Networks	2233.85
8	Anarchism	2023.90
9	Methods of protest	1290.23
10	Social Psychology	1196.92

Table 2: Wikipedia Pages by Centrality

Using Google's PageRank index, it is possible to identify hierarchy in a more straightforward manner. According to this measurement, a node is considered central when other central nodes are connected to it. Of course, the five disciplines are ranked the highest, with a considerable gap between the lowest ranked discipline (anthropology) and the highest ranked terms that are not disciplines. However, there is also a gap within the five disciplines. Sociology is ranked the highest due to its many connections to the other disciplines. Economics is ranked second, probably due to its several indirect connections to other disciplines and political science and psychology are ranked of third and fourth, respectively, and anthropology is ranked last.

Terms, (in contrast to disciplines) that are ranked high in this index are the ones that are strongly connected to several disciplines. Social science is ranked seventh, which is surprising, as one would expect it to be highly connected to all the disciplines assumed to be subordinate to it and therefore expected to be ranked much higher. Wikipedia's writers and editors must not consider some of the five disciplines as part of social science. This might reflect an existing tendency in Israeli academia to separate psychology and economics from social science, as can be seen at Tel-Aviv University where these disciplines are separate schools and are not subordinate to the social science department.

The sixth ranked term is "Stratification and Inequality." Figure 2 presents an ego network of this term. It is clear that this term is highly interdisciplinary and is strongly connected to sociology, economics and political science. One should note that while there is no direct connection between economics and political science, the term "Stratification and Inequality" serves as an indirect link between them, together with sociology.

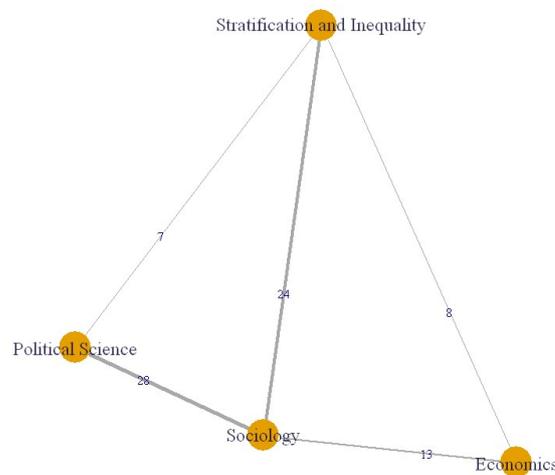


Figure 2: The Ego Network of Stratification and Inequality

The high centrality of stratification and inequality and the specific attribute of its ego network might suggest that it represents a "fractal walk," as suggested by Moody (2004). The articles categorized under this term are associated with separate disciplines with little to no connection between them. In contrast to other terms highly ranked in Google's PageRank index (i.e., political sociology, social psychology, culture), this term is not yet considered a distinct discipline or a distinct interdisciplinary field of study<sup>3</sup>.

<sup>3</sup> For each of the other terms, it is possible to find at least one distinct study program in one of the Hebrew-speaking academic institutes.

## 5. Discussion

### The implication of an implicit hierarchy

The general audience, as well as scholars, are usually aware of the problems and advantages involving the representation of knowledge in Wikipedia. It is therefore generally possible to judge critically the accuracy and quality of a Wikipedia article. However, the actual organization of knowledge in Wikipedia is rarely judged critically. Due to the ease of searching within Wikipedia, people assume that the structure of knowledge within it is much less important.

I argue that this assumption is problematic. Studies of science production and presentation have shown that the organization and categorization of knowledge is neither unimportant nor neutral. Instead, it is used to justify epistemic authority (or the lack of one) by distinguishing between spheres that are within a specific discipline and ones that are not, or are even beyond the scope of science itself. Considering the fact that Wikipedia is one of the most important sources of scientific knowledge for general audiences. The categorization of articles in Wikipedia serves as a way to introduce broader subjects to readers. The way articles are categorized in Wikipedia is important but usually overlooked.

