INTRODUCTION

In the future, technological progress and economic strength, among other aspects of society, will become more and more reliant on patents; thus, it is necessary to understand them more thoroughly. Patents are nebulous entities; the method in which they are classified, the description of the work entailed, and the list of citations of prior art are all markedly flawed (Watson 2003). When attempts are made to utilize or understand patents in some useful manner, such as trying to assign value to a patent, these flaws have proven to be material.

Patents contain many different types of information: inventor’s name, works cited, and description of work, for example. Researchers have used this information to make assumptions about patents. The problem with this approach is that much of the information contained within a patent is erroneous, misleading, or made deliberately vague. In order to make valid decisions using patents, it is necessary to sift through the material contained within the patents to determine what should and should not be used in the decision making process. We developed and tested several hypotheses with respect to the relationship between patents and their citations to determine whether current citation analysis methods are justifiable.

Citation analysis is based on the premise that the number and nature of forward and backward citations are indicators of patent value; specifically, that greater numbers of forward citations indicate greater value (Fleming 2003). A forward citation is a reference to the given patent by another patent. In Figure 1 below, patents A, B, and C are forward citations of patent X.

According to Anthony Brietzman and Patrick Thomas of the Industrial Research Institute, “The idea behind patent citation analysis is that patents cited by many later patents tend to contain important ideas upon which many later
inventors are building," making them more valuable (2002). However, the use of simple citation counts for patent valuation may be flawed because it does not take into account that citation lists are not always complete or valid. In some cases, an incentive exists for patent authors to either not cite all related patents or to include spurious citations. While patent relationships may indicate value, some account must be made for the possibility of invalid citations. We developed three approaches to determine the usefulness of using forward citation counts to value patents.

The premise of the first approach was that forward citation count does not indicate the value of a patent. If we could show that all patents increase in value if they belong to a growing field of innovation, then this would invalidate the premise. This is because as new patents are included in a field, they are usually similar to existing ones. Therefore, the patents lose uniqueness, and have diminished value.

The second approach focused on the size of the portfolio to which the patent belonged. The fundamental question behind this hypothesis is whether the size of the patent portfolio affects the number of forward citations. If characteristics of patent portfolio can predict forward citations, then forward citations are not independent predictors of value.

The third approach postulates that all forward citations are not equal. Citations occur over generations of patents, and two patents that have an equal number of first generation citations may not have the same number of higher generation citations. To gather data for this analysis, we developed a method called a Patent Citation Network to represent patents and their respective citations as a network. Additionally, we developed an implementation tool to determine the number of multigenerational citation relationships by delineating and counting network paths.

2 SIMPLE FORWARD CITATION COUNT AND THE VALUE OF A PATENT

As stated before, many researchers believe that the number of forward citations of a patent indicates value. The hypothesis for the first approach was that the number of forward citations attributable to a patent does not determine its value. We tested this hypothesis by attempting to show that all patents in a field of innovation increase in the number of forward citations.

2.1 Methodology

Our methodology included the use of experimentation to prove by contradiction that the premise is correct.

An important aspect of the experimentation was the use of Morphogenetic sets. Morphogenetic sets are a patent pending technology developed by M-CAM Inc. The Morphogenetic set of a patent X is composed of the relevant two generations of citations related to patent X.

The assumption of this proof was that a Morphogenetic set represents a field of innovation because it includes the patents that are most likely to be similar to the subject patent, or the patent under immediate consideration. We also considered using classification codes to represent a field of innovation but they have been shown to be inconsistent.

We wanted to prove that forward citation count does not indicate the value of a patent through a proof of contradiction. An experiment was performed to determine if there was a significant positive correlation between the number of forward citations of a subject patent and the average number of forward citations of the patents in the subject patent’s Morphogenetic set through time. If the experiment showed that the relationship existed for most investigated subject patents and forward citations did indicate value, then this would mean that the value of patent increases as the size of the of the Morphogenetic set increases. This would then imply that all patents in a field of innovation must increase in value. However, all patents in a field cannot increase in value because as a field grows, the new patents are similar to the existing patents. Consequently, the patents in the set lose uniqueness, which is one of the characteristics that makes a patented technology valuable.

