Startup Tribes: Social Network Ties that Support Success in New Firms

Emergent Research Forum (ERF) paper

Yang Yu

Rochester Institute of Technology

yyu@saunders.rit.edu

Victor Perotti

Rochester Institute of Technology

vperotti@saunders.rit.edu

Abstract

To answer the question how networking plays a role in entrepreneurial success, we have shown through the lens of social network analysis that online social network structure could be considered as efficient indicators to predict startups overall performance. Social network analysis techniques are applied on a big network, which is generated according the following and followers structure on Twitter.com for 644 ICT operating startup companies. A set of network structure related measures are developed to link with overall company performance. The main findings include the better the startups are connected within the startup community "tribe" and occupying the better "position" within the community, the more successful they are. In addition, no matter within or outside the community, the structure of one's friends is closely related to the business success also.

Keywords

Social network, startups

Introduction

Conventional wisdom holds that entrepreneurs should live in one of the central hubs if they wanted their business to thrive. This proved to be good advice, especially for entrepreneurs connected in startup circles, where a simple good idea has high probability to grow. The simple conclusion has historically been that networking is the key or at least one of keys for startup funders that push new ideas and innovation.

To answer the question how networking plays a role in entrepreneurial success, many researchers have explored this potential linkage between network structure and business success. In the literature, a wide variety of relationships among businesses have been studied, either on an organizational level or individual level. For example, on the organizational level, Uzzi (1997) investigated the relationship between suppliers and manufacturers and found that being embedded into a close-knit group of entrepreneurs is good for business performance. Burt (1979) studied the interlocking directorates, which refers to the practice of members of a corporate board of directors serving on the boards of multiple corporations. Consistent with subsequent studies (Pennings 1980; Richardson 1987), Burt found that interlocking directorates could help company to reduce environmental uncertainty and interlocked companies tend to perform better than non-interlocked companies. Some studies focus on relationships on the individual level. Raz and Gloor (2007) found that if a startup's founder has more formal/informal ties, the firm had a higher chance to survive the e-business bubble burst. We noted that most of this analysis so far has been based on "physical" ties and evidence has been provided that "real existing" network structure among entrepreneurs does impact business success.

In the modern environment of Internet and online social networks, entrepreneurial networks are extending beyond the scope of "real" or "face-to-face". Today, startup communities, particularly the high tech ones, can grow anywhere, locally or virtually. Taking advantage of social media and social network

platforms, people can attend all the startup events (from idea sharing, team building, product roadmap designing, to fundraising) without a physical distance limitation. Social network platforms furthermore reduces the distance between entrepreneurs to investors, whether crowdfunders, angel investors or VCs. It is obvious that the relationships built online influence startups. However, to what extent exactly do virtual social ties in online social networks impact the short term or long term of business performance? In this study, we try to determine if certain types of online social networking structure patterns relate to business success. In particular, we analyze the relationship between social network linkage patterns of the startup communities in the field of Information and Communication Technology (ICT) and their overall success.

Method

Data Gathering

We collected data from three major data sources for this study, Seed-DB, CrunchBase and Twitter. Seed-DB is a database of technology accelerators and their resident companies. At end of October, 2013, Seed-DB covers information about 172 accelerators world-wide and 2,921 companies accelerated. Begun in 2007 by TechCrunch.com, CrunchBase is a startup ecosystem database that contains the detailed information for startups such as founders, key employees, financial condition, acquisition news and other important activities. In addition, CrunchBase also includes information of investors.

We gathered all 172 accelerator programs[1] from Seed-DB as well as their accelerated companies. For each company, general funding information is gathered such as company name, website, cohort data, exit status and exit value and funding, etc. A web harvest robot was created to match company and collect detail information about the target company from CrunchBase.com automatically. The target company's physical detail information was parsed and loaded into our database, includes category, employees, location and social media accounts (e.g. Blog and Twitter). The data gathering yielded 731 companies who have Twitter accounts (Table 1), with 23 dead, 64 exited and 644 still operating. An R script based on TwitteR package was used to gather companies' Twitter account following and follower links. We used two main data collecting methods, both relying on the API functions provided by Twitter. First, we gathered detailed information of each user, and then collected the list of users each of them were following and followed with. Relationships in Twitter are directed. A twitter user a interested in the statuses of another user b could sign up as a "follower" of b. In the meanwhile, b's account name will be shown in "following" list of a. In total, there are 2,482,963 following and follower relationship found for those 731 startups, 72.4% of the ties are followers (Figure 1). Both follower and following distributions are skewed, the median follower ties are 823 and following ties is 411.

