SI 608 Project Proposal
Title: The Startup World
Team members: Phoebe(Hui) Liang, Mei Fu, Tong Yin, Poheng Chen

Introduction

In this project, we are interested in the networks between startup companies and their investors. More specifically, we would like to investigate the following questions:

1) The schools the founders of startup companies and investors graduated from
2) the locations of startup companies and the locations of investors
3) the industries of the startup companies and the past companies the investors have invested in
4) Which are the most invested companies in the industry we are interested?
5) On the qualitative side, we could use the depth-first search to find the path in a graph to answer questions like how can one entity get in touch with a particular venture capital through its existing networks in limited time and resources.

Additionally, we plan to focused on startup companies that are focused on big data analytics within the United States.

Data Source

- Crunchbase API:
  We can get a list of Organization Summary for every organization in the Open Data Map. The organization summary will return data in JSON format (https://data.crunchbase.com/v3/docs/organizationsummary).
  Also, we can get a list of Person Summary in JSON format. (https://data.crunchbase.com/v3/docs/personsummary). We can then match the organization with people profiles.

  Below is an example of Organization Summary:

JSON


```
{
  "type": "OrganizationSummary",
  "uuid": "df6628127f970b439d3e12f64f504fbb",
  "properties": {
    "permalink": "facebook",
    "api_path": "organizations/facebook",
    "web_path": "organization/facebook",
    "name": "Facebook",
    "primary_role": "company",
    "short_description": "Facebook is an online social networking service that enables its users to connect with friends and family as well as make new connections.",
    "profile_image_url": "https://res.cloudinary.com/crunchbase-production/image/upload/v1408491700/ypqf483smhnqo0rh6mff.png",
    "domain": "facebook.com",
    "homepage_url": "http://www.facebook.com",
    "facebook_url": "https://www.facebook.com/",
    "twitter_url": "https://twitter.com/facebook",
    "linkedin_url": "http://www.linkedin.com/company/facebook",
    "city_name": "Menlo Park",
    "region_name": "California",
    "country_code": "USA",
    "created_at": "2007-05-25T21:22:15-07:00",
    "updated_at": "2015-04-30T17:19:13-07:00"
  }
}
```


According to the licensing agreement (https://data.crunchbase.com/docs/license), we can apply for the basic access as well as the research access for free. If we are granted the research access, we may be able to get a more complete set of data that are not in the Open Data Map.

- Angel List:
  We plan to retrieve the tags of startups' market from Angel list, and find the related areas connected with big data, such as cloud computing, machine learning, big data analytics, etc. Since we couldn't register an API to get data, we plan to use Selenium or other Python modules to scrape web data.

Networks Analysis Methods

To answer the questions we proposed in the Introduction, we plan to make undirected graph with nodes representing startup companies and investors in our dataset. For the first question, we will use a bipartite graph to show the connections between startup companies and investors. Based on our findings, we will calculate degree of centrality, betweenness centrality, and closeness centrality, and influence range of startup companies and investors. Clustering coefficient is another measure we may use. In conclusion, the proposal is based on what we have learned in the class so far. As we are learning more concepts from the class, we will consider adding more analysis methods to our project and apply more of what we learn to the project.