THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

DIFFERENCES IN SOCIAL NETWORKS BETWEEN CHINESE COMMUNITY AND INDIAN COMMUNITY IN SILICON VALLEY

SHUHAN TIAN
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Mathematics and Economics
with honors in Economics

Reviewed and approved* by the following:

Kalyan Chatterjee
Distinguished Professor of Economics and Management Science
Thesis Supervisor

Russell Chuderewicz
Senior Lecturer in Economics
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

In this paper, I investigate differences in social networks between Chinese community and Indian community in Silicon Valley by referring to Dr. Dossani's paper and Prof. Saxenian's books. I discovered that compared to Chinese immigrants, Indian immigrants are more adaptive to life in the United States (Silicon Valley) because of their language advantage and are more successful in Silicon Valley. To further compare these two communities, I constructed two network models in MATLAB by focusing on dynamic formation of these two communities and made some measurements such as probability distribution and degree distribution to show how these two communities different from each other.

TABLE OF CONTENTS

LIST OF FIGURES iii

LIST OF TABLES iv

ACKNOWLEDGEMENTS v

Chapter 1 Introduction of Background of Chinese and Indian Community in Silicon Valley 1

Chapter 2 Literature Review 4

- 2.1 Literature Review on studies conducted by Prof. Saxenian 4
- 2.2 Literature Review on study conducted by Dr. Dossani 11
- 2.3 Literature Review on Methodology 14

Chapter 3 Model 16

- 3.1 Using random graph to demonstrate social networks 17
- 3.2 Dynamic formation of social networks for Indian and Chinese community 18
  - Scenario 1: 18
  - Scenario 2: 27

Chapter 3 Conclusion 34

Appendix A Code for Scenario 1 36

Appendix B Code for Random Graph 38

Appendix C Code for Scenario 2 40

BIBLIOGRAPHY 45

LIST OF FIGURES

Figure 1. Origins of Engineering and Technology Company Immigrant Founders in Silicon Valley, Calif. 3

Figure 2. When Did You Settle in the United States? 7

Figure 3. Very Important Source of Technology and Business Information 9

Figure 4. Do You Have Plans to Start Your Own Business on a Full-Time Basis? 10

Figure 5. Percentage of Companies with More Than 10 Percent of Its Full-Time Employees from Founder's Country of Birth 10

Figure 6. Example of Community A 19

Figure 7. Probability Distribution of Internal Links (Small Network) 23

Figure 8. Probability Distribution of External Links (Small Network) 23

Figure 9. Chinese Community 24

Figure 10. Indian Community 25

Figure 11. Probability Distribution of Internal Links (Large Network) 26

Figure 12. Probability Distribution of External Links (Large Network) 27

Figure 13. Degree Distribution of Internal Links Case 1 31

Figure 14. Degree Distribution of External Links Case 1 31

Figure 15. Degree Distribution of Internal Links Case 2 32

Figure 16. Degree Distribution of External Links Case 2 33

LIST OF TABLES

Table 1. Immigrants of Indians and Chinese Into Silicon Valley High-Technology Industries, by Year 2

Table 2. Education Level of Indians, Chinese, and Whites in Silicon Valley High-Technology Industries, 1990 2

Table 3. Source of Funds for Those Respondents Involved in Founding Startups 13

Table 4. Difficulties Experiencing in Raising Capital for Startups 14

ACKNOWLEDGEMENTS

I would like to thank Dr. Kalyan Chatterjee for helping me in setting up the topic of my thesis and building models. I would also like to thank Dr. Xiang Wang, my friend Mi JunYang, Yuan Zhenyuan for providing me instructions on how to build models in MATLAB. Finally, I would like to thank my family and friends who have motivated me and supported me.

Chapter 1

Introduction of Background of Chinese and Indian Community in Silicon Valley

Nowadays, when people talk about Silicon Valley, they talk about the large number of Chinese and Indian Engineers. The reason why there are so many Chinese and Indians in Silicon Valley is mainly because of two acts. One is the Hart-Cellar Act of 1965 and the other one is the Immigration and Naturalization Act of 1990. According to Saxenian (2002), the Hart-Cellar Act abolished previous restricted immigration law, allowing people to immigrate to the United States if they possess scarce skills and are family members of American citizen or permanent residence. The Hart-Cellar Act provided opportunities for foreign-born engineers to work in the United States and at that time, new generation of high-technology industry developed rapidly. As more and more immigrants came to the United States and as the demand for high-tech industry increased sharply, according to Saxenian (2002), between 1975 and 1990, Silicon Valley's technology companies created more than 150,000 jobs. As a result, "one third of all scientists and engineers in Silicon Valley's technology industries in 1990 were foreign-born. Of those, almost two thirds were Asians, with the majority of Chinese and Indian descent." (Saxenian, 2002). The Immigration and Naturalization Act of 1990 further facilitated the immigration of engineers because it revised the non-immigrants visas, which helped temporary workers, who owned highly specialized knowledge and skills and also got high education, to stay in the United States.

Table 1. Immigrants of Indians and Chinese Into Silicon Valley High-Technology Industries, by Year

	1980-1989		1970-1979		Before 1970		Native	
	Number	%	Number	%	Number	%	Number	%
Indian	4,367	60	1,963	27	803	11	162	2
Chinese	7,921	41	5,697	30	2,491	13	3,109	16

SOURCE: U.S. Bureau of the Census (1992).

SOURCE: Silicon Valley's New Immigrant High-Growth Entrepreneurs, Saxenian, 2002.

From Table 1, we can clearly see that there is a huge increase in the immigration of Indian and Chinese into Silicon Valley high-technology industries from 1970 to 1990.

Moreover, according to Saxenian's findings, both Indian and Chinese workers in Silicon Valley are highly educated, especially in the technology industry.

Table 2. Education Level of Indians, Chinese, and Whites in Silicon Valley High-Technology Industries, 1990

	Indian		Chinese		White	
	Number	%	Number	%	Number	%
Master's of science to Ph.D.	4,043	55	7,612	40	34,468	18
Bachelor of science	1,581	22	5,883	31	59,861	31
Some university	792	11	3,551	19	64,081	34
High school graduate	600	8	1,002	5	23,488	12
Less than high school	279	4	1,170	6	9,319	5

SOURCE: U.S. Bureau of the Census (1992).

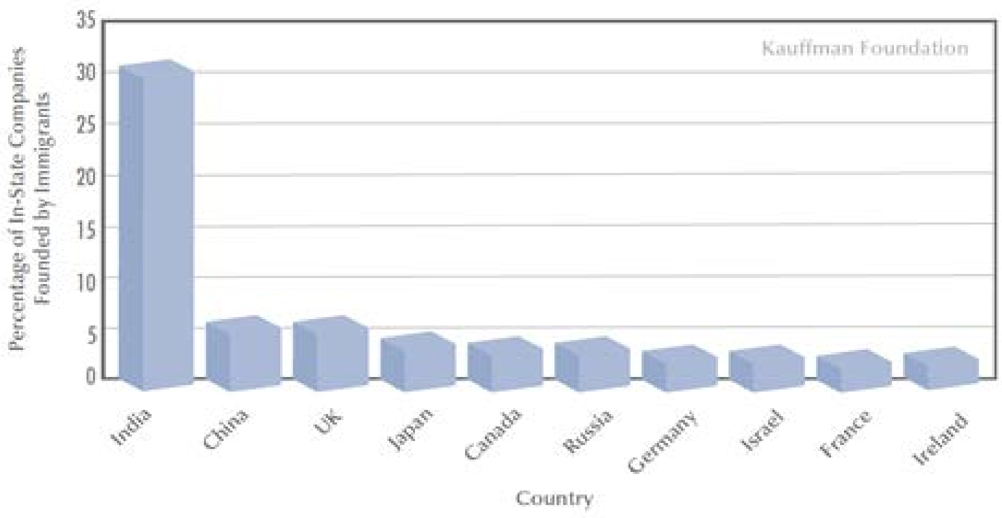
SOURCE: Silicon Valley's New Immigrant High-Growth Entrepreneurs, Saxenian, 2002.

From Table 2, it is clear for us to see that 55% of Indians own degree of Master's of science to Ph.D, and 40% of Chinese own degree of Master's of science to Ph.D while only 18% of Whites own degree of Master's of science to Ph.D.

Beginning in 1980, Indian and Chinese immigrant engineers started their own technology businesses. The difference between Indian-run and Chinese-run companies is that Indian-run companies focus more on software and business service while Chinese-run companies mainly focus on electronic hardware manufacturing and trade. The difference can be explained by the significant language and management advantage that Indian-run firms have but Chinese-run firms lack.

Finally, there is one thing worth noting, which is Indians are dominating Silicon Valley. According to Figure 1 below, from 2006-2012 this time period, we can see that Indians accounted for more than 30% of immigrant-founded companies in Silicon Valley while Chinese only accounted for 5%. The reason behind this might be that compared to Chinese, Indians have language advantage. Compared to other immigrant groups, Indian immigrants have strong academic background. (They have a better mathematical foundation).

Figure 1. Origins of Engineering and Technology Company Immigrant Founders in Silicon Valley, Calif.



SOURCE: Then and Now: America's New Immigrant Entrepreneurs, Part VII, Wadhwa, Saxenian, Siciliano, 2012.

Chapter 2

Literature Review

2.1 Literature Review on studies conducted by Prof. Saxenian

There are lots of researches and studies about the immigrants and their networks in Silicon Valley and most of the studies were conducted at the beginning of the 21st century or the end of the 20th century by Professor Saxenian. Thus, in this part, I will review literature of two of Saxenian's books, *"Silicon Valley's New Immigrants Entrepreneurs"* and *"Local and Global Networks of Immigrant Professionals in Silicon Valley"* and then I will make a summary on studies conducted recently.

In the book, "*Silicon Valley's New Immigrants Entrepreneurs"*, data was primarily derived from the database of high-technology firms between 1980-1990 and interviews with engineers, entrepreneurs and venture capitalists in Silicon Valley. By analyzing the data and responses from interviewees, Saxenian mainly examined how skilled Chinese and Indian immigrants organized ethnic networks in Silicon Valley, how these networks helped them start new technology businesses and how these Chinese and Indian engineers further enhanced entrepreneurial opportunities by building social and economic networks back to their home countries. The reason that Saxenian focused on these two groups is that in Silicon Valley, by 1990, two thirds of foreign-born engineers were from Asian and among those, most were Chinese and Indian immigrants.

