StockRank: An algorithm for automated ranking and selection of stocks in equity markets

Saurabh V. Pendse*[1]

[1] Department of Computer Science and Engineering
The Maharaja Sayajirao University of Baroda, Gujarat, India

July 7, 2010

Abstract

Google Inc. uses its PageRank algorithm to sort internet search queries by their page ranks. This makes the search results more relevant and reliable. In this work, we propose a novel framework which we call StockRank where we apply Google's PageRank algorithm to assign ranks to publicly traded stocks in equity markets. These stock ranks could be useful in identifying key market drivers at any given point of time.

1 Introduction

The name Google has become synonymous with an internet search on the world wide web (WWW). The enormous success of Google Inc. in the search

*To whom correspo dence should be addressed. e-mail: sau2pen@gmail.com

engine business can be attributed to its patented PageRank algorithm [2]. The PageRank algorithm assigns a rank or a relative measure of importance to each web page on the WWW. Once page ranks have been assigned to all web pages, Google displays the results of a user specified search query such that web pages with higher page ranks appear at the top of the results. Because of the technology underlying the PageRank algorithm, this ordering makes the search results much more relevant and trustworthy. Thus the PageRank algorithm acts as a filter that extracts the most important information from the WWW for any user defined query.

In this work, our goal is to adapt Google's PageRank algorithm to assign ranks to stocks in equity markets. We call the resulting algorithm "StockRank". Motivation for this work is the assumption that under any market conditions, the overall behavior or pulse of the market is captured by only a few selected stocks. For instance, a key financial decision by an influential company could dictate the overall stock market behavior whereas another company that is not as influential might not have a significant impact on the overall stock market under a similar situation. Influential stocks such as Google Inc. or Apple Inc. are well known. How can we automatically detect the most influential stocks in the stock market? We also postulate that the degree of importance of each stock is not static but fluctuates as a function of time. Can we track how the relative importance of each stock changes over time in an objective fashion? Investing in stock markets consists of two distinct steps: stock selection and an investment strategy. In this work we address the problem of stock selection using our StockRank algorithm:

1. First, we describe in detail Google's ranking technology.

2. Next, we describe our StockRank algorithm for ranking and selection of stocks. We will show how StockRank assigns an objective measure of time varying importance to each stock in any equity market.

2 Notation

Vectors and matrices will be denoted in bold face possibly with subscripts (e.g. $\boldsymbol{x_1}$, $\boldsymbol{B_1}$). Upper case bold face letters possibly with subscripts (e.g. $\boldsymbol{S_i}$) will also be used to denote objects such as stocks in equity markets. Given a matrix $\boldsymbol{A}$, its transpose will be denoted by $\boldsymbol{A}^T$. The ijth element of a matrix $\boldsymbol{A}$ will be denoted by A_{ij} while the jth element of a vector $\boldsymbol{x}$ will be denoted by x_j. We also use an alternative notation $x(j)$ to indicate the jth element of $\boldsymbol{x}$ when necessary for improving clarity. A vector or matrix of all zeros will be indicated by $\mathbf{0}$. The 2-norm of a vector $\boldsymbol{x}$ will be denoted by $||\boldsymbol{x}||_2 = +\sqrt{\boldsymbol{x}^T\boldsymbol{x}}$. We will use $|x_i|$ to denote the absolute value of x_i. The element-wise absolute value of a vector $\boldsymbol{x}$ will be denoted by $|\boldsymbol{x}|$. For a complex real number $z = a + ib$ where $i = \sqrt{-1}$, we use the absolute value notation to mean $|z| = \sqrt{a^2 + b^2}$.

3 Google's ranking technology

Google's ranking technology consists of two components:

- The ranking methodology, which can be applied to assign ranks to a set of objects.
- A importance measure, which assigns the importance to object i as measured by object j.

Changing the importance measure, leads to a family of ranking algorithms. PageRank is a particular member of this family of ranking algorithms that uses an importance measure suited to assigning ranks to web pages on the WWW. Our proposed StockRank algorithm is also a member of this family of ranking algorithms that uses an importance measure suited to assigning ranks to stocks in equtiy markets.

3.1 Ranking methodology

Suppose we want to assign a measure of relative importance to each of the n objects $\boldsymbol{O}_1, \boldsymbol{O_2}, \ldots, \boldsymbol{O_n}$. Suppose each object $\boldsymbol{O_j}$ "votes" for every other object $\boldsymbol{O_i}$. This "vote" can be simply thought of as the importance of object $\boldsymbol{O_i}$ "in the eyes" of object $\boldsymbol{O_j}$. Let us denote the "vote" by object $\boldsymbol{O_j}$ for object $\boldsymbol{O_i}$ by V_{ij}. Given this information, what is the overall importance of object $\boldsymbol{O_i}$?

3.1.1 Simple vote counting

One simple strategy is to simply count all "votes" for object $\boldsymbol{O_i}$ by all other objects. Suppose x_i is the relative importance assigned to object $\boldsymbol{O_i}$ using this technique. Then we have:

$$x_i \propto \sum_{j=1}^{n} V_{ij} \tag{3.1}$$

In this technique, all voting objects are given equal weight in calculating the relative importance measure x_i for $\boldsymbol{O_i}$.