In this study, I examined a specific case of social sciences in the Hebrew Wikipedia. Some findings seem to present a close representation of the general trends in social science in Israel. These trends are also represented in the structure of Israeli academic institutions.

I claim that overall, the depiction of social science through categorization in Wikipedia has an important role in portraying the “truth” about social science in the Hebrew speaking community. Although the association between sociology and anthropology in Israel is a known fact (Ram 2018, Rotem 2014), the depiction of the weak boundaries between these disciplines, as implied by the multiple direct connections between them in Wikipedia, might have results within Wikipedia itself as well as in the real world. As Lerner and Lomi (2018) suggest, the categorization would affect Wikipedia editors resulting with articles that are less oriented toward one of the distinct disciplines and highlighting the overlaps between them.

Other findings might carry even more severe implications. Both economics and psychology are detached from political science. But since political science is perceived as being related to politics, both economics and psychology might be perceived as “a-political.” This might help these disciplines to gain more epistemic authority as a neutral science and to deny their political implications. In that sense, the lack of direct connections between psychology and economics is also interesting as it might imply that according to the Hebrew Wikipedia, psychology has nothing to contribute to understanding economic behavior and vice versa.

One open question that remains is the position of sociology in the hierarchy reflected by the Hebrew Wikipedia. Sociology serves as a hub to all the other disciplines, with direct connections to all of them to various degrees. This make it ranked the highest on Google’s PageRank but only third on the betweenness measurement. From the perspective of the entire network, it seems that sociology has almost replaced social science as the meta-discipline. On the other hand, this means that there are not many distinct terms in sociology. This might imply that a sociologist would be perceived to have less epistemic authority in the public sphere as they would have to compete with scholars from other disciplines (in contrast to economists and psychologists, who clearly have their own niche). A further study on the impact of sociologists in the public sphere may shed light on the prestige of the discipline and on the value of the centrality measurements in instances of disagreement.

The emergence of a new discipline?

The case of the category “Stratification and Inequality” is interesting by itself. In contrast to other categories found to have a similar ranking, it does not represent a distinguished discipline or an established interdisciplinary field of study. In the Hebrew Wikipedia, it is a hub that connects economics, political science, and sociology in what might be considered as a “fractal walk.” For the moment, the implication of it in the sense of a distinct field shared by scholars from these disciplines who communicate with each other is yet to be seen.

However, it might reflect some recent changes in the global scholarly community and predict the emergence of a distinct niche in the Hebrew-speaking scientific community. Among these changes is the recent gain in popularity of inequality research by economists (Piketty 2015), sociologists (Dorling 2014), and political scientists (Bartels 2016), including conferences on inequality, in which scholars from these disciplines serve together on the organizing committee (Jenkins 2015), and interdisciplinary debates on stratification and inequality in scientific journals (Blanden 2013, Goldthorpe 2013). While the noted evidence is far from conclusive, the findings in this study regarding stratification and inequality may call for further investigation into these processes.

### **Wikipedia and the unintentional production of knowledge**

Gelenter (2013) claims that the construction of Wikipedia as we know it today was through an institutionalization process in which the founders of Wikipedia, Wales and Sanger, used their authority to set up several guidelines and standards, and thus enforced specific ways of constructing knowledge using Wikipedia. For example, a decision made by Wikipedia founders to remove any discussion from the articles themselves to a separate discussion page, made the current content of an article to be more important than the discussion and questions regarding the article, which are less intuitively accessible to the reader.

According to Gelenter (Ibid.), later in Wikipedia’s life, the editors themselves created the guidelines and standards. Through discussions and even well documented intense arguments, they set up hierarchies regarding what should and should not be in Wikipedia. These arguments were also present in the Hebrew Wikipedia, especially regarding the importance of articles on female historical figures (Geva 2013) and events from the near past (Yasseri et al. 2014).

