VISUALIZING STOCK-MUTUAL FUND RELATIONSHIPS THROUGH SOCIAL NETWORK ANALYSIS

Rafael Solis[1]

California State University, USA.

E-mail: rafaels@csufresno.edu

ABSTRACT

In this paper we analyze the network structure of stocks (as actors in a 2-mode/affiliation social network) and their relationship to mutual funds. The analysis reveals a network structure that has both the "hub" and "small world" characteristics of many common social and physical networks thus suggesting that stock affiliation in mutual funds is not a random phenomenon even though the mutual fund selection was done randomly. The data for the analysis is based on a random sample of 18 mutual (stock) funds from the Vanguard and Fidelity family of funds and 99 unique stocks which were part of the 10 top holdings in each fund. This study may also suggest that institutional investors are prone to herd behavior (a social network phenomena), and/or risk aversion shown by the high concentration of (similar) blue-chips in their portfolios.

Key words: Social networks, small world networks, clustering coefficient, mutual funds, herd behavior, risk aversion

I. INTRODUCTION

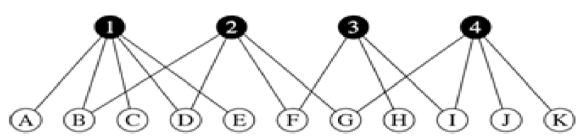
Social Network Analysis has been traditionally used in social and behavioral sciences since their inception a few decades ago. Much of this interest is due to its appealing focus on *relationships* among social entities and on the *pattern* and *implications* of such relationships. The idea is to model the social environment as patterns or

[1] Craig School of Business, California State University, Fresno, 5245 N Backer Ave
Fresno, CA 93740

regularities in relationships between the interacting actors or agents. The presence of such patterns is usually referred as *structure*, which in turn their analysis is given the name *structural analysis*. The purpose of this study is to utilize such methods to expose hidden structures in the relationships between stocks and their affiliation (membership) in mutual funds. Specifically we will use a special type of 2-mode network structure called an affiliation network (Wasserman and Faust, 1994). In these networks, the actors (stocks) are joined together by common membership of groups (in this case the funds). Research utilizing this type of structures includes networks of individuals joined together by their participation on some type of social event (Davis et al., 1941); CEO's of companies joined by their membership to certain clubs (Glaskiewiz and Marsden, 1978); networks of scientist that collaborated in the same paper (Newman et al. 2001); movie actors that appeared in the same film (Watts and Strogatz, 1998) and board of directors that serve on the same board (Mariolis, 2001; Davis et al. 2003). This type of structure utilized in affiliation networks is that of the well-known bipartite graph (Figure 1) where in our case, the nodes {1, 2, 3, 4} represent the mutual fund to which the stocks {A, B, C, …,K} belong. These types of graphs have the characteristic of lacking "loops" among the nodes of the graph. This network can be transformed into a unipartite network (Figure 2) where two stocks are linked if they belong to the same mutual fund. This transformation produces a network that can be analyzed by common network analysis tools like Pajek (Batagelj and Mrvar, 2008).

Central to Social Network analysis are three distinctive features that have been observed in many empirical studies. The first of these is the "small world" effect (Pool and Kochen, 1978; Milgram, 1967) which has been made popular by the book (and play of the same name) "Six Degrees of Separation" (Guare, 1990). This is a characteristic of networks that exhibit a higher degree of connectivity as compared to a random graph (Erdös and Rényi, 1959) Figure 3.

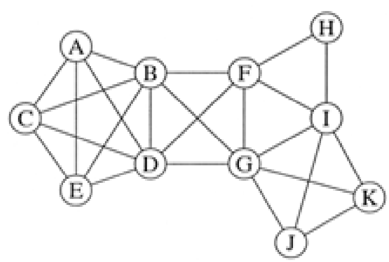
Figure 1
Affiliation (Bipartite) Network



The second property of social network is called "Clustering" (Watts and Strogatz, 1998) which is present when the probability of a tie between two actors (nodes) is much higher if the two actors have one of more mutual acquaintances (links).

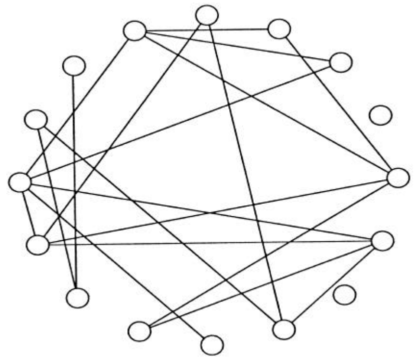
In a simple example, the probability that two of your friends know each other is much higher than that of two randomly selected people in the population know each other. In our context this would imply that two stocks would have a high probability to be in the same mutual fund if they belong to one (or more) common mutual funds. Watts and Strogatz have defined a clustering coefficient (often denoted by C) which is the probability that two nodes (actors) linked to a common node are themselves connected. In many cases, the presence of clustering makes the probability of acquaintance between actors much higher if they have a common link than if they do not. The above authors have shown a value of C from a small percent to 40%-50%, other studies (Newman, 2001, Newman et.al 2001) have shown similar results.

