Where's the Money? The Social Behavior of Investors in Facebook's Small World

Liang Yuxian Eugene
Management Information Systems
National Cheng Chi University
Taipei, Taiwan
ye.eugene@gmail.com

Soe-Tsyr Daphne Yuan
Management Information Systems
National Cheng Chi University
Taipei, Taiwan
yuans@nccu.edu.tw

***Abstract*—Are investing activities dependent on social relationships? In our research, we apply social network analysis to the field of investing behaviors and discover that investors have a tendency to invest in companies that are socially similar to them. While traditional studies on investing behavior tend to focus on factors like psychology, opinions, investing experience etc, they fail to consider social relationship as an important factor. In this paper we provide general rules of thumb that are useful for companies seeking funding from investor. These rules of thumb are generated by analyzing the social relationships between investors and companies found within the small world of Facebook.**

Keywords: social network analysis, investing behavior, Facebook, startup, data mining

I. INTRODUCTION

Do social relationships affect investment behavior? While many studies have been devoted to understanding investment behaviors most lack focus on the aspect of social relationship. We believe that social relationship between an investor and a company is important.

The general intuition is that a positive correlation exists between social relationship and investment activities. Take for example the idea of homophily [16], where "birds of a feather flock together": individuals tend to associate and bond together with others due to similarities. Such social relations are likely found within our small world [13,17] via weak ties [14]. The main research question we want to answer is: Does social similarity lead to investment? Do investors fund companies that are socially similar? To test for such trends, we need to measure similarities between actors of social networks. In the field of social network analysis, there are numerous ways to measure similarities: Common neighbors, shortest path between two actors in a social network, Adamic/Adar, Jaccard Coeffecient, Preferential Attachment are some examples.

Using common measures of similarity in a social network context, our research shows strong evidence that similarity in social relationships has strong correlation with investment behavior.

Our contributions to the literature are as follows:

1) Marriage of social network analyis with investing behavior: We build a social network using data from Crunchbase[1], the largest public database with profiles about companies. Using this social network, we explore how similarity between investors and companies affect investing behavior. To the best of our knowledge, our work is first to use data from Crunchbase as a social network for research purposes.

2) Providing general rules of thumb for companies seeking investment: Our recommendations for companies seeking investment are based on intuitive and common similarity measures to show where companies can find potential investors within their social network. Using these general rules, we hope to increase companies' chances of getting funded from investors.

3) Insights to collective behavior of investors: Our research also provides a collective overview as to how investors invest within a social network.

II. RELATED WORK

We have two parts for related works since our research focuses on the use of social network analysis on investment behavior: previous research on investment behaviors and previous research on social network analysis.

A. Previous Research on Investment Behaviors

Prior studies on investment behaviors can be categorized into six categories based on the type of factors that drive investment behaviors. First: personal opinions [6] such as of finance professors, and found out that it does not affect investment behaviors. Second, investment experience, Hege and Schwienbacher [8] found out that novice investors than to invest slower than experienced ones, but the size and value of investments made by novice investors tend to be larger. Thirdly, geographic location: Grinblatt and Keloharju [9] discovered geographic location affects investing behaviors: foreign investors in Finland tend to purchase past winning stocks and sell past losers. On the other hand, domestic investors sell past winning stocks and purchase losing stocks. Fourth, online and offline communities plays a part too: Tan and Tan [5] explored the roles played by online and offline communities and discovered that offline communities are more influential over investing behaviors. Fifth, psychological

[1] CrunchBase – http://www.crunchbase.com

factors: Bakker, Hare, Khosravi and Ramadanovic [7] on the other hand investigated into psychological factors that impact market evaluation and found out that trust and social influence affects the stability of investment markets. Lastly, genetics: Amir, Henrik and Stephan [4] investigated the relationship between genetics and investment behavior by studying the investment behaviors of identical and fratenal twins. They discovered that "a genetic factor" explains up to a third of twins investing behavior, though not long lasting.

B. Previous Research on Social Network Analysis

There are numerous studies on social network analysis. More importantly in recent years we begin to see the marriage of social network analysis with management science, computer science and other fields, giving rise to what most of us term as "social computing" or "network science".