In this book, Saxenian found that most interviewees attributed the fact that most of Chinese and Indians concentrated on professional and technical positions rather than managerial positions, despite their high level of education attainment to the feeling of being excluded and lack of role model. Most of those interviewees came to Silicon Valley in the mid-20th century, a period of time that there were not many Asians working in Silicon Valley. Working in a world dominated by white men, those people felt that they were isolated so they decided to resign and start their own business. Saxenian concluded that resigning and starting their own businesses was one way that these first-generation engineers responded to the sense of exclusion. The other way they adopted was that as Chinese communities and Indian communities grew during period 1970s to 1980s, immigrants in both communities tended to come together and organize different kinds of activities to promote relationships between members at first. As time moved on, other than just building social networks within their own communities, these immigrants expanded networks to business purposes. They created many professional associations to help people within their own communities build successful businesses. By relying on these networks and taking advantage of resources provided, Silicon Valley's immigrant engineers were able to start technology firms, which stimulated the regional economy and enhanced entrepreneurial opportunities.

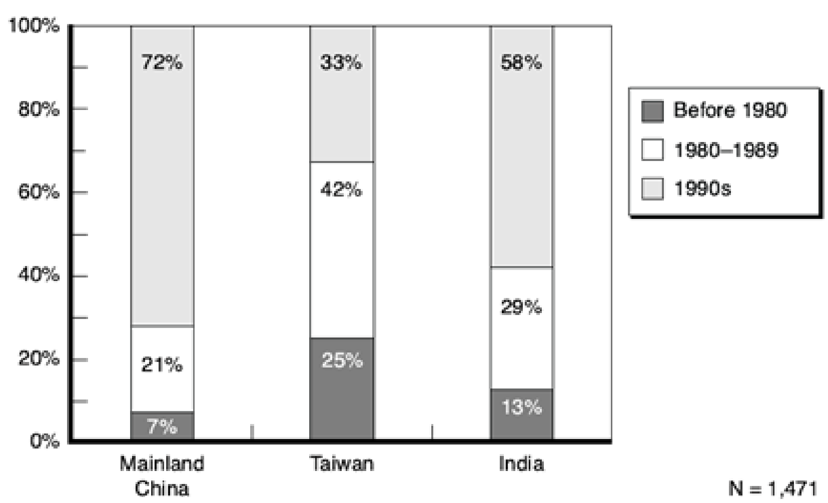
In this book, Saxenian listed several professional and technical associations in Silicon Valley during the1980s and 1990s. Compared to Indian community, which created only two associations, the earliest of which was established in 1991, Chinese community created more than ten associations and more than half of those associations were created before 1990. According to Saxenian, this immobilization of Silicon Valley's Indian immigrants during that period was in part because of not achieving a critical mass in the region. Although there were lots of associations and each association had its own specialization, these associations shared important functions. They all provided first-generation immigrants with professional networks within local technology community, such as information about labor market and recruitment. Moreover, they showed people successful immigrant entrepreneurs, which could be considered as role models. For Chinese associations, they also provided trainings and seminars on English communication skills. Apart from offering professional resources for first-generation immigrants, these associations enabled investment activities among immigrants of different generations. Based on Saxenian's findings, older generation of successful engineers and entrepreneurs often provided guidance and financial support for younger generations. For most of the time, they acted as angel investors by investing their money to promising new ventures. Although it seemed that immigrants were able to live a better life with the support of their own communities, Saxenian suggested that a successful start-up still needed to become part of mainstream to grow. "It appears that the most successful immigrant entrepreneurs in Silicon Valley today are those who have drawn on ethnic resources while simultaneously integrating into mainstream technology and business networks". (Saxenian,1999)

In the last part of this book, Saxenian further checked the expansion of these ethnic networks. She found out that both Chinese and Indian immigrants were building ties back to their home countries. It is understandable because compared to local entrepreneurs, those immigrant entrepreneurs had advantages in language skills and technical and cultural know-how, which enabled them to function better. For Chinese immigrants, especially Taiwanese, the Silicon Valley based entrepreneurs benefited from Taiwan's flexible semiconductor and personal computer manufacturing capabilities. For Indian communities, those entrepreneurs benefited from India's skilled software programming and design talent.

Because of the nature of the research method adopted by Saxenian, this book did not provide direct comparisons between Chinese and Indian communities. However, in "*Local and Global Networks of Immigrant Professionals in Silicon Valley*", data was mainly derived from surveys. With this method, Saxenian was able to compare different immigrant communities more intuitively and further examined some conclusions made in the book *"Silicon Valley's New Immigrants Entrepreneurs"*.

The study described in "*Local and Global Networks of Immigrant Professionals in Silicon Valley"* was based on a survey drawn from the memberships of 17 leading immigrant professional associations in Silicon Valley. Compared to the study in the book *"Silicon Valley's New Immigrants Entrepreneurs"*, this research was conducted in 2001. In this study, Saxenian found out that most of immigrants came to settle in the United States after 1990s. We can see from Figure 2 that among those three immigrant groups, the percentage of immigrants from Mainland China after the 1990s was more than 10 times the percentage of immigrants from Mainland China before 1980. For India, although the increase may not that rapid, the percentage of immigrants from India to the United State before 1980 was relatively small.

Figure 2. When Did You Settle in the United States?

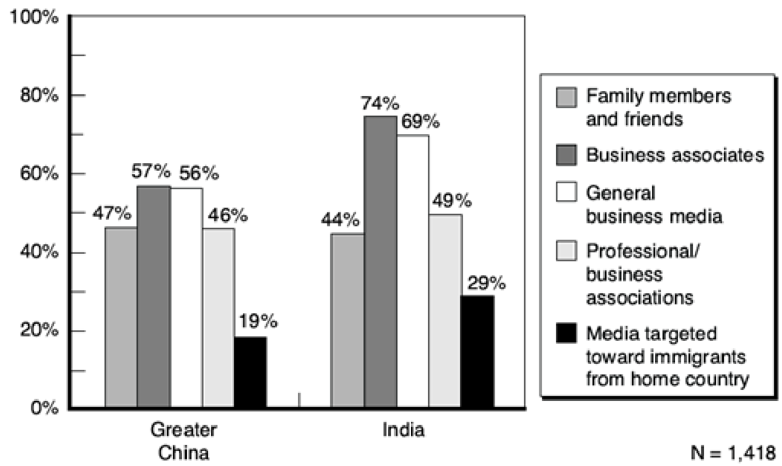


SOURCE: Local and global networks of immigrant professionals in Silicon Valley, Saxenian, Motoyama, & Quan, X, 2002.

Figure 2 further validated what Saxenian stated in her book *"Silicon Valley's New Immigrants Entrepreneurs"*. Since for each immigrant group, there were small numbers of immigrants before 1980, especially for the group of Chinese immigrants, this small number contributed to the first-generation engineers feeling a sense of exclusion, which pushed them to start their own businesses and build social and professional networks within their own communities.

Saxenian also checked the frequency of each immigrant group attending associational activities and she found that most of surveyed took part in associational activities actively, which also means that most of immigrants actively took part in local and professional networks. However, there are some differences between each group. Compared to Chinese and Taiwanese immigrants, Indian immigrants were less active. Saxenian then further investigated how different groups of immigrants would rank sources of information and found out that most of Indian immigrants ranked business associates as very important source of technology and business information and the percentage was 20% higher than that of Chinese immigrants. She thought that this difference might come from the language barriers faced by Chinese immigrants. Moreover, she noticed that for Chinese immigrants, family members and friends played a relatively important role in their social and business life.

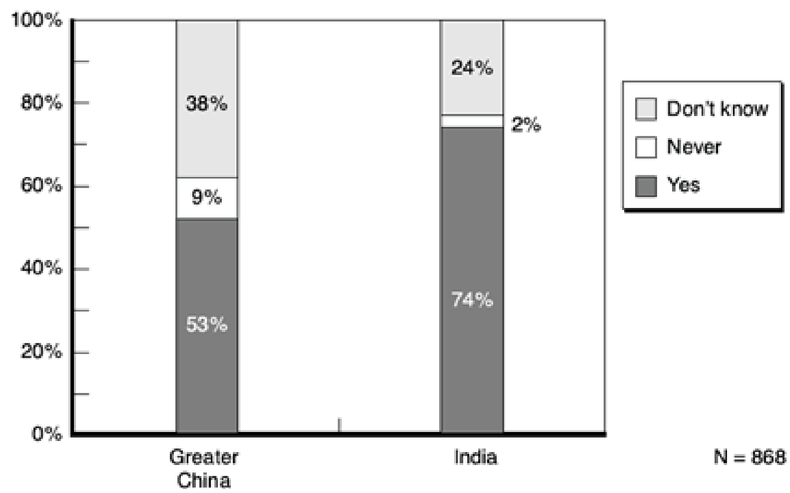
Figure 3. Very Important Source of Technology and Business Information



SOURCE: Local and global networks of immigrant professionals in Silicon Valley, Saxenian, Motoyama, & Quan, X, 2002.

When asked about plans to start your own businesses on a full-time basis, Saxenian found that compared to Chinese immigrants, Indian immigrants seemed to have more ambitions. According to Figure 4, 74% of Indian had plans about starting their own businesses while only 53% of Chinese had ever thought about starting their own businesses. Moreover, those immigrants' entrepreneurs tied back to their home countries, which means they wanted to become transnational entrepreneurs. According to Saxenian's findings, more than 70 percent of Chinese and Indian immigrants considered locating their business in their native countries. Based on her analysis, some of these immigrants actively participated in the investment activities to help start-ups and venture funds in their country of birth. In addition, she found that when asked about important factors that affect decisions of returning to home countries, Chinese immigrants ranked professional opportunities as the most significant factor while Indian immigrants ranked culture and lifestyle as major reasons.

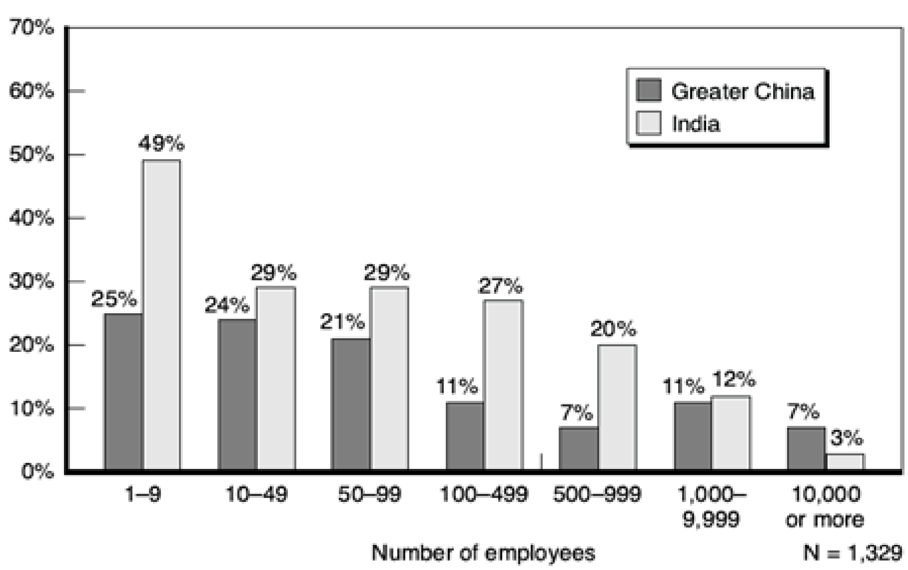
Figure 4. Do You Have Plans to Start Your Own Business on a Full-Time Basis?