Figure 2
Unipartite Projection of an Affiliation Network



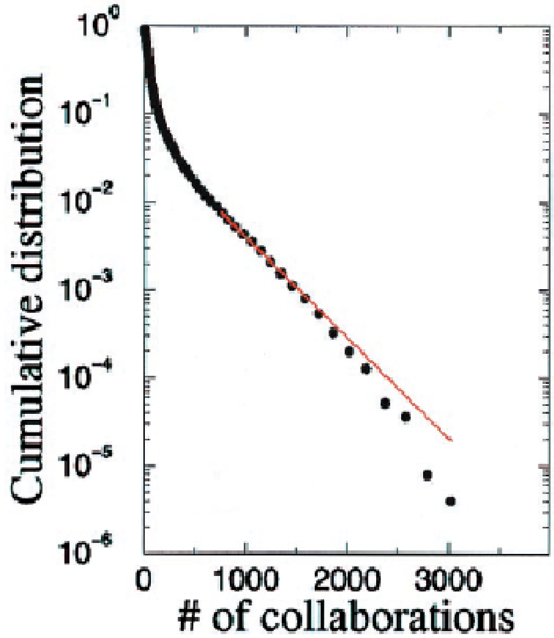
The third characteristic of social networks relates to the long tail distribution of the degrees of separation between the "actors". It has been found in many social networks (e.g. Albert, Barabasi et al 1999) that the probability that two randomly selected actors have a low degree of separation is quite high compared to the same probability for higher degrees (Figure 4 depicts this property, note that in this example, the "actors" are actually movie actors/actresses). If we consider a pure random graph (Figure 3) with probability p that two nodes are linked, the degree distribution can be approximated by a Poisson Distribution with $P_k \approx \lambda^k e^{-\lambda}/k!$ (Erdös and Rényi, 1959). From this result the average degree (of separation) for such networks is (n-1) p where n is the number of nodes/actor. It is well known that this distribution is a poor approximation to the real-world networks' degree distribution which has been shown to be better approximated by a power law

Figure 3
Random Graph N=16 and p=1/7



distribution of the form $P_k=\beta k^{-\alpha}$ (Newman et al, 2002). From this, we can establish the randomness (or lack of it) of a real-world network by investigating the fit of the above two laws. It has been found that many real-world networks follow closely the above well known properties. The fit varies but most show remarkably similar characteristics. In this paper we shall investigate how the above 3 properties characterize the network shown in Figure 1 and discuss the implications of this characterization.

Figure 4
Distribution of # of Collaborations of Movie Actors (Amaral et al 2000)



In a recent paper Kang and Tan (2008) argue that accounting choices are related in firms where their directors interlock in each other boards. Their findings are supported by the social network perspective that argues that managerial actions are embedded in social structures (Granovetter, 1985). These social structures or networks, such as board appointments in other firms, expose top executives and directors to a wide variety of organizational practices and exert some influence on managerial actions and decisions (Geletkanycz and Hambrick, 1997; Wetphal et al., 2001). Social influences from networks of director interlocks are also found to be robust across different settings. For instance, organizational practices, such as donations to non-profit organizations, organizational structures, corporate and business strategies as well as executive compensation have been found to spread across firms connected by director interlocks. (Davies and Greve, 1997; Galaskiewicz and Wasserman, 1989; Wesphal and Frederickson, 2001, Wesphal et al. 2001). Viewing the network of Mutual Funds and Stocks as an affiliation network that follows the small world and the clustering characteristics of non-random networks we'll be able to argue that such networks' behavior also follows that of the human (social) networks discussed above.

II. DATA AND METHODOLOGY

A random sample of 30 mutual funds from the Vanguard fund family produced 18 usable stock-only funds (index, bonds, and other non-stock funds were eliminated from the original sample). From each fund we recorded the top 10 stock holdings. The sample yields a total of 98 different stocks. The sample components (fund and stock symbols) can be found in Appendix A. The data was recorded in a socio-matrix (Wasserman and Faust 1994) where the rows correspond to the mutual funds and the columns the stocks in the sample. A non-empty entry in the matrix represents a stock that is in the top-ten holdings of the (row) mutual fund. This matrix was used as the input to Pajek (Batagelj and Mrvar, 2008) a well know social network analysis software. Figure 5 is the initial visualization of the network. The funds' symbols are in capitals while the stocks' are in lower case.

III. EMPIRICAL FINDINGS

A detail of the above network centered in the fund VGENX is shown in Figure 6. Note that from the graph (which is a 3-d object) we can only see 9 (of its 10) stock components (i.e. the edges from VGENX). The "missing" stock node (Royal Dutch Shell: rds/b) is "behind" one of the displayed nodes. At first glance we can spot several Funds that do not interact with the heavily cluster ones shown at the top of Figure 5. The less overlapping (fund nodes) are the ones shown to the right of the graph (VHCOX and VCVLX), also to some extent VEIEX which have two stocks (ptrbf and lufky) that are part of VGENX and VHGEX respectively.

