

A Comparative Study of Social Network Tools for Analysing Chinese Elites

HeeJeong Jasmine Lee¹, and In Kim^{2*}

¹ Pierson College, PyeongTaek University
PyeongTaek, South Korea
[e-mail: hjlee@ptu.ac.kr]

² Pierson College, PyeongTaek University
PyeongTaek, South Korea
[e-mail: kimin@ptu.ac.kr]

*Corresponding author: In Kim

*Received March 11, 2021; revised June 14, 2021; accepted September 5, 2021;
published October 31, 2021*

Abstract

For accurately analysing and forecasting the social networks of China's political, economic and social power elites, it is necessary to develop a database that collates their information. The development of such a database involves three stages: data definition, data collection and data quality maintenance. The present study recommends distinctive solutions in overcoming the challenges that occur in existing comparable databases. We used organizational and event factors to identify the Chinese power elites to be included in the database, and used their memberships, social relations and interactions in combination with flows data collection methodologies to determine the associations between them. The system can be used to determine the optimal relationship path (i.e., the shortest path) to reach a target elite and to identify of the most important power elite in a social network (e.g., degree, closeness and eigenvector centrality) or a community (e.g., a clique or a cluster). We have used three social network analysis tools (i.e., R, UCINET and NetMiner) in order to find the important nodes in the network. We compared the results of centrality rankings of each tool. We found that all three tools are providing slightly different results of centrality. This is because different tools use different algorithms and even within the same tool there are various libraries which provide the same functionality (i.e., ggraph, igraph and sna in R that provide the different function to calculate centrality). As there are chances that the results may not be the same (i.e. centrality rankings indicating the most important nodes can be varied), we recommend a comparison test using different tools to get accurate results.

Keywords: Chinese elite, power elite, social network analysis, Chinese elite database

A preliminary version of this paper appeared in the Journal of Chinese Studies (94, p83 ~ 111), Korea. This version includes a concrete analysis and supporting implementation results on a comparative study of social network tools.

1. Introduction

Although a minority, China's elite have a great influence on crucial decisions of the country pertaining to numerous areas. Therefore, the analysis of the elite in power – the individuals, their roles and their social interaction has become one of the major areas of research about China. The purpose of this study is to determine the methods and challenges associated with compiling a database containing information of individuals, while also has the capacity to reveal their interconnected paths, based on information. We have collected the structured and unstructured data pertaining to Chinese elites including their academic background, careers, personal relationships, and professional achievements. The 'social tie data' represents a series of social relationships that connect two elites, which indicates the potential avenues of connecting two people. Using the Social Network Analysis (SNA) method that incorporates processes that calculate and highlight interconnecting relationships, the system user can schematize the relationships that could exist among the elites. This is the first step towards development of a SNA system which can be used for identifying the powerful elites who are involved in policy decisions.

While the Communist Party of China monopolizes China's future direction as a country and controls many aspects of the country's politics, economy, society and culture, their governance process remains largely undisclosed. To this end, it is necessary to reconstruct the known information in order to better understand who is in charge of the regime, who will rule, and the ruling process of the Chinese Communist Party. For this purpose, knowledge on the social relations information of the Communist elite of China will be useful in analysing their potential influence in various fields. This study will benefit future researchers who wish to collect and analyse social network data both in and outside of China.

2. Materials and Methods

2.1 The theoretical background

In 1967, the social psychologist, Stanley Milgram, carried out a test on "The Small World Theory" [1], which propositioned that members of large social networks are connected to each other through the short chain of people who know them (i.e. middle acquaintances). Experiments have shown that the average middle acquaintances constituted of about six people, and this discovery has since been known as the "Six Degrees of Separation" [2]. The theory became the foundation for the social network theory, and many experiments have revealed that people can theoretically use six acquaintances to ultimately reach any other person [3-4]. This theory in combination with Milgram's findings has given rise to numerous other interesting research on social networks. For example, in some communities, such as those consisting of mathematicians and actors, the average number of middle acquaintances required for connecting any two of its members was lesser than six [5-6]. Therefore, given that China's central authorities are considered to be a 'small world', the average number of middle acquaintances can also be predicted to be lesser than six. However, in the local states, the relationships between influential individuals could be less pronounced than in the cities due to China's geographical enormity, which could result in a higher degree of separation. Therefore, investigating the relationships of the small world of the Chinese elite could be an interesting future area of research.

