Semantic Hubs and Authorities for Citation Analysis

Jennifer Plank and Subhasish Mazumdar

jen.plank@gmail.com, mazumdar@cs.nmt.edu
New Mexico Institute of Mining and Technology,
801 Leroy Place, Socorro, New Mexico (USA)

Abstract

We describe a method for analyzing the impact of papers on their field. We classify the paper's citations by semantic meaning with numeric encoding and use them as weights on our extension of the *Hubs and Authorities* algorithm to determine the most authoritative papers. We compare our authorities to those determined by straight citation counts. While current common analysis methods only take into account the local citation counts, in part due to the difficulty of classifying large numbers of citations, we rely on graph analysis. We not only use the structure of the links among papers but also the semantics of the citations in order to get a step closer to the actual meanings underlying the citations.

Introduction

Tracking the impact or importance of a given paper, and by extension its author or authors, can be difficult . Existing methods using solely citation counts are considered by many to be inadequate (Adler, Ewing, & Taylor, 2008). For an example of the common pitfalls of straight citation counts, consider a controversial paper which draws a high volume of citations from papers that refute it; clearly high citation count here does not indicate established authority. The increasing tendency to base funding decisions on such metrics therefore is therefore troubling. An analysis method is needed that more closely reflects the actual impact of a paper.

From a research perspective, it can also be difficult to discover which papers are the most important in a field of study, and to determine the relationships between the papers. It would be useful to have an overview saying which papers build on the work of, support, or criticize which other papers. Unfortunately this can currently only be done by reading all of the papers in question, which can be impractical due to the sheer number of papers in some fields. For example, at the ACM digital library the search term "citation analysis" retrieved 56,244 results.

Current methods of analyzing the impact of a paper tend to look only at the citation count. While there are other metrics available, they generally focus on the author or journal rather than the paper itself. One of the most common methods currently used to analyze citations is the *h-index* (Hirsch, 2005), a metric developed by Jorge E. Hirsch in 2005 to quantify the contribution of a given researcher to their field. It is defined as the highest number x such that the researcher has published at least x articles, each of which has received at least x citations. While commonly used, it has a number of weaknesses. For example, it gives a very low score to researchers with a few, very highly cited papers; as Costas and Bordons (Costas & Bordons, 2007) said, "It is better to have 10 papers with 10 citations each than 5 papers with 200 citations each". This and other similar metrics based purely on citation counts are insufficient because they don't take citation types or the larger structure of the citation graph into account.

More complicated methods such as *CiteRank* (Maslov & Redner, 2008), an adaptation of Google's PageRank, still fail to take into account the semantic context of citations. *P-Rank* (Yan, Ding, & Sugimoto, 2011) takes into account the structure of the graph, including author, article, and journal, but not the semantic content of the citations.

In this paper we look at the network of citations as a graph, with each paper being a node and each citation an edge. There are many algorithms for analyzing the structure of a graph; some are easily used or modified for citation graphs. The Modularity Communities (Newman M. E., 2004) algorithm can divide a graph into its inherent communities by detecting tightly-knit clusters within a larger graph; this could be useful for determining related papers. *Betweenness* (Brandes, 2001) could be used in finding papers that cross field boundaries; it calculates the number of shortest paths between nodes that go through a given node. An algorithm could be developed to evaluate the 'similarity' of two papers based on common citations, or to determine the diversity of the papers that cite a given paper, as a measure of how much other research it triggered. We have chosen to look at the *Hubs and Authorities* algorithm (Kleinberg J. M., 1999) because of its relative simplicity, and the immediate applicability of its results. Originally designed to locate authoritative web pages, it is easily modified to determine authoritative papers, based on the full structure of the graph rather than the local citation counts.

While taking graph structure into account is necessary, it is not sufficient. Each citation can have a different meaning, as some papers are essential to the paper that cites them, while others merely indicate work in a related area, or even a criticism of another paper. We need a method of categorizing a citation so as to influence its impact on the ranking of the cited paper.

A number of people have put forward citation classification schemas, ranging from simple '*positive*, *negative*, or *neutral'* schemes to a variety of complex systems involving dozens of categories or several axes of measurement. Garfield (Garfield, 1962) suggested fifteen overlapping categories that citations could fall into; Moravcsik and Murugesan (Moravcsik & Murugesan, 1975) rated citations on each of five aspects. More recently, Simone Teufel (Teufel, Siddharthan, & Tidhar, 2006) has put forward a list of twelve categories. It may also be noted that each academic discipline has its own common citation types, depending on the style of research that is done in the field (Lillquist & Green, 2010). Because of this, a detailed categorization scheme that works well for one field may not be a good fit for another. Despite the availability of classification schemes, current common analysis methods only take into account the local citation counts, in part due to the difficulty of classifying large numbers of citations.