The data for this experiment was provided by M-CAM, through both the M-CAM DOORS™ application and Jason Watson, M-CAM’s Vice-President of Information Systems. The goal in the data collection process was to randomize the subject patents examined by ensuring that their issue dates were varied and that they belonged to a variety of industries.

In the experiment, a regression model was built for each investigated patent. The number of forward citations of the subject patent (S) was the response variable and the average number of forward citations of the patents in the subject patent’s Morphogenetic set (M) was the independent variable. The slope parameter for each model was tested to determine if it was significantly greater than zero, implying that a positive correlation exists.

2.2 Results and Implications

A total of 62 subject patents were investigated in the experiment. The slope parameter for each model was tested at significance levels of .05, .01 and .02. The results of the experiment are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1: Results from experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significance level</td>
</tr>
<tr>
<td>Number of models with positive correlation</td>
</tr>
<tr>
<td>Percentage of models with positive correlation</td>
</tr>
</tbody>
</table>
The results illustrate that 92% of the subject patents experimented had a positive correlation between S and M. The random nature of the data implies that the majority of the patents behave this way. The results of the experiment in conjunction with the proof by contradiction show that forward citations do not accurately reflect the value of a patent.

The accuracy of the models, as measured by the coefficient of determination, was high. However, there were few data points for each model because S and M were collected for every year. More data points could be obtained by shortening the period to months instead of years, although the number of forward citations of patents change at different rates.

To strengthen the validity of this approach, the experiment could be applied to different methods of determining a field of innovation, such as classification codes. If the results are consistent, then there will be a stronger acceptance of the findings.

3 FORWARD CITATIONS AND PATENT PORTFOLIO SIZE

The basis of this approach is the hypothesis that patents that are the property of companies with a large patent portfolio will tend to have more forward citations. This hypothesis is focused on the idea of “self-citation.” A self-citation in this context is simply a citation by one patent of another patent that belongs to the same company. Some research indicates that self-citation may increase a forward citation count because it may be used as a form of advertisement (Johnson and Mareve, 2002). It may also be the case that some self-citation occurs because the owners of patent want to cause an illusion of value for their patents. Either way, excessive self-citation would increase forward citation arbitrarily. If this is true, then it would be difficult to justify the claim that forward citations increase the value of a patent.

3.1 Methodology

Multiple regression was used to determine the strength of the relationship between forward citations and that patent’s portfolio, but first it was necessary to build a dataset. The first step in this process was company selection. Two metrics were used to select companies for the analysis: size, based on current market value, and patent portfolio size. The information for these two metrics came from two different sources: the initial list of 3000 companies with patent portfolio information came from DOORS™, a database of patent information provided by M-CAM®. The financial information on these companies came from S&P’s Research Insight Compustat Database. With this information compiled it was possible to choose a cross-section of companies based on the two metrics. The final dataset included 112 companies totaling 29,785 patents.

For each patent, the dataset contained the following information: forward citations, self-forward-citations, backward citations, backward-self-citations, age, portfolio size, value of company, and “followers.” The “followers” value represents the number of patents owned by the same company as the given patent that have been filed after the given patent. This value combines information from the patent portfolio and age fields, and it is illustrated in Figure 2.

To do this, all patent information for each company was compiled and summarized in another dataset. This dataset contained the following information for each company: average forward citations, average forward-self-citations, average backward citations, average self-backward citations, patent portfolio size, and value.

Again, multiple regression was performed with average forward-self-citations as the response variable. The
results of this model complete second-order model are summarized in Figure 3 below.

<table>
<thead>
<tr>
<th>Multiple R-Squared: 0.7401</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-statistic: 28.76 on 10 and 101 degrees of freedom, the p-value is 0</td>
</tr>
</tbody>
</table>

Figure 4: Results of the modified model

Using this model, over 70% of the variance in average forward-self-citations is explained by the response variable. This model is also statistically significant according to an F-test.

3.3 Implications

The results of the patent data model do not explain enough of the variance of forward citations to support any conclusions concerning forward citations and value. The statistical significance of the model indicates that there is potential for the model to be predictive, but more information needs to be included. One term that was not included in this analysis is the classification codes of patents. They were excluded due to the subjective nature with which they are assigned to patents. However, if is possible to determine the field of study a patent covers, this term may provide further insight into understanding forward citations.