People use social networks to connect with others for the following reasons:

- Meeting (dating, marriage, etc.)
- Shared hobbies, friends and communities
- Occupation-related reasons (for job promotions, job searches, securing partnership, etc.)
- Business interests (the sale of goods and services, for building relationships between organizations, etc.)
- Political or non-profit interests (in sharing political views and agendas, promoting charities, etc.)
- Religious purposes (e.g., missionary work)

The difference between failure and success in business often appears to depend on "who I know" rather than on "what I know." People prefer to meet new people on social networks that are arranged through existing relationships or acquaintances (or match makers) because they consider such meetings to be more comfortable, more efficient, and more likely to lead to favourable outcomes. In comparison, interacting with strangers could involve discomfort, deception, and even physical danger. Such risks are particularly significant when people decide to meet newly acquainted individuals in person following the establishment of online connections, and could discover that their actual personalities are completely different to that reflected in their online profiles. To this end, people who use social networking with the objective of building friendships or developing business relationships are at risk of being exposed to deception and rejection. To prevent such untoward incidences, a verification system is required that guarantees the formal identification of a person. In the absence of such measures, establishing new connections via introductions by mutual friends or their extended contacts is a pragmatic approach as it greatly reduces the associated risks.

2.2 The State of China's Elite DB Establishment and Service

Taiwan, which has an ethereal and special relationship with China, has established a database of the Chinese elite at the National Taiwan University of Political Science, called the "CCP political elite database". The database provides basic information on the political elite in mainland China, including their name, age, place of birth, ethnic group, education, career experience and the timeline of their careers. Currently, it contains the records of around 4,000 Chinese elites, which are limited to officials and political executives who have held various positions of office higher than that of a deputy prime minister since 1966. While the CCP political elite database is beneficial in easily finding the personal information of China's power elite, it does not provide data about the possible correlations that could exist between individuals.

China has also built its own database that contains information pertaining to influential individuals across different entities. These include: the Chinese Communist Party Newspaper Network (<http://cpc.people.com.cn>), the Central Committee of the Communist Party of China, the Executive Committee of the NPC, the Prime Minister and the Head of State Council, and local officials. However, this also does not provide any data on the correlations or networking between the power elites. The users of this database can easily access the personal information of the current leaders in China's major central and provincial areas.

Outside of the political elites, other similar associations also exist such as the Chinese Academy of Sciences (http://casad.cas.cn/ysxx2017/ysmdyjj/qtysmd_124280), which focuses on Chinese scientists. As of July 2019, this organization has compiled the biographies of 778 members, in comparison to a community of only 60 members prior to 2014. Currently, new

members for the Academy are recruited strictly on the recommendation of the existing members or academic organizations, a process in which foreigners and senior members are not permitted to take part. Of the total applicants, only about 20% are recruited to the Academy by way of elections that are held by secret ballot, a membership system that has been criticized to be highly bureaucratic. The membership of the Academy is associated with a range of incentives and entitlements including government benefits that are on par with that of a deputy prime minister, impressive legal grants and large subsidies from local governments. Further, their undue influence in certain areas of research could lead to the monopolization of research grants.

The project, 'Connected China', founded by Thomson Reuters is a prime example of a database of the China elite that is established outside of China, which aims to identify the relationships that exist between the Chinese elite (such as their families, mentors, competition and alliances) and the potential channels of communications associated with them. This project analyses the social networks based on the information gathered by Reuter reporters, and aims to inform the understanding of how vital decisions are made in China. The service web site (<http://china.fathom.info>) examines the professional network of China's power elites in depth, including their social, political, economic and social aspects. In this scheme, relationships between the elites were defined as '1 step' or '2 steps', which denotes that their link is direct or through a middleman, respectively. The collection of data represents the interactive social relationships of 399 Chinese power elites. The locations of the elites are determined by the Force Atlas layout algorithm implemented in Gephi [7], an open source network visualization software that is used in numerous fields (such as social network analysis, biology, genetics, etc.) [8], which enables the visualization of graphs and networks. The project demonstrates the social relationships of Chinese elites in an interactive manner, along with their personal profile data [9].

The Force Atlas layout algorithm represents the data in aesthetically superior graphs despite the speed at which they are generated, with little or no cross lines between them. **Fig. 1** illustrates the social network relations of the Chinese President, Xi Jinping, as an example, as represented by the Force Atlas layout algorithm.