We first use graph analysis methods such as *Hubs and Authorities* to draw upon the structure of the graph as a whole to better represent the impact of a paper. Next, we add the semantics of citation types to take another step towards the actual meaning inherent in the citation graph. We compare the 'Authoritative' papers found by our method to the papers that receive high citation counts.

Background Information

In this section we go over the definitions and algorithms used later in the paper.

Citation Classification

There have been many citation ontologies put forward (Garfield, 1962) (Moravcsik & Murugesan, 1975) (Teufel, Siddharthan, & Tidhar, 2006), but none have been widely accepted as a general standard for classifying citations. Our goal for this application is a simple classification into broad categories, so we do not need to have separate definitions for each possible meaning of a citation; identifying the general level of impact of the citation is sufficient.

We will examine two categorization schemes; first, the 5-category scheme of *Essential, Considerable, Limited, Minor Criticism* and *Major Criticism*. This scheme is taken in large part from [8]. Next, we consider a 3-category scheme created by combining the *Essential* and *Considerable* categories into a single *Positive* category, and combining *Major* and *Minor Criticisms* into a single *Negative* category. Table 1 shows the citation types and how the 3-

category and 5-category schemes relate. We will refer to citations by their 5-category classification, as these categories can be unambiguously mapped to the 3-category scheme.
The 5-category division focuses on how much support the citing paper draws from the cited paper, or if the citation is critical, whether it criticizes all or only a part of the cited paper. For our purposes, this is sufficient. There are more complex schemes (Bornmann & Daniel, 2008), but these were judged more complex than was needed here, as most of them try to include the author's motivation for citing as well as fine-grained analysis of what is said about the cited paper or article.

Table 5: Citation Types

3-category	5-category
Positive	Essential
	Considerable
Neutral	Limited
Negative	Minor Criticism
	Major Criticism

Hubs and Authorities Algorithm

We use the *Hubs and Authorities* algorithm, designed in 1999 by Kleinberg (Kleinberg J. M., 1999). Its original purpose was to identify 'Authoritative' web pages using the patterns of hyperlinks to them, particularly from 'Hub' or directory sites, but it has been adapted to many applications, both in web analysis and elsewhere (Cohn & Chang, 2000) (Mizzaro & Robertson, 2007) (Greenberg, 2009). We adapt this to identify authoritative papers, using citations instead of hyperlinks, and with survey papers in place of hubs. In the standard Hubs and Authorities algorithm, each hyperlink has the same weight, and contributes the same amount to the values computed.

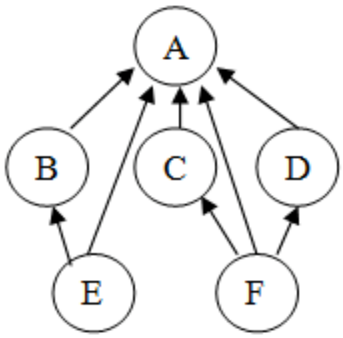

Figure 1: Hubs and Authorities Example Graph

The Hub and Authority values for each paper are calculated iteratively. Every paper starts with values of 1 for both its Hub and Authority values; during each iteration of the algorithm, a paper's authoritativeness is calculated as the sum of the Hub value of the papers that refer to the paper (Kleinberg, Kumar, Raghavan, Rajagopalan, & Tomkins, 1999); that is, for every paper p using the notation that $q \rightarrow p$ means that q cites p,

$$auth_p = \sum_{q \text{ such that } q \rightarrow p} hub_q$$

where $auth_p$ is the authority value of paper p, hub_q is the hub value of paper q, and q varies over the set of papers that cite p. The Hub value is similarly calculated as the sum of the authorities of the papers this paper refers to:

$$hub_p = \sum_{q \text{ such that } p \to q} auth_q$$

Note that the direction of citation is changed from the authority calculation. At the end of each iteration of the algorithm, the computed Hub and Authority values are normalized. We used a sum-of-squares method in which the normalized value is the square of the computed value divided by the sum of the squares of all *n* computed values, corrected to keep the sign:

$$norm(V) = sign(V) * \frac{V^2}{\sum_{i=1}^{n} V_i^2}$$

As a consequence, values of 1.0 only appear if a single value has completely overshadowed all other Hub or Authority values; sign correction allows for the possibility of negative Hub or Authority values.