However, the use of categories in the Hebrew Wikipedia went almost under the radar. As can be seen from the discussion page regarding categories, Hebrew Wikipedia editors almost never follow the rules and guidelines for creating categories and categorizing articles, even though they are aware of the importance of categories. Moreover, there are almost no discussions about categorization itself, at least for categories sampled in this study. The work of categorization of articles in social sciences in Wikipedia is done by numerous Wikipedia editors, without a discussion, peer review, or effective guidelines. As a result, these Wikipedia editors are mirroring their own image of the structure of social science on Wikipedia and thus projecting it to a much greater audience. This is a striking contrast to the classic Encyclopedia. If in the first Encyclopedia, as Darnton (1984) claims, the categorization was a process of elite scholars deciding on the hierarchy of disciplines, in the contemporary Wikipedia, at least in Hebrew, we have a process with almost no interference from the scientific community, is much less centralized and lacks formal or informal regulation. Nevertheless, the results are similar, with boundaries and hierarchy between disciplines.

In that sense, the boundary-work, reflected in Wikipedia categorization, resembles less the process described by Wallerstien (2000) than the one suggested by the Burawoy’s critique of it (2008). While Wallerstien, like most of scholars of scientific boundary work, perceives it as a result of the struggle between central stakeholders in the academic field, Burawoy suggested that much of the formation of a scientific discipline is due to interaction between the academic stakeholders and the general

public. Instead of a top to bottom approach, Burawoy (2008) claimed that the process in the social sciences, especially at the semi-peripheral countries, are “not retreating behind the walls of academia, but advancing through the trenches of civil society” (p. 146). The findings presented in this study are pointing to an even more complex system, even where to prominent stakeholders are not directly involved, the boundary-work is still reflected and reproduced.

## 6. Future perspectives

The present study demonstrated the importance of knowledge structure in Wikipedia. This research exploited a unique context, namely a medium-sized Wikipedia with a high association between its users and a specific community. However, in order to understand better the implications of knowledge categorization on Wikipedia, we need to investigate and compare more case studies. Examination of different Wikipedias in different languages and other scientific disciplines, such as in the natural sciences and humanities will be necessary.

Recently, several enterprises have sought to automate the way Wikipedia is written (Tran & Cao 2013, Khalatbari & Mirroshandel 2015). In parallel, scholars, especially in the field of computer science, use Wikipedia as a source of data for various studies in order to understand human language (Hirschberg & Manning 2015). For these scientists and developers, it is crucial to understand that the way Wikipedia represents reality is in fact subject to certain social constructs, not only through the articles themselves, but also through the way they are organized.

## References

- Abbott, A. (2001). *Chaos of disciplines*. University of Chicago Press.
- Bartels, L. M. (2016). *Unequal democracy: The political economy of the new gilded age*. Princeton University Press.
- Ben-David, J., & Collins, R. (1966). Social factors in the origins of a new science: The case of psychology. *American sociological review*, 451-465.
- Blanden, J. (2013). Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology. *Journal of Economic Surveys*, 27(1), 38-73.
- Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, 92(5), 1170-1182.
- Burawoy, M. (2008). Open the social sciences: To whom and for what?. *Portuguese Journal of Social Science*, 6(3), 137-146.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7), 107-117.
- Burack, J. A., Iarocci, G., Bowler, D., & Mottron, L. (2002). Benefits and pitfalls in the merging of disciplines: The example of developmental psychopathology and the study of persons with autism. *Development and psychopathology*, 14(2), 225-237.