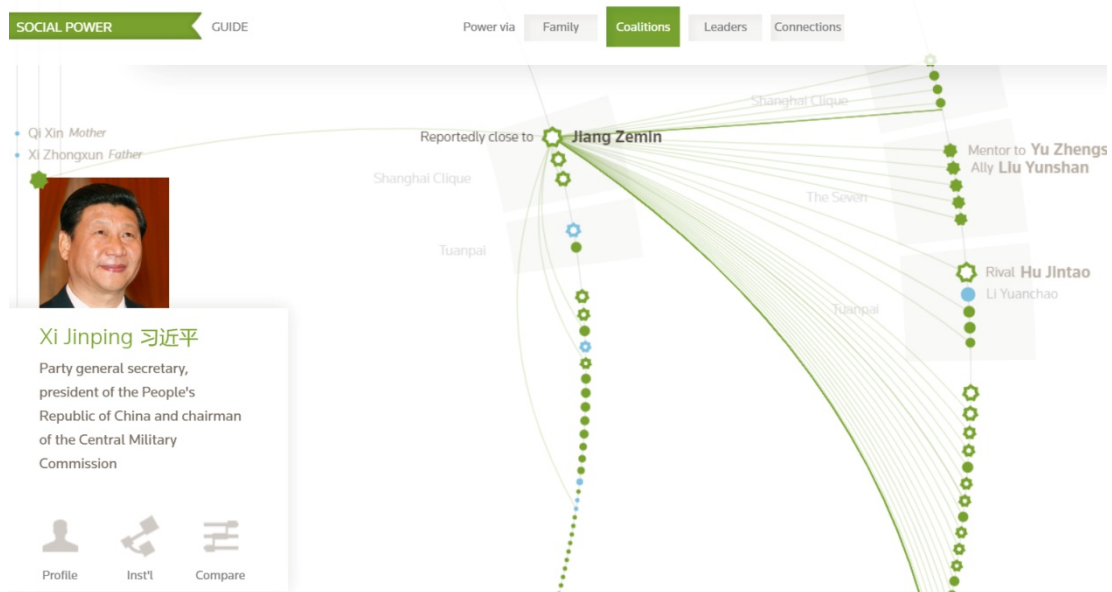


Fig. 1. The Social Network Relations of China's Xi Jinping (<http://china.fathom.info>).

The Global Database of Events, Language and Tone (GDELT)(<https://gdeltproject.org>) uses a series of software tools to crawl the world's broadcast, print and online news in real time, extract the events that are being reported, as well as the location of these events along with sentiment information [10]. Researchers explored and coded to create the social relations between individuals (i.e. actors) that are engaged in these events by doing network analysis. These actor relations are combined together with Tone (i.e. sentiment) and other factors to create a new dataset which was subsequently used in the prediction tasks [11].

3. Network Analysis Methodology

3.1 The definition of 'elite'

In order to determine a method for the development of a database that incorporates information on the elites of China, it is important to define what constitutes as the 'power elite' in Chinese society. To achieve this, we utilized organizational and event factors where we have assumed that individuals who belong to important organizations (such as the National Congress of the Communist Party of China) or are engaged in hosting or participating in significant events of the country are considered as elites. **Table 1** summarises the factors that were considered.

Table 1. Structural and Behavioural Approach to Extracting Information from the China elite.
Source: Korea Silk Road Research Institute, Proposal for Building a Governing Elite DB in China's Reform and Opening Period (2017).

Organizational Factors	Event factors
State institutions, Governmental organizations, Economic organizations, Enterprises, Social organizations, Schools, NGOs, Cultural and artistic organizations, Religious organizations, etc.	Important meetings, Policies, Political changes, Theories, Debates, Demonstrations, Social movements, Political meeting, Academic conferences, Cultural events, Artistic ideas, Artistic movements, Contemporary issues, etc.

Organizational factors are clearly represented by elite groups of people that serve as high-ranking officials such as the Central Committee of the Communist party, who are the top decision-makers that guide the direction of the party, the state, and the society in general. For example, according to this approach, the following individuals could all be considered to be Chinese elites: the President and Vice-president of an organization that is directly under the Central Committee of the Communist Party; the Secretary, the Director and Deputy director of the State Council; the Chief and Deputy chief of the Party Committee; the Chief and Deputy chief of the Hasty administrative unit; the County District Chief and Deputy chief of staff; and the Head of the National mass organization and the Deputy chief of staff [12]. In economic organizations, the details of the CEOs of major companies and representatives of various social organizations are extracted through existing organizational factor reviews (e.g., from the 'top 50 Chinese companies').

Existing research on the Chinese elite has largely adopted a horizontal or a vertical method for data collection in order to provide personal or profession information of an individual. Therefore, the data obtained constitutes only fragmentary and distributed information (e.g., "Elite A worked at a specific time or region" or "Elite B came from that school"). This approach has several disadvantages, such as the excessive costs of

reconstruction as a result of the high access speeds required when data originated from different fragments are combined, and the potential for the database to be ineffective if there was site failure due their lack of back-up copies for data.