The algorithm continues until there is no change for two consecutive iterations, although for simplicity most implementations stop at a given number of iterations. Previous work has been done showing that the number of iterations required for the algorithm to converge is polynomial on the number of nodes in the graph, both for weighted and un-weighted graphs (Peserico & Pretto, 2009) (Mizzaro & Robertson, 2007). Since the values assigned to any given paper can vary with changes in other parts of the network, it is not the values themselves that matter as much as the relative differences between values. A paper with an authority value that appears low may still be the most authoritative paper in its field, if that field has lower citation rates in general than other fields.

As a small example, we take the graph in Figure 1 and show the hub and authority values after each iteration in Table 6. In this example, it takes fourteen iterations to converge to an accuracy of 3 decimal places. We see that the h(B), h(C), and h(C) sequences are the same, because the nodes B, C, and D in the graph cite the same papers; similarly, a(C) and a(D) are the same because they are cited by the same papers.

Table 6: Hub and Authority values at each iteration

Iter-ation	h(A)	a(A)	h(B)	a(B)	h(C)	a(C)	h(D)	a(D)	h(E)	a(E)	h(F)	a(F)
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	0.000	0.893	0.063	0.036	0.063	0.036	0.063	0.036	0.250	0.000	0.563	0.000
2	0.000	0.590	0.191	0.037	0.191	0.187	0.191	0.187	0.206	0.000	0.222	0.000
3	0.000	0.876	0.147	0.037	0.147	0.043	0.147	0.043	0.166	0.000	0.392	0.000
4	0.000	0.749	0.189	0.021	0.189	0.115	0.189	0.115	0.205	0.000	0.228	0.000
5	0.000	0.872	0.173	0.037	0.173	0.045	0.173	0.045	0.183	0.000	0.297	0.000
6	0.000	0.827	0.189	0.028	0.189	0.073	0.189	0.073	0.205	0.000	0.230	0.000
7	0.000	0.871	0.183	0.037	0.183	0.046	0.183	0.046	0.196	0.000	0.254	0.000
8	0.000	0.857	0.188	0.033	0.188	0.055	0.188	0.055	0.205	0.000	0.230	0.000
9	0.000	0.871	0.187	0.036	0.187	0.046	0.187	0.046	0.201	0.000	0.238	0.000
10	0.000	0.867	0.188	0.035	0.188	0.049	0.188	0.049	0.204	0.000	0.230	0.000
11	0.000	0.871	0.188	0.036	0.188	0.046	0.188	0.046	0.203	0.000	0.233	0.000
12	0.000	0.870	0.188	0.036	0.188	0.047	0.188	0.047	0.204	0.000	0.230	0.000
13	0.000	0.871	0.188	0.036	0.188	0.046	0.188	0.046	0.204	0.000	0.231	0.000
14	0.000	0.871	0.188	0.036	0.188	0.047	0.188	0.047	0.204	0.000	0.231	0.000

Semantic Hubs and Authorities

We propose multiplying the hub or authority value being summed by a weight on the citation, so that the authoritativeness will be calculated by

$$auth_p = \sum_{q \text{ such that } q \rightarrow p} (hub_q * weight_{qp})$$

where $auth_p$ is the new authority value of paper p, hub_q is the hub value of a paper that cites p, $weight_{qp}$ is the weight of the citation from q to p, and q varies over the set of papers that cite p. Similarly, a paper's Hub value is calculated by

$$hub_p = \sum_{q \text{ such that } p \rightarrow q} (auth_q * weight_{pq})$$

Because we have found that *Limited* is by far the most common citation type, citations that have not yet been categorized are assumed to be *Limited* and share the same weight. We have set this weight to 1 for both of the possible weighting schemes we examine. *Essential* and *Considerable* citations should have greater positive influence on the calculated values than *Limited* citations, while both types of *Negative* citations should have corresponding negative influence.

We consider two possible weighting methods. The first is to manually select the weights using knowledge of the algorithm and the desired behavior. Since the *Limited* or *Neutral* citation type is the standard, we set its weight to 1. The more strongly positive citation types of *Considerable* and *Essential* are set to 2 and 3 respectively, and *Major Criticism* and *Minor Criticism* to -1 and -2. In the 3-category scheme, *Positive* has a weight of 2, and *Negative* a weight of -1.