Startup category	Count
Consumer Web	154
Software	79
Mobile/Wireless	70
Enterprise	65
Advertising	45
eCommerce	42
Games/Entertainment	29
Analytics/Big Data	25
Education	22

---

[1] The 172 programs include almost all the accelerator in information and communications technology (ICT) sector, some famous ones such as Y Combinator, Techstars, etc.

Other[2]	188
Total	731

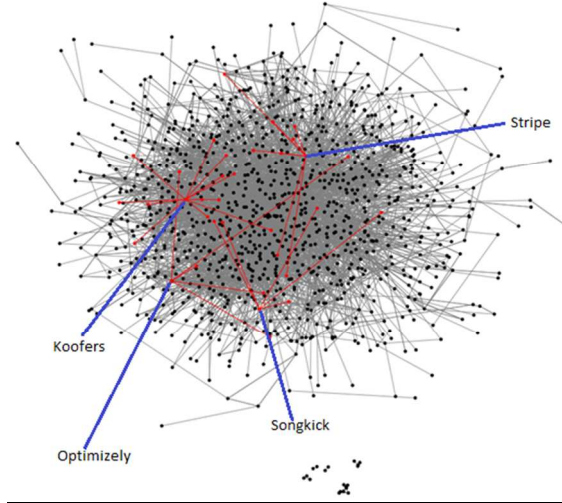**Table 1. Startup categories**

Figure 1. Follower network of startups

Measuring Startup Success

In this study, we focus on 644 operating startups only and sought to develop a simple measure to proxy the success of startups. To make the observations comparable, we removed those non-operating startups, including 23 dead and 64 exited ones. Because all the firms are still in operating, we could not take the cash out value (exit) as the proxy of success. There are several efficient and proven metrics using to measure the operating startups success in business world. Obviously, some metrics such as profit/deal flow, company stability, and sustainability are easy to "measure success". However, those metrics are not always available for startups. Because all of the startups in this project belong to ICT industry, we considered the total raised funding as a proxy of performance for each company. However, those startups that have been operating longer and have bigger size would have a clear advantage for fund raising. So, in addition we incorporate the length of the startup life and size of the company. Thus, we developed a measure similar to fundraising capability, $P_i = \frac{F_{i,t_0}}{L_i \times E_{i,t_0}}$ (1), in which P_i is the fundraising capability for company i, F denotes the total fund in the life cycle, L is the life length of startup company (by month, start counting from cohort date), E denotes the current (t_0) employee number. Although ignore the evolution curve of each startup on the timeline, this proxy still capture the businesses' overall capability of fundraising.

Social Capital, Online Network Ties and Twitter

Recent discussions related to social capital and social media usage have begun to distinguish ties into different categories such as ties for bridging and ties for bonding (Norris 2004; Woolcock and Narayan 2000). According to the literature, the bridging ties facilitate the dissemination of information and solidarity within rather weak ties and are likely to lead to social participation. However, the bonding ties are found within homogenous networks of like-minded people and are likely to result in support for the individual. Therefore, in order to obtain a clear picture of social effects of online communication via social network sites, one should take into account both forms of social capital. Hofer and Aubert (2013) argued

---

[2] The other category includes social networking, health/fitness, finance/venture, travel, search, fashion, hospitality/food, sports, etc.

that through its directed friendship model, Twitter use affects bridging and bonding online social capital differently. In their study, they consider bridging social capital as a result of following other Twitter users and bonding social capital as a result of being followed on Twitter. In this study, we differentiate both bridging and bonding ties within and out tribe, which could be defined as a homogenous group. We believe this classification could help people understand the ties pattern and roles better.