The purpose of this study is to overcome these weaknesses and to establish a database that ascertains the personal relationships of the power elites in China. In collecting data, various meetings and events of the Chinese Communist Party was used as one of the main data collection methods, which was used in combination with other justified methods that were previously applied by the 10 co-researchers for collecting their existing data. All co-researchers are subject-matter-experts (SME) and university academics with PhDs, who are engaged in the field of Chinese studies, and are highly literate in Chinese. The data has been collected with the aim of conducting a comprehensive social network analysis including the identification of the key power elites within a particular social network and community (such as a clique or a cluster). The data has been collected since September 2017 and continued until the completion of the project in February 2021.

3.2 The definition of the Elite Relation

Borgatti et. al identified four categories of factors that determine the types of power elites: organizational factors, social relations, interactions and flows (described further in [Table 2](#)) [13]. In this scheme, if two persons belong to one common category, they can be considered to have relations.

Table 2. Structural and Behavioural Approach to Extracting Information from China Elite.
Source: Borgatti et. al, Network analysis in the social sciences, p.894. Reconfigured by authors.

Type	Description
Membership	This denotes instances where two elites belong to one or more groups or participate in one or more events together. For example, when two elites belong to the same membership (state, party, government, economic, social, school, cultural and artistic organizations, religious organizations, etc.) or attend the same event (important meetings including policy meetings, demonstrations, social movements, academic conferences, cultural events, etc.) in the same location (at the same time).
Social relations	This includes kinship (family, friends, professional or academic) and cognitive relationships (knowing). These types of relationships can be used in personal community research.
Interactions	A connection based on the act of talking, providing help or advice.
Flows	A type of relationship based on exchanges between the elites, such as the flow of resources, goods, information, and influence in social networks.

3.3 Basic Data Collection Method

First, we focused on how to extract, organize, and analyse the primary data of the elites in order to build the DB. The goal was to overcome the limitations of being able to obtain an elite's profile from portal sites and to establish a higher level of data collection. The data pertaining to individual elites will be collected using all available information from databases across China and overseas and portal sites. Most of the existing data has been collected by the 10 co-researchers who are deeply knowledgeable SMEs in the research area (with a particular emphasis on China), and using methods based on firm foundations of research. **Fig. 2** illustrates the approach applied for collecting the basic profile data of an elite. In collecting data, we have included the information of not only the currently active elites but also of those who have died, and those who have officially retired but continue to have influence on major issues in China, such as Mao Zedong, Deng Xiaoping, and Jiang Zemin.

The process of entering data into the database is as follows: firstly, the official information that is available on the Internet was collected; secondly, integrated filtering was carried out; and, finally, the validation process was conducted to correct any missing or incorrect information. As some of the records relate to the elites who are still active, their data will be continuing to be updated depending on the availability of new information.

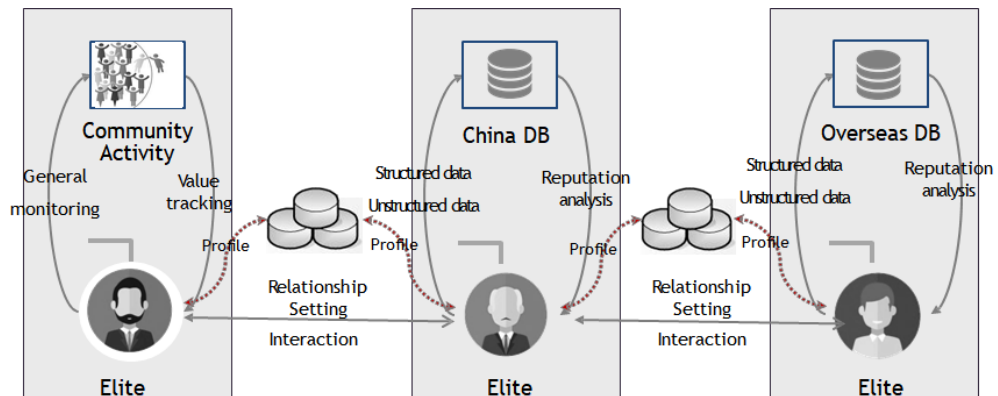


Fig. 2. Basic Data Collection Method.

To assess the relationships between the individuals, we carefully collect both inductive and deductive data. The inductive data collection method enables the collection of data through informal routes, and is used for qualitative data analysis. A deductive data collection method is used to collect data that is disclosed to the general public in newspapers and portals, and is used for quantitative data analysis [9].