The manually selected weights were designed to give only slightly more emphasis to supporting citations, and a slight negative impact to refuting citations. In the 3-category scheme, *Positive* citations have twice the impact of *Neutral* citations. Similarly, *Negative* citations have the opposite effect of *Neutral* citations. In the 5-category scheme, the weights from *Positive, Neutral* and *Negative* are applied to *Considerable, Limited*, and *Minor Criticism*, while *Essential* and *Major Criticism* receive weights of 3 and -2, respectively.

The second weighting method we consider is to assign the weights relative to the inverse of the type's frequency of occurrence, where frequency is defined as the probability that a randomly selected citation from our sample would be of the given type. We manually classified 445 citations from eleven papers published by the ACM to gather frequency data, then calculated a *normalized inverse frequency* (NIF) so that the Neutral or Limited type would have a weight of 1, using the following formula:

$$NIF(type) = \frac{1 - freq(type)}{1 - freq(Limited)}$$

Table 3 shows the count and frequency of each citation type along with their manual and normalized inverse frequency. The supportive citation types of *Essential*, *Considerable*, and *Limited* use the normalized value as a positive number, the critical citations as negative. As no *Major Criticism* citations were found in the sample, the frequency of *Minor Criticism*

Table 3: Citation Type Frequencies and Two Weighting Schemes

	Count	Freq.	NIF	Manual		Count	Freq.	NIF	Manual
					Ess.	21	.047	5.74	3
Pos.	63	.142	5.17	2	**Cons.**	42	.094	5.46	2
Neut.	371	.834	1	1	**Lim.**	371	.834	1	1
Neg.	11	.025	-5.87	-1	**Min. Crit.**	11	.025	-5.87	-1
					Maj. Crit.	0	.025*	-5.87*	-2

citations was also used for *Major Criticism* citations, denoted by an asterisk.

The Inverse Frequency weighting method is designed to put a larger emphasis on the rarer citation types; an *Essential* support citation has more than five times the effect of a *Limited* citation, yet a single minor criticism carries enough weight to more than offset it. The difference in the 3-category scheme is even greater.

Citations containing direct criticisms are very rare; we find only 2 to 4% negative citations in the frequency findings of researchers (Bornmann & Daniel, 2008) (Teufel, Siddharthan, & Tidhar, 2006) who use their own various categorization schemes. One explanation for this finding is that most authors will go out of their way not to offend any of their peers. Instead of stating a refutation or criticism outright, they will simply present their findings without directly contradicting a previous paper.

Results

In this section we go over some synthetic examples of the algorithm, then move on to a larger graph of real-world data.

Scenario 1: Supporting vs. Refuting Citations

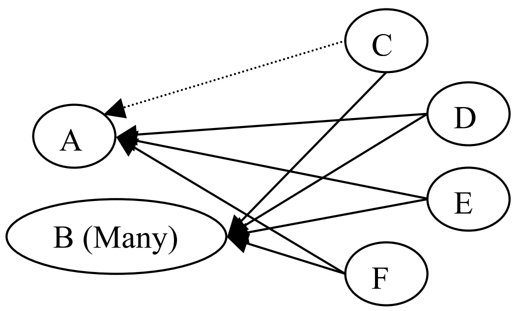

Figure 2: One Categorized Citation

This synthetic data set shows how the modified algorithm differs from the standard algorithm when dealing with citations received from later papers. Four papers all cite both Paper A and a collection of other papers represented by B. One of the citations of A is varied through both Positive and Negative citation types, all other citations are Neutral. The large group of B papers is necessary to keep the Hub values of the four citing papers relatively constant, in order to observe the effect of the citations on A.

It would be natural to assume that paper A would be more authoritative if one citation is supportive, and less authoritative if it is critical. However, both straight citation counts and the standard Hubs and Authorities algorithm give papers A and B exactly the same rating, as they are cited by all the same papers. Table shows the authoritativeness of A and B using the

Table 5: Authoritativeness for Scenario 1, with a supportive citation

	Manual	IF(3-cat)	IF(5-cat)
Paper A	.128	.213	.221
Paper B	.102	.101	.101

Table 6: Authoritativeness for Scenario 1, with a critical citation

	Manual	Inv. Freq.
Paper A	.052	-.081
Paper B	.103	.103

Manual weights and the Inverse Frequency weights. Using the manually selected weights, paper A is somewhat more authoritative than paper B. When we used the inverse frequency weights, however, paper A became more than twice as authoritative as paper B. Table shows the authoritativeness of papers A and B as calculated using the modified algorithm with manual weights and weights calculated from inverse frequency, when one of the citations of

A is critical. Only one value is shown for Inverse Frequency weights, because the lack of any *Major Criticism* citations in the sample set removes any difference between the 3-category schema and the 5-category schema for this example. Once again, both paper A and B have the same citation count and authority using the un-weighted algorithm. In the manual case the authority of paper A is halved, but in inverse frequency the Authority value becomes negative.

