An Empirical Study of Software Market Share: Diversity and Symbiotic Relations

Yu, Liguo

To cite this article: Yu, Liguo. "An Empirical Study of Software Market Share: Diversity and Symbiotic Relations." *First Monday*, vol. 17, no. 8-6, Aug. 2012, https://doi.org/10.5210/fm.v17i8.3409.

PEER-REVIEWED JOURNAL ON THE INTERNET

An empirical study of software market share: Diversity and symbiotic relations

by Liguo Yu

Abstract

With the increasing use of software products, software ecosystems have emerged. Software ecosystems not only include the same type of products, but also include other related products that they support or depend on. Software marketplace diversity and the symbiotic relations between software products are important properties of a software ecosystem. They have great impact on the popularity and evolution of a software product. This paper presents an empirical study of a software marketplace ecosystem, which is formed by operating systems, Web browsers, and Web servers: (1) Using the concept of market share entropy, we analyze the diversity of the marketplace; and, (2) Using correlation tests, we analyze the symbiotic relations between products. Based on the results of these two studies, we analyze the relation between marketplace diversity and symbiosis in a software ecosystem.

Contents

1. Introduction

Software products are being used everywhere in our society. On one hand, software systems are not used independently, but interact with other software systems. For example, an application software system needs to be used with operating, database, and network systems. On the other hand, software systems are closely dependent on hardware; the evolution of computer hardware always results in the rise or the decline of software systems. Moreover, software systems have direct relations with different stakeholders, such as owners, developers, and users. Therefore, software products together with their interdependent communities form an ecosystem.

Software ecosystems have been studied in recent years. Messerschmitt and Szyperski (2003), in their fundamental monograph *Software ecosystem*, studied software in the context of users, developers, society and economics. They found that the use and sales of software strongly influences software development and evolution. Since the publication of this monograph, the notion of software ecosystems has attracted more attention.

In a software ecosystem, market share is an important element. It refers to the proportion of the total available market that is being serviced by a given software product (Kress and Snyder, 1994). It is represented as the percentage of a software product unit sales volume over the total volume of units of the same type of product in the marketplace. Traditionally, market share is considered to be dependent on internal factors, such as the quality and price of a product and the service and the marketing strategies of a company. With the emergence of software ecosystems, external factors, such as marketplace diversity and symbiotic relations between software products, are considered equally important in determining market share (Jensen and Scacchi, 2005; Scacchi, *et al.*, 2006).

In this paper, we present an empirical study of a software ecosystem — a marketplace formed by operating systems, Web browsers, and Web servers. Based on market share, we analyze marketplace diversity and symbiotic relations between software products in this ecosystem. The objective of this study is to understand how software ecosystems might affect a software market.

2. Market share entropy and software marketplace diversity

The concept of entropy originates in thermodynamics (Maxwell, 2001), where it is used to measure the disorder of particles, such as atoms, molecules, and plasma, in a closed system. the notion of entropy has been extended to information theory, where it is used to measure the amount of information contained in a program unit, such as variable, expression, function, and query (Shannon, 1948).

Information entropy has been used in many fields, including business and software engineering. For example, Sandroni (2005) used market entropy to represent the accuracy of agent beliefs and found that belief accuracy is the critical factor in determining success in trading. McCauley (2003) used the empirical market distribution to represent an asset's

ABOUT THE AUTHOR
Liguo Yu
http://www.cs.iusb.edu/~yul
Indiana University South Bend

Liguo Yu is an assistant professor of Indiana University South Bend, Computer Science and Informatics.

entropy and found that financial markets are unstable and accordingly do not behave thermodynamically. On the other hand, because information plays a key role in financial market, information entropy theory is also utilized in financial market analysis. For example, Theil (1967) applied information theory to economics, such as the measurement of income inequality, industrial concentration, and international trade concentration. Maasoumi and Racine (2002) used entropy metrics to examine the predictability of stock market returns and found that the entropy metric is capable of detecting nonlinear dependency within the returns and nonlinear affinity between the returns. Chen (2005a; 2005b) showed that most empirical evidence about investor performance and market behavior can be explained by an entropy –based information theory.

In software engineering, information entropy is also widely used. For example, Hassan and Holt (2003) used entropy to represent the complexity of the software development process. They studied source control repositories of six open source projects and found that complexity entropy is a good indicator of software quality (Hassan, 2009). Krein, *et al.* (2009) proposed the concept of language entropy, which represents the diversity of programming languages used in a software project. Taylor, *et al.* (2008) defined author entropy and used it to represent authorship diversity for a software program. They found that large files are more likely to have a dominant author rather than smaller files (Casebolt, *et al.*, 2009).

Hence, the application of entropy has gone beyond thermodynamics and information theory. However, to the our knowledge, this concept has not been applied specifically to the study of software ecosystems. In this paper, it is used to study software marketplace diversity.

Consider a closed marketplace (ecosystem) M, formed by n products, m_1, m_2, ..., m_n. At a certain time t, each product has market share $p_1(t)$, $p_2(t)$, ..., $p_n(t)$, respectively. The market share entropy of marketplace M at time t is defined as

$$E(M,t) = -\sum_{i=1}^{n} p_i(t) \log_2 p_{i,}(t) \qquad (1)$$

Equation (1) is the entropy function. It is first introduced in thermodynamics and later applied to information theory (Shannon, 1948). The market share entropy (E) in Equation (1) represents the diversity of a marketplace. For example, if a marketplace is dominated by a small number of products, *i.e.*, a few products occupy the major market share, the entropy (E) value of this marketplace is low, which means the diversity of this marketplace is low; if a marketplace is evenly shared by a large number of different products, the entropy (E) value of this marketplace is high, which means the diversity of this marketplace is high.

Consider a marketplace with four products, Product 1, Product 2, Product 3, and Product 4. Table 1 illustrates five different distributions of market share at a certain time. Distribution D1 has the least diverse market; market share is solely occupied by Product 1. Distribution D2 is more diverse than Distribution D1, the market is shared by Product 1 and Product 2, although the distribution is uneven. Distribution D3 is more diverse than Distribution D2 in that the market is evenly shared by Product 1 and Product 2. Distribution D4 is even more diverse then Distribution D3 in that the market is shared by three products, Product 1, Product 2, and Product 3. Distribution D5 is most diverse: the market is evenly shared by all four products. In Table 1, the market share entropy of each distribution is calculated and it shows that market share entropy increases as the diversity of the marketplace increases. Therefore, we can use market share entropy (E) to represent and measure marketplace diversity.

Table 1: Possible entropy values of a software ecosystem with four products (Yu, *et al.*, 2009).

Distribution	Market share				Entropy
	P_1	P_2	P_3	P_4	
D_1	100%	0	0	0	0.00
D_2	75%	25%	0	0	0.81
D_3	50%	50%	0	0	1.00
D_4	50%	25%	25%	0	1.50
D_5	25%	25%	25%	25%	2.00

3. Software ecosystem symbiotic relations

Biologists use symbiosis (symbiotic relation) to represent close ecological relationships between two different species. Some biologists (Sapp, 1994) consider that Darwin's concept of natural selection, driven by competition, is by itself incomplete, and claim that selection is also strongly based on the outcomes of symbiotic relations among organisms. According to this theory, species that are selected by environmental changes are not just those that are successful in combative competition but those that know how to cooperate with others (Yu and Ramaswamy, 2006; Yu, *et al.*, 2008).

In our previous research, three major types of symbiotic relations were identified in software ecosystems (Yu, *et al.*, 2008).

- Mutualism: both products benefit from their relations.
- Competition: both products are harmed by their relations.
- Neutralism: both products are unaffected by their relations.

Most of these symbiotic relations are likely to be reflected in their business relationships. For example, the relationship between Linux (http://www.linux.org/) and GNU (http://www.gnu.org/) might be considered as mutualism, in which Linux is the kernel of an operating system and GNU provides system utilities. They are mutually beneficial and dependent. In contrast, the relation between FreeBSD (http://www.freebsd.org/) and Linux might be considered competitive because of the operating system marketplace. The relation between OpenOffice (http://www.openoffice.org/) and Apache Web Server (http://httpd.apache.org/) might be neutralism, because there are no direct dependencies between them.