3.4 Data Collection Scope

The collected data represented an array of highly diverse information about the elites that could be entered into the database. These include: names, photos, dates of birth, occupations, gender, current position, educational background, career, family, close associates, related organizations, awards, books, papers, works, religion, hobbies, professional fields, marriage status, children and preferences (including impressive books, favourite movies and songs, etc.).

After entering the basic information for the individual elites, the relational information between two elites were collected. In a previous study, Keller systematically investigated the

relationship between sponsors and beneficiaries by analysing the sponsor networks using social network analysis tools. In addition, the author applied various theories to assess the best approach to measure the sponsor networks that were associated with the members of the Central Committee of the Communist Party of China, from 1982 to 2012 [14]. In 2016, Keller introduced two methods that could be used for data collection with regard to informal networks and the factional relationships of Chinese elites. The exploratory approach traces the alliances through a qualitative evaluation based on the information obtained from an insider, whereas the structured approach infers the factional relationship using publicly available data [15].

In order to examine the relationship information of two elites, various aspects in their lives were analysed such as their careers, associated organizations, educational background, family ties, and colleagues. For political elites, holding positions of office in an influential organization was sufficient to be identified as an elite, such as the members the Central Committee and State Council of the Communist Party of China, the main officials of the Chinese city and state, and the members of the Chinese People's Congress. The Connected China Project [8] identified the indicators of vital relationships that warrant an individual to be considered as an elite, which include the 'golden' sons-in-law, past and current leaders, princelings and allies. The connections are often boosted by the complex relationship of in-laws, mentored by a 'mishu' – which means relative, to whom the elites report and are close. Within this indicator, the network of each elite is completed using the information given not only by an individual's family members but also by guardians and secretaries that are connected with their families. Therefore, current positions and past experiences of the elites were taken into account in order to collate this information, which is vital to analysing the relationships between the elites.

The relationship indicators appear to be independent of each other, but the 'guanxi mechanism', which refers to the special nature of Chinese relationships, operates superficially. Another database called the 'Bang' exists, which comprises of the data of the major power elites in China. It comprises predominantly of information of the elites from Shanghai and their princelings. 'Princelings' refer to the children who are the 2nd or 3rd generation descendants of members of the revolutionary circle of the Chinese Communist Party, one of the fractions that constitute the core power of China. In addition to the direct descendants, men who are connected to the group members through marriage – called 'sons-in-law' – are also incorporated to the inner circle of the princelings. They are currently active in China's political and business circles and considered as the most influential power elite group in the country.

Apart from blood ties, another important indicator is 'reported close to'. There are three basic factors in the process of building Chinese guanxi, namely, humanity, disgrace and power [16]. These three factors have played a significant role in safeguarding positions of privilege through successive generations in powerful Chinese families, honed through their close relationships and affiliations between the parents, grandparents, the people who have been involved with them, and their childhood peers.

According to existing knowledge, if the relationship between two elites that is considered to be 1 step (direct link) and 2 steps (link through middleman) is extended to their princelings and relatives, links between most of China's leaders can be formed.

Once the data is collected, it is possible to use the system to conduct an integrated search by sector (e.g., political, economic, social or cultural), name, department, position, occupation, birthplace, school, origin, date of birth, family name, etc. The system is designed to allow the incorporation of "and" or "or" search parameters, which enables refining the search of elites further by combining two or more search criteria.

3.5 Developing the China Elite DB System

This study provides a system that enables probing basic information about China's elites using an online database, with associated graphs for ease of visualization. The system can be accessed by designated users at the web address: <http://koreasilkroad.kr>. We plan to extend this access to the general public in the future through a membership-based scheme. The system displays the basic and relationship data for the elites, and thereby portrays a broad personal profile for each of them. The relationship data reveals existing relationship between the elites; for example, if the personal profile indicates the university an elite graduated from, the relationship data would identify other elites who attended the same institute, and therefore, are potential friends.

The system of the elite database combines number of functions such as compiling data from various sources, analysing relationships, determining the patterns of connections and analysing their influence. SNA is a methodology that organizes social structures into objects (nodes, actors) and their relationships (links, edges), models them as networks, and analyses their structures to identify important embedded information such as network degree, density and others (see the Fig. 3). Depending on the type of data and the purpose of the analysis, various SNA methods can be applied. To this end, SNA has given rise to possibilities that allow users to analyse the 'relationships' and 'interactions' that were difficult to identify from traditional statistical analyses, and to identify the macroscopic and micro relations of relationships, which could be consequently used to predict the potential formations, structures or changes that could occur in existing relationships.