A similar example from real data can be found by examining the paper '*Elvin has left the building: a Publish/Subscribe service with Quenching'*. This paper has ten papers citing it, with one of the citations being in the *Considerable* category and the others *Limited*. We look at the authority value and rank in the set, for the standard algorithm and for each of our weighting schemes, as well as straight citation counts. The large set of 'B' papers was not necessary, as the paper was part of a much larger graph.

The minor changes from the manual weights raised the paper's Authority value slightly, and moved it up one in the ranking. The inverse frequency weights, however, resulted in a dramatic rise in authority value, and a continued rise in rank, shown in Table 7. In this example we only changed the weights of the references citing this one paper; in an example where all citations are categorized, the effect will be less pronounced as other papers are also pushed upward.

Table 7: Relative Authoritativeness and rank in the set, for a single paper

	Authority Value	Rank
Cite Count	-	23
Standard	.077	11
Manual	.087	10
IF(3-cat)	.142	9
IF(5-cat)	.150	8

Scenario 2: Referencing an Authoritative Paper

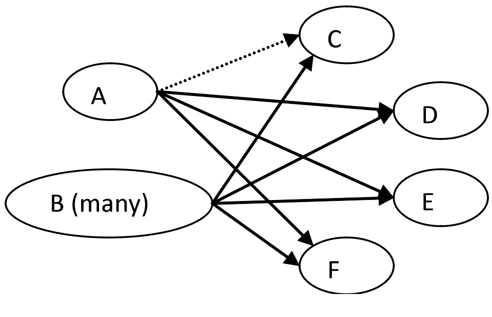

Figure 3: Scenario 2: Citing Authorities

The concept of Authority is straightforward; an authoritative paper is one that has been well received and is influential in the field. The Hub value computed by the algorithm is less easy to grasp. A hub paper is defined as one that refers to a lot of authorities. An example of this is a survey or 'State of the Field' paper; in general, papers with high Hub values organize and categorize existing knowledge rather than add new research.

The semantics with which these survey papers cite other papers is also important, from either the hub or authority's perspective. If a hub cites a paper with high authoritativeness, the influence of that citation on the citing paper's Hub value is increased. If the citation is

supportive, the effect is multiplied. If the citation is critical, however, the same effect is negated, reducing rather than increasing the citing paper's Hub value.
This example shows the effect of a citation type on the citing hub. Paper A cites four others. The type of one of the citations is varied, while the others remain *Limited*. Many other papers also cite the same four, all with *Limited* citation types. This holds the Authority values of the four papers relatively constant, so that the effect on just paper A can be observed. We look at the effect using all five citation types, although the *Limited* type shows no difference between paper A and the others.

Table 8: Hub values for Scenario 3 – Varying Types

	Manual	IF(3-cat)	IF(5-cat)	Type
Paper A	.154	-	.229	Essential
Others	.102	-	.100	
Paper A	.128	.213	.221	Considerable
Others	.102	.101	.101	
Paper A	.103	.103	.103	Limited
Others	.103	.103	.103	
Paper A	.052	-.081	-.081	Minor
Others	.103	.103	.103	Criticism
Paper A	-.027	-	-.081	Major
Others	.103	-	.103	Criticism

As expected, *Considerable* and *Essential* citations have increased the hub value of the citing paper, and *Minor* and *Major Criticism* citations have decreased it. In the case of *Major Criticism*, the hub value becomes negative for both Manual and Inverse Frequency weightings, while a *Minor Criticism* only results in a negative value for the Inverse Frequency weights. The effect of the 3-category weighting scheme is again slightly less pronounced than the 5-category scheme.

Scenario 3: A Larger Graph

Here we examine a larger graph, consisting of 284 nodes and 348 edges. This graph includes most of the citations we have categorized, excluding one paper whose citations did not connect to this graph. Uncategorized citations are not included. The papers centered in a star formation of citations in Figure 4 have had all of their citations categorized, while the ones with only one or a few connections have only had some of their citations categorized. Thicker lines indicate a heavier-weighted citation. Removing uncategorized citations as we did here will give the fully categorized papers an artificial boost in hub value. We ran our modified Hubs and Authorities on this data and observed the top hub and authority values using citation types, and both standard and weighted hubs and authorities.
Table shows the top ten hubs for each method. Due to the nature of the sample, the top ten are the ten fully categorized papers for all ranking methods, though the ranking order changes. The un-categorized Hubs and Authorities and straight citation count rankings are very similar, with 3 and 4 exchanging places and 8-10 changing, but the rest remain the same, so that the average movement is only 0.6. The difference between un-categorized and manual weights is greater, with an average movement of 1.2 ranks. Going to the 3-category Inverse Frequency system produces a small change of .8 ranks on average, and the average change between the 3-category and 5-category systems is very small, with an average change of only 0.2 ranks. The total average change from un-typed to 5-category Inverse Frequency is 1.6 ranks.