However, some symbiotic relations are not apparent. Consider Java and Javascript; it is difficult to understand their relationships without supporting data and supporting analysis. Another example is IBM and HP, both supporters of Linux project. Does that mean they have a common interest and accordingly their relationship represents mutualism or are they competing for the same market and accordingly competitive? Moreover, a software ecosystem is so complicated that we might see two different relationships from different perspectives. Therefore, we need a quantitative approach to studying software symbiotic relations.

In this paper, we use statistical methods to study symbiotic relations between software products. More specifically, Spearman's rank correlation test (Hogg and Craig, 1995) is utilized to analyze correlations between market shares of software products.

Consider a closed marketplace (ecosystem) M, formed by n products, m_1, m_2, ..., m_n. At certain time t, each product has market share $p_1(t)$, $p_2(t)$, ..., $p_n(t)$, respectively. A null hypothesis is formulated to describe the relations between the market share of two software products p_i and p_j in a certain time range of observation.

Social media were considered important information sources due to the variety of information available on the sites (Jung and Moro, 2012). Individuals, whether experts or laypersons, uploaded information about the earthquake and nuclear accident on social media platforms. Many local newspapers in earthquake–affected areas set up an account either on commercial social media outlets, such as Twitter, or used local social networking sites to send out information to local residents and newspaper subscribers (Nihon Shinbun Kyokai, 2011). Social media also served as a dissemination channel for the mass media. Within several days of the earthquake, NHK streamed its programs on social media such as Ustream and Nico Nico Douga (Niconico). Major news agencies sent out real–time news on Twitter several times a day. As social media were used widely by individuals, organizations and mass media to produce, consume and exchange information, the information utility of social media in the emergency situation increased. In order to examine goals and media dependency in a disaster situation, five research questions are proposed.

> *H_{01}: There is no correlation between the market shares of Product m_i and Product m_j*

The correlation coefficient (α) returned by the Spearman test is in the range [-1, 1]. A value of α in the range of [-1, 0) indicates a negative correlation; a value of α in the range of (0, 1] indicates a positive correlation. A value of α=0 is rarely seen and indicates no correlation between m_i and m_j.

Another parameter returned by the Spearman test is the significance (p value) of the correlation. For either a positive correlation or a negative correlation, if the p value is less than or equal to 0.001 (significance is at the 0.001 level), we will reject the null hypotheses and conclude that strong positive (or negative) correlation exists between Product m_i and Product m_j. Accordingly, a strong positive correlation is referred as possible mutualism between m_i and m_j; a strong negative correlation is referred as possible competition between m_i and m_j; and all other results are referred as possible neutralism (no relation) between m_i and m_j. These analysis principles are summarized in Table 2.

Table 2: Spearman's test and possible symbiotic relations between two products.

Correlation coefficient (α)	**>0**		**<0**		**=0**
Correlation	positive		negative		no
Significance (p value)	≤0.001	>0.001	≤0.001	>0.001	—
Strength of correlation	strong	weak	strong	weak	—
Possible symbiotic relation	mutualism	neutralism	competition	neutralism	neutralism

Using one Spearman's test to determine symbiotic relations might not be convincing, because Spearman's rank correction only considers the values of market share. Two market shares might be accidentally correlated even there is no symbiotic relations between them. To improve the accuracy of Spearman's test on symbiotic relations, we also study changes of market share, *i.e.*, differential data of market share.

Consider a closed marketplace (ecosystem) M, formed by n products, m_1, m_2, ..., m_n. At certain time t, each product has market share $p_1(t)$, $p_2(t)$, ..., $p_n(t)$, respectively. The market share differential of product m_i at time t is defined as

$$Dif(m_i, t) = (p_i(t)-p_i(t-1))/p_i(t-1) \qquad (2)$$

In Equation 2, $p_i(t)$ is the market share of Product m_i at time t, $p_i(t-1)$ is the market share of Product m_i at time $t-1$. $Dif(m_i, t)$ is therefore the measure of the degree of market share changes of product m_i from time $t-1$ to time t. Similarly, a null hypothesis is formulated to describe the relations between market share changes of two software products m_i and m_j in a certain time range of observation.

> H_{02}: There is no correlation between the market share differential of product m_i and product m_j.

Spearman's test is performed again on differential data of software market shares. Following the same principles shown in Table 1, a strong positive correlation of market share differential is referred as possible mutualism between m_i and m_j; a strong negative correlation of market share differential is referred as possible competition between m_i and m_j; and all other results are referred as possible neutralism between m_i and m_j. The underlying logic for these principles is that (1) if two products have mutualism relations, the increase of the market share of one product will result in the increase of the market share of another product; (2) if two products have competition relations, the increase of the market share of one product will result in the decrease of the market share of another product.

In this study, a symbiotic relation between two products is confirmed only if it is both determined through Spearman's tests on Hypothesis H_{01} and on Hypothesis H_{02}, *i.e.*, market share correlation and market share differential correlation. Keep in mind that statistical correlation does not necessary indicate a symbiotic relation between software products. We must interpret it together with properties of the software product itself.

First, because mutualism is rarely seen between products of the same category and competition is rarely seen between products of different categories, in this study, positive correlations are only going to be used to infer a mutualism relationship between products of different categories and negative correlations are only used to infer the competition relationship between products of the same category. This means, for products of the same category, a strong positive correlation is not considered as mutualism; for products of different category, a strong negative correlation is not considered as competition.

Second, neutralism means two products have no relations. However, as with any other research, we can only prove the existence of some relations, we cannot prove the non–existence of some relations. For two products, even we could not statistically detect mutualism or competition, we still cannot conclude that they do not have these relationships. Mutualism or competition might be too weak to be observed in our study. Therefore, in this research, we only determine mutualism and competition. Any unconfirmed mutualism or competition is not considered as neutralism without further investigation.

Based on these two principles, the scheme to statistically detect symbiotic relations between software products is illustrated in Table 3 and Table 4, in which symbol "—&edquo; indicates a non–confirmed relationship. We remark here that (1) these two principles are applied in this study because of the clear software categories used in our case study. In other studies, different principles might be considered; and, (2) in this paper, the symbiotic relations between two software products are determined by the statistical test (Spearman's test), which might be different from other observations.

Table 3: Confirmation of the symbiotic relations between products of the same category through hypothesis tests H_{01} and H_{02}.

Possible symbiotic relations inferred through hypothesis		***H_{01}*** mutualism	neutralism	competition
H_{02}	mutualism	—	—	—
	neutralism	—	—	—
	competition	—	—	competition

Table 4: Confirmation of the symbiotic relations between products of different categories through hypothesis tests H_{01} and H_{02}.

Possible symbiotic relations inferred through hypothesis		***H_{01}*** mutualism	neutralism	competition
H_{02}	mutualism	mutualism	—	—
	neutralism	—	—	—
	competition	—	—	—

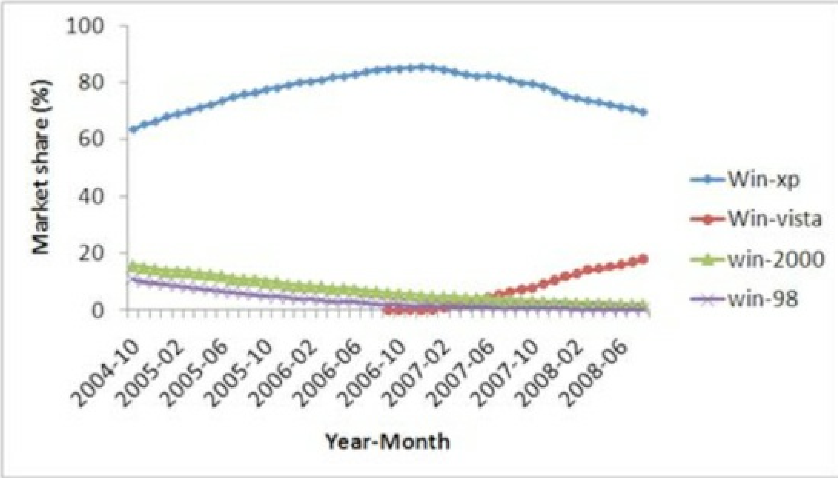
4. Case studies of software marketplace ecosystems

In this section, we study market share of various operating systems, Web browsers, and Web servers. The market share data for operating systems and Web browsers was collected from Net Applications (2010). The market share data for Web servers was collected from the Web Server Survey (Netcraft, 2010). Products that have a short evolution history and market share history were not included in this study, such as the Windows 7 operating system, Chrome Web browser, and Google web server. Although these products occupy a certain market share at this point, their short evolution histories (less than three years) could not provide enough data for this study. On the other hand, some products, such as the Mozilla Web browser are included in this study because of their relative long evolution history and market place history.