The above 2-mode network was transformed to a 1-mode (as shown in Figure 7). This network is an "all-stock" network which shows connections between stocks in such a way that if two stocks are in the same fund there is a direct link between them. The degree of a vertex (stock) is the number of edges (links) emanating from that vertex. Also from Figure 7 we can see that several stocks form "hubs". These hubs (or important vertices) have been reported in many real-life networks (like the actors network discussed above). The hub effect doesn't manifest in purely random networks like the one shown in Figure 3. Thus hubs are the most connected stocks which are the ones that appear in several funds and therefore have a degree larger than 9 (stocks that appear in the top-ten holdings in only one fund have a degree of 9, that is they are connected to only 9 other stocks within the fund). Figure 9 and Table 1 show the most connected stocks (degree 24 or higher).

Figure 5
Vanguard Funds and their Top Ten Holdings

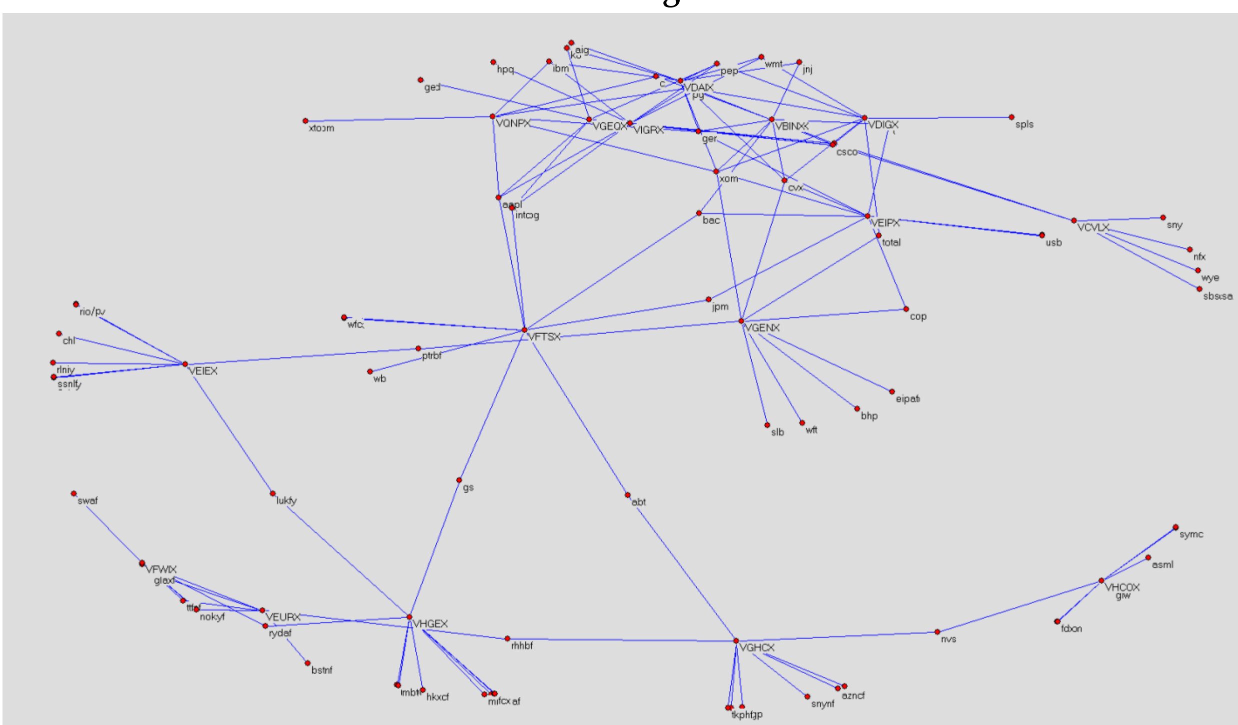


Figure 6
Detail of Network in Figure 5