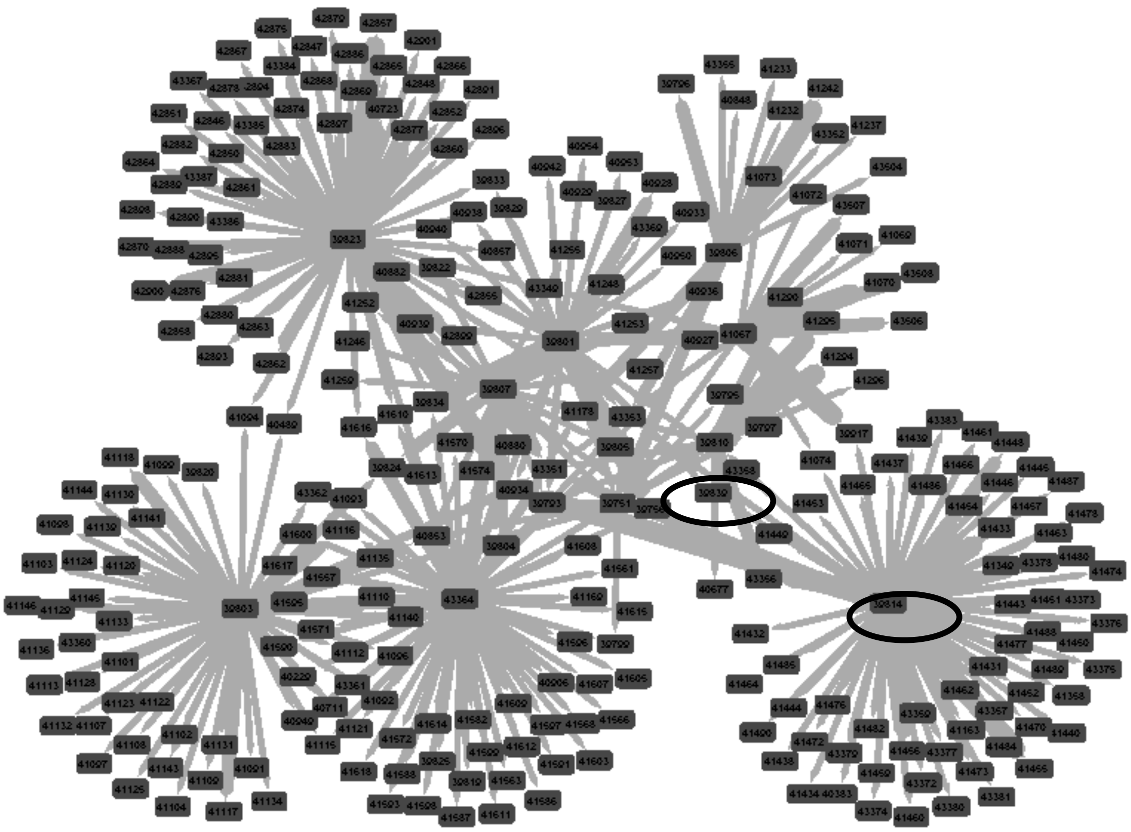

Figure 4: Structure of the larger graph

Table shows the top ten authority values for the categorized graph, using each method. There is more variance in the members of the top ten, as the ten fully categorized papers are not given an artificial boost in authority rating. The citation counts have a lot of overlap, but it is clear that the papers with higher citation counts also score well in the uncategorized Hubs and Authorities algorithm. As we move on to the Manually Weighted method, only half of the top ten remain in the top ten; five lower-cited other papers are also introduced. This trend continues through the 3- and 5-category Inverse Frequency methods, until only one of the original top ten remains. These papers with fewer citations achieve high authority because their citations are more heavily weighted. Looking at the data, all five 2nd place papers in the