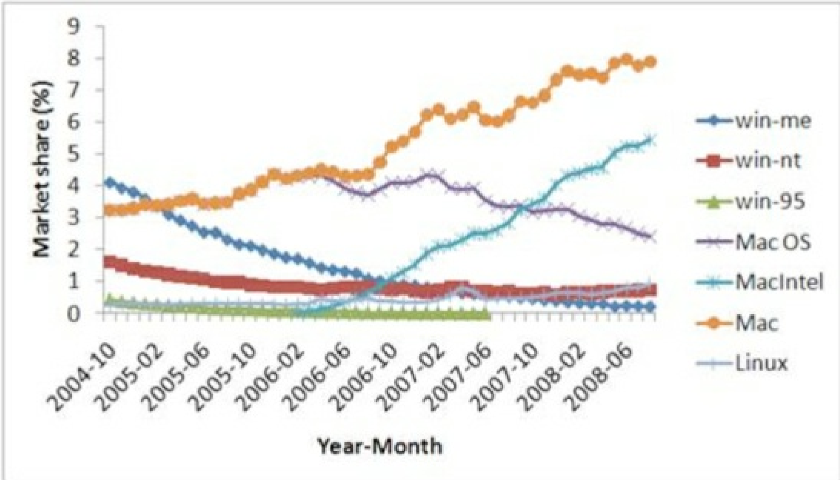
To satisfy these conditions, we choose to study data in the range of October 2004 to August 2008, during which, major operating systems, Web browsers, and Web servers have certain observable market share. The data was collected monthly, which means the market share of each product for each month from October 2004 to August 2008 was recorded.

4.1. General results

Figure 1, Figure 2, and Figure 3 illustrate the evolution of market share of major operating systems, Web browsers, and Web servers, respectively. It should be noted that market share is measured as the percentage of an observed market that is being serviced by one product. The accumulation of the market share of all the products shown in one figure represents the entire (100 percent) market place.



(a)



(b)

Figure 1: The evolution of market share of operating systems:
(a) Part 1; and, (b) Part 2 (Yu, *et al.*, 2009).

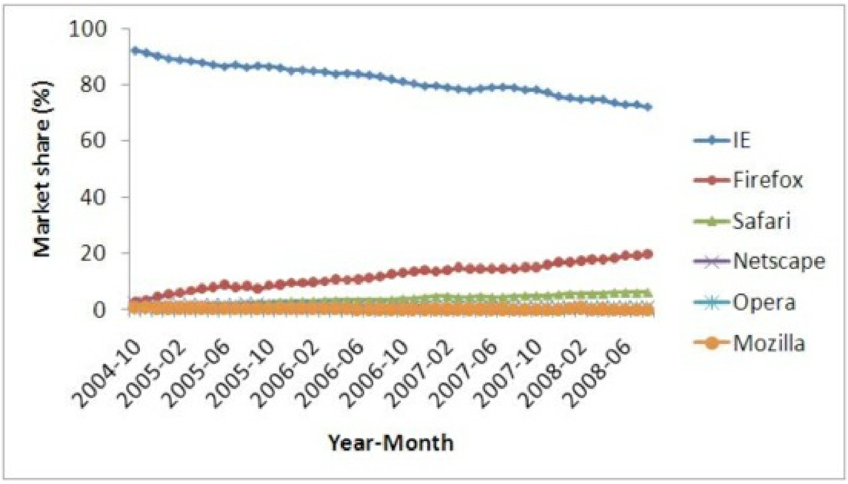


Figure 2: The evolution of market share of Web browsers (Yu, *et al.*, 2009).

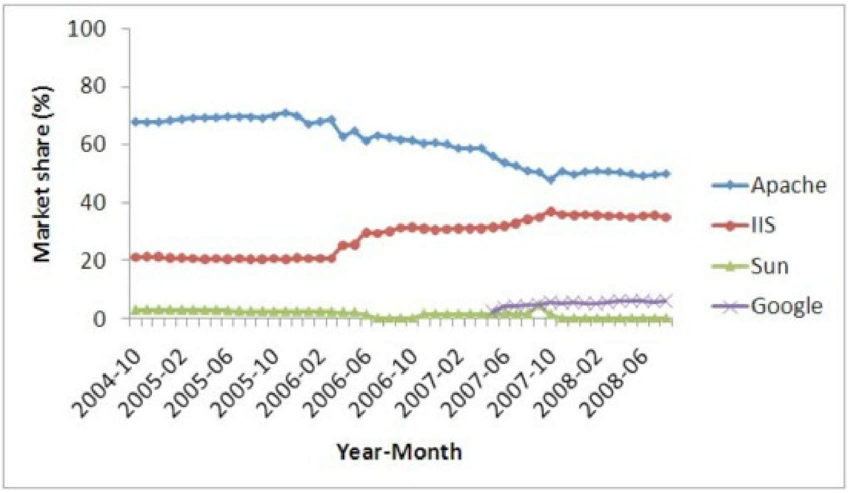


Figure 3: The evolution of market share of Web servers (Yu, *et al.*, 2009).

For operating systems (Figure 1), Windows in general has decreasing trends. For Web browsers (Figure 2), IE (Internet Explorer)'s market share is decreasing while others are increasing. For Web servers (Figure 3), the market share of Apache is decreasing while the market share of IIS (Internet Information Services) is increasing.

4.2. Marketplace diversity

From Figures 1 to 3, we can see that in general the market share of a certain software marketplace is becoming more and more evenly distributed among different products. To quantitatively measure this trend, the evolution of market share entropy for each marketplace (operating systems, Web browsers, and Web servers) was studied. The evolution of market share entropy for operating systems, Web browsers, and Web servers is illustrated in Figure 4, Figure 5, and Figure 6, respectively.

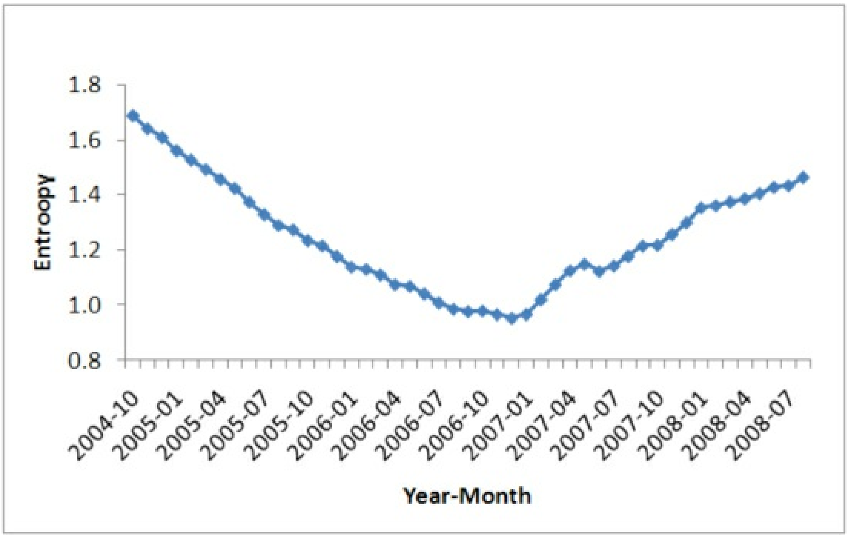


Figure 4: The evolution of market share entropy for operating systems (Yu, *et al.*, 2009).