SOURCE: Local and global networks of immigrant professionals in Silicon Valley, Saxenian, Motoyama, & Quan, X, 2002.

In this book, Saxenian also mentioned that as time moved on, immigrants' reliance on ethnic networks would go down. This declination was especially obvious in the Indian community. Figure 5 shows that there was a huge declination when number of employees increased from 1-9 to 10-49. And when number of employees exceeded 10,000, only 3% of companies had more than 50 percent of employees from founders' country of birth.

Figure 5. Percentage of Companies with More Than 10 Percent of Its Full-Time Employees from Founder's Country of Birth



SOURCE: Local and global networks of immigrant professionals in Silicon Valley, Saxenian, Motoyama, & Quan, X, 2002.

Saxenian's findings also showed the pattern of financing of start-ups for immigrants. Immigrants usually started their own companies by relying on personal savings and angel investors. This reliance slowly transferred to venture capital for subsequent rounds of funding.

In these two books, Saxenian made comprehensive analyses about immigrants and their networks in the Sillicon Valley. However, the research in these two books were conducted more than 10 years ago, some analyses might not be able to use to explain what happened in the 21st century. So I also looked into several more studies that followed up Saxenian's researches.

Most of follow-up studies were described in an article named "America's New Immigrant Entrepreneurs: Then and Now." This article made a summary on these studies and researches and concluded that there was an accelerated growth of Indian entrepreneurship during 2006 – 2012, and Indians would continue to dominate groups in rates of entrepreneurship.

2.2 Literature Review on study conducted by Dr. Dossani

I also reviewed on Dossani's paper, "Chinese and Indian Engineers and their Networks in Silicon Valley". In this paper, although Dr. Dossani used the same data as Professor Saxenian did, he made a more detailed comparison between Chinese (including Taiwan) community and Indian community. He mainly focused on the differences in the percentage of sources of capital that people would use during the initial funding round and the subsequent funding rounds among three groups of community, People Republic of China, Taiwan and India.

In his comparison for sources of funds for those respondents involved in founding startups, he divided the sources of capital into three big sources, personal saving, informal source of funds, which included personal savings, family/friends in United States, family/friends outside United States and angel investors, and formal source of funds, which included venture capital firm in and outside United States and banks, government loans etc. According to Dossani, informal source of funds is typically contacted directly through ethnic professional associations while formal source of funds is contacted indirectly through ethnic professional associations. In order to make an accurate comparison, he divided China into People Republic of China and Taiwan. After comparing the percentage of each source, he found that in the initial stage of funding, angel investors and personal savings are the two most important sources of funds for PRC, Taiwan and India. However, in subsequent rounds, venture capital becomes a more important source of funds. Thus, Dossani concluded that in the initial funding round, people rely heavily on informal sources other than formal sources and in the subsequent funding rounds, there is a shift towards formal sources. The reason is that subsequent funding is based on the performance of the company, which is more credible to any financer, thus it is easier to get larger and cheaper funds in the mainstream.

What’s more, Dossani found out that Taiwanese respondents rely more heavily on venture capital firms outside United States than do Chinese respondents and Indian respondents. The reason is that the venture capital industry in Taiwan is more developed than in China and India. For Mainland China, although it prefers venture capital firms in U.S in initial stage, it seems that in the subsequent stages, Mainland China tends to work with venture capital firms outside the United States. However, unlike Taiwan and Mainland China, no matter in initial funding round or in subsequent funding rounds, India relies heavily on U.S. based funding source.

The other thing Dossani noticed was that there are not much differences in percentage of People Republic of China-respondents and India-respondents relying on venture capital firms in the United States, which made him confused because Indians have advantage in language and management training compared to Chinese. He then asked respondents questions about difficulties when raising capital, and found out that the most difficult thing is the access to investors. 61.8% of People Republic of China-respondents and 69.9% of India-respondents thought it is very hard to get access to investors. The large percentage of disadvantage in access to investor might help explain the little difference in reliance on venture capital in the United States between People Republic of China-respondents and India-Respondents. However, the paper did not provide explanation about why People Republic of China-respondents heavily rely on U.S based venture capital firms both in initial and subsequent stage since Chinese have disadvantage in language, it is not easier for them to attract venture capital firms in the United States.

Table 3. Source of Funds for Those Respondents Involved in Founding Startups

	Initial Funding Round			Subsequent Funding Rounds		
Sources of Capital	PRC	Taiwan	India	PRC	Taiwan	India
Personal savings	52.7	42.0	61.5	19.5	19.6	25.1
Family/friends in United States	16.4	26.0	23.1	9.8	21.7	10.0
Family/friends outside United States	20.0	20.0	5.1	12.2	19.6	4.2
Angel investors	41.8	38.0	46.5	26.8	26.1	25.5
Venture capital (VC) firm in United States	34.5	18.0	35.9	58.5	34.8	61.1
VC firm outside United States	10.9	30.0	5.1	31.7	47.8	14.6
Banks, government loans, other	12.7	8.0	16.5	26.8	30.4	28.5

SOURCE: Chinese and Indian Engineers and their Networks in Silicon Valley, Dossani, 2002.

Table 4. Difficulties Experiencing in Raising Capital for Startups

	PRC	Taiwan	India
Access to investors	61.8	64.7	69.9
Language difficulties during presentation	11.8	11.8	1.8
Inadequate business plan	26.5	41.1	23.5
Inadequate technical skills	2.9	5.9	2.4
Inadequate management skills	26.5	41.2	15.7

SOURCE: Chinese and Indian Engineers and their Networks in Silicon Valley, Dossani, 2002.

2.3 Literature Review on Methodology

In this part, I briefly reviewed two random graph-based models of networks, one is Erdös-Rényi model and the other one is Barabasi-Albert model by referring to Professor Jackson's book, "*Social and Economic Networks*". These two models, one static and the other one growing, are important references for building my own models.

Erdös-Rényi model is a simplest random graph model because the number of nodes is fixed and the formation of links is not complicated. It describes an N nodes network with fixed probability P of formation of links between any two nodes. Moreover, it requires that the formation of links is independent of each other. Thus, in Erdös-Rényi model, link formation follows binomial distribution. According to this characteristic, Jackson made some basic calculations, such as the probability of forming two links given three people in this network, the probability of empty network etc. He also provided the general formula of probability of forming m links on n nodes for any given network, which is $P^m * (1-P)^{\binom{n}{2}-m}$. Apart from basic probability calculations, Jackson further checked the degree distribution of Erdös-Rényi model. Since the link formation follows binomial distribution, for any given node i, the probability of it has d links is $\binom{n-1}{d} * P^d * (1-P)^{n-1-d}$ because for one node, it can only form links with other $n-1$ nodes.

Compared to Erdös-Rényi model, Barabasi-Albert model is a relatively complex model because the number of nodes and the probability of formation of links are changing over time. Barabsi-Albert model can also be called as preferential-attachment model because of the characteristic of probability of formation of link. Probabilities of newly added nodes connecting to existing nodes are proportional to number of links that existing nodes have. Therefore, an already heavily linked node is more likely to be attached by newly added nodes. Based on this characteristic, Jackson calculated the frequency distribution of links and showed that expected number of links follows power distribution with degree -3, which means network generated by this model is scale free. The most notable feature of this scale free network generated by Barabasi-Albert model is that some nodes have relatively more links than others and these nodes are relatively older nodes.

Chapter 3

Model

Based on the literature review of Professor Saxenian's book and Dr Dossani's paper, what I found interesting are that firstly, both two communities share similarities but also have differences in looking for funds and starting their own business. Also, compared to Chinese immigrants, it seems that Indian immigrants are more adaptive to life in the United States (Silicon Valley) because of their language advantages and it seems that Indian immigrants are more successful in Silicon Valley then Chinese immigrants. Thirdly, Taiwanese and Indians immigrants are more likely to connect to their home country compared to Chinese immigrants.

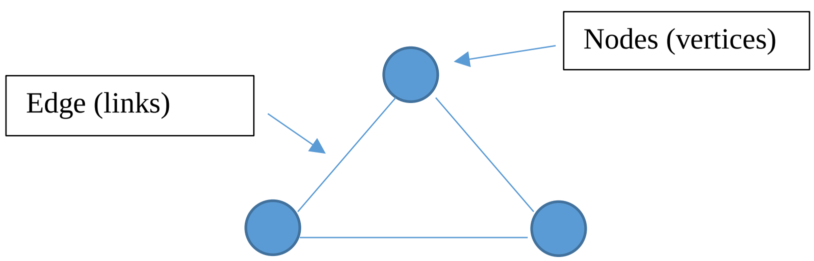
In order to further understand what I am interested in, which are differences in social networks between Chinese and Indian communities, I need to build models to compare each group of people. However, data for different communities starting their own business and their tie to their home countries are hard to acquire, I decide to focus on the fact that compared to Chinese immigrants, Indian immigrants are more likely to jump out of their own communities to expand their social networks, which means the probability of Indian immigrants looking for outside resource is bigger than that of Chinese immigrants. For Chinese immigrants, because of their language disadvantage, the probability for them staying in their own community and relying on inside resource is bigger than that of Indian immigrants.

In this part, I will simulate two scenarios to demonstrate these two communities and try to calculate some basic characteristics of networks to compare these two communities.

3.1 Using random graph to demonstrate social networks

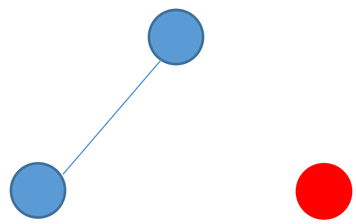
Since what I am interested in is about social networks, or networks, I look into the random graph model for networks. Firstly, I will talk about using nodes (vertices) and edges (links) to demonstrate networks.

Assume there are three people in a community and they share similar characteristics. Thus, we can use three same nodes to demonstrate these three people and we connect these three nodes to show their links between each other. The graph below shows what I described above and it is a simple network.



However, when people have different characteristics, we should use different nodes. For example, we can use red nodes to present Indian people within their own community and use black nodes to present people outside Indian community. Also, the links can take different forms. Sometimes, we will see the link is directed and sometimes the link is undirected. For example, in worldwide web network, links are directed because when we put a link in our own website, by clicking that link, we will be directed to certain page. However, from that page, probably there is no link that can be used to direct to our own website. In this paper, I assume that links, the connection between people, are undirected.