3.1.2 Weighted vote counting

This is also a vote counting technique. But in this case, V_{ij}, the vote for $\boldsymbol{O_i}$ by $\boldsymbol{O_j}$ is weighted by the relative importance measure of $\boldsymbol{O_j}$. Thus the effective contribution of $\boldsymbol{O_j}$ to the total vote count for $\boldsymbol{O_i}$ is $V_{ij}\, x_j$. The total weighted vote count for $\boldsymbol{O_i}$ is proportional to its relative importance x_i. Hence we can write:

$$x_i \propto \sum_{i=1}^{n} V_{ij} x_j \tag{3.2}$$

Introducing a proportionality constant λ on the left hand side and re-writing 3.2 in vector form we get:

$$\lambda \boldsymbol{x} = \boldsymbol{V} \boldsymbol{x} \tag{3.3}$$

It is clear from 3.3 that $\boldsymbol{x}$ is an eigenvector of the matrix $\boldsymbol{V}$ with eigenvalue λ. The rank of the ith object $\boldsymbol{O_i}$ is simply the ith element x_i of vector $\boldsymbol{x}$. Ideally, we would like the computed page ranks to be positive and unique. Under what conditions can we select a particular eigenvector $\boldsymbol{x}$ of $\boldsymbol{V}$ that satisfies these properties?

3.2 Uniqueness and positivity of ranks

In this section, we show that there are certain mild conditions under which the computed ranks are unique and positive. In brief, if the elements of matrix $\boldsymbol{V}$ are strictly positive then the Perron-Frobenius theorem guarantees the uniqueness and positivity of ranks. We refer the reader to [1] for further details on this and related topics. We give below the statement and proof of the Perron-Frobenius theorem.

Theorem 3.1. *Perron-Frobenius Theorem*
Consider an $n \times n$ matrix $\boldsymbol{V}$. If all elements of $\boldsymbol{V}$ satisfy $V_{ij} > 0$ then the following statements are true:

1. *$\boldsymbol{V}$ has a real eigenvalue $\lambda > 0$ such that the corresponding eigenvector $\boldsymbol{x}$ has all strictly positive elements, i.e. $x_i > 0$.*

2. *If ψ is another eigenvalue of $\boldsymbol{V}$ with eigenvector $\boldsymbol{h}$ distinct from $\boldsymbol{x}$ then $\lambda \geq |\psi|$.*

3. *If $\boldsymbol{y}$ is an eigenvector of $\boldsymbol{V}$ with eigenvalue λ, then $\boldsymbol{y} = k\,\boldsymbol{x}$ with $k \in \mathbf{R}$. In other words, the eigenvector corresponding to the largest eigenvalue λ is unique.*

Proof. Suppose $\boldsymbol{h}$ is a possibly complex eigenvector of $\boldsymbol{V}$ with possibly complex eigenvalue ψ. Equating the ith elements on both sides:

$$\sum_{j=1}^{n} V_{ij}\, h_j = \psi\, h_i \tag{3.4}$$

Taking absolute values on both sides and using the triangle inequality we get:

$$\sum_{j=1}^{n} |V_{ij}|\, |h_j| \geq |\sum_{j=1}^{n} V_{ij}\, h_j| = |\psi|\, |h_i| \tag{3.5}$$

Since each element $V_{ij} > 0$, we have $|V_{ij}| = V_{ij}$ and hence we can write 3.5 as:

$$\sum_{j=1}^{n} V_{ij}\, |h_j| \geq |\psi|\, |h_i| \tag{3.6}$$

In matrix form, we can write:

$$\boldsymbol{V}\, |\boldsymbol{h}| \geq |\psi|\, |\boldsymbol{h}| \tag{3.7}$$

Therefore, for each eigenvalue ψ, there exists a non-negative real number $|\psi|$, and a non-negative vector $|\boldsymbol{h}|$ such that 3.7 is true. Motivated by this, suppose we solve the following optimization problem:

$$\max_{\lambda \in \mathbf{R}} \;\; \lambda \tag{3.8}$$

$$\text{with} \;\; \boldsymbol{V}\boldsymbol{x} \geq \lambda\, \boldsymbol{x}, \; \text{for some } \boldsymbol{x} \geq \boldsymbol{0} \text{ and } \boldsymbol{x} \neq \boldsymbol{0} \tag{3.9}$$

The problem is clearly feasible. For instance, the absolute value of the largest magnitude eigenvalue is a candidate solution. Suppose λ is the solution to 3.8 with $\boldsymbol{x} \geq 0$ satisfying $\boldsymbol{V}\boldsymbol{x} \geq \lambda\, \boldsymbol{x}$. We will try to show that $\boldsymbol{x}$ must in fact satisfy $\boldsymbol{V}\boldsymbol{x} = \lambda\,\boldsymbol{x}$. Suppose this is not the case. Then there must be at least one element i for which:

$$\sum_{j=1}^{n} V_{ij} x_j = \lambda\, x_i + \delta, \text{ where } \delta > 0 \tag{3.10}$$

while for all other elements $k \neq i$:

$$\sum_{j=1}^{n} V_{kj} x_j \geq \lambda\, x_k, \text{ where } k \neq i \tag{3.11}$$

Now consider another vector $\boldsymbol{z}$ with elements:

$$z_i = x_i + \theta, \text{ where } \theta > 0 \tag{3.12}$$

$$z_k = x_k, \text{ if } k \neq i \tag{3.13}$$

Therefore for $k \neq i$ we have from 3.12 that:

$$\sum_{j=1}^{n} V_{kj} z_j = \sum_{j=1}^{n} V_{kj} x_j + V_{ki}\,\theta \tag{3.14}$$

Since each element $V_{ki} > 0$ and $\theta > 0$, we get from 3.11 and 3.14 that:

$$\sum_{j=1}^{n} V_{kj} z_j > \lambda\, x_k, \text{ where } k \neq i \tag{3.15}$$

From 3.15 and 3.12 we can re-write the above equation as:

$$\sum_{j=1}^{n} V_{kj} z_j > \lambda\, z_k, \text{ where } k \neq i \tag{3.16}$$