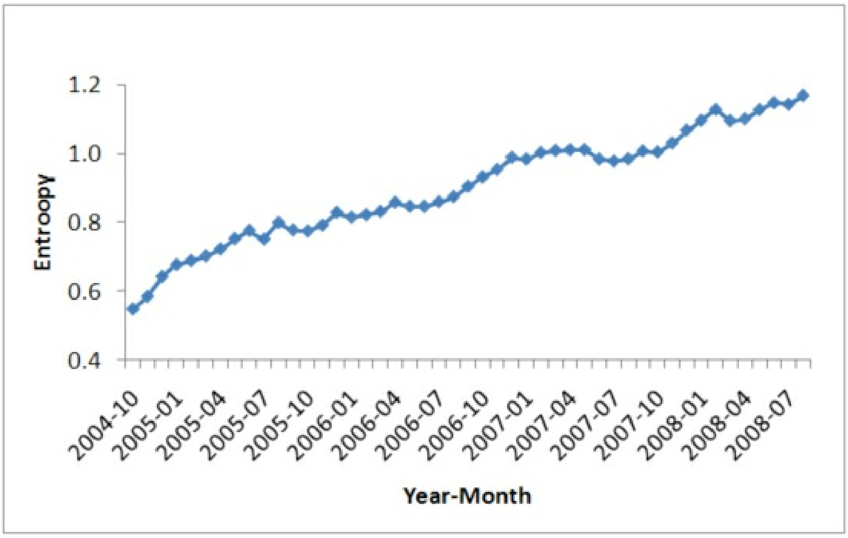


Figure 5: The evolution of market share entropy for Web browsers (Yu, *et al.*, 2009).

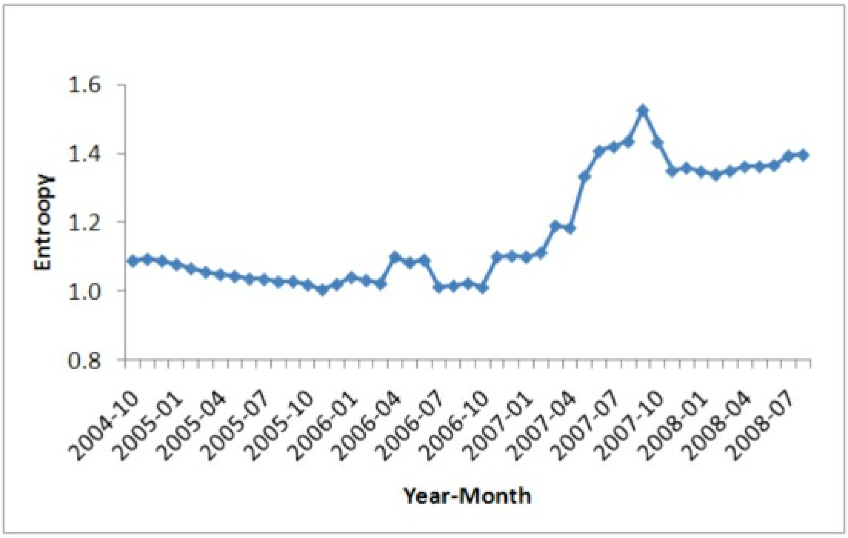


Figure 6: The evolution of market share entropy for Web servers (Yu, *et al.*, 2009).

From Figure 5 and Figure 6, it is clear that market share entropies of Web browsers and servers are increasing. Therefore, the market place for Web browsers and servers is becoming more diverse. For operating systems (Figure 4), market share entropy decreased since October 2004 and started to increase by the end of 2006. This behavior can be explained by examining operating system market share evolution in Figure 1. Since 2004, the market share of Windows XP had been increasing until the end of the 2006, when Windows Vista is released. After that, increased market share of Windows Vista and decreased market share of Windows XP resulted in an increase of operating systems market place diversity (market share entropy).

4.3. Symbiotic relations

In the software marketplace ecosystem we studied, each major Web browser or server belonged to one software organization. There are few organizations that have two different products on the same market, except Mozilla, which has both the Mozilla and Firefox browsers. In contrast, the market place for operating systems is different: There are several products from Microsoft, such as Windows XP, Windows Vista, and Windows NT. To understand the symbiotic relations between software products in this ecosystem, we organize products into two groups. Group 1 is a sub ecosystem, which contains all Microsoft operating systems. Group 2 is the whole ecosystem, which contains nine representing products of three categories, operating systems, Web browsers, and Web servers. The two groups and their products are listed below.

- Group 1 (Windows sub ecosystem): Win–Vista, Win–XP, Win–2K, Win–98, Win–NT, and Win–ME.
- Group 2 (entire ecosystem): Windows, Linux, Mac, IE, Firefox, Safari, Apache, IIS, and Sun.

In Group 2, three representing products were selected for each category. Windows is the Windows sub ecosystem, which contains all Windows operating systems and Mac contains both MAC OS and MAC Intel.

4.3.1. Windows sub ecosystem

To study the symbiotic relations of various Windows operating systems, two null hypotheses (H_{01} and H_{02}) are formulated for each pair of Windows products listed in Group 1. Spearman's rank correlation tests were then performed to examine correlations between market shares of products (H_{01}) and correlations between market share differentials of products (H_{02}). The results are summarized in Table 5 and Table 6.

Table 5: Correlation between market shares of Windows operating systems.
Note: *Correlation is significant at the 0.001 level.
Duration of data set: 24 months.

	Win –Vista	Win –XP	Win –2K	Win–98	Win –NT	Win –ME
Win –Vista	—	**-.983***	**-.998***	**-1.000***	-.385	**-.996***
Win–XP		—	.981*	.983*	.334	.979*
Win–2K			—	.998*	.394	.995*
Win–98				—	.485	.996*
Win–ME					—	.393

Table 6: Correlation between market share differentials of Windows operating systems.
Note: *Correlation is significant at the 0.001 level.
Duration of data set: 23 months.

	Win –Vista	Win –XP	Win –2K	Win –98	Win –NT	Win –ME

Win –Vista	—	.322	-.059	.000	-.211	-.118
Win–XP		—	.038	.067	-.273	.551*
Win–2K			—	.121	.622*	-.138
Win–98				—	.-168	.345
Win–ME					—	-.359

As described earlier, for products of the same category, Spearman's tests are only used to detect competition, which have significant negative correlations at the 0.001 level. The corresponding relations in Table 5 and Table 6 are in bold. Study of the correlation between market shares of Windows operating systems, four significant negative correlations were found between Vista and Win–XP, Vista and Win–2000, Vista and Win–98, and Vista and Win –ME. Therefore, hypothesis test H_{01} shows that Vista is a possible competitor of Windows XP, 2000, 98, and ME.

However, the Spearman's test of correlation between differentials of market share of Windows operating systems (Table 6) could not confirm our speculations. No significant negative correlations are found. Therefore, based on hypothesis test H_{02}, we cannot confirm the possible competition relationships noted in Table 5.

Therefore, combining Spearman's tests on market share correlation and market share differential correlation and following the scheme described in Table 3, we statistically find no competition relationships among different products representing the Windows operating system. We remark here that we do not use positive correlations to infer mutualism between products of the same category.

4.3.2. Whole ecosystem

To study the symbiotic relations of the nine products in the three different categories listed in Group 2, two null hypotheses (H_{01} and H_{02}) were formulated for each pair of products. Spearman's rank correlation tests were then performed to study correlations between market shares of products (H_{01}) and correlations between market share differentials of products (H_{02}). The results are summarized in Table 7 and Table 8.

Table 7: Correlation coefficient between market shares of products in the whole ecosystem.
Note: *Correlation is significant at the 0.001 level.
Duration of data set: 47 months.

	Operating systems			Web browsers			Web servers		
	Windows	Linux	Mac	IE	Firefox	Safari	Apache	IIS	Sun
Windows	—	**-.917***	**-.996***	**.991***	-.986*	-.995*	**.899***	-.881*	**.823***
Linux		—	.894*	-.930*	**.919***	**.906***	-.859*	**.831***	-.788*
Mac			—	-.984*	**.978***	**.995***	-.887*	**.868***	-.814*
IE				—	**-.994***	**-.990***	**.898***	-.873*	**.843***
Firefox					—	.985*	-.901*	**.877***	-.840*
Safari						—	-.900*	**.875***	-.828*
Apache							—	**-.953***	.729*
IIS								—	**-.736***
Sun									—

Table 8: Correlation coefficient between market share differentials of products in the whole ecosystem.
Note: *Correlation is significant at the 0.001 level.
Duration of data set: 46 months.