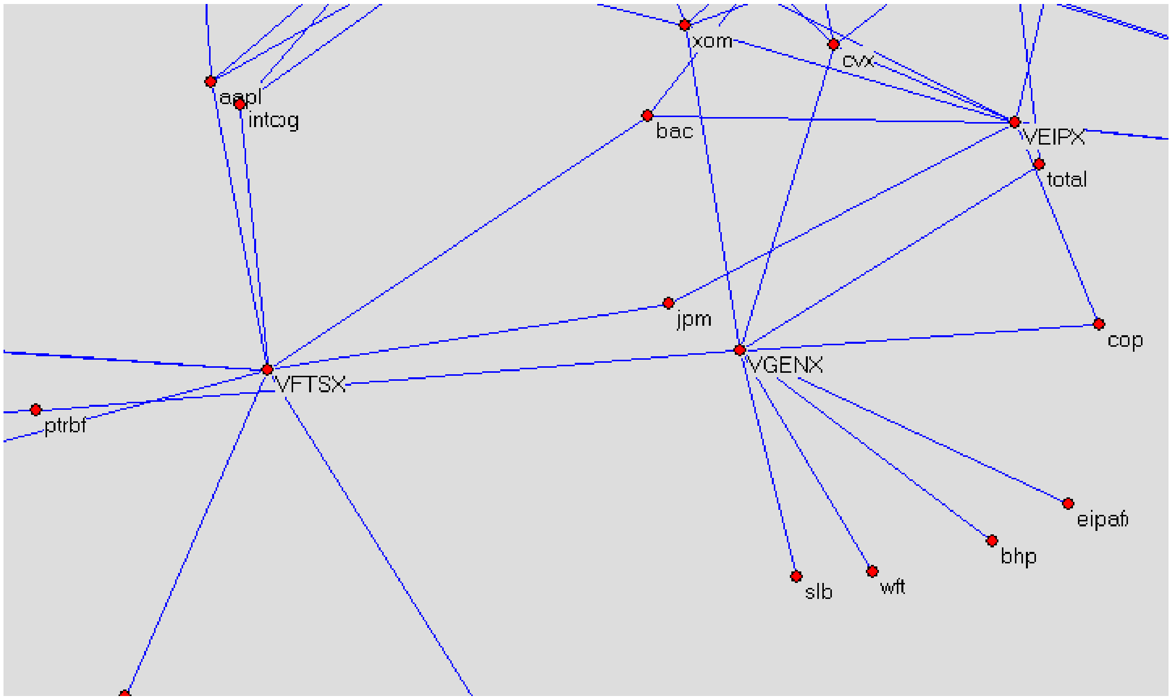


Figure 7
1-Mode Projection of the Network in Figure 5

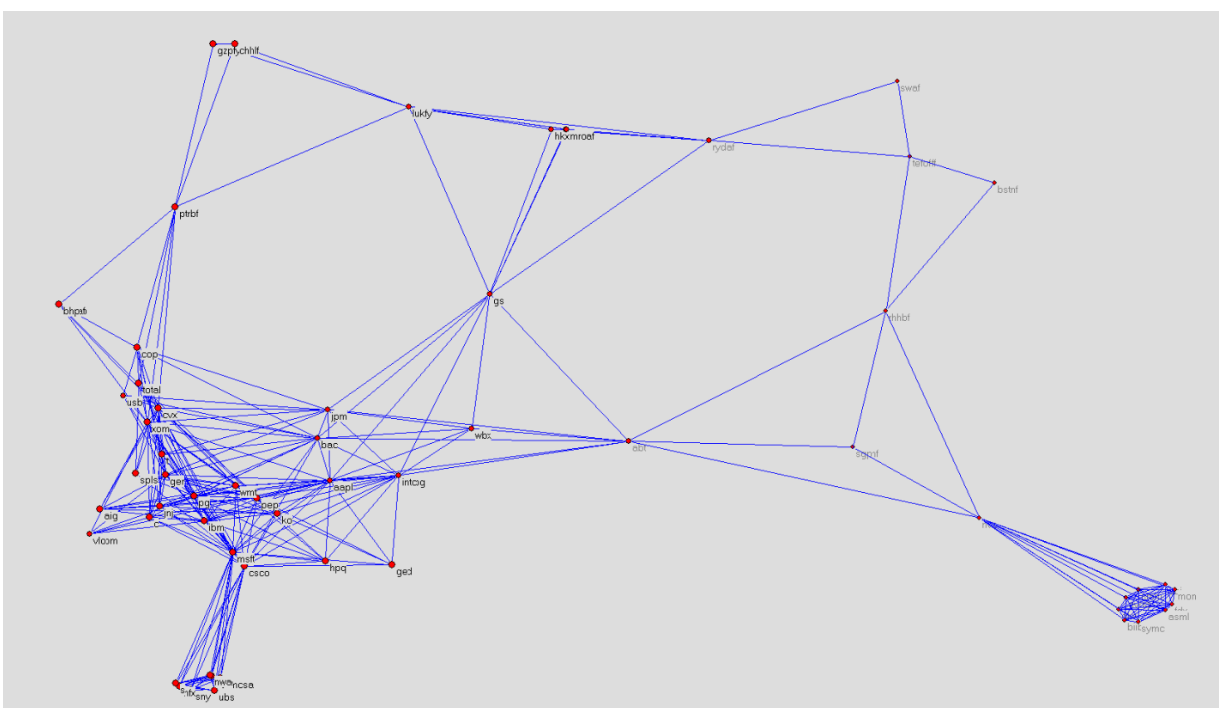


Figure 8
1-Mode Network Detail

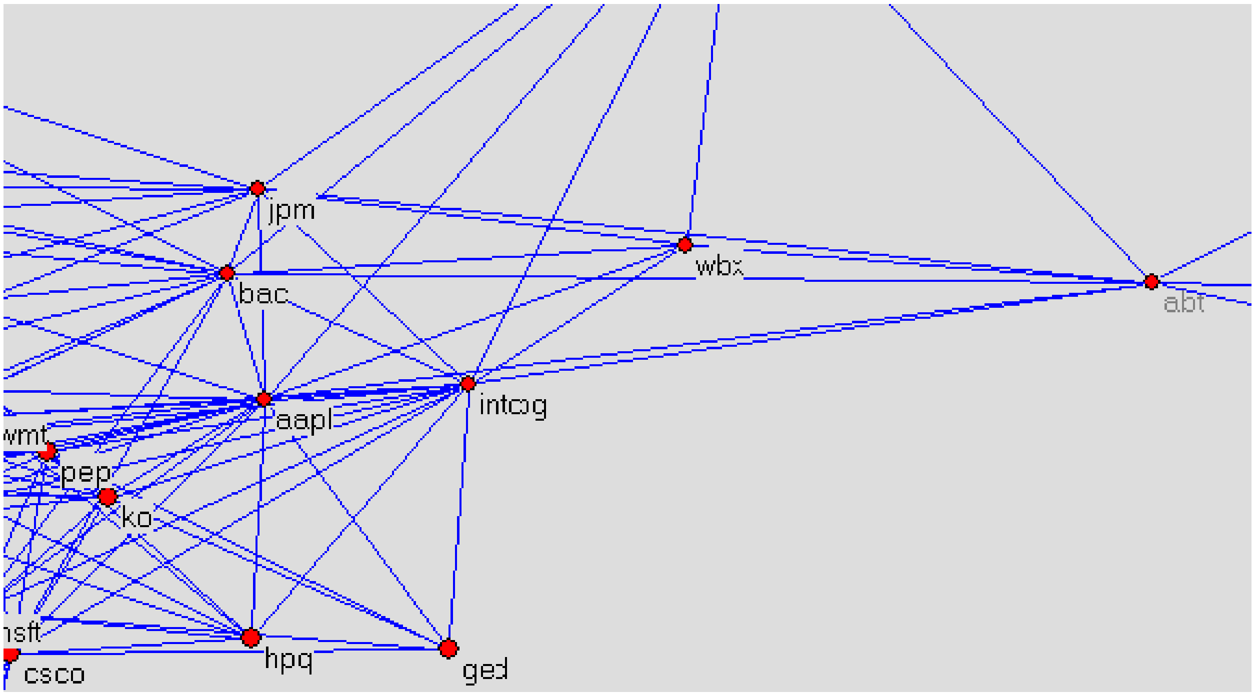


Figure 9
Most Connected Stocks Detail
(Circle Diameter Proportional to Vertex Degree)

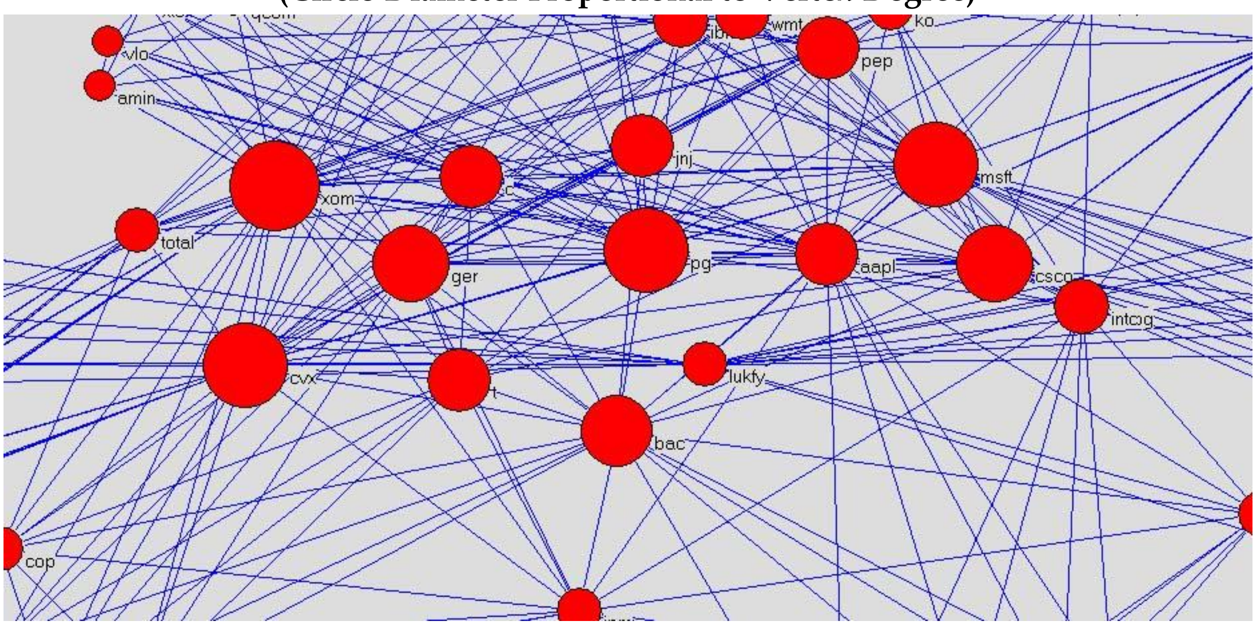


Table 1
Most Connected Stocks

Symbol(Name)	Degree
Xom (Exxon)	34
Bac (Bank of America)	30
Cvx (Chevron)	29
Csco (Cisco)	28
Aapl (Apple Computer)	27
Pg (Procter & Gamble)	26
Ger (General Electric)	24
Msft (Microsoft)	24