Similarly

$$\sum_{j=1}^{n} V_{ij} z_j = \sum_{j=1}^{n} V_{ij} x_j + V_{ii}\,\theta \tag{3.17}$$

From 3.10 and 3.17 we get:

$$\sum_{j=1}^{n} V_{ij} z_j = \lambda\, x_i + \delta + V_{ii}\,\theta \tag{3.18}$$

Substituting x_i in terms of z_i from 3.12 we have:

$$\sum_{j=1}^{n} V_{ij} z_j = \lambda\,(z_i - \theta) + \delta + V_{ii}\,\theta \tag{3.19}$$

Rearranging

$$\sum_{j=1}^{n} V_{ij} z_j = \lambda\, z_i + \delta + (V_{ii} - \lambda)\,\theta \tag{3.20}$$

We can choose a sufficiently small θ such that $\delta + (V_{ii} - \lambda)\,\theta > 0$ and consequently we get:

$$\sum_{j=1}^{n} V_{ij} z_j > \lambda\, z_i \tag{3.21}$$

From 3.16 and 3.21 it is possible to find a positive vector $\boldsymbol{g} > 0$ such that:

$$\boldsymbol{V}\boldsymbol{z} = \lambda\,\boldsymbol{z} + \boldsymbol{g} \tag{3.22}$$

Choose ϕ such that:

$$\phi \leq \frac{g_i}{z_i} \Rightarrow \boldsymbol{g} \geq \phi \boldsymbol{z} \tag{3.23}$$

Since $g_i > 0$ and $z_i \geq 0$, we must have $\phi > 0$. Consequently 3.22 and 3.23 give us:

$$\boldsymbol{V}\boldsymbol{z} \geq (\lambda + \phi)\boldsymbol{z} \tag{3.24}$$

Thus we have found a vector $\boldsymbol{z} \geq 0$ such that $\boldsymbol{V}\boldsymbol{z} \geq (\lambda + \phi)\boldsymbol{z}$ with $(\lambda + \phi) > \lambda$ since $\phi > 0$. Thus λ cannot be the solution to the optimization problem 3.8 which is a contradiction. Hence 3.10 cannot hold for any element i. Since $\boldsymbol{V}\boldsymbol{x} \geq \lambda\,\boldsymbol{x}$, we must have:

$$\boxed{\boldsymbol{V}\boldsymbol{x} = \lambda\,\boldsymbol{x}} \text{ with } \boldsymbol{x} \geq \boldsymbol{0} \text{ and } \boldsymbol{x} \neq \boldsymbol{0} \tag{3.25}$$

In other words, λ is an eigenvalue of $\boldsymbol{V}$ with eigenvector $\boldsymbol{x}$. From 3.7 for every eigenvalue ψ of $\boldsymbol{V}$, there exists $|\psi| \geq 0$ and $|\boldsymbol{h}| \geq 0$ such that the feasibility condition in 3.8 is satisfied. Since λ maximizes the objective function in 3.8, we must have:

$$\boxed{\lambda \geq |\psi| \geq 0 \text{ for any eigenvalue } \psi \text{ of } \boldsymbol{V}} \tag{3.26}$$

This proves part 2 of the theorem. Now suppose the ith element on the right hand side of 3.25 is 0 i.e., $\lambda x_i = 0$. However, since $\boldsymbol{x} \geq 0$ and $\boldsymbol{x} \neq 0$ with $V_{ij} > 0, \forall i, j$ the ith element on the left hand side of 3.25 is strictly positive (> 0) and hence we have a contradiction. Therefore $\lambda x_i \neq 0\ \forall i$. Since $\lambda \geq 0$ and $\boldsymbol{x} \geq 0$ this implies:

$$\boxed{\lambda > 0 \text{ (strictly positive eigenvalue)}} \tag{3.27}$$

$$\boxed{\boldsymbol{x} > 0 \text{ (strictly positive eigenvector)}} \tag{3.28}$$

This proves part 1 of the theorem. Next, we prove part 3 of the theorem. Suppose there is another eigenvector $\boldsymbol{y}$ of $\boldsymbol{V}$ distinct from $\boldsymbol{x}$ with eigenvalue λ. This means $\boldsymbol{y} \neq \boldsymbol{0}$ and $\boldsymbol{y} \neq k\boldsymbol{x}$ for any $k \in \mathbf{R}$. Thus

$$\boldsymbol{V}\boldsymbol{y} = \lambda\,\boldsymbol{y} \tag{3.29}$$

From 3.29 and 3.25 we can write for any $\beta \in \mathbf{R}$ that

$$\boldsymbol{V}(\boldsymbol{x} - \beta\boldsymbol{y}) = \lambda\,(\boldsymbol{x} - \beta\boldsymbol{y}) \tag{3.30}$$

Thus $(\boldsymbol{x} - \beta\boldsymbol{y})$ is also an eigenvector of $\boldsymbol{V}$. Hence using 3.7 we can write:

$$\boldsymbol{V}|(\boldsymbol{x} - \beta\boldsymbol{y})| \geq \lambda\,|(\boldsymbol{x} - \beta\boldsymbol{y})| \tag{3.31}$$

Note that since by assumption $\boldsymbol{y} \neq k\boldsymbol{x}$ for all $k \in \mathbf{R}$, we have $|(\boldsymbol{x} - \beta\boldsymbol{y})| \geq 0$ and $|(\boldsymbol{x} - \beta\boldsymbol{y})| \neq 0$. Using arguments similar to those leading to 3.25 from 3.10 we can show that the inequality in 3.31 must in fact be an equality given that λ is the solution to 3.8:

$$\boldsymbol{V}|(\boldsymbol{x} - \beta\boldsymbol{y})| = \lambda\,|(\boldsymbol{x} - \beta\boldsymbol{y})| \tag{3.32}$$