	Operating systems			Web browsers			Web servers		
	Windows	Linux	Mac	IE	Firefox	Safari	Apache	IIS	Sun
Windows	—	-.222	**-.927***	**.603***	-.344	-.917	.057	-.024	-.004
Linux		—	-.005	-.455*	.251	-.043	.100	-.076	-.218
Mac			—	-.527*	.402	**.955***	-.020	-.019	-.021
IE				—	**-.793***	**-.467***	-.073	.150	.123
Firefox					—	.336	.001	-.038	-.107
Safari						—	-.076	.040	.010
Apache							—	**-.510***	-.055
IIS								—	.005
Sun									—

Table 7 shows the correlation coefficient between market shares of nine represenitive software products. Table 8 shows the correlation coefficient between market share differentials of nine software products. In these two tables, the data is divided into two type of zones: green zones are for correlations of products belong to different categories; yellow zones represent correlations of products belong to the same category. Accordingly, we will look for significant negative correlations in yellow zones and significant positive correlation in green zones. The corresponding coefficients are bolded in each zone.

From the yellow zones of Table 7, we can detect the following possible competition relationships: Windows and Linux; Windows and Mac; IE and Firefox; IE and Safari; IIS and Apache; and, IIS and Sun. From the yellow zones of Table 8, we can detect the following possible competition relationships: Windows and Mac; IE and Firefox; IE and Safari; and, IIS and Apache. Combing the results of these two tests and following the scheme illustrated in Table 3, we statistically detect four competition relationships in the whole ecosystem: Windows and Mac; IE and Firefox; IE and Safari; and, IIS and Apache.

From the green zones of Table 7, we detect the following possible mutualism relationships: Windows and IE, Linux and Firefox, Linux and Safari, Mac and Firefox, Mac and Safari, Windows and Apache, Windows and Sun, Linux and IIS, and Mac and IIS, IE and Apache, and IE and Sun, Firefox and IIS, and Safari and IIS. From the green zones of Table 8, we can detect the following possible mutualism relationships: Windows and IE; and, Mac and Safari. Combing the results of these two tests and following the scheme illustrated in Table 4, we statistically detect two mutualism relationships in the whole ecosystem: Windows and IE; and, Mac and Safari.

These six symbiotic relations are illustrated in Figure 7, in which products of the same categories are grouped together. Figure 8 through Figure 13 show scatter plots of market share and market share differential between each pair of partners (mutualism) or competitors (competition) in these six symbiotic relations. As noted in Figures 8 to 13, positive correlations are found between partners and negative correlations observed between competitors.

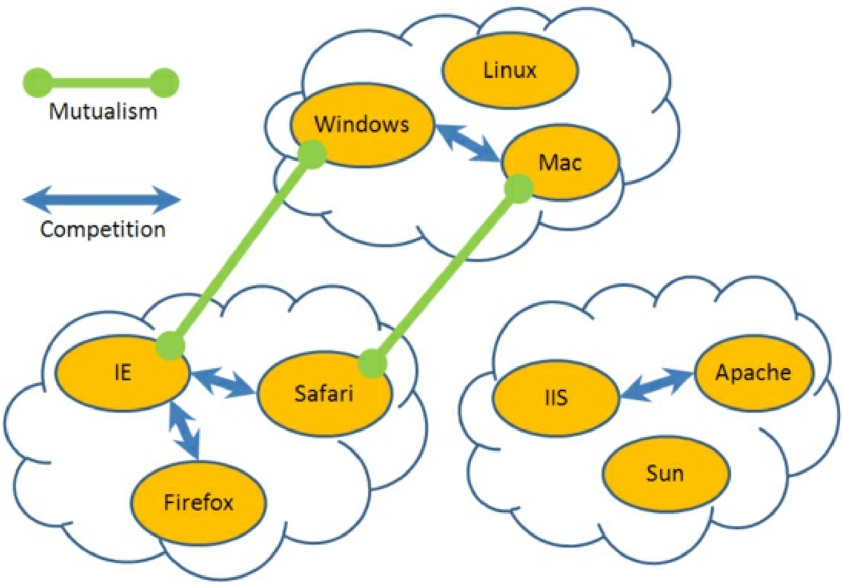
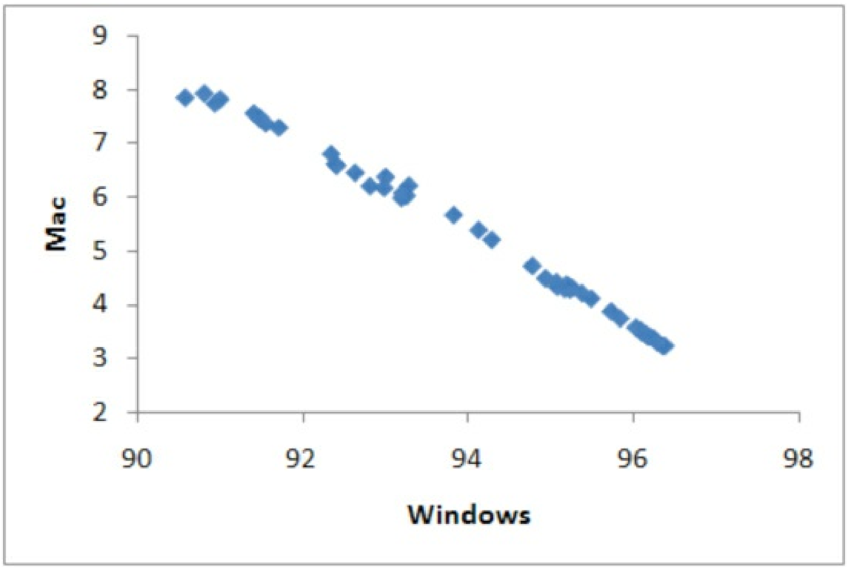
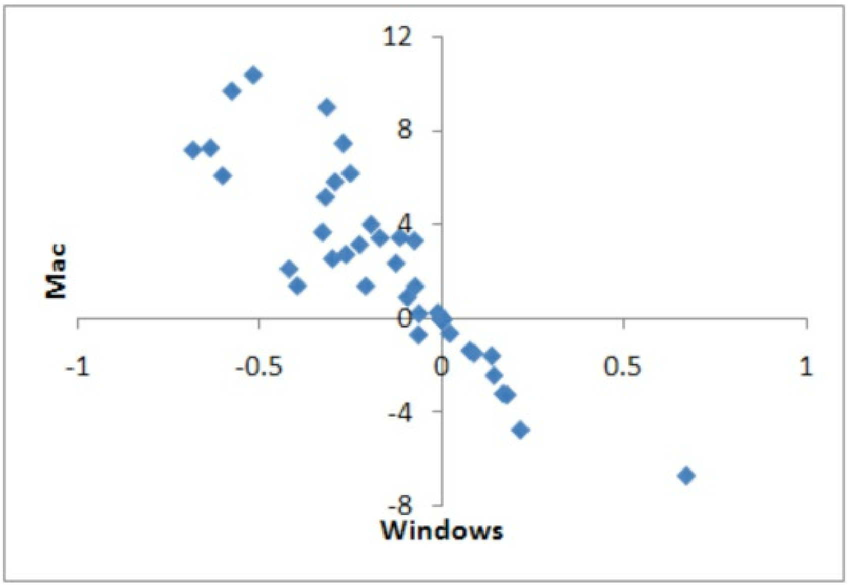


Figure 7: Symbiotic relationships among software products in the whole ecosystem.