Common social network analysis topics and its relevant techniques and applications are, but not limited to centrality analysis [18], community detection [21,22,23], link prediction [11,29], label prediction [25,26], information diffusion [23,24] and team formation [28,29].

Other related work includes statistical features of networks [1, 3] such as information networks, collaboration networks, biological networks and social networks.

The similarities of the above applications is that the use of social network analysis techniques often improve the performance of the solution for the given problem domain. We see the use of algorithms or similarity measures ranging from Common Neighbors, shortest paths, Katz, PageRank, Jaccard Coeffecient, Adamic/Adar etc or its variants to help provide measures.

III. SOCIAL BEHAVIOR OF INVESTORS

While various factors have been proposed for explaining investing behaviors, little or no studies have incorporated the use of social network analysis of investment behaviors. Our intuition is that investing is an act based on social relationships, be it personal or professional. We propose that investors are more likely to invest in a company where they are of greater similarity in a social network context. The notion of similarity between two vertices in a social network has been studied extensively, more notably in the area of social sciences. Algorithms and methodologies have also been developed to measure similarity.

For the purposes of our research, we use the measures covered in Section II extensively to compute the similarities of Investors and Companies within a social network. As such, we propose that Investors are more likely to invest in Companies when they are of greater similarity in a social network context.

A. CrunchBase Dataset

The CrunchBase dataset is TechCrunch's open database with information about startups, investors, trends, milestones, companies etc. and it relies on the web community to edit most of its pages. It consists of a rich source of companies, people, and financial organization information of the technology world in the United States. As of 1st April 2012, the profiles found at CrunchBase consisted of 86,224 companies, 114,406 persons, 7600 financial organizations, 4,171 service providers, 27,290 funding rounds and 6453 acquisitions.

The CrunchBase dataset allows public access of its data via a JSON API, and we've collected a local copy of the data on February 2012, which consisted of about 95% of the dataset compared to 1st April 2012.

While it may not be obvious, the CrunchBase dataset represents a rich multi-modal social network of investors and companies. For instance, each company shows a list of people who are currently (or previously worked) for a company; drilling further we get to see the person's profile which states his list of companies (or financial organizations) which he is involved in. For the purposes of our research, we gathered data related to companies, persons and financial organizations.

B. Data Selection

Since we are interested in the investing behaviors of investors within a small world context, we chose Facebook as the seed node, and gathered People, Companies and Financial Organizations found within its social, funding and investment relationships within 4 degrees of separation from Facebook.

We selected Facebook as the seed node due to the company's meteoric rise in the social network industry and its recent IPO. We chose 4 degrees of separation as a cutoff point as opposed to 6 degrees of separation due to the fact that recent advances in technology has somewhat reduced the degrees of separation between people as shown in [13]. In addition, there are limits to the "Horizon of Observability" [10] from the viewpoint of using Facebook as a seed node.

We identified 11916 companies, 1122 financial organizations and 12127 people within 4 degrees of separation from Facebook if we take in account of both Investment and Social relationships. However, there are only 1152 people, 922 financial organizations and 7745 companies if we take into account entities via social relationships. Similarly, if we take into account of entities via investment relationships only, there are 11756 people, 6634 financial organizations, 756 companies. There are overlap between entities found via social and investment relationships.

C. Definitions and Examples

1) People: People refer to founders, executives and other persons working for a particular company or organizations. Examples from our dataset include Mark Zuckerberg and Peter Thiel.

2) Companies: Some examples of *Companies* include Google, Facebook and Microsoft.

3) Financial Organization: Financial Organizations are organizations that typically perform the act of investment on Companies. Prominent examples in our dataset include Accel Partners and Digital Sky Technologies.

4) Investors: Investors consists of *People*, *Companies* and *Financial Organizations*. This is due to duality of roles played by People, Companies and Financial Organizations in the CrunchBase dataset. For example, companies like Microsoft plays the role of a *Company*, yet performed an act of investment on other companies such as Facebook in the early days. Similarly, Peter Thiel is a Person entity, yet invested in Facebook.