SNA & Measures

Basic concepts of graph

The basic idea of social media is simple. A network is a set of actors (also can be called elements or nodes) that are related with each other n some way. Given a finite set N of actors: $N = \{1, 2, \ldots, n\}$, and a finite number of relations $g_{ij} \in \{0,1\}$:

$$g_{ij} = \begin{cases} 1 & if\ there\ is\ a\ link\ between\ i\ and\ j \\ 0 & otherwise \end{cases}$$

A social network g is defined as a set of nodes N with links between them. By $N_i(g)$ we denote the neighborhood of node i in the network g, i.e., the set of nodes with which node *i* has a link:

$$N_i(g) = \{j \in N : g_{ij} = 1\}$$

The degree $d_i(g)$ of a node *i* in g is the number of *i*'s neighbors in g, i.e., $d_i(g) = |N_i(g)|$

Relations g_{ij} can represent any type of relationship, such as the husband-wife relationship, the employer-employee relationship or the friendship, etc.

Social network structure measures

One important feature of networks is the relative position of individuals in them. The position of a node always is considered as a key to understanding its role within a network(Wasserman 1994). In the literature, several metrics are proposed to describe the position an actor insider a network. In particular, the centrality position metrics measure the importance of the node(Wasserman 1994). Specifically, because the relations/links in this study represent communication channels between actors, these metrics measure the prestige and potential impact of the node on the distribution of information across the network.

Centrality is a structural characteristic of individuals in the network, meaning a centrality score tells you something about how that individual fits within the network overall (Ibarra 1993). Individuals with high centrality scores are often more likely to be leaders, key conduits of information, and be more likely to be early adopters of anything that spreads in a network. Low centrality individuals can be termed peripheral. Being peripheral can have advantages as well by protecting individuals from negative contagion and influence (think of the flu spreading or the ricocheting effects of that really deflating coworker) (Daly and Haahr 2007). Sometimes lower centrality is associated with less work overload in an organization.

Social network scientists have invented literally dozens of centrality measures to characterize slightly different aspects of structural positions in networks. In this study, we will import several the most commonly used centrality measures, Degree Centrality and Betweenness Centrality (Freeman 1977; Girvan and Newman 2002; Hofer and Aubert 2013; White and Borgatti 1994). In addition, we develop a new measure Tribe Centrality, to capture the link features within the startup club.

Degree Centrality

An important node is involved in large number of interactions. The degree centrality indicates how well a node is connected in terms of direct connections, i.e., it keeps track of the degree of the node (Girvan and Newman 2002). This measure can be seen as an index of the node's *communication activity*.

The *degree centrality $C_d(i; g)$ of node i in network g* is given by

$C_d(i; g) = \frac{d_i(g)}{n-1} = \frac{|N_i(g)|}{n-1}$ (2)

Following the study of Hofer and Aubert (2013), we measure the bridging social capital of a given startup by calculating out-degree, which counting how many following tie in its account to link other Twitter users, and measure the bonding social capital by calculating out-degree, which counting how many follower it has.

Betweenness Centrality

Centrality measures are linked closely to the notion of power. One very prominent representative is betweenness centrality(Freeman 1977), which measures the extent to which an actor is on the shortest path between any other pair of actors. Thus, the more connections depend on a given actor, the more power associated with that actor. In other word, an important node will lie on a high proportion of paths between other nodes in the network. The betweenness centrality is based on how important a node is in terms of connecting other nodes. It is useful as an index of the potential of a node for *control of communication(White and Borgatti 1994)*. In this study scenario, betweenness centrality is considered as a measure for social capital of brokerage.