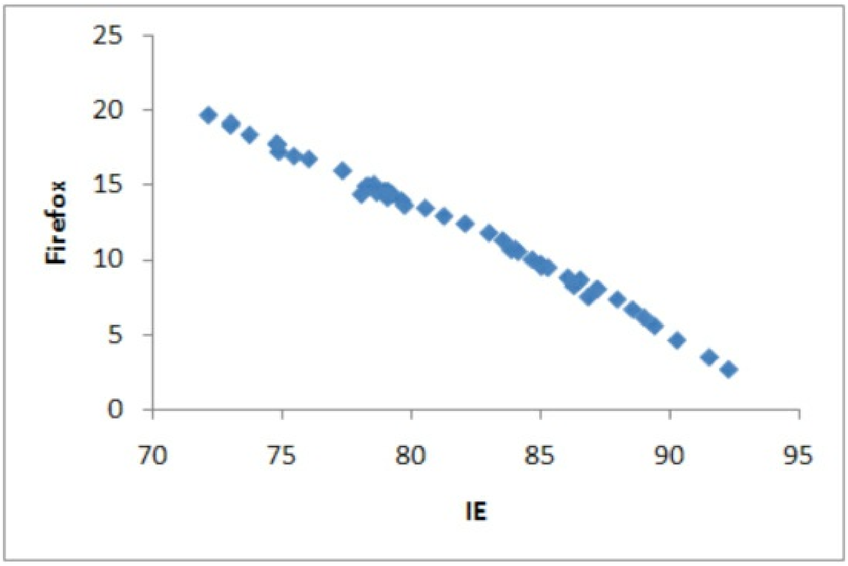


(a)

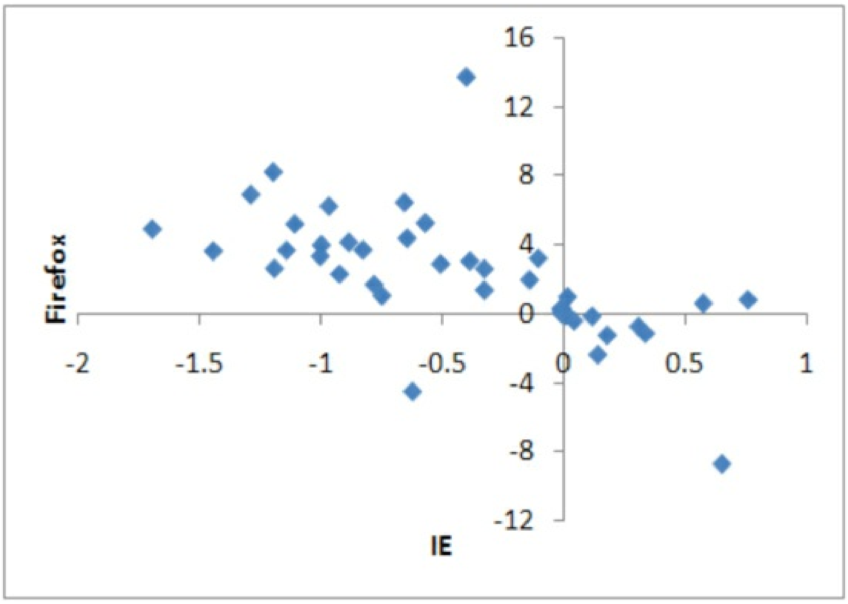


(b)

Figure 8: Scatter plots of Windows and Mac: (a) market share; and, (b) market share differential.

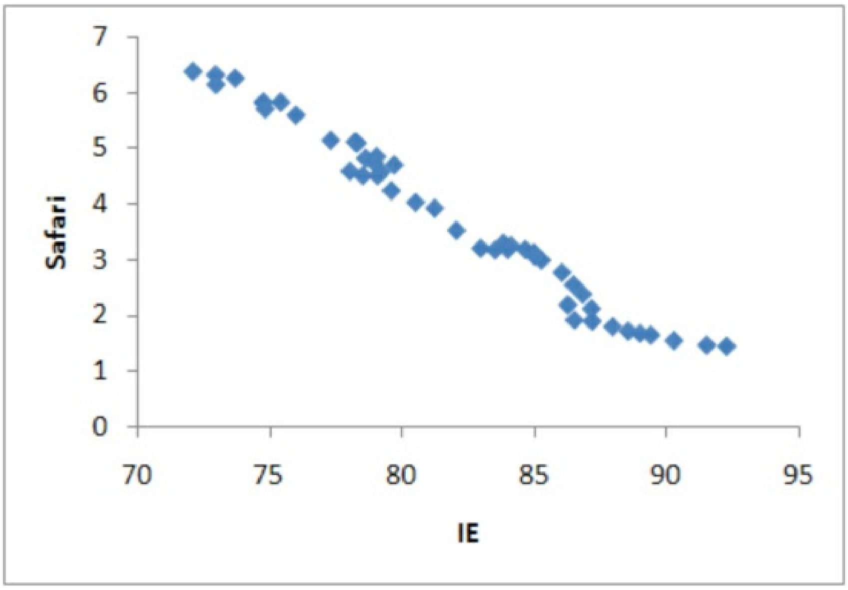


(a)

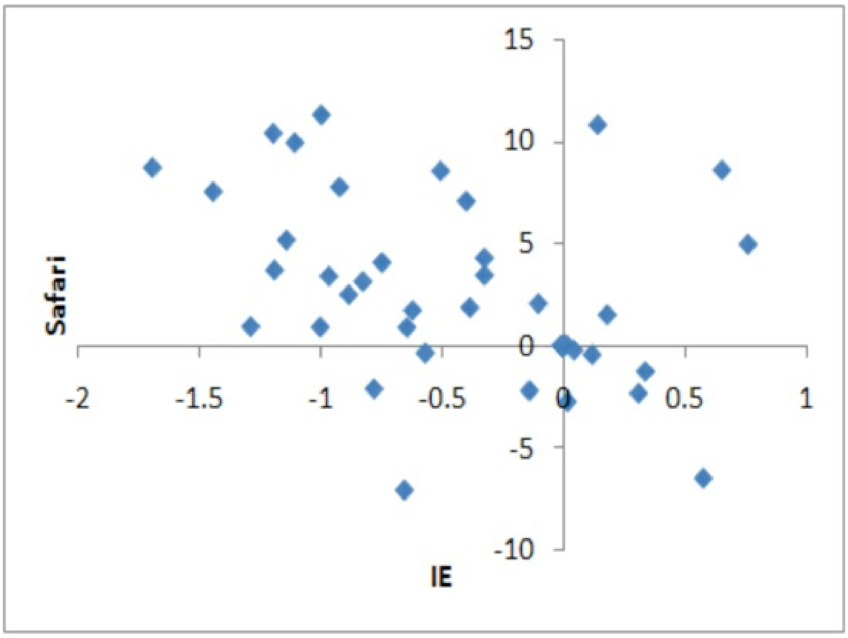


(b)

Figure 9: Scatter plots of IE and Firefox: (a) market share; and, (b) market share differential.

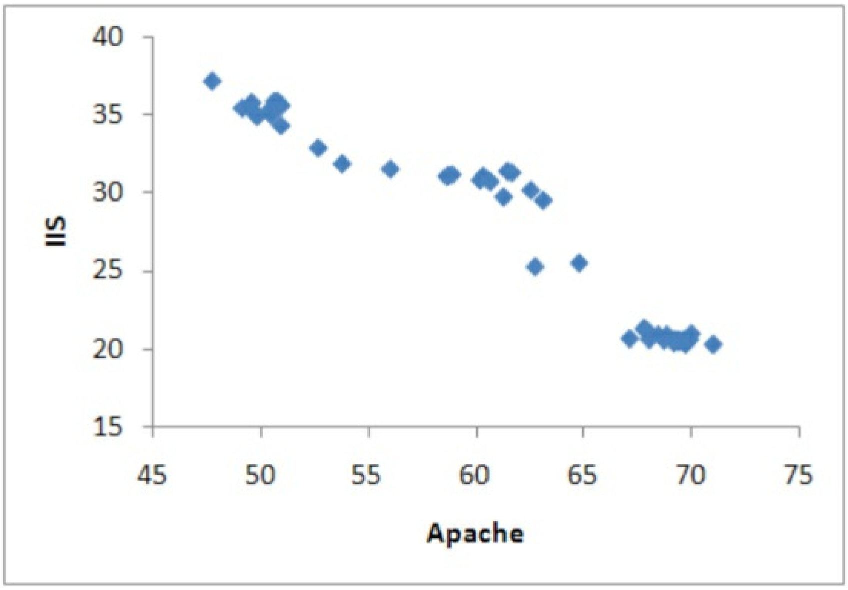


(a)