5) Social Graph, G_{social}: We define G_{social} (V_{ocial}, E_{social}) as a undirected graph, where vertices can be made up of either a *People*, *Companies* or *Financial Organizations*, while E_{social} is formed when a particular Person has a relationship (such as employment) with a *Company* or *Financial Organization*.

6) Investment Graph, $G_{investment}$: We define $G_{investment}$($V_{investment}$, $E_{investment}$) as a directed graph, where vertices can be made up of either a Investors or Companies, while $E_{investment}$ is formed when an Investor invests in a particular Company.

D. Methodology, Algorithms and Intuition

The algorithms used for analysis are generally adapted from graph theories and social network analysis. The general intuition is that the greater the similarity between an Investor and a Company, the greater the likelihood that the Investor will invest in that particular Company.

We want our research to benefit companies who may or may not have strong background in social network analysis. Hence, we selected the following 5 methods for our analysis due to its simplicity and ease of understanding: shortest paths, Adamic/Adar, Jaccard Coeffecient, Common Neighbors and Preferential Attachment. We initially conducted experiments on G_{Social} and $G_{Investment}$ based on the metrics in [1] in order to compare the graphs on a global scale, but we find that the metrics provided limited information as to how social relationships can affect investments. Therefore, we chose to use graph distance (shortest paths) and methods based on node neighborhoods (Jaccard Coeffecient, Adamic/Adar, Common Neighbors and Preferential Attachment) as it allows us to do a one-on-one comparisons between each *Investor* and *Companies*. In addition, since we are interested in the small world of investing behaviors, methods based on node neighborhoods should provide us with insights as to how social similarity can affect investment behavior. All methods used for our analysis assign a score (x, y) to pairs of nodes <x, y>, based on the input graph G_{social}. Nodes X and Y are defined as follows: Node X represents *Investor*, while node Y denotes *Company*. This is because we want to compare the similarities of Investors and Companies for the purposes for our research. No comparisons are made when node X equals node Y. We define the set of neighbors of node x to be $\Gamma(x)$.

In general, the greater the similarity based on the scores, the greater the likelihood of investment. The algorithms used for comparing similarities are as follows:

1) Shortest Path: We simply consider the shortest path between *Investors* and *Companies*. The general intuition is that the "closer" *Investors* are to *Companies* (and vice-versa) the more likely that *Investors* will invest in such *Companies.* We define score (x, y) to be the length of the shortest path between an *Investor* and a *Company.*

2) Adamic/Adar: Adamic and Adar [3] considered similarity between two personal homepages by computing features of the pages and defining the similarity between two pages to be:

$$\sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{\log|\Gamma(z)|} \qquad (1)$$

We consider the similarity feature to be the common neighbors.

3) Jaccard Coeffecient: The *Jaccard Coefficient* measures the probability that both x and y have a feature f, for a randomly selected feature f that either x or y has. Here, we take f to be neighbors in G_{social}, leading us to the measure score

$$\frac{|\Gamma(x)\cap\Gamma(y)|}{|\Gamma(x)\cup\Gamma(y)|} \qquad (2)$$

4) Common Neighbors: It is considered as the most direct implementation. According to Newman, the general intuition is that the number of common neighbors of node X and node Y has a correlation with the probability that they will collaborate in the future, under the context of a collaboration network. The Score(x,y) for common neighbors is defined as follows:

$$|\Gamma(x)\cap\Gamma(y)| \qquad (3)$$

5) Preferential Attachment: Preferential Attachment suggests that the probability that a new edge has node x as an endpoint is proportional the current number of neighbors of x [17]. The Score(x, y) for preferential attachment is defined as follows:

$$|\Gamma(x)|\cdot|\Gamma(y)| \qquad (4)$$

We use these algorithms to compare each pair of *Investor* and *Company* node in the G_{Social}. For each pair, we take note of the score for each algorithm and mark if the *Investor* invested in the *Company* or not.