Equation 3.32 holds for any choice of β. In particular, it also holds for a choice that makes the first element of the vector on the RHS of 3.32 0. Suppose we choose β such that $x_1 - \beta y_1 = 0$. Then the first element on the RHS will be 0. Since $|(\boldsymbol{x} - \beta\boldsymbol{y})| \geq 0$ and $|(\boldsymbol{x} - \beta\boldsymbol{y})| \neq 0$ with $V_{ij} > 0\ \forall\, i, j$ the first element on the LHS of 3.32 $\sum_{j=1}^{n} V_{1j}|x_j - \beta y_j| > 0$ and hence we have a contradiction. This means that our assumption that $\boldsymbol{y} \neq k\boldsymbol{x}$ for any $k \in \mathbf{R}$ must be false. Thus we must have

$$\boxed{\boldsymbol{y} = k\,\boldsymbol{x} \text{ for some } k \in \mathbf{R}} \tag{3.33}$$

But this means that $\boldsymbol{y}$ is not an eigenvector of $\boldsymbol{V}$ that is distinct from $\boldsymbol{x}$. In fact, $\boldsymbol{y}$ can be obtained simply as a scalar multiple of $\boldsymbol{x}$. Hence the eigenvector corresponding to the largest and real eigenvalue λ is unique. This proves part 3 of the theorem. □ □

Corollary 3.2. *If $\boldsymbol{V}$ is a symmetric $n \times n$ matrix with all positive elements $\boldsymbol{V} > \mathbf{0}$ then the largest eigenvalue $\lambda > 0$ with eigenvector $\boldsymbol{x} > \mathbf{0}, \boldsymbol{x} \neq \mathbf{0}$ satisfies $\lambda > |\psi|$ where ψ is another eigenvalue of $\boldsymbol{V}$ with eigenvector $\boldsymbol{h}$ distinct from $\boldsymbol{x}$.*

Proof. The fact that $\lambda \geq |\psi|$ as well as $\lambda > 0$ with $\boldsymbol{x} > \mathbf{0}$ follows from the Perron-Frobenius theorem for non-symmetric matrices 3.1. If $|\psi| < \lambda$ there is nothing to prove. So suppose $|\psi| = \lambda$. Since $\boldsymbol{V}$ is symmetric, all its eigenvalues are real numbers and so the only way $|\psi| = \lambda$ is if $\psi = \lambda$ or $\psi = -\lambda$. Because $\boldsymbol{V}$ has real elements and the eigenvalues of $\boldsymbol{V}$ are real the eigenvector $\boldsymbol{h}$ for eigenvalue ψ also has real elements.

Case 1: $\psi = \lambda$: If $\psi = \lambda$ with eigenvector $\boldsymbol{h}$ then from the Perron-Frobenius theorem (part 3) 3.1 it follows that $\boldsymbol{h} = k\,\boldsymbol{x}, k \in \mathbf{R}$ where $\boldsymbol{h}$ is the eigenvector for ψ. Since by assumption $\boldsymbol{h}$ is an eigenvector distinct from $\boldsymbol{x}$ we have a contradiction. Therefore $\psi \neq \lambda$.

Case 2: $\psi = -\lambda$: Suppose $\psi = -\lambda$ with eigenvector $\boldsymbol{h}$. Noting that in this case $|\psi| = \lambda > 0$, then from 3.7 we get:

$$\boldsymbol{V}|\boldsymbol{h}| \geq \lambda\,|\boldsymbol{h}| \tag{3.34}$$

Since $|\boldsymbol{h}| \geq 0$ using a similar logic to that used after 3.8 we see that the above inequality must in fact be an equality. That is

$$\boldsymbol{V}|\boldsymbol{h}| = \lambda\,|\boldsymbol{h}| \tag{3.35}$$

Since $\boldsymbol{h}$ is an eigenvector with eigenvalue $\psi = -\lambda$ we also have

$$\boldsymbol{V}\boldsymbol{h} = -\lambda\,\boldsymbol{h} \tag{3.36}$$

Adding equations 3.35 and 3.36 we get

$$\boldsymbol{V}(\boldsymbol{h} + |\boldsymbol{h}|) = \lambda\,(|\boldsymbol{h}| - \boldsymbol{h}) \tag{3.37}$$

Now consider the following 3 cases:

Case A: $\boldsymbol{h}$ is non-negative: In this case $\boldsymbol{h} \geq \mathbf{0}$ and so $|\boldsymbol{h}| = \boldsymbol{h}$. Now $\boldsymbol{V}$ has all positive elements and $\boldsymbol{h} \neq \mathbf{0}$ since it is an eigenvector. Thus $|\boldsymbol{h}| \neq \mathbf{0}$ and $\boldsymbol{V}(\boldsymbol{h} + |\boldsymbol{h}|) = 2\boldsymbol{V}|\boldsymbol{h}| > \mathbf{0}$. Therefore the left hand side of 3.37 becomes $2\boldsymbol{V}|\boldsymbol{h}| > \mathbf{0}$ while the right hand side of 3.37 becomes $\lambda(|\boldsymbol{h}| - \boldsymbol{h}) = \mathbf{0}$ which is a contradiction.

Case B: $\boldsymbol{h}$ is non-positive: In this case $\boldsymbol{h} \leq \mathbf{0}$ and so $|\boldsymbol{h}| = -\boldsymbol{h}$. Thus $\lambda(|\boldsymbol{h}| - \boldsymbol{h}) = 2\lambda|\boldsymbol{h}|$. Since $\boldsymbol{h}$ is an eigenvector $\boldsymbol{h} \neq \mathbf{0}$ and so $|\boldsymbol{h}| \neq \mathbf{0}$. However, since $|\boldsymbol{h}| \geq \mathbf{0}$ there is an index i such that $|h_i| > 0$. In addition, $\lambda > 0$ from Perron-Frobenius theorem 3.1 part 1. Therefore the ith element on the right hand side of 3.37 is $2\lambda|h_i| > 0$ while every element including the ith element on the left hand side of 3.37 is 0 since $(\boldsymbol{h} + |\boldsymbol{h}|) = \mathbf{0}$. Thus we have a contradiction.