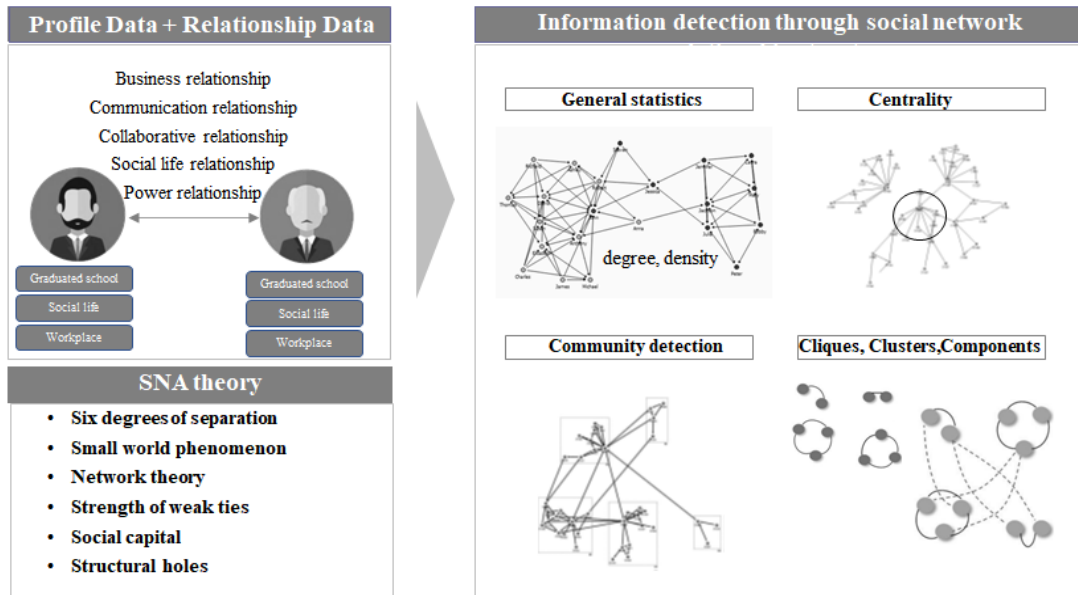


Fig. 3. Social Network Analysis.

The system developed by us uses a database to store the basic and relational data of China's elites. This study comprised of four steps (as illustrated in Fig. 4). During the first step, we focused on the system design and implementation, in which the database schema was designed and the applications of Apache and MySQL were incorporated into the database. In step two, we compiled the basic data of individual elites, which involved collecting as much

information as possible including their academic background, career, family relationships and achievements. Step three incorporated social network analysis, where the relationships between elites were entered as nodes and links. During the final step, the SNA models were developed, visualised and optimised.

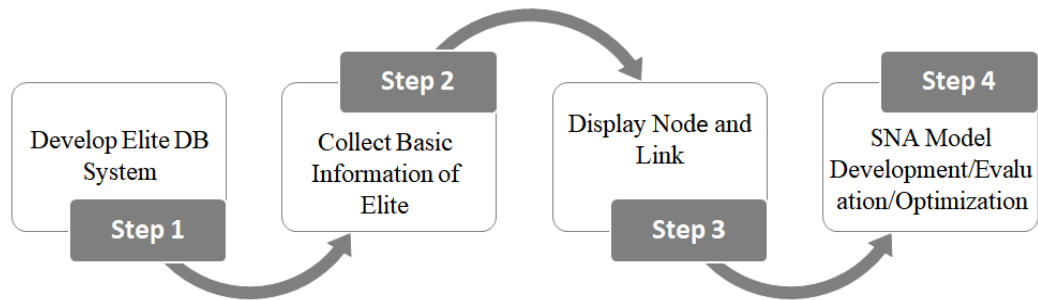


Fig. 4. The Social Network Analysis Process.

3.5.1 List of elites and their social relationships

Fig. 5 illustrates a list of elites whose data has been collected, which include the data pertaining to their relationships. As of June 2021, data records of 4,521 elites have been stored in the database.

Elite ID	Name(Kr)	Name(Ch)	Name(Pi)	Current Org	Post	Action
1	쉬자인	许家印	Xu JIayin	당기관(party organ)	상무위원	View & Edit Delete
3	장이머우	张艺谋	Zhang Yimou		영화감독	View & Edit Delete
4	왕치산	王岐山	Wang Qishan	국가기관(state organ)		View & Edit Delete
5	리커창	李克强	Li Keqiang	당기관(party organ)		View & Edit Delete
6	리젠홍	李建红	Li Jianhong	당기관(party organ)	위원	View & Edit Delete
7	쉬녜사	徐念沙	Xu Niansha	당기관(party organ)		View & Edit Delete
10	저우쥘칭	周俊卿	Zhou Junqing	기업(enterprise)		View & Edit Delete
11	판강	潘刚	Pan Gang	기업(enterprise)	이사장	View & Edit Delete


Fig. 5. List of elites.