The second half of this analysis indicates a considerably stronger relationship. Known information concerning a company and its patents allows for prediction of self-citations for this company. This information can allow for a general understanding of a patent based on the characteristics of its owner. This can be of great use when building predictive models, since it provides knowledge concerning the citation habits of companies.

4 EQUAAL CITATION THEORY

If we believe that the number of forward citations indicates the relative value of two patents, then we believe that two patents with the same number of forward citations are of equal worth. When we consider patents observed over two or more generations, this primary assumption of patent citation counts becomes suspect (Fleming 2003). We questioned this assumption by developing a theory of Unequal Citations.

When multiple generations of forward citation counts are considered, patents that are presumed to be of equal value when only direct, or single generation, forward citation counts are used, become less identical. One patent may have significantly more forward citations in its second generation than another. In such a case, the patents that cite one subject patent may have more forward citations than the patents that cite another subject patent. This is problematic because the single generation citation counts advo-

cated by patent citation analysis still consider two such patents equally valuable. Essentially, then, the single generation forward citation count method is in a state of contradiction. The method dictates that more forward citations constitute greater patent value while ignoring distinctions in citation relationships beyond one generation.

If the Unequal Citation Theory is applicable, we must conclude that single generation forward citation counts alone cannot be used as accurate indicators of relative patent value because all forward citations are not equally valuable. We believe that the fundamental problem with using citation counts as patent value indicators is that citation counts fail to account for the multigenerational nature of patent citations.

4.1 Methodology – Patent Citation Networks

In order to test the validity of the Unequal Citation Theory, we sought to determine whether or not a correlation exists between the number of first and second generation citations. To facilitate the process of obtaining data on second generation citation relationships, we developed the concept of a Patent Citation Network (PCN).

A Patent Citation Network (PCN) is a method of visualizing the citation relationships between various patents at a particular time as a network. In a PCN, patents are represented as nodes and citations are represented as directed edges. An edge exists from node $i$ to node $j$ if patent $i$ cites patent $j$. An example citation network is illustrated in Figure 5.

Figure 5: A Patent Citation Network

One of the advantages of a PCN is that it can account for multigenerational relationships between patents. A network path is a set of edges that connect two patents. In the Figure X above, a multiple paths exist from patent A to patent Z. For example, A is connected to Z via a path made of the edges connecting A to B, B to C, and C to Z. A path of length $n$ indicates the existence of an $n$th generation citation relationship. The expediency and ease of counting paths within a network was improved by applying the concept of adjacency matrices (Thomas 2003).
To test our hypothesis we compiled a list of 40 patents from various fields of study. For each of these patents, we calculated the number of first and second generation forward citations. We then used linear regression to determine whether there exists a correlation between first and second generation forward citation counts.

4.2 Results

The scatter plot of first and second generation citation counts, shown in Figure 6, below, seems to indicate a linear relationship between the two counts. The results of linear regression, summarized in Figure 7, below, confirmed that there exists a statistically significant relationship between the number of first and second generation forward citations.

The exact values of $\alpha$ and $\beta$ are not important in this analysis. What is important is that the value of $\beta$ is not zero. Moreover, the p-value associated with $\beta$ indicates that $\beta$ is not equal to zero at all significance levels. Thus, a significant relationship exists between the first and second generation forward citation counts. The R-Squared value of .82 indicates that this model accounts for over 80% of the variance, signifying a model strength which permits conclusions to be drawn from the observed relationship.

Even so, the Unequal Citation Theory should not be dismissed based upon this analysis alone. Our data set was comprised of 40 patents from varying fields. As such, our results may hold true on average, but not for all fields. Further analysis should be performed on data sets comprised of patents that are all a part of the same industry. Secondly, our analysis only considered the relationship between first and second generation citations; future analyses should consider higher generation citations as well. Finally, in the current implementation, patent citation networks are bounded by a methodology that includes the filtering of irrelevant citation relationships. This bound on the network was included to improve the efficiency of the path counting tool by restricting citation networks to a manageable size. However, by restricting the network first or second generation citations that could impact the analysis may have been removed.

5 CONCLUSION

The purpose of the research presented in this paper was to investigate the validity of patent citation analysis; specifically to evaluate the use of forward citation counts as a measure of relative patent value. In pursuit of this objective, we considered three conditions which, if true, would indicate that forward citation counts are not an acceptable measure of relative patent value.