Case C: $\boldsymbol{h}$ has at least 1 positive and 1 negative element: In this case $(|\boldsymbol{h}| - \boldsymbol{h}) \neq \mathbf{0}$ and $(\boldsymbol{h} + |\boldsymbol{h}|) \neq \mathbf{0}$. Note that since $\boldsymbol{h}$ has real elements $(\boldsymbol{h} + |\boldsymbol{h}|) \geq \mathbf{0}$. Suppose i is an index for which $h_i > 0$. For this index i we have $|h_i| - h_i = 0$. Since $\boldsymbol{h} \neq \mathbf{0}$, then $(\boldsymbol{h} + |\boldsymbol{h}|) \neq \mathbf{0}$. Given that $\boldsymbol{V}$ has all positive elements, all elements on the left hand side of 3.37 including the ith element are > 0. However, the ith element on the right hand side of 3.37 is $\lambda(|h_i| - h_i) = 0$ which is a contradiction.

Therefore, when $\psi = -\lambda$ then no matter what the structure of the corresponding eigenvector $\boldsymbol{h}$ is we get a contradiction. Hence $\psi \neq -\lambda$.

Combining the results of Case 1 and Case 2, we can write $\lambda \geq |\psi|$ with $\lambda \neq |\psi|$ and therefore we must have $\lambda > |\psi|$ which proves the corollary. □ □

3.3 Rank computation and power iteration

In this section, we describe how to compute the eigenvector $\boldsymbol{x}$ corresponding to the largest eigenvalue λ.

Theorem 3.3. *Suppose $\boldsymbol{V}$ is a symmetric $n \times n$ matrix with all positive elements $\boldsymbol{V} > \mathbf{0}$. Let $\lambda > 0$ be the largest eigenvalue of $\boldsymbol{V}$ with eigenvector $\boldsymbol{x} > \mathbf{0}$. Then the unit norm eigenvector $\boldsymbol{x_N} = \frac{\boldsymbol{x}}{||\boldsymbol{x}||_2}$ corresponding to λ can be computed as follows:*

1. *Choose any vector $\boldsymbol{x_0}$ such that $\boldsymbol{x}^T\boldsymbol{x_0} \neq 0$.*

2. *Compute* $\boldsymbol{g_m} = \frac{\boldsymbol{V}^m \boldsymbol{x_0}}{||\boldsymbol{V}^m \boldsymbol{x_0}||_2}$ *for increasing values of* m.

3. *As* $m \to \infty$, $\boldsymbol{g_m} \to \pm \boldsymbol{x_N}$.

Proof. Since $\boldsymbol{V} > \mathbf{0}$ then by Corollary 3.2 there exists a real eigenvalue $\lambda > 0$ with a real eigenvector $\boldsymbol{x} > 0$. Suppose $\psi_1, \psi_2, \ldots, \psi_{n-1}$ are the other $(n-1)$ eigenvalues of $\boldsymbol{V}$ different from λ and let $\boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_{n-1}}$ be the corresponding eigenvectors. Now $\lambda > |\psi_i|$ for $i = 1, 2, \ldots, (n-1)$ from Corollary 3.2. Since $\boldsymbol{V}$ is symmetric, the set of n eigenvectors $\boldsymbol{x}, \boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_{n-1}}$ form an orthogonal basis in $\mathbf{R}^n$, i.e., $\boldsymbol{x}^T \boldsymbol{h_i} = 0, i = 1, 2, \ldots, (n-1)$ and $\boldsymbol{h_i}^T \boldsymbol{h_j} = 0$ if $i \neq j$. Hence any vector $\boldsymbol{x_0}$ in $\mathbf{R}^n$ can be written as a linear combination of the n orthogonal vectors $\boldsymbol{x}, \boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_{n-1}}$:

$$\boldsymbol{x_0} = c\,\boldsymbol{x} + \sum_{i=1}^{n-1} d_i\,\boldsymbol{h_i} \tag{3.38}$$

Pre multiplying both sides of 3.38 by $\boldsymbol{x}^T$ and using the orthogonality property of eigenvectors we get:

$$\boldsymbol{x}^T \boldsymbol{x_0} = c\,\boldsymbol{x}^T \boldsymbol{x} \tag{3.39}$$

By condition 1, in Theorem 3.3, $\boldsymbol{x}^T \boldsymbol{x_0} \neq 0$ and so from 3.39 we get:

$$c \neq 0 \tag{3.40}$$

Since λ is an eigenvalue with eigenvector $\boldsymbol{x}$, $\boldsymbol{V}\boldsymbol{x} = \lambda \boldsymbol{x}$. Now $\boldsymbol{V}^2 \boldsymbol{x} = \boldsymbol{V}\boldsymbol{V}\boldsymbol{x} = \boldsymbol{V}\lambda\boldsymbol{x} = \lambda\boldsymbol{V}\boldsymbol{x} = \lambda\lambda\boldsymbol{x} = \lambda^2\boldsymbol{x}$. Similarly for each eigenvalue ψ_i, $\boldsymbol{V}^2 \boldsymbol{h_i} = \psi_i^2 \boldsymbol{h_i}$. Using the same logic, for any integer m we can write:

$$\boldsymbol{V}^m \boldsymbol{x} = \lambda^m \boldsymbol{x} \tag{3.41}$$

$$\boldsymbol{V}^m \boldsymbol{h_i} = \psi_i^m \boldsymbol{h_i} \tag{3.42}$$

Pre multiplying both sides of 3.38 by $\boldsymbol{V}^m$ we get:

$$\boldsymbol{V}^m \boldsymbol{x_0} = c\,\boldsymbol{V}^m \boldsymbol{x} + \sum_{i=1}^{n-1} d_i\,\boldsymbol{V}^m \boldsymbol{h_i} \tag{3.43}$$