IV. FINDINGS AND DISCUSSION

The findings matched our intuition in general: the greater the similarity, the more investments are likely to occur. This is the case based on results using shortest path, Adamic/Adar and Jaccard Coefficient. There are also counter-intuitive findings, such as less investment activities occurring where an *Investor* and a *Company* have more common neighbors. Here we discuss our findings in details:

A. The greater the similarity between an Investor and a Company, the more likely that investments will occur

In the case of our experiment using Shortest Path on G_{Social}. We find that 100% of the investments occurred at shortest path lengths at 14 and below. Approximately 49% of the investment occurred at Shortest Path of 7 and below. This means that no investment activities are found when the shortest path is greater than 14. In Figure 2, we can see that bulk of the investment activities concentrates between shortest path lengths of 2 to 7. Investment activities decreases when shortest path length is greater than 7; investment activities decreases as the shortest path length increases. This is a stark contrast if we take into account of both *Investment* and *Social* relationships where the the distance between any 2 pairs of entities will range from 1 to

8; G_{Social} is a subset of both social and investment relationship of the dataset and paths between an Investor and a Company may turn out to be NULL or greater than 8 as certain companies, financial organization and people are not found in $G_{Social.}$

The evidence presented here shows that investments are very likely to occur within the small world, and that the shorter the shortest path (greater similarity) the greater the likelihood of investments.

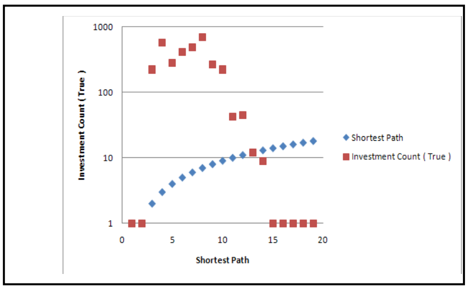

Figure 1. Relationship between Shortest Paths and Occurences of Investments

We see similar trends in our experiments using Adamic/Adar and Jaccard Coeffecient. In general, investment occurences increases as the score increases, this is shown in Figure 3 and 4 respectively:

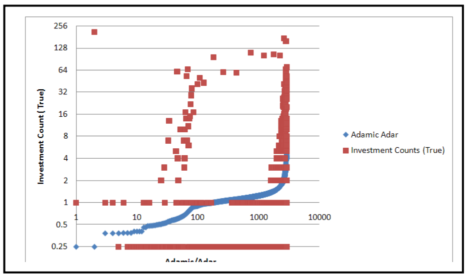

Figure 2. *Relationship between Adamic Adar and Occurences of Investments*

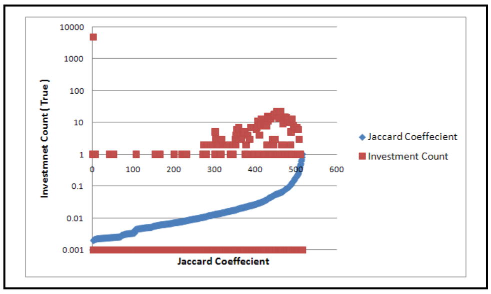

Figure 3. Jaccard Coeffecient vs Occurences of Investments

B. Investments can still occur even when an Investor has nothing in common with the Company.

While experiments using *Shortest Paths*, *Adamic/Adar* and *Jaccard Coefficient* have shown intuitive results, there are outlier cases where investments occurred despite similarity scores of zero or when shortest paths do not exist between an Investor and a Company.

We analyzed deeper into such cases and discovered there are 2 overwhelming trends which: the *Company* in question have received at least 1 round of funding (seed or angle funding) and that the company belong to a popular industry, such as the Web, Software and Mobile industries.

For example in the case of Shortest Paths we noticed that bulk of the investments occurs in funding round A, signifying that companies receiving investments from investors with shortest path of NULL have at least received one round of investment. Similar findings were found in our experiments for Adamic/Adar and Jaccard Coeffecient approximately 48% of the investments occurs in Round A and or later rounds, if we exclude unattributed funding rounds in the case for Adamic/Adar. Investment activities in this case peaked at Round C. For experiment using Jaccard Coefficient, approximately 57% of all investments occurred in Round A and later rounds, excluding unattributed funding rounds.