Table 9: Top Ten Hub values for categorized graph

HUBS	Outgoing Citations		Un-categorized		Man-ual		IF 3-cat		IF 5-cat	
Distributed computing research…	75	1	0.581	1	0.875	1	0.898	1	0.908	1
the Many Faces of Publish/Subs…	65	2	0.554	2	0.181	4	0.085	6	0.067	7
On objects and events	56	4	0.372	3	0.153	5	0.098	4	0.081	4
The JEDI Event-Based Infrastruct…	62	3	0.327	4	0.332	2	0.372	2	0.364	2
Design and evaluation of a wide…	31	5	0.227	5	0.198	3	0.139	3	0.126	3
Publish/Subscribe in a mobile en…	19	6	0.164	6	0.104	6	0.080	7	0.068	6
Mercury: a scalable publish-sub …	15	7	0.117	7	0.102	7	0.091	5	0.081	5
Matching events in a content-b …	10	9	0.082	8	0.039	10	0.017	10	0.014	10
Exploiting event stream interpr…	9	10	0.078	9	0.057	8	0.047	9	0.042	9

Table 10: Top Ten Authority values for categorized graph

AUTHS	Incoming Citations		Un-categorized		Manual		IF 3-cat		IF 5-cat	
Elvin has Left the Building: A Publish…	9	1	0.267	1	0.182	3	0.125	3		
Matching Events in a content-based…	8	2	0.232	2	0.151	5				
An Efficient Multicast Protocol for …	8	2	0.227	3	0.16	4				
Achieving scalability and expressiveness…	4	4	0.163	4	0.149	6				
Exploiting an event-based infrastructure…	4	4	0.161	5	0.338	1	0.394	1	0.405	1
A design framework for Internet-scale…	3	5	0.118	6						
the process group approach to reliable…	5	3	0.118	6						
the JEDI Event-base Infrastructure…	3	5	0.101	7						
A reliable multicast framework for …	2	6	0.099	8						
Bimodal Multicast	2	6	0.099	8						
Exploiting IP Multicast in Content-Based…	1	7			0.223	2	0.269	2	0.281	2
Locating Data in (Small-World?) Peer-to-…	1	7			0.223	2	0.269	2	0.281	2
The Anatomy of the Grid: Enabling Scala…	1	7			0.223	2	0.269	2	0.281	2
The GriPhyN Project: towards Petascale	1	7			0.223	2	0.269	2	0.281	2

Manual weighting method receive *Essential* citations from *"Distributed Computing Research Issues…"* (circled near the right in Figure 4), which consistently has the highest hub value in the graph.

The one paper that remains in the top ten using every method*, "Exploiting an Event-Based Infrastructure…"* (circled near the middle in Figure 4) actually increases in rank when using categorized Hubs and Authorities, implying that one or more of its citations is strongly positive, and from a high hub-valued paper. Looking at the citations it has received, we saw that indeed 2 of its 4 citations were *Essential*, and 2 Neutral, and one of its *Essential* citations was also from *"Distributed Computing Research Issues…"*.

The highest-cited paper, "*Elvin has left the Building…"*, did not remain in the top ten. As we saw in Scenario 1, this paper has no *Essential* citations, and only one *Considerable*. One of its citations was un-categorized, and so not included in the large data set. It remains at the top in the un-categorized Hubs and Authorities rankings, but moves down to third in the categorized ones. This downward movement is due to the other papers positive citations from higher-ranked hubs; as shown in Table , the relative prominence of the highest ranked hub goes up dramatically in the categorized weighting methods

The last two papers listed in Table, *"An Experimental Open Architecture…"* and *"DERPA: A Generic Distributed Event…"*, are among the top ten papers only in the 5-category Inverse Frequency method. In the Manual method, they have rank 10 (papers 16 and 17). They each received one *Essential* citation from *"The JEDI Event-Based Infrastructure…",* which is ranked 2nd in the weighted Hubs and Authorities algorithm. None of the papers in the top ten had received any *Critical* citations, so the reduction in rank seen in several papers in Table is due to other papers receiving more *Essential* or *Considerable* citations, rather than any *Critical* citations of these papers.

Using the Hubs and Authorities algorithm allows infrequently-cited yet potentially important papers to rise to the top-ranked spots, bringing attention to papers that might be overlooked using a citation-count method.

Discussion and Future Work

Current widely-used methods of assessing the value of a paper are insufficient as they do not take into account either the structure of the citation graph as a whole or the semantic content of citations. In order to remedy this, we use a simple citation classification scheme to take into account the type of reference, e.g., background information, supporting work, or refutation; associate numeric weights with those classes; and modify the Hubs and Authorities algorithm to make use of those weights.