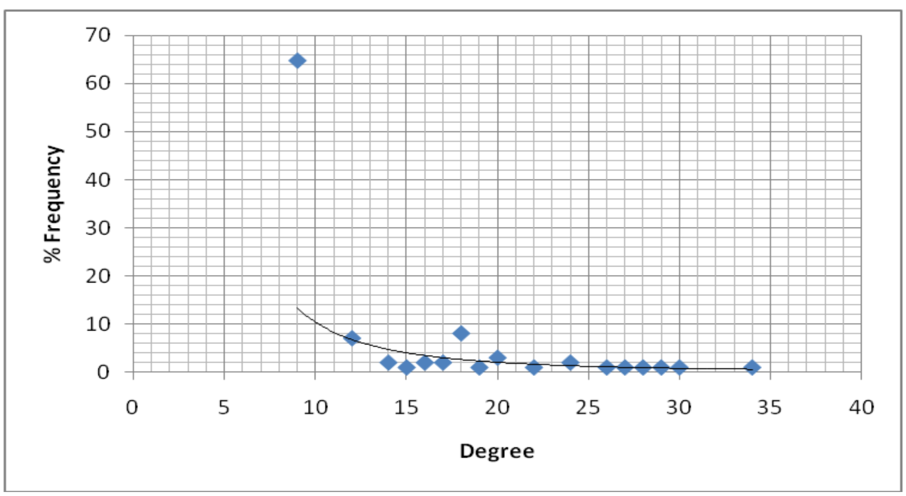
Keeping in mind that the sample of funds was random, it is surprising to see a handful of stocks (out of 99) to be of high importance. This characteristic has been observed in (human) social networks and it has been given the tag "the rich-get-richer" (Reka and Barabasi, 2001) which refers to the characteristic that certain hubs (individuals) in social networks don't have any problem in adding more links, the authors label these networks "aristocratic". So in some sense, the stocks in Table 1 can be considered an example of magnet/aristrocratic/important hubs which happen to be in several of the randomly selected funds in the sample and therefore having more connections with other stocks.

Table 2 .Degree Distribution

Degree	Freq	Freq%	CumFreq	CumFreq%	Representative
9	64	64.6465	64	64.6465	rimm
12	7	7.0707	71	71.7172	bp
14	2	2.0202	73	73.7374	t
15	1	1.0101	74	74.7475	c
16	2	2.0202	76	76.7677	total
17	2	2.0202	78	78.7879	ko
18	8	8.0808	86	86.8687	nvs
19	1	1.0101	87	87.8788	wmt
20	3	3.0303	90	90.9091	ibm
22	1	1.0101	91	91.9192	pep
24	2	2.0202	93	93.9394	ger
26	1	1.0101	94	94.9495	pg
27	1	1.0101	95	95.9596	aapl
28	1	1.0101	96	96.9697	csco
29	1	1.0101	97	97.9798	cvx
30	1	1.0101	98	98.9899	bac
34	1	1.0101	99	100	xom

The other important characteristic is that of clustering. The network of stocks has a clustering coefficient of .8598 which is high in comparison to an Erdös & Rényi random network with an average degree of each vertex equals that of the Stocks network (p=.12), which has a clustering coefficient of .1229. Another measure of how "actor importance" arises in a network is that of "centrality". Centrality is defined as a measure of "prominence" or "involvement" among actors (Wasserman and Faust). Again we compare the stock network against the random network. The centrality coefficient for the stock network is .30259 while the random network is .4830, thus indicating that the stock network's few "prominent" hubs are not that crucial for the connection between two nodes. The other characteristic mentioned in the introduction is that of "Diameter". This has to do with the "small-world" behavior of many social and/or physical networks and also related to "cohesiveness". A complete network is a network with maximum density (e.g. all nodes are connected). We defined the network density by the proportion of the number of lines in the network to the total possible number of lines (thus a complete network will have density 1). Again, the higher the density, the "smaller the world". Relating this to the "6 degrees of separation" book by Guare (1990), a complete network will have 1-degree of separation (that is you can go from any vertex to any other vertex in one step). Again for comparison purposes the cohesiveness of the Erdös & Rényi network proposed above is .1214 while the stock network in this study has a cohesiveness coefficient of .1261. A similar measure is the network diameter which relates to the longest shortest path from two vertices in the network. The random network has a diameter of 3 (it takes 3 or less jumps to get from any vertex to any other vertex). The stock network has a diameter of 6 (the longest shortest path is from "   im" to "teva") with an average degree distance between two vertices of 2.91.