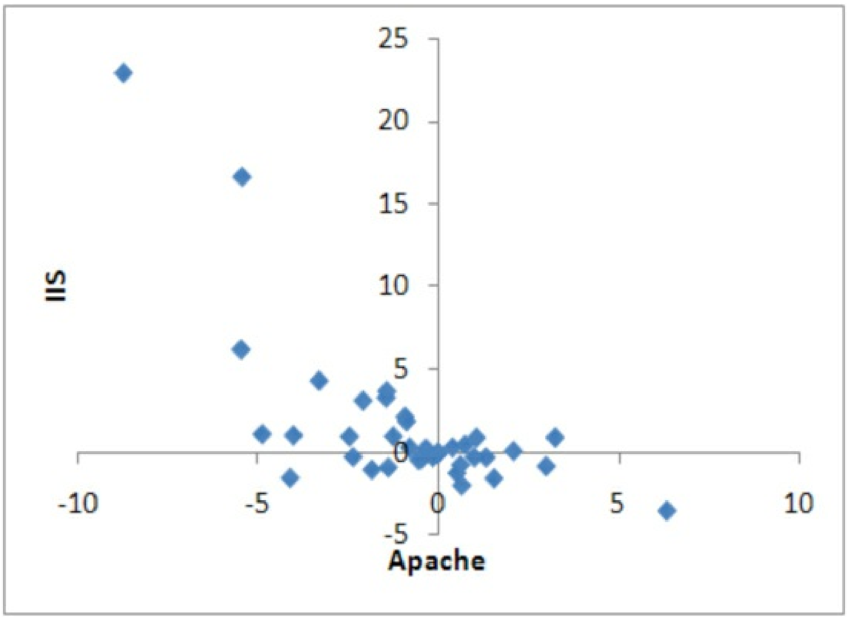


(b)

Figure 10: Scatter plots of IE and Safari: (a) market share; and, (b) market share differential.

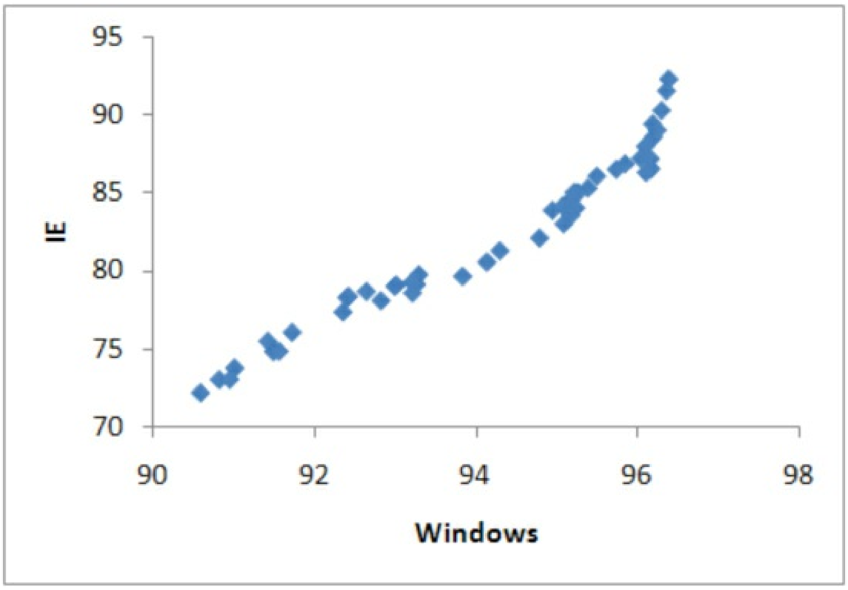


(a)

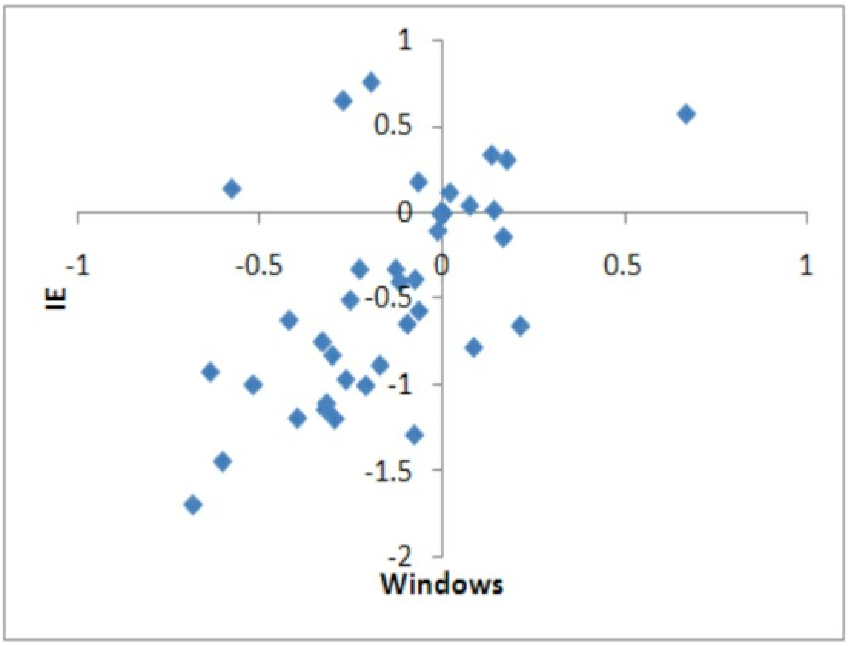


(b)

Figure 11: Scatter plots of IIS and Apache: (a) market share; and, (b) market share differential.

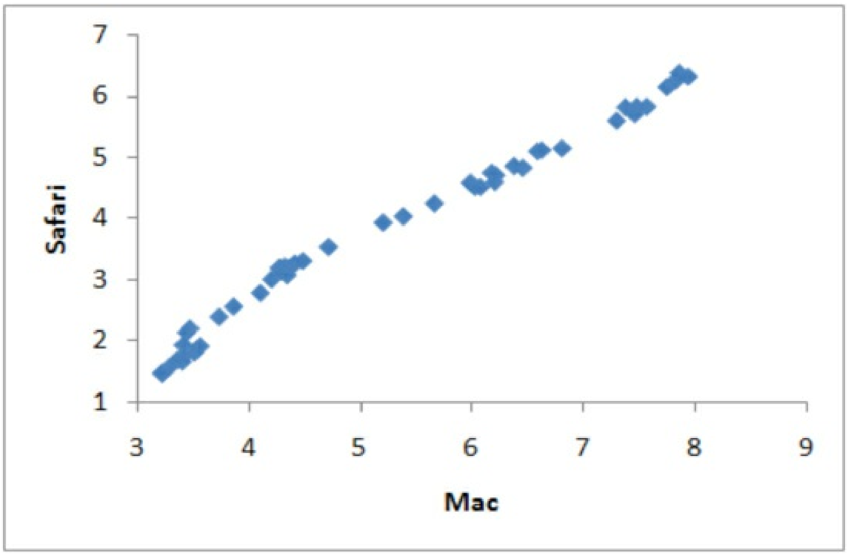


(a)

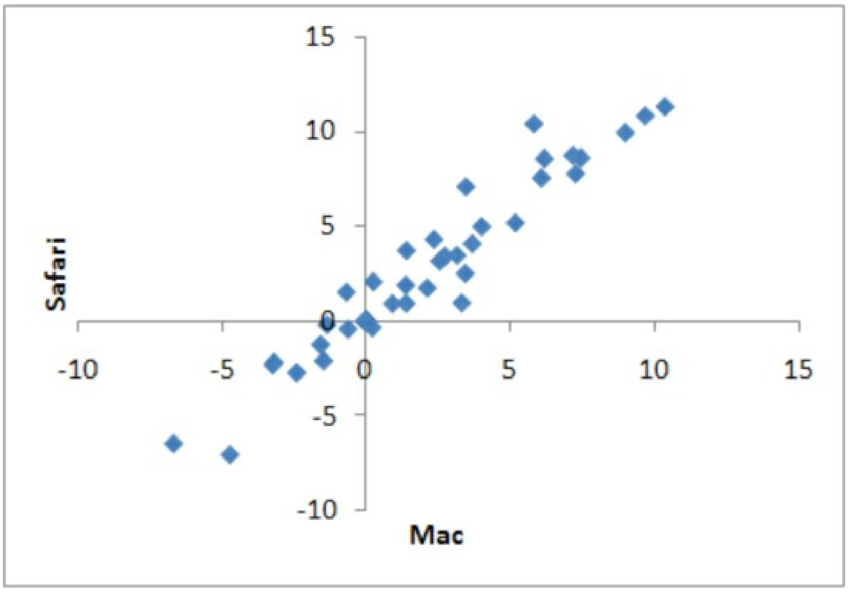


(b)

Figure 12: Scatter plots of Windows and IE: (a) market share; and, (b) market share differential.