Assigning *Considerable* and *Essential* citations double and triple the weight of *Neutral* citations, and *Minor* and *Major Criticism* citations weights of -1 and -2 respectively, has a significant, but not overwhelming, effect on the final rankings of a citation network. Use of a weight that is the inverse of the frequency of the citation type has more effect, in the most extreme case nearly wiping out the top ten authorities obtained using un-categorized hubs and authorities.

The critical citations also have a much greater effect than seems desirable, removing all authoritativeness with very few citations. It may be better to halve the weight of Negative citations in both Manual and Inverse Frequency weighting schemes.

We have seen that modifying Hubs and Authorities to include semantic weights produces results more intuitively appropriate for the papers in specific situations. Further work needs to be done to determine the best balance of weights, and to determine the most appropriate categorization scheme or schemes. Different academic fields may require different categorization schemes, as they exhibit altered citation patterns.

Applying weights to citations as a modification of the Hubs and Authorities algorithm allows the impact of various citations to be more accurately modelled. While the exact weights and types optimal for any given field remain to be found, a knowledgeable guess can lead to improvement in the algorithm's output.

One potential area for future work is developing a method of automatically typing citations. As long as all citations must be categorized manually, we will have only partially categorized graphs to work with. There are some promising avenues of exploration in the area of semantic analysis (Colbaugh & Glass, 2010) that could be pursued.

Another interesting facet of the citation network that we have not examined here is time. Each paper is inserted into the graph at a specific point in time, and research has shown that the time of publication in relation to the age of the field is very influential on the number of citations that a paper will receive (Newman M. E., 2009).

Bibliography

Adler, R., Ewing, J., & Taylor, P. (2008, June). Citation Statistics. *Joint Committee on Quantitative Assessment of Research* .

Bornmann, L., & Daniel, H.-D. (2008). What do Citation Counts measure? A Review of Studies on citing Behavior. *Journal of Documentation , 64* (1), 45-80.

Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology* 25, 163–177.

Cohn, D., & Chang, H. (2000, June). Learning to Probailistically Identify Authoritative Documents. *International Conference on Machine Learning 2000* .

Colbaugh, R., & Glass, K. (2010, May). Estimating Sentiment Orientation in Social Media for Intelligence Monitoring and Analysis. *IEEE Conference on Intelligence and Security Informatics* .

Costas, R., & Bordons, M. (2007). The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics , 1(2)*, 193–203.

Garfield, E. (1962). Can citation indexing be automated? *Essays of an information scientist , 1*, pp. 84-90.
Greenberg, S. A. (2009). *BMJ* , 339:b2680.
Hanney, S., Grant, J., Jones, T., & Buxton, M. (2005). Categorising citations to trace research impact. In P. &. INGWERSEN (Ed.), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics.* Stockholm, Sweden: Karolinska University Press.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, 102* (46).
Kleinberg, J. M. (1999, December). Hubs, Authorities, and Communities. *ACM Computing Surveys* 31(4).
Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. S. (1999). The web as a graph: measurements, models, and methods. *Proceedings of the 5th Annual international Conference on Computing and Combinatorics*, (pp. 1-17). Tokyo, Japan.
Lillquist, E., & Green, S. (2010). The discipline dependence of citation statistics. *Scientometrics , 84*, 749–762.
Maslov, S., & Redner, S. (2008, October 29). Promise and Pitfalls of Extending Google's PageRank Algorithm to Citation Networks. *The Journal of Neuroscience* , 28(44):11103-11105.
Mizzaro, S., & Robertson, S. (2007, July). HITS Hits TREC - Exploring IR Evaluation Results with Network Analysis. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007.
Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Soc. Stud. Sci , 5*, pp. 86-92.
Newman, M. E. (2004, April). Coauthorship Networks and Patterns of Scientific Collaboration. *Proceedings of the National Academy of Sciences.* April 6, 2004. vol. 101 no. 1 5200-5205.
Newman, M. E. (2009). The first-mover advantage in scientific publication. *Europhysics Letters* 86, 68001.
Peserico, E., & Pretto, L. (2009). HITS Can Converge Slowly, but Not Too Slowly, in Score and Rank. *Lecture Notes in Computer Science , 5609*, pp. 348-357.
Teufel, S., Siddharthan, A., & Tidhar, D. (2006). An annotation scheme for citation function. *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, (pp. 80–87). Sydney.
Yan, E., Ding, Y., & Sugimoto, C. R. (2011). P-Rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the American Society for Information Science and Technology , 62* (3), 467-477.