- Clauson, K. A., Polen, H. H., Boulos, M. N. K., & Dzenowagis, J. H. (2008). Scope, completeness, and accuracy of drug information in Wikipedia. *Annals of Pharmacotherapy*, 42(12), 1814-1821.
- Count, A. (1830). *Course of Positive Philosophy (Complete)(Vol.1)*.
- Csardi, G., & Nepusz, T. (2015). *igraph: Network analysis and visualization*. R package version, 1(0).
- Darnton, R. (1984). Philosophers trim the tree of knowledge: the epistemological strategy of the Encyclopédie. The great cat massacre and other episodes in French cultural history, 191-215.
- Deflem, M. (2003). The sociology of the sociology of money: Simmel and the contemporary battle of the classics. *Journal of Classical Sociology*, 3(1), 67-96.
- Dorling, D. (2014). *Inequality and the 1%*. Verso Books.
- Fabian, J. (1975). Taxonomy and Ideology: on the boundaries of concept classification. *Linguistics and Anthropology*, Lisse, 183-197.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215-239.
- Gal, S., & Irvine, J. T. (1995). The boundaries of languages and disciplines: How ideologies construct difference. *Social research*, 967-1001.
- Garfunkel, S. L. (2008). Wikipedia and the meaning of truth: Why the online encyclopedia's epistemology should worry those who care about traditional notions of accuracy. *Technology Review*, 111(6), 84-86.
- Gaziano, E. (1996). Ecological metaphors as scientific boundary work: Innovation and authority in interwar sociology and biology. *American Journal of Sociology*, 101(4), 874-907.
- Gelernter, L. *A New Epistemic Culture. Wikipedia as an Arena for the Production of Knowledge in Late Modernity*.
- Geva, S. (2013) This is how They are Making History (too): Feminist Practice in the Academia [Hebrew]. *Gilui Da'at* (4), 127-136
- Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American sociological review*, 781-795.
- Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line*. University of Chicago Press.
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070), 900-902.

- Goldthorpe, J. H. (2013). Understanding—and misunderstanding—social mobility in Britain: The entry of the economists, the confusion of politicians and the limits of educational policy. *Journal of Social Policy*, 42(3), 431-450.
- Gulbenkian Commission on the Restructuring of the Social Sciences. (1996). *Open the social sciences: Report of the Gulbenkian Commission on the restructuring of the social sciences*. Stanford University Press.
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266.
- Hoover, R. (2016) *Bias In Wikipedia: Different Links, Different Stories*. CS 229 Final Report
- Jenkins, S. P. (2015). To what extent is inequality increasing? Some data issues. LIS Lecture presented at the 6th ECINEQ conference, Luxembourg, 13 July 2015.
- Jullien, N. (2012). *What We Know About Wikipedia: A Review of the Literature Analyzing the Project (s)*.
- Khalatbari, S., & Mirroshandel, S. A. (2015). Automatic construction of domain ontology using wikipedia and enhancing it by google search engine. *J Inf Syst Telecommun*, 3, 248-258.
- Lamont, M., & Molnár, V. (2002). The study of boundaries in the social sciences. *Annual review of sociology*, 28(1), 167-195.
- Lerner, J., & Lomi, A. (2018). Knowledge categorization affects popularity and quality of Wikipedia articles. *PloS one*, 13(1), e0190674.
- Leshnick, A., & Livio, O. (2017). *Technology Meets Postmodern Epistemology: The Discourse of Hebrew Wikipedia Deletion Pages*. *AoIR Selected Papers of Internet Research*, 6.
- Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., & Lanamäki, A. (2015). “The sum of all human knowledge”: A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology*, 66(2), 219-245.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American sociological review*, 69(2), 213-238.
- Nastase, V., & Strube, M. (2008, July). *Decoding Wikipedia Categories for Knowledge Acquisition*. In *AAAI (Vol. 8, pp. 1219-1224)*.
- Nov, O. (2007). What motivates wikipedians?. *Communications of the ACM*, 50(11), 60-64.
- Piketty, T. (2015). About capital in the twenty-first century. *American Economic Review*, 105(5), 48-53.
- Ponzetto, S. P., & Navigli, R. (2009, July). *Large-Scale Taxonomy Mapping for Restructuring and Integrating Wikipedia*. In *IJCAI (Vol. 9, pp. 2083-2088)*.