3.5.2 Elite Basic Information Input Screen User Interface

Fig. 6 shows the user interface of the database that is used to insert the basic information of the elites.

Basic Information

From PC
URL Link



Name(Korean) 시진핑 Name(Chinese) 习近平 Name(Pinyin) Xi, Jinping

산시(山西) Place of Birth(City) Place of Birth(County) 푸핑(富平)

Date of Birth 1953 Date of Death Gender Male

한족(Han Nationality) Religion 기타(Others) Religion(Other)

Save

Family

Relation Type Relation Type(Other)

Name(Korean) Name(Chinese) Name(Pinyin)

+ Save & Add More

Relationship Type	Name(Korean)	Name(Chinese)	Name(Pinyin)	Action
아버지(father)	시중신	习仲勋	Xi, Zhongxun	View & Edit Delete
어머니(mother)	지신	齐心	Qi, Xin	View & Edit Delete
배우자(spouse)	펑리위안	彭丽媛	Peng, Liyuan	View & Edit Delete
자녀(children)	시밍궈	习明泽	Xi, Mingze	View & Edit Delete

Fig. 6. Elite Basic Information Input Screen – An Example.

3.6 Quality of Data

The quality of data is defined as the state or fitness of the data for its intended uses in operations. Therefore, data can be considered to be of high quality if it meets the criteria of: completeness, uniqueness, validity, consistency and accuracy (as shown in Table 3) [15].

Table 3. Definitions of the general data quality standards.

Criteria	Definition
Completeness	Mandatory items are not omitted.
Uniqueness	Data items should be unique and not duplicated.
Validity	The data item should exist within the specified valid range and its domain.
Consistency	The structure, values and forms of data should be defined consistently.
Accuracy	The values of objects should reflect and refer to those that exist in the real-world.

Detailed standards of quality were derived based on five data quality parameters, as shown in Table 4, which also illustrates examples for their application with regard to the present study.

Table 4. Defining general data quality standards.

Quality Standard	Detailed Quality Standard	Quality Criteria Description	Example of utilization
Completeness	Individual completeness	Mandatory columns shall not contain any missing values.	The names of the elite cannot be left blank (e.g., Chinese name and Pinyin).
Effectiveness	Range effectiveness	The value shall exist within the given effective range.	Ethnic group information (e.g., Han is one of the 55 Ethnic groups).
Accuracy	Up to date	Information generation and update of information should be maintained.	Record the date of the last change of the data to inform the user of the latest update.
Uniqueness	Unique Uniformity	Column(s) shall have a unique value without duplicate values.	Elite IDs cannot be duplicated.
Consistency	Reference Integrity	Reference integrity should be maintained in a reference relationship between the elites.	The Organization Type is determined by reference to the Organization Category.

Following the compilation of data, we noticed that the photos of some of the elites were unavailable on the database due to copyright issues. Therefore, we updated the system to provide users with two options in uploading the photos: ‘from the PC’ or providing an associated URL link. If the URL is used, it bypasses any copyright issues as the photo can be accessed directly from the original Internet portal site.

In order to guarantee the accuracy of the data quality criteria, information pertaining to the elites needs to be updated constantly. In meeting this requirement, we modified the system to display the date of the latest update with the corresponding search results. This minimises the risk of information inconsistency in the system.

3.7 Social Network Analysis

There are numerous tools available in the market for analysing network including R, UCINET and NetMiner among others. Three tools offer the same functionality to calculate network centrality. These three software has differences in UI (e.g. command base vs GUI based), processing speed, number of handling nodes and licencing fee.

R (<https://www.r-project.org>) is an open-source project. It is not specifically designed for network analysis but R packages provide powerful features for network analysis. The strength of R in comparison to other software is threefold. First, R allows reproducible analysis that maybe difficult with GUI applications. Second, R provides powerful data analysis for manipulating data to prepare network analysis. Lastly, there is an increasing variety of packages created to make R an entire network analysis tool including igraph and ggraph [18]. R can easily process one million records.

