Analyzing Market Trends and Startups through Money Flow

Samhita Karnati
Advisor: Andrea LaPaugh
Independent Work Seminar 03: Information Discovery Through Relationships

Motivation and Goal

Since the dot com boom, the barrier to entry to start a company has gone down significantly. Now, someone with an idea can make a website or an app to provide their services. But not every marketspace has been as profitable others, and some startups have gone so far as to create new marketspaces. Regardless of the eventual success or failure of these startups, they all had to start somewhere – namely, funding. And once the successful startups become big enough, they usually get acquired or go public. So, at every stage of a startup, there is some distinct flow of money.

The goal of this project is to use this flow of money to try and identify interesting market trends and particularly pivotal startups. Understanding market trends and patterns of revolutionary startups over time is very valuable for both current startups and for future-founders.

Problem Background and Related Work

Past work on market trends has mainly been in relation to economic trends on a macro-level, and not related to interactions between funding entities, the actual companies, and the public. Firms like PwC and Strategy& have published strategy reports like the "2016 Technology Industry Trends" report,[1] looking at such economic trends and their impacts on the technology industry. In terms of past work that has used datasets related to individual entities, like information from TechCrunch, most published papers relate to prediction of events. An example is "A Supervised Approach to Predict Company Acquisition with Factual and Topic Features Using Profiles and New Articles on TechCrunch"[2] from Carnegie Mellon. This work focuses on M&A prediction and primarily utilizes qualitative characteristics of the organizations of interest (like founder profiles and company mission statements) instead of relationships between these organizations.

Approach

While there has been research into market trends, this work has mainly been on an economic level. Given that many new markets are created as a result of startups and that there is so much money that flows in and out of startups, I think that using data on startups and the relationship between them and entities that put money and take money out of startups is a new and interesting approach to look at market trends.

On the other hand, crunchbase data has been used mainly for prediction purposes. This information is most useful to financial institutions in determining where to invest their money, but not as useful for the actual startups themselves. Thus, using crunchbase's quantitative data for trend and pattern detection is a novel approach that is of interest to a different group of people than the current work serves.

---

[1] Surr, Pierre-Alain. "2016 Technology Industry Trends." *Strategy&*. Strategy& and PwC, n.d. Web. 3 Oct. 2016.
[2] Xiang, Guang, et al. "A Supervised Approach to Predict Company Acquisition with Factual and Topic Features Using Profiles and News Articles on TechCrunch." *ICWSM*. 2012.

Plan

Part 1: Dataset acquisition
The datasets for this project will be acquired from crunchbase, which can be easily accessed and used via their API endpoints. Crunchbase provides a zip file called the Daily CSV Export, which has up-to-date information on each of over 466 thousand companies.[3] While this export is seemingly just a snapshot on the day it was accessed, it also contains historical data, including funding rounds. Thus, this dataset is enough to construct digraphs for money flow for the present moment in time, as well as before and after pre-selected acquisitions, mergers, funds, etc. There are separate CSV files for acquisitions, investments, IPOs, funds, and funding rounds.

Part 2: Making the maps
Each map I will make will have a node for each VC firm/investor, each startup included in the dataset, and the public. I will use NetworkX – "a Python language software package for the creation, manipulation, and study of the structure, dynamics, and function of complex networks"[4] – to construct the digaraphs (most likely using the graph generation functions from NetworkX), where edges represent money flow and the weight of each edge is the amount of money involved in the transaction.

I will then identify important events that have happened in the last 10 years in the startup world and construct maps from before and after these events. Events of interest include the funding of Google, the funding of Twitter and Facebook (advent of massively successful social networking platforms), funding of Uber (creation of a whole new marketspace), funding of Airbnb, the acquisition of LinkedIn, and the IPO of Twilio, to name a few. The CSV files include dates with each event, so to create the before maps, I will only scrape up until right before the event of interest. In order to make the maps, I will combine acquisition, funding/investment, and IPO data. When there is an acquisition, the node will disappear from the graph (so before/after graphs will be used to explore these events).

Part 3: Analysis
In order investigate clusters in the maps that might correspond to marketspaces, I will use functions from NetworkX like those that compute the clustering coefficient and centrality. After I make the maps, I will have more insight into which NetworkX methods will be of use. To analyze how events affected the structure of the maps over time, I will isolate affected nodes and compute the same measures as before, and compare.

Evaluation

The central questions that are being investigated in this project relate to market trends and the impact of changes in money flow on market structure. One way to evaluate "pivotal" events is to check if they have tangible affects to the maps' structures. For example, many people have said that Uber pioneered the peer-to-peer marketspace, but are there actually quantifiable effects? Does the money flow show this new marketspace creation? The measures computed before and after supposedly pivotal events will give a quantitative way to evaluate this importance. On the flip side, if we assume that these events are important, then is money flow the best way to quantify importance? This latter question is really an evaluation of the methods and approach used in this project.

---

[3] "Daily CSV Export." *Crunchbase Data*. Crunchbase, n.d. Web. 3 Oct. 2016.
[4] *NetworkX 1.11 Documentation*. NetworkX, n.d. Web. 3 Oct. 2016.