Such outlier investments happen in popular industries, such as the Web, Software and Mobile industries. This trend holds across our experiments for Shortest Paths, Adamic/Adar and Jaccard Coeffecient. We noticed that the Web, Software and Mobile industries happen to be in the top 3 of the investment activities. This implies that being in popular industries can help mitigate the lack of social connections.

C. Having too much similar neighbors can result in less investment opportunities.

We expected that having greater number of common neighbors or having greater Preferential Attachment score results in greater occurrences of investments. However, our experiment results showed counter-intuitive results: Firstly, the greater the number of common neighbors, the less likely investments will occur: For a start, consider Figure 5*:*

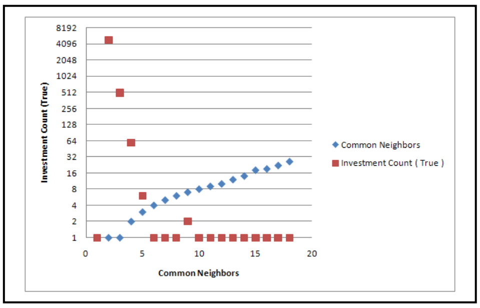

Figure 4. Common Neighbors vs Investment Occurences

We can see that as common neighbor increases, investment occurrence actually decreases. While this may seem counter intuitive, this makes sense if we consider the type of neighbors the Investors and Companies have in common. For instance, if the neighbors are in fact competitive, having more common neighbors can mean greater amount of competition. Given greater amount of competition, there might be a chance where the industry provides limited growth for individual companies, and hence limited returns on investment from the Investors point of view. We see a similar trend in the case of Preferential Attachment:

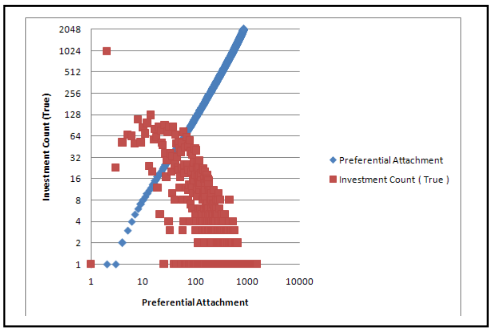

Figure 5. Preferential Attachment vs Investment Occurences

V. SUMMARY AND FUTURE WORK

We have presented and discussed about social behaviors of investors in a small world context. It turns out that investment activities and hence potential investors are likely found within our social networks. Using easy to understand and intuitive similarity measures, we find that the greater the similarity between an *Investor* and a *Company*, the greater the likelihood that the *Investor* will invest in that *Company*. To summarize, our findings are as follows:

1) Being "closer" to an investor leads to greater occurrence of investment: Our results show that bulk of the investment activities occur within shortest path lengths of 7 and in general, investing activites occur where similarity scores are higher.

2) Being in a popular industry helps if there is limites social relationship: We find that investment activities can still occur when there are no social relationships. But this happens when the companies in question are in popular industries and when they have largely received at least 1 round of funding.

3) Having too much common neighbors can result in less investment activities: Our research also find that investment occurrences tend to decrease as number of common neighbor increases. This could signify that having too many similar neighbors can mean greater competition and hence less room for growth.

To conclude, we hope that our research can help companies enhance their chances of getting funded and to inspire new applications of social network analysis on more and different domains, which in our case, investing behaviors. Lastly, we hope to provide a fresh take on factors affecting investing behaviors. In the future, we hope to develop a prediction model base on these findings.

REFERENCES

[1] David Liben-Nowell, Jon Kleinberg. The Link Prediction Problem for Social Networks. Journal of the American Society for Information Science and Technology, 58(7):1019 - 1031, May 2007.