After defining using nodes to represent people and links to represent connections between people, we still need to think about the probability of formation for each link. In the example above, all three links are connected. However, it is not always the case that all nodes will connect to each other. For example, among these three people, two of them are good friends. Thus, it is more likely that these two people build a connection with each other and the one left become isolated, which means there is no link connecting him to other two people. So it is more likely that final graph would be:



（Blue nodes denote two people who are good friends with each other and red node denotes the one who are strangers to these two people)

3.2 Dynamic formation of social networks for Indian and Chinese community

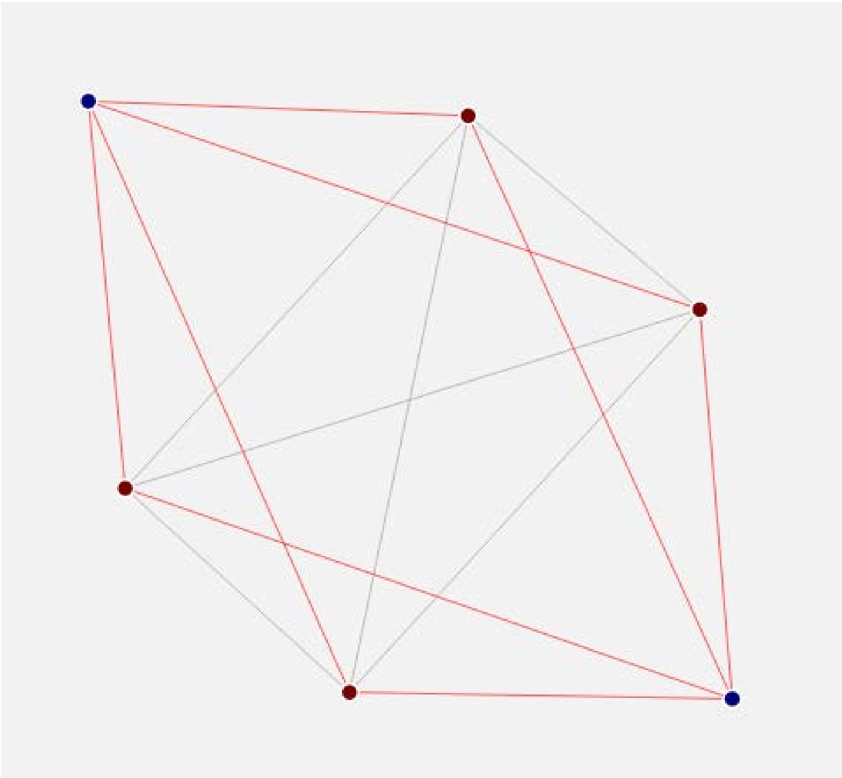
In this part, I will focus on the dynamic formation of social networks and I will build two models based on Erdös-Rényi model and Barabasi-Albert model. In the first scenario, number of people inside and outside Chinese community is equal to number of people inside and outside Indian community and the number of people is fixed. In the second scenario, for each period, there will be one person added into each community.

Scenario 1:

In this scenario, firstly, I will give a simple example to demonstrate my model and then I will extend to general situation.

Assume there is a community, named A. There are four people inside A, and two people outside A. People inside A share same characteristics and will connect with each other. They will also connect with two people outside A, but two people outside A will not connect with each other. Therefore, if we do not consider the probability of formation of each connection, in total, there will be $\binom{4}{2} + 4 * 2 = 6 + 8 = 14$ links, six internal links and eight external links. (Link is formed between two people, so we just choose two people out of four people to get six internal links. For external links, each of the four people inside community A builds connection with two people outside community A, so there are eight external links). Thus, if I use nodes and edges to demonstrate connections within community A, the graph would be:

Figure 6. Example of Community A



Red nodes represent people inside A and four grey lines represent internal links. Blue nodes represent people outside A and eight red lines represent external links.

However, this network with total fourteen links is not immediately formed. In real-life network, connections between people are built step by step. Thus, in this simple model, I assume that at the beginning, there is no connection between people inside and outside community A, which means all fourteen links are inactivated at the beginning. Then I assume that for each period, link has probability to be activated. For internal links, each one of them has probability P to be activated. For external links, each one of them has probability R to be activated. Thus, the probability of being inactivated is $1 - P - R$. To make it easy for analyzing, I make several assumptions. Firstly, if link will be activated, only one link will be activated each time. Secondly, each link, no matter internal or external, can only be activated once and activation of each link is independent of each other. Lastly, what happened in the previous period would not affect what will happen in the next period, which means all periods are independent of each other.

Based on above assumptions, it is easy to find out that if we look at internal and external links respectively, their formation follows binomial distribution. For example, if we only focus on internal links, activation of internal links have binomial distribution $B \sim (N, P)$. If we only focus on external links, activation of external links have binomial distribution $B \sim (M, R)$. N, M represents number of periods repeated. However, in this simple example, there are at most six internal links, and eight external links. For activation of internal links, if the number of periods repeated exceeds six, it is impossible for us to find out the probability of forming more than six internal links. Similarly, for activation of external links, if the number of periods repeated exceeds eight, it is impossible for us to find out the probability of forming more than eight external links. Thus, once number of periods repeated exceeds the largest number of links this process generates, activation of links does not follow binomial distribution. This is understandable because if we repeat this process for infinite many times, at last, all links will be activated. In addition, since I need to consider both internal links and external links at the same time in my model, in this simple example, number of periods repeated cannot exceed six, which is the minimum number of six internal links and eight external links.

If I use A to represent Chinese and Indian community respectively, P_{India} will be smaller than P_{China} and R_{India} will be greater than R_{China} based on my findings from literature review. In this simple example, let P_{China} =0.6, P_{India}=0.3, R_{China}=0.3, R_{India}=0.6 and number of periods repeated be 6. With all parameters set properly, we are able to compare two communities by calculating the probability of forming internal and external links and expected number of internal and external links.

For internal links, let random variable X be number of internal links. Since number of period repeated is six, X can be 0,1,2,3,4,5,6. Since activation of internal links is a binomial distribution, the probability mass function for internal link is $P(X=x)=\binom{6}{x}P^x(1-P)^{6-x}$. For external links, similarly, the probability mass function is $P(X=x)=\binom{6}{x}R^x(1-R)^{6-x}$. Thus, for Chinese community, the probability mass function for internal link is $P(X=x)=\binom{6}{x}0.6^x(1-0.6)^{6-x}$. So $P(X=0)=0.4^6=0.0041$, $P(X=1)=6*0.6^1*0.4^5=0.0369$, $P(X=2)=15*0.6^2*0.4^4=0.1382$, $P(X=3)=20*0.6^3*0.4^3=0.2765$, $P(X=4)=15*0.6^4*0.4^2=0.3110$, $P(X=5)=6*0.6^1*0.4^5=0.1866$, $P(X=6)=0.4^6=0.0467$. Thus, the expected internal links for Chinese community are $0*0.0041+1*0.0369+2*0.1382+3*0.2765+4*0.3110+5*0.1866+6*0.0467=3.6$, which are expected

Chinese immigrants' connections within their own community. For Indian community, the probability mass function for internal link is $P(X = x) = \binom{6}{x}0.3^x(1 - 0.3)^{6-x}$. So $P(X = 0) = 0.7^6 = 0.1176$, $P(X = 1) = 6 * 0.3^1 * 0.7^5 = 0.3025$, $P(X = 2) = 15 * 0.3^2 * 0.7^4 = 0.3241$, $P(X = 3) = 20 * 0.3^3 * 0.7^3 = 0.1852$, $P(X = 4) = 15 * 0.3^4 * 0.7^2 = 0.0595$, $P(X = 5) = 6 * 0.3^5 * 0.7^1 = 0.0102$, $P(X = 6) = 0.3^6 = 0.000729$. Thus, the expected internal links for Indian community are $0 * 0.1176 + 1 * 0.3025 + 2 * 0.3241 + 3 * 0.1825 + 4 * 0.0595 + 5 * 0.0102 + 6 * 0.000729 = 1.8$, which are expected Indian immigrants' connections within their own community. With the same method, we can also calculate the expected external links for Chinese and Indian community. After calculation, I found out that the expected Chinese immigrants' connections with outside are 1.8 while the expected Indian immigrants' connections with outside are 3.6. In addition, on average, for there will be $3.8 + 1.8 = 5.6$ links in total for Chinese community and there will be $1.8 + 3.8 = 5.6$ links in total for Indian community.

For expectation of internal and external links, the results are obvious to us because compared to Indian immigrants, Chinese immigrants prefer to relying on building connections in their own community. Thus, there is no doubt that on average, there will be more internal links for Chinese community than for Indian community. However, Indian immigrants are more willing to connect with outside, so on average, there will be more external links for Indian community than for Chinese community.

For probabilities of forming internal and external links, figures below show differences of these two communities more clearly.

Figure 7. Probability Distribution of Internal Links (Small Network)