By $p_i(kj)$ and $p(kj)$ we denote the number of geodesics between k and j containing $i \notin \{k, j\}$, and the total number of geodesics between k and j, thus,

The *betweenness centrality*

$$C_b(i; g) = \frac{2}{(n-1)(n-2)} \sum_{k \neq j; i \notin \{k,j\}} \frac{p_i(kj)}{p(kj)} \quad (3)$$

Tribe Centrality

Among the network ties for a startup's twitter account, some ties link to peers' twitter accounts and some ties link to unrelated accounts. There are 5,112 follow or following relationship were found among these 644 startups. In this study, we explore whether links within and across homogenous networks have different roles. To differentiate them from the overall centrality (including all ties), we design a new measure named **tribe centrality** to measure both degree centrality and betweenness centrality within the tribe. Here we define the tribe as the network composed by all the ICT startups. Figure 2 shows an example (a startup company named Sportlyzer) of links among startup tribe. Compare to outside following and follower, links among tribe has relative higher density.
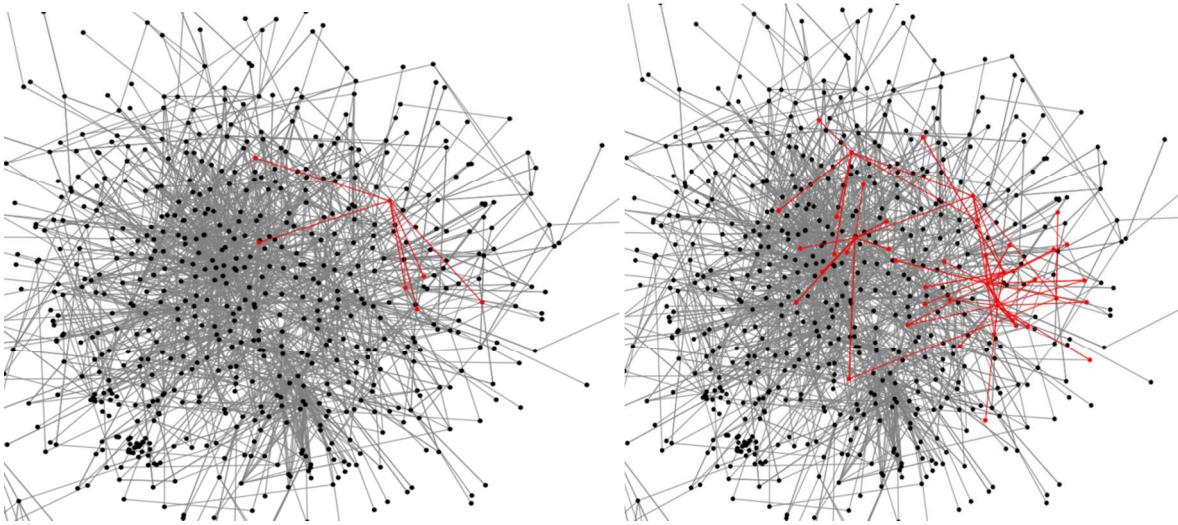

Figure 2. Sportlyzer's Tribe Links

Model and Results

A simple linear regression model was calculated to predict the overall startup success based on online social network structure features.

$$P_i = \alpha + \beta_1 C_{d_{in}}(i; g_o) + \beta_2 C_{d_{out}}(i; g_o) + \beta_3 C_{d_{in}}(i; g_t) + \beta_4 C_{d_{out}}(i; g_t) + \beta_5 C_b(i; g_o) + \beta_6 C_b(i; g_t) + \varepsilon_i$$
(4)

Model description

P_i represents the overall fundraising capability for a startup i *(formula 1)*.

$C_{d_{in}}$ represents the in-degree centrality (follower pattern) of startup i.

$C_{d_{out}}$ represents the out-degree (following pattern) centrality of startup i.

C_b represents the betweenness centrality of startup i.

g_o represents the overall network, which includes all the linked twitter accounts.

g_t represents the tribe network, which includes those 644 ICT startups only.

ε_i represents the startup-specific effect that capture the idiosyncratic characteristics associated with each startup.

Variable	Coefficient	Std. Err.	P
Constant	41.56	30.13	0.66
$C_{d_{in}}(i; g_o)$	-.17	0.11	0.35
$C_{d_{out}}(i; g_o)$	0.09	0.04	0.00***
$C_{d_{in}}(i; g_t)$	0.17	0.07	0.00***
$C_{d_{out}}(i; g_t)$	0.08	0.03	0.03**
$C_b(i; g_o)$	1.38	1.22	0.74
$C_b(i; g_t)$	1.93	1.17	0.00***

Note:

*** $p<.01$.

** $p<.05$.

* $p<0.10$.