(a)



(b)

Figure 13: Scatter plots of Mac and Safari: (a) market share; and, (b) market share differential.

In this study, we verified six symbiotic relations among operating systems, Web browsers, and Web servers. These relations are further explained below.

1. Mutualism was found between Windows and IE and Mac and Safari, because IE and Safari are major Web browsers for Windows and Mac, respectively.
2. Competition was found between Windows and Mac. Because both Windows and Mac are targeted to personal computers, we believe they are largely competitors. In contrast, because Linux is mainly targeted to servers, we did not find a competitive relationship between Linux and Windows and Linux and Mac.
3. For Web browsers, competition was found between IE and Firefox and IE and Safari, because they all can be used on the Windows operating systems. However, competition was not found between Safari and Firefox, although both of them can be used on Macs. We speculate that Firefox has not been widely used on Macs so that it is not a competitor to Safari.
4. For Web servers, competition was found between IIS and Apache. Competition was not found between Sun and others. Examining the market share trend of Sun in Figure 3, we speculate that the market share of Sun Web servers is relatively small and stable. It does not fluctuate dramatically with market changes which could explain why we did not detect competition between Sun and other Web servers.
5. Apache is a Web server mainly used on Linux. However, we did not detect mutualism between Apache and Linux. There could be several reasons for this observation. First, Linux is not the only platform for Apache. For example, Apache can also run on Windows.
6. IIS is a Web server mainly running on Windows. However, we did not detect mutualism between IIS and Windows. It could be possible that only a small number of Windows

devices are installed with IIS and used as Web servers. In contrast, IE is installed on almost all Windows machines, explaining why we found mutualism between IE and Windows, but not between IIS and Windows.

4.3.3. Discussion

In our first study, we analyzed market place diversity of operating systems, Web browsers, and Web servers. We found their market share entropy values are around 1.2–1.4, in which operating systems and Web servers are slightly more diverse than Web browsers. We speculate these similarities are related to their symbiosis.

Considering mutualism between Windows and IE and between Mac and Safari, an increase of market share for one product leads to an increase of the market share for the another. Most of these popular products belong to a few organizations so close symbiotic relations might exist between these entities. For example, one developer might participate in several open source projects, such as Linux and Apache. Their integrated relationships might affect market share of these products.

Suppose a new product from a different organization enters the market place. If it can successfully occupy a certain market share, market share entropy and market place diversity will increase. However, we did not see this in the market place of operating systems, Web browsers, and Web servers. We speculate that this is largely due to symbiotic relations. Without setting up mutualism with a leading product, it is difficult to compete with products that have already dominated the market place for years. Hence, we see a relatively stable market place diversity (market share entropy) for operating systems, Web browsers, and Web servers.

5. Conclusions

In this paper, we studied software ecosystems formed by operating systems, Web browsers, and Web servers. Using market share data, we analyzed the diversity of the market place, based on information entropy, as well as symbiotic relationships among various products, based on statistical analysis.

Our study found relatively stable market place diversity for operating systems, Web browsers, and Web servers. We also detected six symbiotic relationships between software products. We speculate that the stability of this ecosystem is related to symbiotic relationships among software products.

About the author

Liguo Yu is an assistant professor of Indiana University South Bend, Computer Science and Informatics.
E–mail: ligyu [at] iusb [dot] edu

Acknowledgements

An initial version of this paper appeared at the Eighth Annual Conference on Applied Research in Information Technology, Acxiom Laboratory for Applied Research, University of Central Arkansas in 2009. Errors in the initial version were corrected in this paper.

References

J.R. Casebolt, J.L. Krein, A.C. MacLean, C.D. Knutson, and D.P. Delorey, 2009. "Author entropy vs. file size in the gnome suite of applications," *Proceedings of the Sixth IEEE International Working Conference on Mining Software Repositories*, pp. 91–94, and at http://sequoia.cs.byu.edu/lab/files/pubs/Casebolt2009.pdf, accessed 31 July 2012.

J. Chen, 2005a. "Information theory and market behavior," *ICFAI Journal of Behavioral Finance*, volume 2, number 4, pp. 25–45.

J. Chen, 2005b. *The physical foundation of economics: An analytical thermodynamic theory*. Hackensack, N.J.: World Scientific.

A.E. Hassan, 2009. "Predicting faults using the complexity of code changes," *ICSE '09: Proceedings of the 31st International Conference on Software Engineering*, pp. 78–88.

A.E. Hassan and R.C. Holt, 2003. "The chaos of software development," *IWPSE '03: Proceedings of the Sixth International Workshop on Principles of Software Evolution*, p. 84.

R.V. Hogg and A.T. Craig, 1995. *Introduction to mathematical statistics*. Fifth edition. Englewood Cliffs, N.J.: Prentice Hall.

C. Jensen and W. Scacchi, 2005. "Process modeling across the Web information infrastructure," *Software Process: Improvement and Practice*, volume 10, number 3, pp. 255–272.

J.L. Krein, A.C. MacLean, D.P. Delorey, C.D. Knutson, and D.L. Eggett, 2009. "Language entropy: A metric for characterization of author programming language distribution," *Proceedings of the Fourth Workshop on Public Data about Software Development*, at http://sequoia.cs.byu.edu/lab/files/pubs/Krein2009.pdf, accessed 31 July 2012.

G. Kress and J. Snyder, 1994. *Forecasting and market analysis techniques: A practical approach*. Westport, Conn.: Quorum Books.

E. Maasoumi and J. Racine, 2002. "Entropy and predictability of stock market returns," *Journal of Econometrics*, volume 107, numbers 1–2, pp. 291–312.

J.C. Maxwell, 2001. *Theory of heat*. Mineola, N.Y.: Dover.

J.L. McCauley, 2003. "Thermodynamic analogies in economics and finance: Instability of markets," *Physica A: Statistical Mechanics and its Applications*, volume 329, numbers 1–2, pp. 199–212.

D. Messerschmitt and C. Szyperski, 2003. *Software ecosystem: Understanding an indispensable technology and industry*. Cambridge, Mass.: MIT Press.

Net Applications.com, 2010. "Market share for mobile, borwsers, operating systems and search engines," at http://marketshare.hitslink.com, accessed 31 July 2012.

Netcraft, 2010. “Index of articles: Web server survey 2010,” at http://news.netcraft.com/fullindex.html, accessed 31 July 2012.

A. Sandroni, 2005. “Market selection when markets are incomplete,” *Journal of Mathematical Economics*, volume 41, numbers 1–2, pp. 91–104.

J. Sapp, 1994. *Evolution by association: A history of symbiosis*. New York: Oxford University Press.

W. Scacchi, J. Feller, B. Fitzgerald, S. Hissam, and K. Lakhani, 2006. “Understanding free/open source software development processes,” *Software Process: Improvement and Practice* volume 11, number 2, pp. 95–105.

C. Shannon, 1948. “A mathematical theory of communication,” *Bell System Technical Journal*, volume 27, pp. 379–423, 623–656.

Q.C. Taylor, J.E. Stevenson, D.P. Delorey, and C.D. Knutson, 2008. “Author entropy: A metric for characterization of software authorship patterns,” *Proceedings of the Third International Workshop on Public Data about Software Development*, pp. 42–47.

H. Theil, 1967. *Economics and information theory*. Chicago: Rand McNally.

L. Yu and S. Ramaswamy, 2006. “Software and biological evolvability: A comparison using key properties,” *SE ’06: Proceedings of the Second International IEEE Workshop on Software Evolvability*, pp. 82–88.

L. Yu, S. Ramaswamy, and J. Bush, 2008. “Symbiosis and software evolvability,” *IT Professional*, volume 10, number 4, pp. 56–62.

L. Yu, S. Ramaswamy, and R.B. Lenin, 2009. “Entropy studies of software market evolution,” *Proceedings of the Eighth Annual Conference on Applied Research in Information Technology*, pp. 87–91.

Editorial history

Received February 15 2011; accepted 20 July 2012.

An empirical study of software market share: Diversity and symbiotic relations
by Liguo Yu
First Monday, Volume 17, Number 8 - 6 August 2012
http://firstmonday.org/ojs/index.php/fm/article/view/3409/3286
doi:10.5210/fm.v17i8.3409