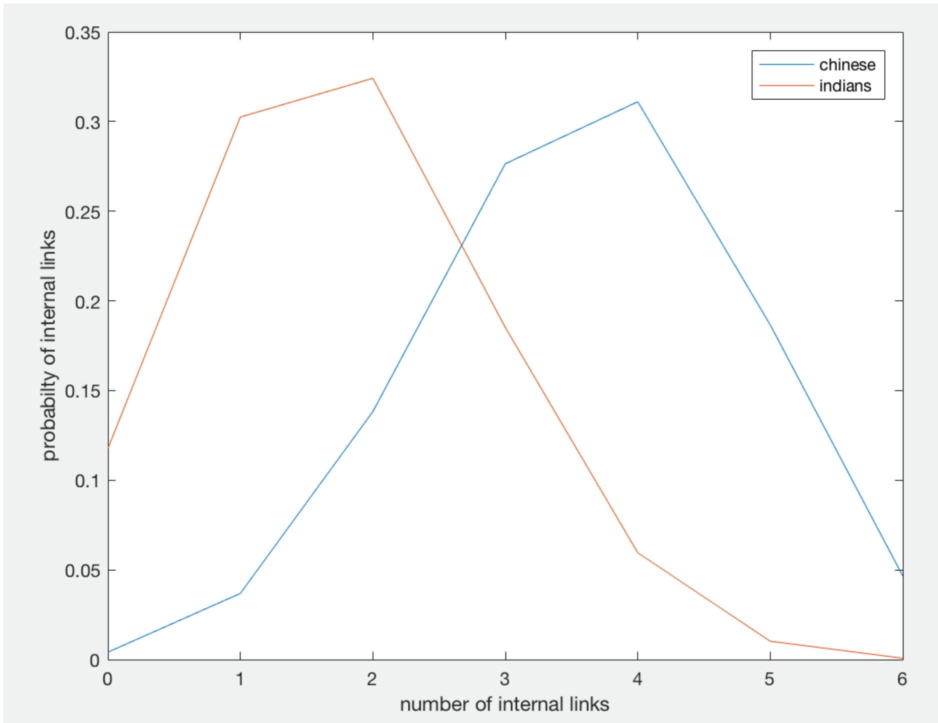
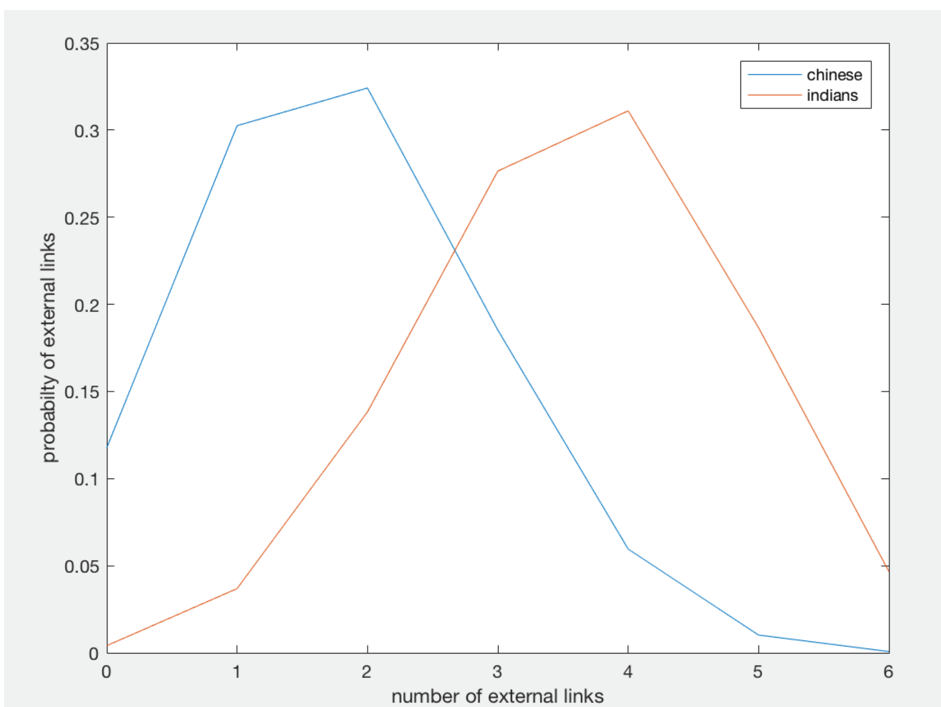


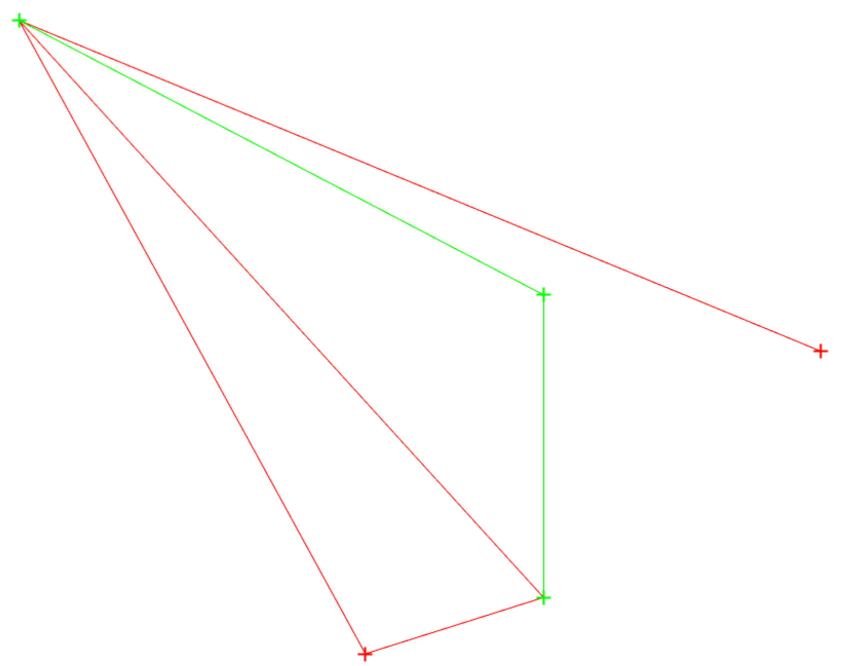
Figure 8. Probability Distribution of External Links (Small Network)



From figures above, we can see that for internal links, Chinese community has greater probability of forming more than three links, which means in Chinese community, the probabilities of forming relatively more internal links are greater. For external links, Indian community has greater probability of forming more than three links, which means in Indian community, the probabilities of forming relatively more external links are greater.

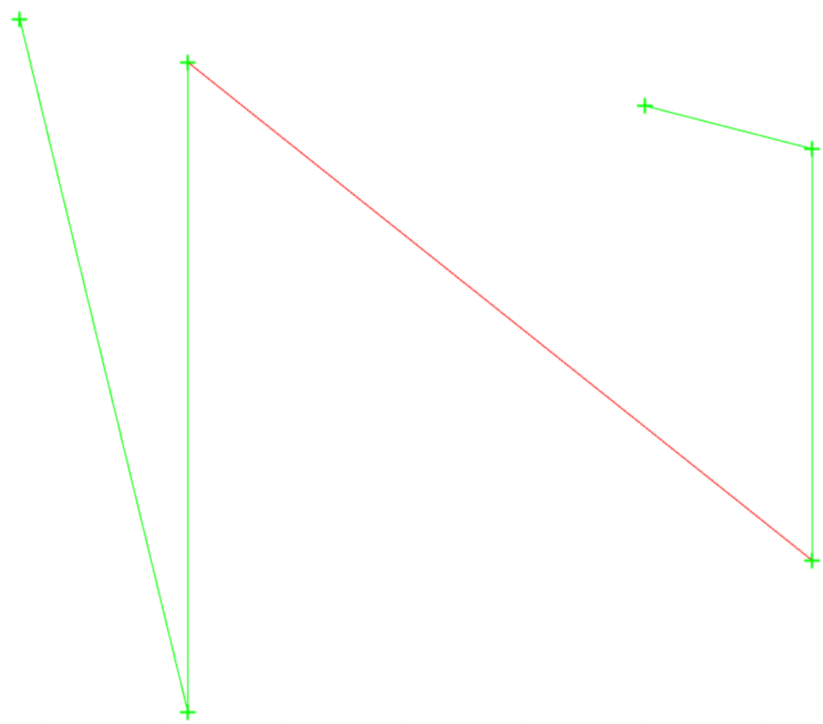
In addition, for this simple example, random graph is also a good way to represent differences of these two communities. Figures below show possible random graphs for both two communities. Red cross represents people inside community and green cross represents people outside community while red line represents internal link and green line represents external link.

Figure 9. Chinese Community



In this random graph, we can see that four internal links are formed while only two external links are formed. What’s more, one person outside Chinese community becomes isolated. (Isolated node will not show on the graph)

Figure 10. Indian Community



In this random graph, we can see that only one internal link is formed while four external links are formed. Moreover, there are two people inside Indian community becoming isolated. (Isolated node will not show on the graph)

However, in this simple example, the number of people inside and outside community A is too small, which means the number of periods repeated is too small. Thus, what we found out may not be representative.

After comparing Chinese and Indian community by focusing on small network, next I will focus on general situation and I will compare these two communities by focusing on large network.

Assuming that there are n people inside community A and m people outside community A, in total, there will be $\binom{n}{2} + nm$ links in total, $\binom{n}{2}$ internal links and nm external links. From what I discussed above, number of periods repeated should not be greater than minimum number of internal and external links. However, it can't be too small, either. If I let this number be N, internal link has binomial distribution $B\sim(N, P)$ and external link has binomial distribution $B\sim(N, R)$. Thus, according to property of binomial distribution, there are on average $N * P$ internal links and $N * R$ expected links.

Then, I still let P_{China} =0.6, P_{India}=0.3, R_{China}=0.3, R_{India}=0.6. For n and m, I let it be 50 and 40. Therefore, N cannot exceed $\min(\binom{50}{2}, 50 * 40)$, which is 1225. Figures below show probability distributions of these two communities with N=500.

Figure 11. Probability Distribution of Internal Links (Large Network)