From 3.41 and 3.43 we get:

$$\boldsymbol{V}^m \boldsymbol{x_0} = c\,\lambda^m \boldsymbol{x} + \sum_{i=1}^{n-1} d_i\,\psi_i^m \boldsymbol{h_i} \tag{3.44}$$

Taking the squared norm on both sides of 3.44 and noting that $\boldsymbol{x}, \boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_{n-1}}$ are orthogonal to each other we get:

$$||\boldsymbol{V}^m \boldsymbol{x_0}||_2^2 = (c^2\, \lambda^{2m})||\boldsymbol{x}||_2^2 + \sum_{i=1}^{n-1}(d_i^2\, \psi_i^{2m})||\boldsymbol{h_i}||_2^2 \tag{3.45}$$

Therefore $\boldsymbol{g_m}$ as defined in item 3 of Theorem 3.3 is given by:

$$\boldsymbol{g_m} = \frac{\boldsymbol{V}^m \boldsymbol{x_0}}{||\boldsymbol{V}^m \boldsymbol{x_0}||_2} = \frac{c\, \lambda^m \boldsymbol{x} + \sum_{i=1}^{n-1} d_i\, \psi_i^m \boldsymbol{h_i}}{\sqrt{(c^2\, \lambda^{2m})||\boldsymbol{x}||_2^2 + \sum_{i=1}^{n-1}(d_i^2\, \psi_i^{2m})||\boldsymbol{h_i}||_2^2}} \tag{3.46}$$

Now $c \neq 0$ from 3.40 and $\lambda > 0$ from Theorem 3.1. Thus $c\, \lambda^m \neq 0$ for all integers m. Simplifying 3.46 we get:

$$\boldsymbol{g_m} = \frac{\boldsymbol{V}^m \boldsymbol{x_0}}{||\boldsymbol{V}^m \boldsymbol{x_0}||_2} = \frac{c\, \lambda^m}{|c\, \lambda^m|} \frac{\boldsymbol{x} + \sum_{i=1}^{n-1}(d_i/c)\, (\psi_i/\lambda)^m \boldsymbol{h_i}}{\sqrt{||\boldsymbol{x}||_2^2 + \sum_{i=1}^{n-1} \{(d_i/c)^2\, (\psi_i/\lambda)^{2m}\}\, ||\boldsymbol{h_i}||_2^2}} \tag{3.47}$$

Since $\lambda > 0$ we can write 3.47 as

$$\boldsymbol{g_m} = \frac{\boldsymbol{V}^m \boldsymbol{x_0}}{||\boldsymbol{V}^m \boldsymbol{x_0}||_2} = \frac{c}{|c|} \frac{\boldsymbol{x} + \sum_{i=1}^{n-1}(d_i/c)\, (\psi_i/\lambda)^m \boldsymbol{h_i}}{\sqrt{||\boldsymbol{x}||_2^2 + \sum_{i=1}^{n-1} \{(d_i/c)^2\, (\psi_i/\lambda)^{2m}\}\, ||\boldsymbol{h_i}||_2^2}} \tag{3.48}$$

Since $|\psi_i| < \lambda$ for all i we have

$$\left(\frac{\psi_i}{\lambda}\right)^m \to 0 \text{ as } m \to \infty \tag{3.49}$$

Therefore from 3.48 and 3.49 we have

$$\boldsymbol{g_m} = \frac{\boldsymbol{V}^m \boldsymbol{x_0}}{||\boldsymbol{V}^m \boldsymbol{x_0}||_2} = \frac{c}{|c|} \frac{\boldsymbol{x}}{\sqrt{||\boldsymbol{x}||_2^2}} = \frac{c}{|c|} \frac{\boldsymbol{x}}{||\boldsymbol{x}||_2} \tag{3.50}$$

Since $\frac{\boldsymbol{x}}{||\boldsymbol{x}||_2} = \boldsymbol{x_N}$ is the normalized eigenvector for eigenvalue λ we can write 3.50 as

$$\boldsymbol{g_m} = \frac{c}{|c|} \boldsymbol{x_N} \tag{3.51}$$

It is clear from 3.51 that if $c > 0$ then $\boldsymbol{g_m}$ converges to $\boldsymbol{x_N}$ and if $c < 0$ then $\boldsymbol{g_m}$ converges to $-\boldsymbol{x_N}$. This proves the theorem. □

□

The algorithm detailed in Theorem 3.3 is called the power iteration. Although Theorem 3.51 deals with symmetric matrices for simplicity, it can be shown that the power iteration also converges for matrices $\boldsymbol{V}$ with a dominant eigenvalue i.e., when there is one eigenvalue that is larger in magnitude than all other eigenvalues provided that $\boldsymbol{V}$ is diagonalizable. $\boldsymbol{V}$ is diagonalizable if it has a linearly independent basis of eigenvectors. Most matrices are in fact diagonalizable and it can be shown that any matrix with distinct eigenvalues is diagonalizable.