The first approach taken examined the relationship between the number of forward citations attributable to an individual patent and the average number of forward citations in its innovation space. We hypothesized that if a positive correlation between these two variables exists and forward citations indicate value then it must be true that the value of a patent is positively correlated with the value of its innovation space. We hypothesized that if a positive correlation between these two variables exists and forward citations indicate value then it must be true that the value of a patent is positively correlated with the value of its innovation space. This conclusion, however, cannot be true, so either forward citations do not accurately reflect value or the number of forward citations of an individual patent cannot be positively correlated to the number of generations, the results of this analysis seem to indicate that the theory is invalid.

<table>
<thead>
<tr>
<th>*** Linear Model ***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation: 2ndGen = $\alpha + \beta$*1stGen</td>
</tr>
<tr>
<td>Coefficients: Value p-value</td>
</tr>
<tr>
<td>$\alpha$ -43.1513006 0.0556338</td>
</tr>
<tr>
<td>$\beta$ 12.0375723 0.0000000</td>
</tr>
<tr>
<td>Multiple R-Squared: 0.8210853</td>
</tr>
<tr>
<td>F-statistic: 174.3917 on 1 and 38 degrees of freedom, the p-value is 8.881784e-016</td>
</tr>
</tbody>
</table>

Figure 7: Summary of Linear Regression Results

The existence of a significant relationship between first and second generation citation counts indicates that the Unequal Citation Theory may not constitute a valid challenge to current patent citation analysis methods. The observed correlation reveals that two patents with equal numbers of forward citations will often have approximately equal numbers of second generation forward citations. As the Unequal Citation Theory assumes that no significant relationship exists between citation counts of different
forward citations for its innovation space. By using empirical evidence, we showed that a positive correlation does indeed exist between a patent’s forward citations and the forward citations of its innovation space. We conclude that according to this proof by contradiction, forward citation counts cannot be an accurate reflection of patent value.

The second approach taken focused on the relationship between the number of forward citations attributable to a patent, and the size its patent portfolio. If a relationship could be shown to exist, this would suggest that forward citation counts are not independent of company size and that they therefore cannot objectively measure patent value. No such relationship could be found among 112 patent portfolios comprised of 29,785 total patents. The results of this experiment were therefore inconclusive in evaluating patent citation counts.

In the final approach, we examined the relationship between first and second generation forward citations. We theorized that if no such relationship existed, then patents understood to be of equal value when only one generation of citations is considered would appear to have different values when further generations of patent citations are considered. Essentially, if no such relationship existed, forward citation counts of only one generation would be an incomplete and therefore unacceptable measure of patent value. A correlation was shown to exist between the first and second generations of citations for 40 randomly selected patents. While the data set should be expanded before this theory is dismissed entirely, the preliminary results are inconclusive in terms of evaluating patent citation counts.

While two of our approaches were inconclusive, our analysis of the relationship between a patent’s forward citation count and the average forward citation count for the patents in its innovation space demonstrated that the use of forward citations as a valuation method may be problematic. This finding could be strengthened by further development of all three approaches.

ACKNOWLEDGEMENTS

We would like to acknowledge the contributions of M•CAM®, Inc.

REFERENCES


AUTHOR BIOGRAPHIES

MICHAEL W. BRINN is a fourth-year Systems and Information Engineering major from Virginia Beach, VA concentrating in Management Systems. When not working diligently on this project, Mike was preparing for the World Chair Racing Championship by honing his chair racing skills in Olsson Hall.

JULIA M. FLEMING is a fourth-year Systems and Information Engineering and Economics major from Annandale, VA concentrating in Economic Systems. After a thorough analysis of the factors contributing to patent value, Julia is pleased to announce that a synonym for “weird” is “unique.”

FERNANDO M. HANNAKA is a fourth-year Systems and Information Engineering major from Sao Paulo, Brazil concentrating in Computer Information Systems. Learning about patent analysis enhanced his understanding of the possibilities of inventions. This tropical student plans to one day create the most valuable patent in the world, called “Warm snow.”

COLIN B. THOMAS is a fourth-year Systems and Information Engineering major from Stafford, VA concentrating in Management Systems. During the course of this project, Colin has gone gaga over patents and has decided to pursue patent analysis as a life goal.