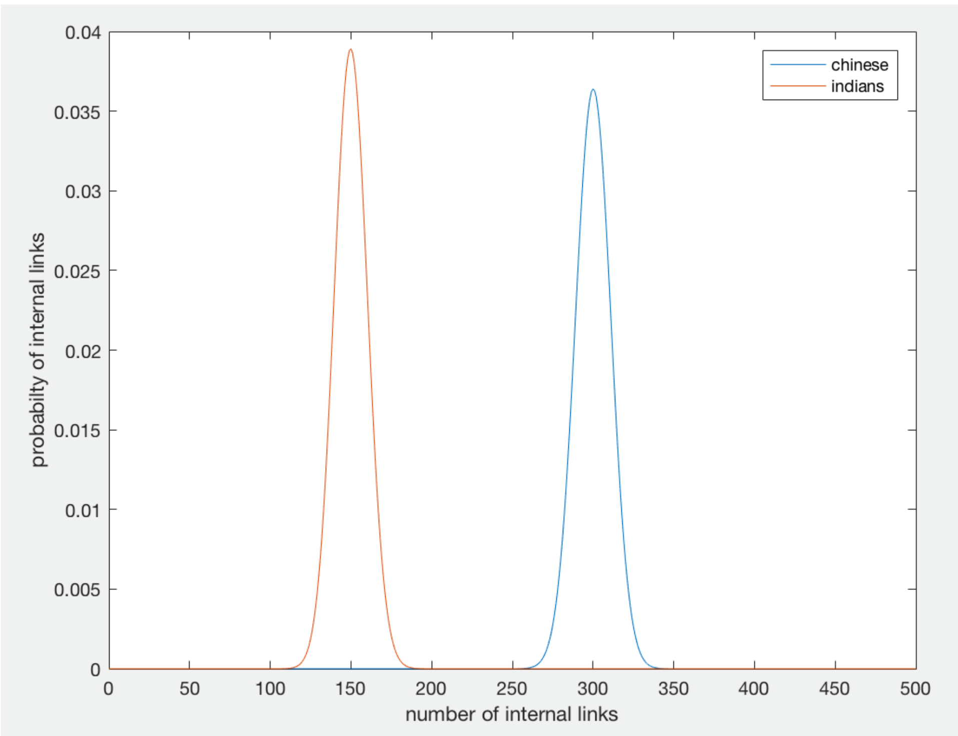
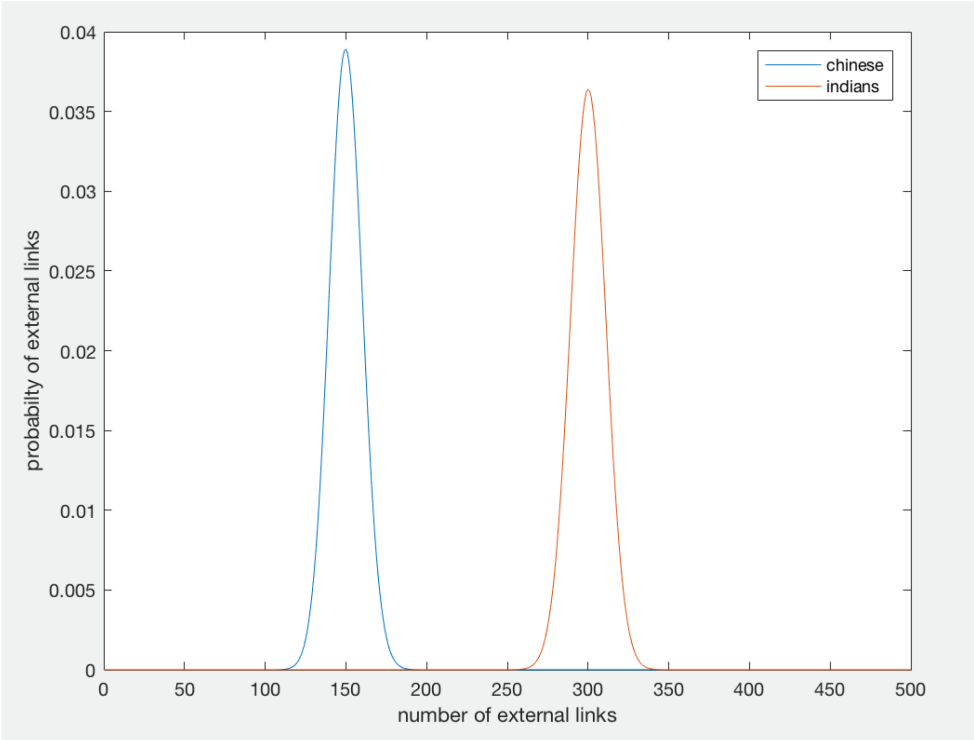


Figure 12. Probability Distribution of External Links (Large Network)



From these two figures above, it is clear to see that Chinese community has the greatest probability to form 300 internal links while Indian community has the greatest probability to form 150 internal links. Moreover, for Indian community, probability of forming 300 links is almost 0. However, for external links, Indian community dominates Chinese community because it has greater probability to form relatively more external links.

Scenario 2:

In this scenario, I assume that number of people inside community A is not fixed. Assume there are only 5 people inside community A and 300 people outside community A at the beginning. For connections of these 5 people and 300 people, I assume that there is at least one link for each person inside and outside community A. For example, for 5 people inside community A, there are at least 3 links between them. For 300 people outside community A, there are at least 300 links between them and 5 people inside community A. Assume there is no connection between people outside community A. Then, at each period, I assume that there is a newcomer joining in community A.

After this newcomer moves into community A, he needs to consider how he can build connections with people inside and outside community A. For example, when I came to the United States for the first time, what I thought about was how many Chinese friends I wanted to make and how many American friends I wanted to make.

After this newcomer making up his mind on building connections with people inside and outside community A, it is time to think about the probability that he connects himself to each person inside and outside community A.

Based on Barabási-Albert network model, the probability that a newcomer connecting to existing people i is $p_i = \frac{D_i}{\sum_j D_i}$, where D_i denotes the degree of person i, and $\sum_j D_i$ means the summation of degree over all pre-existing people j.

However, in this scenario, I assume that the probability of a newcomer connecting to each person inside community A does not totally depend on this probability for the reason that this probability only shows the preference attachment of a new comer, which means a heavily linked person in a community is more likely to be chosen when a newcomer joins this community. We still need to think about the probability that this newcomer randomly chooses a people inside community A to build connections. For example, it is reasonable that a newcomer of a community starts building his connections with people who are keen on making friends but it is also reasonable that he just randomly attends some parties and makes some friends.

Thus, in my model, in each period, the probability of a newcomer connecting to people already inside community A is a combination of preference attachment and randomness, which means this probability is comprised of a relation-based probability and location-based probability.

$$p_i = \frac{D_i}{\sum_j D_i} * \alpha_1 + (1 - \alpha_1) * p^*,$$

Where p^* is $1/n$, representing the probability that a newcomer randomly builds connections with others. n represents the number of people inside community A. Since for each period, a newcomer will join community A, n changes as time moves. As more and more people moves into community A, p^* becomes smaller and smaller, which means a newcomer will be more likely to connect himself to already heavily linked people instead of randomly connecting to someone. α_1 is a parameter used to balance p_i and it is from 0 to 1.

Also, for connections outside of community A, I use the same expression of p_i. Thus, the probability that a newcomer connecting to people outside of community A is $p_i = \frac{D_i}{\sum_j D_i} * \alpha_1 + (1 - \alpha_1) * p^*$, where i denotes people outside of community A. In this case, D_i denotes the number of external links between person i and people inside community A and j is from 1 to 300 and p^* equals to 1/300.

When it comes to differences between these two communities, I think there are two major differences. The first difference lies in number of connections a newcomer builds. If this newcomer is Chinese immigrant, he will build more connections with people inside Chinese community. If this newcomer is Indian immigrant, he will build more connections with people outside Indian community.

The second difference is that the value of α_1 is different. α_1 should be greater for Chinese community than for Indian community because inside their own community, instead of randomly talking to someone, Chinese are more likely to connect to people who has lots of internal resources, which means Chinese immigrants value inside resources more than Indian immigrants. However, for connections outside of community, α_1 should be greater for Indian community than for Chinese community because Indians value outside resources more. Thus, when building connections with people outside of their own community, they are more likely to go to someone who already has lots of connections with people inside the community rather than randomly building connections with someone.

Although there are two major differences, I only consider one difference each time. Thus, if both two communities have the same number of connections with people inside and outside their own community, their difference lies in α_1. If both two communities already have difference in number of connections, I keep α_1 the same for both communities.

When building my model in MATLAB, I let number of people added into each community be 300 and number of simulations repeated be 1000. Results of these 1000 simulations will be average degree distribution of internal and external links over these 1000 simulations. In the study of networks, degree means how many links that one node has and degree distribution $d(k)$ means the distribution of fraction of nodes with degree k. By analyzing degree distribution of these two communities, we are able to better understand the differences of these two communities.

Case 1. Number of connections is different for two communities, but α_1 is the same.

Figure 13. Degree Distribution of Internal Links Case 1

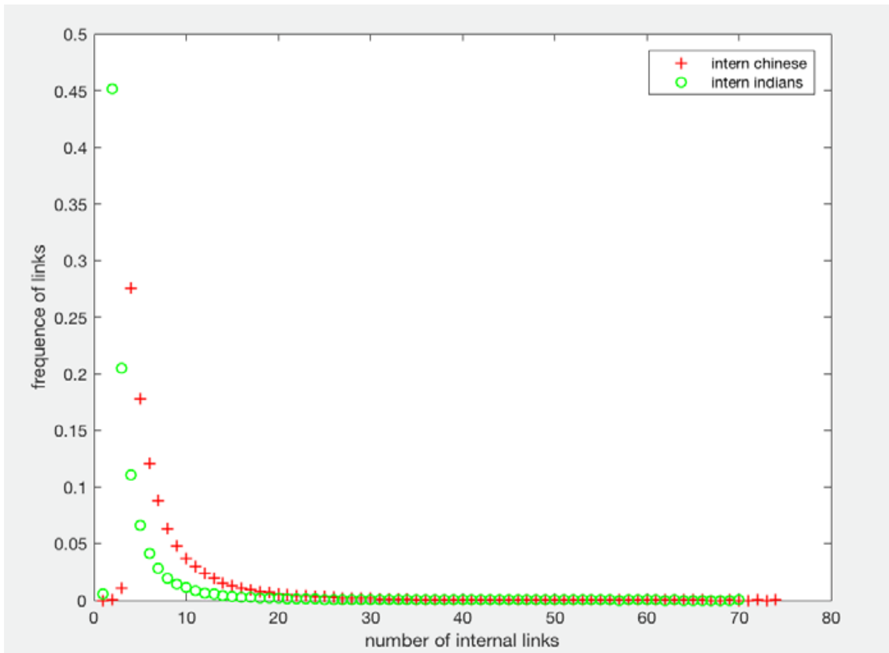
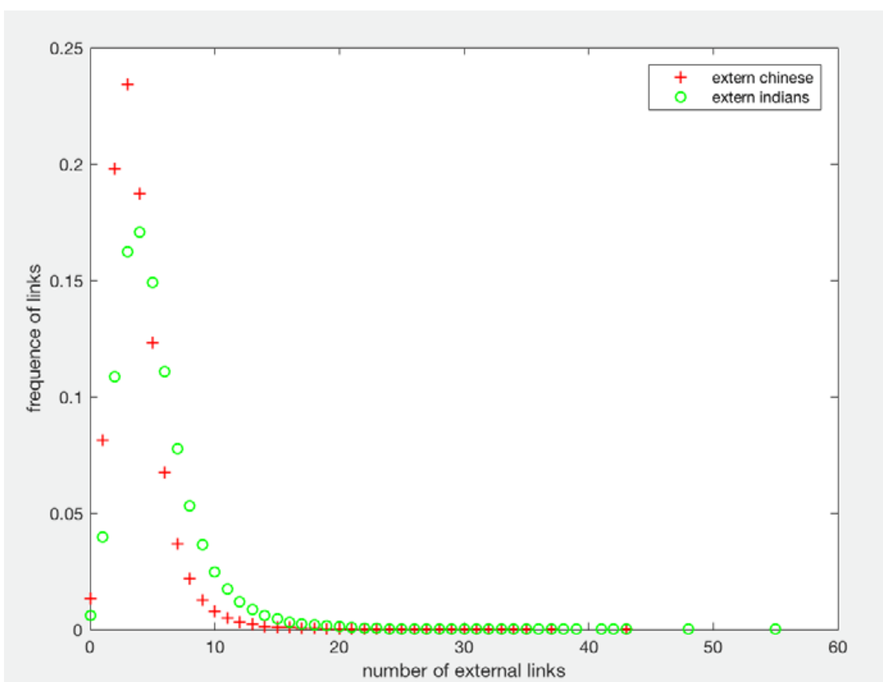


Figure 14. Degree Distribution of External Links Case 1



In this case, for Chinese newcomers, I let number of connections they make with people inside be four and number of connections they make with people outside be two. For Indian newcomers, I let number of connections they make with people inside be two and number of connections they make with people outside be four. From these two figures, we can see that for internal links, the degree distribution of Chinese community is almost the degree distribution of Indian community shifting to the right, which means in most cases, compared to Indian community, there are more Chinese immigrants having relatively more internal links. If we look at external links, it seems that the degree distribution of Indian community is almost the degree distribution of Chinese community shifting to the right, which means there are more people outside Indian community having relatively more links with Indian immigrants.