Figure 10
Degree Distribution of the Stock Network (Solid Line is a Power Law Fit)



The network in Guare's book is much larger than the stock network, and compared to the Erdös-Rényi network with a diameter of 3 and average degree distance of 2.07. From this and from the power-law distribution of the vertex degrees (Figure 10) we can conclude that the stock network under study is a small-world network.

IV. CONCLUSION

In this paper we attempt to provide a new approach to visualizing the way stocks are affiliated to mutual funds. The analysis is based on well known Social Networks research in which the "actors" in the network are the stocks and the events to which they are affiliated (and therefore connected) are the mutual funds to which they belong. Data for the analysis consists of a random sample of 18 mutual funds from the Vanguard family of funds which yield 99 unique stocks (we considered the top ten holdings in each fund). The sample can be found in Appendix A. The data analysis was performed with Pajek (Batagelj and Mrvar, 2008) which is a well known Social Network analysis and visualization tool. The findings indicate that the stock network has a high coefficient of clustering (indicating "prominent" of "hubs" of stocks). These hubs of stocks indicate that the fund managers' stock selections are not made independently as if the network would had been that of a purely random graph (Erdös & Rényi, 1959) with similar expected number of links (n*p where p is the probability that a given vertex (stock) will be connected to another, p=.12, for this size of a network). Furthermore, the stock network has both a higher diameter and average degree distance between two vertices (6 and 2.91 .vs. 3 and 2.01 for the random graph) thus suggesting a small-world behavior, which in turn can be interpreted as a highly connected network of stocks and thus also suggesting a small number of stocks (mostly blue-chips) being part of the randomly selected mutual funds.

Even though the main trust of this paper is to encumber structure in a fairly complex network, we also found that what we expected to be a low clustered network even though the selection procedure of the funds was not. This may also suggest that mutual funds are not necessarily a tool for diversification and/or a superior investment strategy unless there is high turnover on the composition of the holdings (Wermers, 2000). Furthermore, the formation of supernodes may also suggest, to a great extent, that fund managers also experience herd behavior most commonly seen at the individual investor level.

REFERENCES

1. Amaral, L. A. N., Scala, A. and Barthe´ le´my M. and Stanley, H. E. (2000). Classes of Small-World Networks. Proc. Natl. Acad. Sci. USA. 97, 11149–11152.
2. Barthe´le´my, M. and Amaral, L. A. N. (1999). Phys. Rev. Lett. 82, 3180–3183.
3. Batagelj V. and Mrvar A. (2003). Pajek - Analysis and Visualization of Large Networks. In Juenger, M., Mutzel, P. (Eds.): Graph Drawing Software. Springer (series Mathematics and Visualization), Berlin.

4. Batagelj V. and Mrvar, A. (2008). Pajek – Program for Large Network Analysis. http://pajek.imfm.si/doku.php .
5. Davis, A., Gardner, B. B. & Gardner, M. R. (1941). *Deep South.* Univ. of Chicago Press, Chicago.
6. Davis, G. F., Yoo, M. and Baker, W. E. (2003). The Small World of the American Corporate Elite, 1982-2001. *Strategic Organization*; 1, 301-326.
7. de Nooy W., Mrvar, A. and Batagelj, V. (2005). Exploratory Social Network Analysis with Pajek, Structural Analysis in the Social Sciences 27, Cambridge University Press.
8. Erdös, P. and Rényi, A. (1959). On Random Graphs. Publ. Math. 6, 290-297.
9. Galaskiewicz, J. (1997). An Urban Grants Economy Revisited: Corporate Charitable Contributions in the Twin Cities, 1979–81, 1987–89. Administrative Science Quarterly 42,445–71.
10. Mark Granovetter. (2005) The Impact of Social Structure on Economic Outcomes. Journal of Economic Perspectives 19, 33-50.
11. Guare, J. (1990). Six Degrees of Separation: A Play (Vintage, New York).
12. Mariolis, P. (2001). Interlocking Directorates and the Control of Corporations. Soc. Sci. Quart. 56, 425-439.
13. Newman, M.E.J., Strogatz, S.H. and Watts, D.J. 'Random Graphs with Arbitrary Degree Distributions and Their Applications', Physical Review E 64, 026118. (2001).
14. Park, H. W. and Nam, I. Hyperlink-affiliation network structure of top web sites: examining affiliates with hyperlink in Korea. J. Am. Soc. Inf. Sci. Technol. 53-7, 592-601. (2002).
15. Pool, I. & Kochen, M. (1978). Soc. Netw. 1, 1-48.
16. Reka, A., and Barabasi, A-L. (2001). Emergence of Scaling in Random Networks. Science 286, 509-512.
17. Wasserman, S. and Faust, K. (1994). Social Network Analysis. Cambridge University Press.
18. Watts D.J. and Strogatz, S.H. (1998). Small Worlds. Nature, 393, 440-442).
19. Watts, D. J. (1999). Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton Univ. Press, Princeton, NJ.
20. Wermers, R. (2000). Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses. Journal of Finance, 55-4, 1655-1703.