- Ram, U. (2018). The State of Sociology: Some Contemporary Concerns. In Israeli Sociology (pp. 147-159). Palgrave Macmillan, Cham.
- Reigersberg, S., & Elsbeth, M. (2010). Applied ethnomusicology, music therapy and ethnographically informed choral education: The merging of disciplines during a case study in Hopevale, Northern Queensland.
- Rodgers, D. M. (2013). Insects, instincts and boundary work in early social psychology. *History of the Human Sciences*, 26(1), 68-89.
- Rotem, N. (2014). The Differences between Sociology and Anthropology [Hebrew]. Pickpook. Retrieved from <http://www.pickpook.org/2014/05/13/%D7%94%D7%94%D7%91%D7%93%D7%9C-%D7%91%D7%99%D7%9F-%D7%A1%D7%95%D7%A6%D7%99%D7%95%D7%9C%D7%95%D7%92%D7%99%D7%94-%D7%9C%D7%90%D7%A0%D7%AA%D7%A8%D7%95%D7%A4%D7%95%D7%9C%D7%95%D7%92%D7%99%D7%94/>
- Ruhnau, B. (2000). Eigenvector-centrality—a node-centrality?. *Social networks*, 22(4), 357-365.
- Small, M. L. (1999). Department conditions and the emergence of new disciplines: Two cases in the legitimation of African-American Studies. *Theory and Society*, 28(5), 659-707.
- Staub, T., & Hodel, T. (2016). Wikipedia vs. Academia: An Investigation into the Role of the Internet in Education, with a Special Focus on Wikipedia. *Universal Journal of Educational Research*, 4(2), 349-354.
- Swedberg, R. (1999). Max Weber as an economist and as a sociologist: Towards a fuller understanding of Weber's view of economics. *American Journal of Economics and Sociology*, 58(4), 561-582.
- Teplitskiy, M., Lu, G., & Duede, E. (2017). Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology*, 68(9), 2116-2127.
- Tran, T., & Cao, T. H. (2013). Automatic Detection of Outdated Information in Wikipedia Infoboxes. *Research in Computing Science*, 70, 211-222.
- Vivaldi, J., Cabrera-Diego, L. A., Sierra, G., & Pozzi, M. (2012, May). Using Wikipedia to Validate the Terminology found in a Corpus of Basic Textbooks. In LREC (pp. 3820-3827).
- Wallerstein, I. (2000). Open the social sciences. *Asia-Pacific Social Science Review*, 1(1), 1-1.
- Wikipedia [Hebrew] retrieved from: <https://he.wikipedia.org/wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94> (28.2.2018)

- Wikipedia contributors. (2018, August 30). Wikipedia:Categorization. In Wikipedia, The Free Encyclopedia. Retrieved 08:27, September 4, 2018, from <https://en.wikipedia.org/wiki/Wikipedia:Categorization>
- Wikipedia contributors. (2018, July 5). Wikipedia:Categorization. In Wikipedia, The Free Encyclopedia. Retrieved 08:27, September 4, 2018, from <https://he.wikipedia.org/wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94:%D7%A7%D7%98%D7%92%D7%95%D7%A8%D7%99%D7%94> [Hebrew]
- Wikipedia contributors. (2018, August 15). Wikipedia talk:Categorization. In Wikipedia, The Free Encyclopedia. Retrieved 08:27, September 4, 2018, from [https://he.wikipedia.org/wiki/%D7%A9%D7%99%D7%97%D7%AA\\_%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94:%D7%A7%D7%98%D7%92%D7%95%D7%A8%D7%99%D7%94](https://he.wikipedia.org/wiki/%D7%A9%D7%99%D7%97%D7%AA_%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94:%D7%A7%D7%98%D7%92%D7%95%D7%A8%D7%99%D7%94) [Hebrew]
- Wright, E. O. (2011). Real utopias. *Contexts*, 10(2), 36-42
- Yam, S. C. J. (2012). A sociology of knowledge approach to cooperative knowledge generation: ideology and structural tension of Nupedia. *West East journal of social sciences*, 1(1), 48-54.
- Yasseri, T., Spoerri, A., Graham, M., & Kertész, J. (2014). The Most Controversial Topics in Wikipedia. *Global Wikipedia: International and cross-cultural issues in online collaboration*, 25.
- Zhang, Y., Sun, A., Datta, A., Chang, K., & Lim, E. P. (2010, June). Do Wikipedians follow domain experts?: a domain-specific study on Wikipedia knowledge building. In *Proceedings of the 10th annual joint conference on Digital libraries* (pp. 119-128)