[2] Lada A Adamic and Eytan Ada. Friends and Neighbors on the Web

[3] M E. J Newman. The Structure of Collaborative Network. www.pnas.orgycgiydoiy10.1073ypnas.021544898

[4] Amir Barnea, Henrik Cronqvist and Stephan Siegel. "Nature or Nurture: what determines investor behavior ?". Journal of Financial Economics 98 (2010) 583–604

[5] Tan, W.-K., Tan, Y.-J. "An exploratory investigation of the investment information search behavior of individual domestic investors." Telemat. Informat. (2011), doi:10.1016/j.tele.2011.09.002

[6] James S.Doran, David R.Peterson, Colby Wright. "Confidence opinions of market efficiency,and investment behavior of finance professors." Journal of Financial Markets13 (2010) 174–195

[7] L. Bakker, W. Hare,, H. Khosravi and B. Ramadanovic. "A social network model of investment behaviour in the stock market." Physica A 389 (2010) 12231229.

[8] Pierre Giot, Ulrich Hege, Armin Schwienbaher. "Expertise of Reputation? The Investment Behavior of Novice and Experienced Private Equity Funds."

[9] Mark Grinblatt and Matti Keloharju. "The Investment Behavior and performance of various investor types: a study og Finland's unque data set."

[10] Friedkin, Noah. "Horizon of Observability and Limits of Informal Control in Organizations", Social Forces 1981

[11] J. Leskovec, D. Huttenlocher, J. Kleinberg.Predicting Positive and Negative Links in Online Social Networks". ACM WWW International conference on World Wide Web (WWW), 2010

[12] M. E. J Newman, "The Structure and function of complex networks", SIAM Review 45, pp 167 – 256, 2003

[13] Lars backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, Sebastiano Vigna. " Four Degrees of Seperation"

[14] Mark S. Granovetter. "The Strength of Weak Ties", American Journal of Sociology, Volume 78, Issue 6 (May, 1973)

[15] M. E. J Newman, "Clustering and preferential attachment in growing networks", Physical Review E, Volume 64, Issue 2, 11 April 2001

[16] Miller McPherson, Lynn Smith-Lovin, James M Cook, "Birds of a Feather: Homophily in Social Networks", Annual Review of Sociology, Volume 27: 415-444, August 2001

[17] Jeffrey Travers and Stanley Milgram, "An experimental study of the small world problem", Sociometry, Volume 32 No 4 December 1969, Page 425 – 443

[18] Linton C Freeman, "Centrality in Social Networks Conceptual Clarification", Social Networks (1978/79) 215 – 239

[19] Woochang Hwang, Taehyong Kim, Murali Ramanathan and Aidon Zhang, "Bridging Centrality: Graph Mining from Element Level to Group Level", Knowledge Discovery and Data Mining Conference 2008

[20] E. Ravasz, A. L. Somera, D. A Mongru, Z.N Oltvai and A. –L. Barabasi, "Hierarchical Organization of Modularity in Metabolic Networks", Science Volume 297, 30 August 2002

[21] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", PNAS, Volume 99, No. 12, 11 June 2002

[22] M. E. J. Newman, "Modularity and community structure in networks", PNAS, Volume 103, Non 23, 6 June 2006

[23] Jure Leskovec, Anreas Krause, Carlos Guestrin, Chrisos Faloutsos, Jeanne VanBriesen and Natalie Glance, "Cost-effective outbreak detection in networks", KDD 2007

[24] David Kempe, Jon Kleinberg and Eva Tardos, "Maximizing the spread of influence through a social network", SIGKDD 2003

[25] Brian Gallagher, Hanghang Tong, Tina Eliassi-Rad and Christos Faloutsos, "Using ghost edges for classification in sparesly labeled networks", KDD 2008

[26] Tomasz Kajdanowicz, Przemyslaw Kazienko and Piotr Doskocz, "Labnel-dependent featuer extraction in social networks for node classification", SocInfor 2010, LNCS 6430, page 89 – 102

[27] Theodoros Lappas, Kun Liu and Evimaria Terzi, "Finding a Team of Experts in Social Networks", KDD 2009

[28] Mehdi Kargar and Aijun An, "Discovering Top-k Teams of Experts with/without a Leader in Social Networks", CIKM 2011

[29] L. Backstrom, J. Leskovec ,"Supervised Random Walks: Predicting and Recommending Links in Social Networks", ACM International Conference on Web Search and Data Mining (WSDM), 2011.