4 The StockRank algorithm

Consider an equity market with n publicly traded stocks. These can include composite indices such as DOW and NASDAQ. In the StockRank algorithm, we are interested in ranking stocks in equity markets. In the previous sections, we have already described Google's ranking methodology as well as the power iteration for computing ranks. Now the question is, how should we measure the votes cast by object j for object i: V_{ij}? In StockRank we want to extract a set of primary stocks which influence the market as a whole. In other words, we would like to capture a set of stocks such that their price fluctuation drives the prices of other stocks. Mathematically, the price fluctuation over time of stocks with high StockRank will be highly correlated with the price fluctuation of many other stocks in the market. Let object $\boldsymbol{O_i}$ represent the ith stock in the marktet. We associate with $\boldsymbol{O_i}$ a vector $\boldsymbol{f_i}^{(t)}$ of length t such that:

$$\boldsymbol{f_i}^{(t)}(r) = \text{price of stock } \boldsymbol{O_i} \text{ on the } r\text{th day in the past}, r = 1 \dots t \tag{4.1}$$

Therefore $\boldsymbol{f_i}^{(t)}(1)$ is yesterdays price for stock $\boldsymbol{O_i}$, $\boldsymbol{f_i}^{(t)}(2)$ is day before yesterdays price for stock $\boldsymbol{O_i}$ and so on. We look at a snapshot of the stock market t days into the past and associate with it a voting matrix $\boldsymbol{V}^{(t)}$ for the purposes of computing stock ranks. Since the objective is to extract key market drivers, the ijth element of $\boldsymbol{V}^{(t)}$ is simply a measure of the correlation between the length t past stock price vectors $\boldsymbol{f_i}^{(t)}$ and $\boldsymbol{f_j}^{(t)}$ associated with stocks $\boldsymbol{O_i}$ and $\boldsymbol{O_j}$. Mathematically,

$$V_{ij}^{(t)} = 1 + \text{corrcoef}\left(\boldsymbol{f_i}^{(t)}, \boldsymbol{f_j}^{(t)}\right) + \varepsilon, \text{ where } \varepsilon > 0 \tag{4.2}$$

The definition given in 4.2 ensures that $V_{ij}^{(t)} > 0, \forall i, j$. We normally chose ε to be a small constant such as $\varepsilon = 10^{-4}$. Note that the matrix $\boldsymbol{V}^{(t)}$ with elements defined by 4.2 satisfies $\boldsymbol{V}^{(t)} > \mathbf{0}$. Hence by the Perron Frobenius Theorem 3.1 there exists a real eigenvalue $\lambda^{(t)} > 0$ with a real eigenvector $\boldsymbol{x}^{(t)}$ with positive elements i.e., $\boldsymbol{x}^{(t)} > \mathbf{0}$. Also from Corollary 3.2, since $\boldsymbol{V}^{(t)}$ is symmetric, $\lambda^{(t)}$ is strictly larger than any other eigenvalue of $\boldsymbol{V}^{(t)}$. Morever, we can use the power iteration described in 3.3 to compute $\lambda^{(t)} > 0$ and $\boldsymbol{x}^{(t)} > \mathbf{0}$. Once $\boldsymbol{x}^{(t)}$ is computed, we normalize it to give StockRank values between 0 and 1 as follows:

$$\boldsymbol{SR}^{(t)} = \frac{\boldsymbol{x}^{(t)}}{\sum_{i=1}^{n} \boldsymbol{x}^{(t)}(i)} \tag{4.3}$$

By construction, the sum of StockRank values over the n stocks is unity i.e.,

$$\sum_{i=1}^{n} SR^{(t)}(i) = 1 \tag{4.4}$$

The StockRank for any stock can be estimated as:

$$SR^{(t)}(i) = \text{StockRank of stock } \boldsymbol{O_i} \text{ using past data of length } t \tag{4.5}$$

If t is small then we only look at stock price realizations in the immediate past - the short sighted view. On the other hand, when t is large then we look at long term stock prices - the long sighted view. In our current formulation, we get a StockRank value for each stock as a function of the length of past data used t. Stocks that are the key market drivers are expected to show up consistently with a high StockRank regardless of the length of past data t used. On the other hand, market anomalies or new strong market players will show a high StockRank in the short run, i.e., for small values of t.

5 Experiments and Results

Our StockRank algorithm has been implemented in MATLAB `www.mathworks.com`. We tested our StockRank algorithm on real stock market data from the National Stock Exchange (NSE) `http://www.nseindia.com`. We obtained the data for $n = 50$ stocks from NSE over a period of $p = 100$ days from Google finance `http://www.google.com/finance` using an automated program written in Java. This data was then used as an input to the StockRank algorithm. We used the following approach:

- For each stock i, we created a length t vector $\boldsymbol{f_i}^{(t)}$ containing the past price of stock i for the last t days.
- Using these past prices, we created a voting matrix $\boldsymbol{V}^{(t)}$ based on length t data for each stock.
- $\boldsymbol{V}^{(t)}$ was used to compute stock ranks $\boldsymbol{SR}^{(t)}$.
- In our experiments, we used past data of various lengths $t = t_1, t_2, \ldots, t_m$. Here we used $m = 90$ approximately equally spaced values between 10 and 100 for t_i using the command `floor( linspace(10,100,90) )` in MATLAB.

Once we calculate $\boldsymbol{SR}^{(t)}$ using past data of various lengths, we compute the mean StockRank regardless of t as follows:

$$\boldsymbol{SR}_{mean} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{SR}^{(t_i)} \tag{5.1}$$

6 Discussion

In this work, we first described Google's ranking technology which is in fact a family of ranking algorithms of which PageRank is a special case. Next, we describe in detail the mathematical theory behind Google's ranking technology including a proof of Perron-Frobenius theorem and the actual computation of ranks using the power iteration. Finally, we show how Google's ranking technology can be applied to rank stocks in equity markets using our StockRank algorithm. StockRank is similar in spirit to PageRank but uses a different voting metric. The idea in StockRank is to extract key market drivers and hence we used the metric defined in 4.2. Many other metrics similar to 4.2 are possible and the definition of an optimal metric would be the object of future investigation.