APPENDIX

A: Affiliation Network

VAAPX	bac	c	csco	cvx	ger	jnj	msft	pg	t	xom
VBINX	bac	c	csco	cvx	ger	jnj	msft	pg	t	xom
VCVLX	cmcsa	csco	sgp	nfx	nwa	qcom	bac	f	ubs	wye
VDAIX	aig	cvx	ger	ibm	jnj	ko	pep	pg	wmt	xom
VDIGX	adp	cvx	mdt	msft	pep	pg	spls	total	wmt	xom
VEIEX	amov	chl	gzpfy	pbr	ptrbf	rio/p	rlniy	ssnlf	teva	tsm
VEIPX	bac	cop	cvx	ger	jpm	mo	t	usb	vz	xom
VEURX	bp	bstnf	glaxf	hbcyf	nok	nsrgf	rhhbf	tefof	ttfnf	vdfof
VFINX	bac	c	csco	cvx	ger	jnj	msft	pg	t	xom
VFTSX	aapl	abt	bac	goog	gs	intc	jpm	mrk	wb	wfc
VFWIX	bp	hbcyf	nok	nsrgf	rydaf	swaf	tefof	tm	ttfnf	vdfof
VGENX	bhp	cop	cvx	eipaf	ptrbf	rds/b	slb	total	wft	xom
VGEQX	aapl	csco	de	ge	gild	goog	intc	ko	msft	pep
VGHCX	abt	azncf	frx	lly	mck	nvs	rhhbf	sgp	snynf	tkphf
VHCOX	abi	asml	biib	fdx	glw	mon	nvda	nvs	rimm	symc
VHGEX	eonaf	fcx	gs	hkxcf	ing	lukfy	mbt	mro	ntdof	rydaf
VIGRX	aapl	csco	goog	hpq	ibm	intc	msft	pep	pg	wmt
VQNPX	aapl	amin	c	ger	ibm	pg	qcom	vlo	xom	xto

B: Least Connected Stocks

Symbol(Name)	No. Edges
Rimm (Research in Motion)	9
Mon (Monsanto)	9
Nvda (Nvidea)	9
Bib (Biogen)	9
Asml (ASML Intl)	9
Glw (Corning)	9
Fdx (FedEx)	9
Abi (Applera)	9
Symc (Symantec)	9
Wye (Wyeth)	9
Cmcsa (Comcast)	9
Nwa (North Western Airlines)	9
Ubs (UBS Ag)	9
S (Spring Nextel)	9

Nfx (Newfield Exploration)	9
Aig (American International Group)	9
Adp (Automatic Data Processing)	9
Mdt (Medtronic)	9
Spls (Staples)	9
Gzpfy (Gazprom)	9
Chl (China Mobile)	9
Amov (America Movil)	9
Ssnlf (Samsung)	9
Rlniy (Reliance Industries)	9
rio/p (Companhia Vale Ads)	9
Rio (Companhia Vale do Rio Doce)	9
rds/b (Royal Dutch Shell plc)	9
Eipaf (Eni Spa Roma)	9
Slb (Schlumberger Limited)	9
Wft (Weatherford International Ltd.)	9
Bhp (BHP Billiton Ltd.)	9
Mo (Motorola)	9
Vz (Verizon Communications Inc.)	9
Usb (US Bancorp)	9
Glaxf (Glaxosmithkline Plc)	9
Bstnf (Banco Santander Sa)	9
Mrk (Merck & Co. Inc.)	9
Wfc (Wells Fargo & Co)	9
Wb (Wachovia Corp.)	9
Amin (American International Industries Inc.)	9
Qcom (Qualcomm)	9
Vlo (Valero Energy)	9
Xto (Xto Energy)	9
Hpq (Hewlett-Packard Co.)	9
Eonaf (E ON AG)	9
Mro (Marathon Oil Corp.)	9
Ing (ING Groep NV)	9
Ntdof (Nintendo Co Ltd)	9
Fcx (Freeport-McMoRan Copper & Gold Inc.)	9
Mbt (Mobile Telesystems OJSC)	9
Hkxcf (Hong Kong Exch & Cle)	9
Gild (Gilead Sciences Inc.)	9
De (Deere & Co.)	9
Ge (General Electric)	9

Tm (Toyota Motor Corp.)	9
Swaf (Sekt Wachenheim)	9
Sgp (Schering-Plough Corp.)	9
Lly (Eli Lilly & Co.)	9
Frx (Forest Laboratories Inc.)	9
Snynf (Sanofi-Synthelabo Sa)	9
Mck (McKesson Corp.)	9
Azncf (Astrazeneca Plc Ord)	9
Tkphf (Takeda Pharmaceutica)	9