Case 2. α_1 is different for two communities, but number of connections is the same.

Figure 15. Degree Distribution of Internal Links Case 2

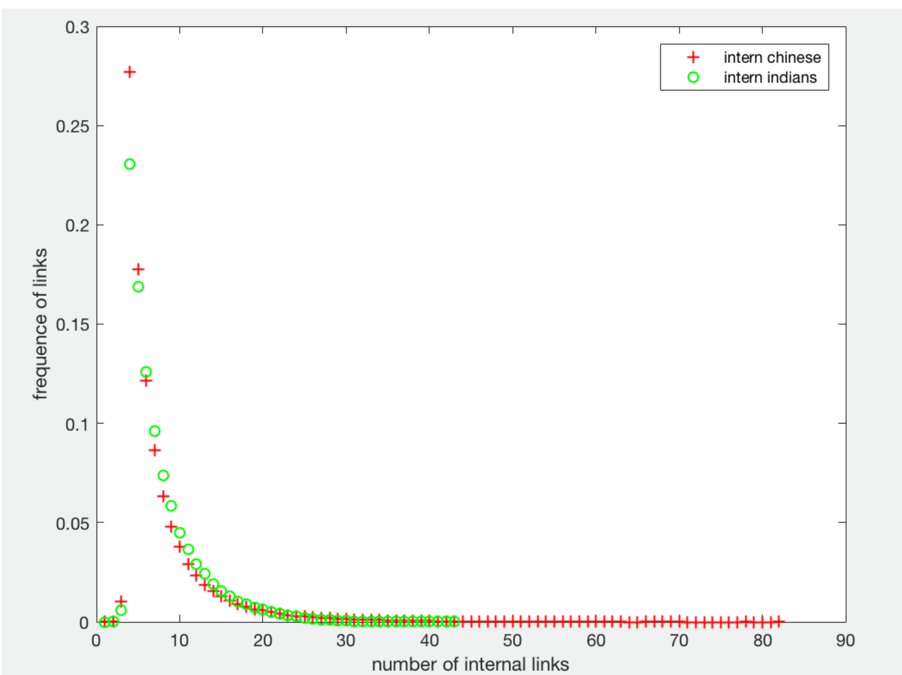
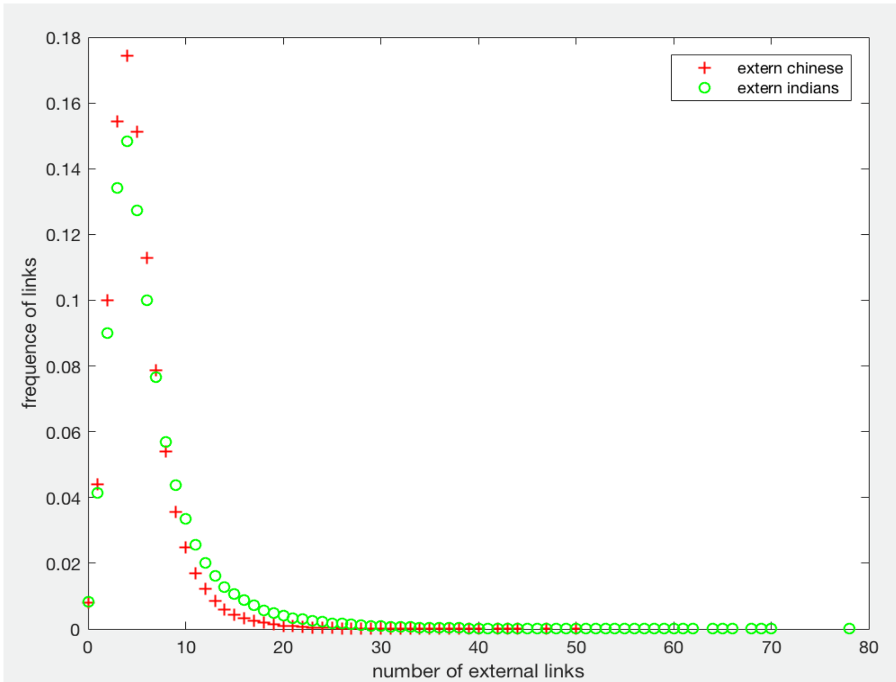


Figure 16. Degree Distribution of External Links Case 2



In this case, α_1 for Chinese internal connections is 0.6, α_1 for Chinese external connections is 0.3, α_1 for Indian internal connections is 0.3, and α_1 for Indian external connections is 0.6. From these two figures, we can see that the value of α_1 changes the tail of degree distribution. For internal links, since α_1 for Chinese internal connections is greater than α_1 for Indian internal connections, it is more likely that there is someone inside Chinese community having lots of internal links, which means for degree distribution of internal links, Chinese community has long tail. For external links, since α_1 for Indian external connections is greater than α_1 for Chinese external connections, it is more likely that there is someone outside Indian community having lots of external links with Indians inside, which means for degree distribution of external links, Indian community has long tail.

Chapter 4

Conclusion

Based on literature reviews of Prof. Saxenian's books and Dr. Dossani's paper, I found out that compared to Indian immigrants, Chinese immigrants prefer communicating and building connections with people inside their own community. While for Indian immigrants, because of their language advantage, they are more likely to jump out of their own community and connect with people outside of their community.

To better understand the difference between these two communities, I build two models based on Erdös-Rényi model and Barabasi-Albert model. Both of my two models demonstrate the dynamic formation of connections inside and outside Chinese community and Indian community. The first model is a relatively simple model but it still enables us to see the big difference between Chinese and Indian community. By looking at the probability distribution, we are able to see that the probabilities of forming relatively more internal links are greater for Chinese community while the probabilities of forming relatively more external links are greater for Indian community.

Although this simple model provides us a clear view of difference between these two communities, it is not realistic. Firstly, it is impossible that number of people inside Chinese community or Indian community remain fixed over time. Secondly, it is impossible that any two people connect with each other with same probability.

For my second model, the number of people inside community changes as time moves on and the probability of connections between people also changes, which, compared to the first model, matches actual situation more. Moreover, the degree distributions of internal links and external links show significant difference of these two communities, further confirming that Chinese immigrants' reliance on internal links and Indian immigrants' reliance on external links.

Appendix A

Code for Scenario 1

```
clear all
close all
m = 50;%number of people inside
n = 40;%number of people outside
Pc = 0.6;%probability for internal link for china
Pi = 0.3;%probability for internal link for india
Rc = 0.3;%probability for external link for china
Ri = 0.6;%probability for external link for india
A = 500;%number of period A has upper and lower bound,upper bound min(M,N)
M = (m*(m-1))/2;%total number of internal link
N = m*n;%total number of external link



E1=0;
E2=0;
E3=0;
E4=0;
for i = 0:1:A
    P1(i+1,:) = nchoosek(A,i)*(Pc^i)*((1-Pc)^(A-i));%probabilty of internal
links for chinese
    P2(i+1,:) = nchoosek(A,i)*(Pi^i)*((1-Pi)^(A-i));%probabilty of internal
links for indian
    j(i+1,:)=i;
    E1= P1(i+1,:)*i+ E1;%expectation of internal links for chinese
    E2= P2(i+1,:)*i+ E2;%expectation of external links for indian

end
for i = 0:1:A
    P3(i+1,:) = nchoosek(A,i)*(Rc^i)*((1-Rc)^(A-i));%probabilty of external
links for chinese
    P4(i+1,:) = nchoosek(A,i)*(Ri^i)*((1-Ri)^(A-i));%probabilty of external
links for indian
    k(i+1,:)=i;
    E3= P3(i+1,:)*i+ E3;
    E4= P4(i+1,:)*i+ E4;
end
Ec = E1+E3;%expectation of total links for chinese
Ei = E2+E4;%expectation of total links for indians

figure(1);
plot (j,P1);
hold on
plot (j,P2);
xlabel('number of internal links');
ylabel('probabilty of internal links')
legend('chinese','indians');
```

```
figure(2);
plot (k,P3);
hold on
plot(k,P4);
xlabel('number of external links');
ylabel('probabilty of external links')
legend('chinese','indians');
```

Appendix B

Code for Random Graph

```
clear variables
close all
m = 4; %number of internal notes
n = 2; %number of external notes
p = randperm(100);
q = randperm(100);
x = p(1:m);
y = q(1:m);
A (1,:)= x;
A (2,:)= y;%condition of internal notes
p1 = randperm(100);
p2 = randperm(100);
x1 = p(1:n);
y1 = p(1:n);
B (1,:)= x1;
B (2,:)= y1;%condition of external notes
Z1 = randperm(m);
Z2 = randperm(m);
z1 = Z1(1:m);
z2 = Z2(1:m);
a = 6;%period
b1 = 0.3;%probability
b2 = 0.6;
k = 0;
l = 0;
for i = 1:a
    c = linspace(1,100,100);
    d = randi([1,100],1,1);
    if d <= b1*100
        k = k+1;
    elseif d >= 100-100*(b2)
        l = l+1
    else
    end
end

cc = combnk(1:m,2)';
dd = (m*(m-1))/2;

for i = 1:k
    AA = randperm(dd);
    aa = AA(1,1);
    Tx =[A(1,cc(1,aa)) A(2,cc(1,aa))
        A(1,cc(2,aa)) A(2,cc(2,aa))]';
    plot (Tx(1,1:2),Tx(2,1:2),'r-+');
    cc(:,aa)= [];
    dd = dd -1;
    hold on
```

```
end

for i=1:l
    V = randperm(m*n);
    v = V(1,i);
    v1 = mod(v,m);
    v2 = mod(v,n);
    if v1 == 0
        v1 = v1 + m;
    else
    end
    if v2 == 0
        v2 = v2 + n;
    else
    end
    Q(1,i) = v1;
    Q(2,i) = v2;
end

qq = l;
AA = randperm(qq);
for i = 1:l

    aa = AA(1,i);
    Tx =[A(1,Q(1,aa)) A(2,Q(1,aa))
        B(1,Q(2,aa)) B(2,Q(2,aa))]';
    plot (Tx(1,1:2),Tx(2,1:2),'g-+');
    hold on

end
```