We applied our StockRank algorithm to past 100 days worth of data from the NSE. It was found that the top 2 stocks with the highest StockRanks were DLF and RELIANCE. These two stocks have been doing particularly well in the past few months and are considered to be solid stocks. It is interesting that StockRank can pick these up automatically. Low ranked stocks such

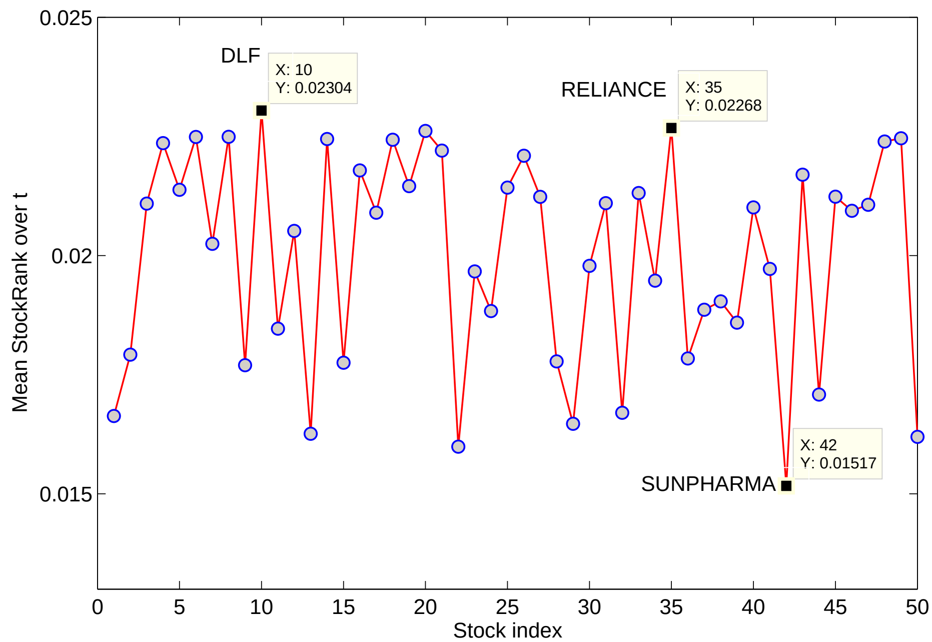

Figure 1: Figure shows the mean StockRank values $\boldsymbol{SR}_{mean}$ for $n = 50$ stocks from NSE. More influential stocks are expected to have a higher mean StockRank. The top two stocks with the highest mean StockRanks are DLF and RELIANCE. The stock with the lowest StockRank was SUNPHARMA.

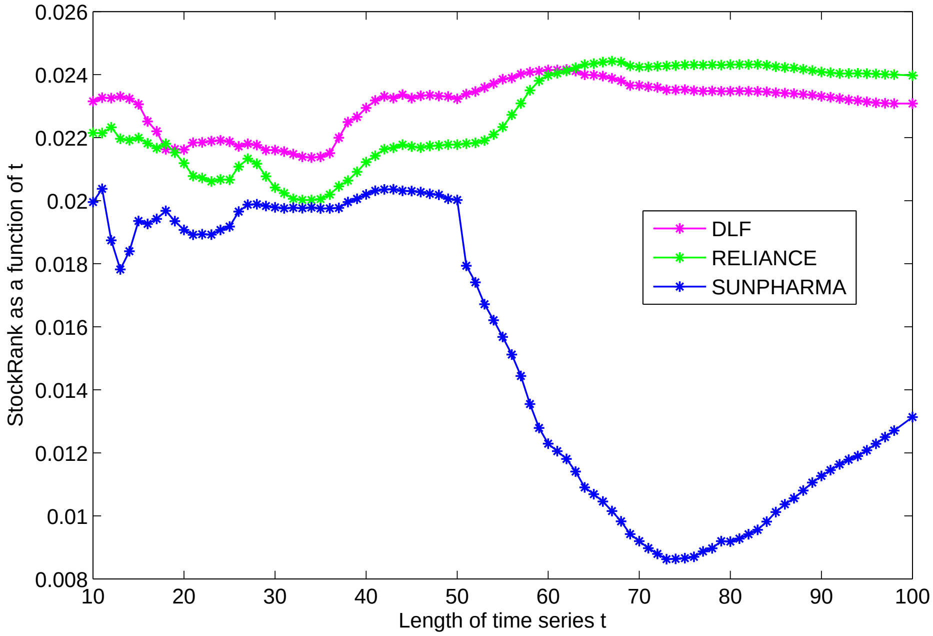

Figure 2: We can view $\boldsymbol{SR}^{(t)}$ as a function of the stock market snapshot of length t. Figure shows the StockRank as a function of t for the stocks: DLF, RELIANCE and SUNPHARMA. It can be seen that DLF and RELIANCE have consistently high StockRanks whereas SUNPHARMA has consistently low StockRanks regardless of t.

as SUNPHARMA in our test case can be considered to be outliers or non-influential in the market. A suggested investment strategy is to invest in highly ranked stocks, however a more detailed testing of this approach is warranted.

7 Conclusion

In summary, we have applied Google's ranking technology to devise StockRank, a ranking algorithm for stocks in equity markets similar to Google's PageRank algorithm for ranking webpages on WWW. StockRank is expected to be useful in extracting highly influential stocks from an equity market as a function of time. This could be beneficial in devising profitable investment strategies.

References

[1] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[2] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.