Table 2. Linear regression estimation results for network structure features.

A significant regression equation was found $F_{(5, 638)}=20.28$, with an adjusted R^2 of 0.17. Table 2 shows the estimation results using linear regression for each twitter network structure features. Consistent with prior social network analysis and entrepreneurship literature, it was found that online social network structure <u>does</u> influence startup success. Specifically, across all studied companies, the structure of whom is followed significantly contributed to startup success. Furthermore, all network centrality measures within the tribe were positively associated with startup success: both the in-degree and betweenness centrality within the tribe had a positive impact on the overall performance. In other word, who your friends are matters, whether within or outside the tribal community. These structure of these friends are closely related to the business success. Also, who you are following in the community and the brokerage capability within the tribe are also important indicators for overall success.

An interesting fact is the lack of evidence that followers from outside the community influences startup success, neither brokerage in overall network. This finding suggests that who follows you from outside of the tribe has very little or will not affect the business success.

Discussion

In this study, we have shown through the lens of social network analysis that online social network structure could be considered as efficient indictors to predict startups overall performance. We have been able to show that the better the startups are connected within the startup community "tribe" and occupying the better "position" within the community, the more successful they are. We did not provide evidence showing direct relationships between overall followers number, overall betweenness centrality and startup performance.

While our analysis and results are still preliminary, there are a couple of limitations of the current study. For instance, although we differentiated the tie within and out tribe in the analysis, we still count all the links the same weight. In addition, in this study, we select ICT related startups only, which might have the same pattern of social network and cause generalizability issue. In addition, we still have not conclusively answered the question of causality between networking structure features and business success. We speculate the link of causality goes both ways but still need future study to detect.

References

Burt, R. S. 1979. "A Structural Theory of Interlocking Corporate Directorates," *Social Networks* (1:4), pp. 415-435.

Daly, E. M., and Haahr, M. 2007. "Social Network Analysis for Routing in Disconnected Delay-Tolerant Manets," *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*: ACM, pp. 32-40.

Freeman, L. C. 1977. "A Set of Measures of Centrality Based on Betweenness," *Sociometry*), pp. 35-41.

Girvan, M., and Newman, M. E. 2002. "Community Structure in Social and Biological Networks," *Proceedings of the National Academy of Sciences* (99:12), pp. 7821-7826.

Hofer, M., and Aubert, V. 2013. "Perceived Bridging and Bonding Social Capital on Twitter: Differentiating between Followers and Followees," *Computers in Human Behavior* (29:6), pp. 2134-2142.

Ibarra, H. 1993. "Network Centrality, Power, and Innovation Involvement: Determinants of Technical and Administrative Roles," *Academy of management journal* (36:3), pp. 471-501.

Norris, P. 2004. *The Bridging and Bonding Role of Online Communities*. Thousand Oaks, California: Sage Publications.

Pennings, J. M. 1980. *Interlocking Directorates:[Origins and Consequences of Connections among Organizations' Boards of Directors]*. Jossey-Bass San Francisco, Washington, London.

Raz, O., and Gloor, P. A. 2007. "Size Really Matters-New Insights for Start-Ups' Survival," *Management Science* (53:2), pp. 169-177.

Richardson, R. J. 1987. "Directorship Interlocks and Corporate Profitability," *Administrative science quarterly*), pp. 367-386.

Uzzi, B. 1997. "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness," *Administrative science quarterly*), pp. 35-67.

Wasserman, S. 1994. *Social Network Analysis: Methods and Applications.* Cambridge university press.

White, D. R., and Borgatti, S. P. 1994. "Betweenness Centrality Measures for Directed Graphs," *Social Networks* (16:4), pp. 335-346.

Woolcock, M., and Narayan, D. 2000. "Social Capital: Implications for Development Theory, Research, and Policy," *The world bank research observer* (15:2), pp. 225-249.