Appendix C

Code for Scenario 2

```
clear all
m = 5;%number of internal people
n = 300;%number of external people
A=3;%randi([0 4],1,m);% internal chinese
B=1;%randi([0 3],1,n);% external chinese
C=1;%randi([0 4],1,m);% internal indians
D=3;%randi([0 3],1,n);% external indians
alphaic = 0.6;% alpha of internal chinese
alphaec = 0.3;% alpha of external chinese
alphaii = 0.6;% alpha of internal indians
alphaei = 0.3;% alpha of external indians
Z = randi(m+n,m+n);
Z = mod(Z,2);
k =300;
kk = 1000;
for x =1:kk
for i = 1:m+n
    Z(i,i)=0;
end
for i = 1:m
    if sum(Z(i,1:m)) == 0
        M = randi(100,m);
        Z(i,1:m) = M(i,:);
        Z(i,1:m) = mod(Z(i,1:m),2);

    else
    end
    if sum(Z(i,m+1:m+n)) == 0
        N = randi(100,n);
        Z(i,m+1:m+n) = N(1,:);
        Z(i,m+1:m+n) = mod(Z(i,m+1:m+n),2);

    else
    end
end
for i = 1:m+n
    for j = i:m+n
        Z(j,i) = Z(i,j);
    end
    %initial the social network
end

for i = 1:m+n
    Z(i,i)=0;
end

Zfinal_c = [Z(1:m,1:m) zeros(m,k),Z(1:m,m+1:m+n)
          zeros(k,m+k+n)
```

```
          Z(m+1:m+n,1:m),zeros(n,k),Z(m+1:m+n,m+1:m+n)];
Zfinal_i = Zfinal_c;
connected = 0;
picked = 0;
for i = 1:k
    choosed=zeros(2,m+i+1);
    for j = 1:A
    ph = rand(1);
    if ph <= alphaic
        picked = randperm(m+i-1,1);
        g = 0;
        for h = 1:m+i-1
            if Zfinal_c(picked,h) == 1
                g = g + 1;
                choosed(2,g) = h;
                choosed(1,g) = g;
            end
        end
           if g ~= 0
            connected = randperm(g,1);
            Zfinal_c(m+i,choosed(2,connected)) = 1;
            Zfinal_c(choosed(2,connected),m+i) = 1;
           else
           end
     elseif ph > alphaic
         picked = randperm(m+i-1,1);
         Zfinal_c(m+i,picked) = 1;
         Zfinal_c(picked,m+i) = 1;
     end

    end
end

    for e = 1:m+n+k
        for f = e:m+n+k
        Zfinal_c(e,f) = Zfinal_c(f,e);
        end

    end


connected = 0;
picked = 0;
picked1 = 0;
for i = 1:k
    choosed=zeros(2,m+n+1);
    for j = 1:B
    ph = rand(1);
    if ph <= alphaec
        picked = randperm(m+i-1,1);
        g = 0;
        for h = m+k+1:m+k+n
            if Zfinal_c(picked,h) == 1
```

```
                g = g + 1;
                choosed(2,g) = h;
                choosed(1,g) = g;
            end
        end
        if g ~= 0
            connected = randperm(g,1);
            Zfinal_c(m+i,choosed(2,connected)) = 1;
            Zfinal_c(choosed(2,connected),m+i) = 1;
        else
        end
     elseif ph > alphaic
         picked1 = randperm(n,1);
         Zfinal_c(m+i,picked1+m+k) = 1;
         Zfinal_c(picked1+m+k,m+i) = 1;
     end

    end
end

    for e = 1:m+n+k
        for f = e:m+n+k
        Zfinal_c(e,f) = Zfinal_c(f,e);
        end

    end


connected = 0;
picked = 0;
for i = 1:k
    choosed=zeros(2,m+i+1);
    for j = 1:C
    ph = rand(1);
    if ph <= alphaii
        picked = randperm(m+i-1,1);
        g = 0;
        for h = 1:m+i-1
            if Zfinal_i(picked,h) == 1
                g = g + 1;
                choosed(2,g) = h;
                choosed(1,g) = g;
            end
        end
        if g ~= 0
            connected = randperm(g,1);
            Zfinal_i(m+i,choosed(2,connected)) = 1;
            Zfinal_i(choosed(2,connected),m+i) = 1;
        else
        end
     elseif ph > alphaii
         picked = randperm(m+i-1,1);
         Zfinal_i(m+i,picked) = 1;
```

```
            Zfinal_i(picked,m+i) = 1;
        end

    end
end

    for e = 1:m+n+k
        for f = e:m+n+k
        Zfinal_i(e,f) = Zfinal_i(f,e);
        end

    end

connected = 0;
picked = 0;
picked1 = 0;
for i = 1:k
    choosed=zeros(2,m+n+1);
    for j = 1:D
    ph = rand(1);
    if ph <= alphaei
        picked = randperm(m+i-1,1);
        g = 0;
        for h = m+k+1:m+k+n
            if Zfinal_i(picked,h) == 1
                g = g + 1;
                choosed(2,g) = h;
                choosed(1,g) = g;
            end
        end
        if g ~= 0
            connected = randperm(g,1);
            Zfinal_i(m+i,choosed(2,connected)) = 1;
            Zfinal_i(choosed(2,connected),m+i) = 1;
        else
        end
     elseif ph > alphaii
         picked1 = randperm(n,1);
         Zfinal_i(m+i,picked1+m+k) = 1;
         Zfinal_i(picked1+m+k,m+i) = 1;
     end

    end
end

    for e = 1:m+n+k
        for f = e:m+n+k
        Zfinal_i(e,f) = Zfinal_i(f,e);
        end

    end

for i =1: m+k
```

```
    Tchinese_i(1,i) = sum(Zfinal_c(i,1:m+k));

    Tchinese_final_i(1,i+(x-1)*(m+k)) = Tchinese_i(1,i);
end

for i =1: m+k
    Tindians_i(1,i) = sum(Zfinal_i(i,1:m+k));

    Tindians_final_i(1,i+(x-1)*(m+k)) = Tindians_i(1,i);
end

for i = m+k+1:m+k+n
    Tchinese_e(1,i-m-k) = sum(Zfinal_c(i,1:m+k));

    Tchinese_final_e(1,i-m-k+(x-1)*(n)) = Tchinese_e(1,i-m-k);
end

for i = m+k+1:m+k+n
    Tindians_e(1,i-m-k) = sum(Zfinal_i(i,1:m+k));

    Tindians_final_e(1,i-m-k+(x-1)*(n)) = Tindians_e(1,i-m-k);
end
end

Tchinese_internal = tabulate(Tchinese_final_i);
Tindians_internal = tabulate(Tindians_final_i);
Tchinese_external = tabulate(Tchinese_final_e);
Tindians_enternal = tabulate(Tindians_final_e);

figure(1);
plot(Tchinese_internal(:,1),Tchinese_internal(:,3)/100,'r+')
hold on
plot(Tindians_internal(:,1),Tindians_internal(:,3)/100,'go')
legend('intern chinese','intern indians')
xlabel('number of internal links')
ylabel('frequence of links')


figure(2);
plot(Tchinese_external(:,1),Tchinese_external(:,3)/100,'r+')
hold on
plot(Tindians_enternal(:,1),Tindians_enternal(:,3)/100,'go')
legend('extern chinese','extern indians')
xlabel('number of external links')
ylabel('frequence of links')
```

BIBLIOGRAPHY

Dossani, R. (2002). Chinese and Indian Engineers and their Networks in Silicon Valley. Retrieved from https://cddrl.fsi.stanford.edu/sites/default/files/Dossani_Survey.pdf

Jackson, M. O. (2011). Social and economic networks. Princeton, NJ: Princeton University Press.

Saxenian, A. (2002). Silicon Valley's New Immigrant High-Growth Entrepreneurs. Economic Development Quarterly, 16(1), 20-31. doi:10.1177/0891242402016001003

Saxenian, A. (1999). Silicon Valley's new immigrant entrepreneurs. San Francisco, CA: Public Policy Institute of California.

Saxenian, A., Motoyama, Y., & Quan, X. (2002). Local and global networks of immigrant professionals in Silicon Valley. San Francisco, CA: Public Policy Institute of California.

Wadhwa, V., Saxenian, A., & Siciliano, F. D. (n.d.). (2012). Then and Now: America's New Immigrant Entrepreneurs, Part VII. SSRN Electronic Journal. doi:10.2139/ssrn.215987

ACADEMIC VITA
Shuhan Tian

(814) 777-8161 | sqt5285@psu.edu | Address: 601 Vairo Blvd Apt 337 PA 16803

EDUCATION

Pennsylvania State University (PSU) --University Park State College, PA
B.S. Mathematics *B.S. Economics (Schreyer Honors College)* *Minor in Statistics* Class of 2017
Honors: Dean's List awarded all semesters (Top 1% at PSU); B.A. Economics (*Schreyer Honors College*)

Thesis Title: Differences in social networks between Chinese community and Indian community in Silicon Valley
Thesis Supervisor: Kalyan Chatterjee

ACCOMPLISHMENT

The President's Freshman Award	2014	The Evan Pugh Scholar Award	2016
The President's Sparks Award	2015		

ACTUARIAL EXAMINATIONS

SOA Probability/CAS 1	Jul 2015	Completing coursework for VEE credit
SOA Financial Mathematics/CAS 2	Dec 2015	

WORK EXPERIENCE

Bank of China Insurance Company Beijing, China
Online Intern in Actuarial Department Sept 2016 – present
Refined Generalized Linear Model by adjusting existed parameters and adding new parameters
Participated and assisted in the reserve valuation by creating run-off triangle spreadsheet

Bank of China Insurance Company Beijing, China
Actuarial Intern in Actuarial Department Jul 2016 – Aug 2016
Updated the auto policy dataset by classifying and merging relevant data in SAS
Built a Generalized Linear Model to price auto insurance in SAS

Penn State Economics Department State College, PA
Research Assistant & Teaching Assistant Aug 2015 – May 2016
Met on a weekly basis for the project named "Negative vs. Positive Campaign Advertisements"
Digitized political election datasets: presidential, governor & congress candidate profiles using Excel
Assisted in compiling DMA (Designated Market Area) datasets, whereby bordering counties are counted and recorded
Constructed regression models of political variables to obtain empirical results using STATA
Assisted professor during classes, proctored and graded exams, taking full responsibility of collecting homework

SKILLS & OTHERS

Language: Chinese (native), English (fluent)
Computer skill: C/C++, R, Minitab, SAS, STATA, MS Excel, MS Word, MS PowerPoint