UCINET (<https://sites.google.com/site/ucinetsoftware/home>) is an extensive package for the analysis and visualisation of social network data and is primarily used in academia. UCINET can read and write differently formatted documents, as well as Excel files. It offers a wide range of analytics methods that include centrality measures, role analysis and subgroup identification. However, it can handle only up to 32,767 nodes (with some exceptions) [19]. Site licenses are available for corporate customers.

NetMiner (<http://www.netminer.com>) is commercial software for the exploratory analysis and visualisation of large network data on a social network. It is used primarily in general research and it allows users to draw their network data visually and interactively, and understand the underlying patterns and structures of the network. Some of its top features are data transformation, statistics, network analysis, visualisation of network data and a programming language based on Python. With appropriate licenses, NetMiner's computing power is capable of handling over 1,000,000 nodes.

The data collected in the present study will be used to analyse the social networks of China's elites. We intend to follow-up on the analysis of social networks of elites in future research, where their contributions could be utilized as an avenue for creating value in various sectors in the country. We have analyzed 667 nodes with 3314 links.

3.7.1 Degree Centrality

Degree centrality is defined as a direct number of ties that a node has. In order to evaluate degree centrality measurement, we computed centrality for the 667 China's elites. We then ranked the elites in descending order in values and compared the first 10 elites. In the analysis, we have found that the top 3 persons are *Xi Jinping*, *Li Keqiang*, and *Wang Qishan* upon which 3 tools (i.e., R, UCINET and NetMiner) show consensus. However, we also have found that there is some difference in centrality score between R, UCINET versus NetMiner. The top 10-degree centrality nodes are shown in [Table 5](#).

Table 5. Top 10 nodes by out-degree centrality.

	R		UCINET 6		NetMiner 4	
	Elite	Out Degree	Elite	Out Degree	Elite	Out Degree
1	Xi Jinping	102	Xi Jinping	102	Xi Jinping	0.095310
2	Li Keqiang	81	Li Keqiang	81	Wang Qishan	0.083207
3	Wang Qishan	66	Wang Qishan	66	Li Keqiang	0.074130
4	Wang Huning	50	Wang Huning	50	Xu Qiliang	0.057489
5	Xu Qiliang	49	Xu Qiliang	49	Liu Yandong	0.054463
6	Liu Yunshan	45	Liu Yunshan	45	Wang Huning	0.054463
7	Liu Yandong	45	Liu Yandong	45	Liu Yunshan	0.051437
8	Wang Yang	41	Wang Yang	41	Sun Chunlan	0.051437
9	Sun Chunlan	41	Sun Chunlan	41	Wang Yang	0.046899
10	Zhao Leji	33	Zhao Leji	33	Yang Xiaodu	0.043873

3.7.2 Betweenness Centrality

Betweenness centrality can be a way of measuring centrality within a graph based upon shortest paths. A node with greater betweenness centrality could have more control across the community because more information can pass through that node. The betweenness of a specific node is equivalent to the number of shortest paths coming from all pairs of nodes in the network that passes through that node. In China's elites' network, a person who has high betweenness centrality would play this bridging role in a network. In the analysis, we have found that three tools are agreeing on the high rankings although there are some differences in a lower-ranking (i.e., from rank 7 to rank10) in centrality score between R, NetMiner versus UCINET. As a result of the analysis of betweenness centrality, top 3 persons are *Xi Jinping*, *Peng Liyuan*, and *Wang Qishan* upon which 3 tools show consensus. Top 10 betweenness centrality nodes are shown in [Table 6](#).

Table 6. Top 10 nodes by out betweenness centrality.

	R		UCINET 6		NetMiner 4	
	Elite	Betweenness	Elite	Freeman Betweenness	Elite	Betweenness
1	Xi Jinping	2650	Xi Jinping	2562	Xi Jinping	0.0060
2	Peng Liyuan	1632	Peng Liyuan	1608	Peng Liyuan	0.0037
3	Wang Qishan	1602	Wang Qishan	1543	Wang Qishan	0.0036
4	Li Keqiang	1382	Li Keqiang	1348	Li Keqiang	0.0031
5	Dong Mingjun	1303	Dong Mingjun	1146	Dong Mingjun	0.0030
6	Zhao Kezhi	1188	Zhao Kezhi	1144	Zhao Kezhi	0.0027
7	Hu Jintao	1075	Zhao Leji	941	Hu Jintao	0.0024
8	Wang Shoujun	1014	Hu Jintao	929	Wang Shoujun	0.0023
9	Zhao Leji	1003	Han Zheng	923	Zhao Leji	0.0023
10	Yu Jianfeng	966	Wen Jiabao	872	Yu Jianfeng	0.0022

3.7.3 Closeness Centrality

The closeness centrality of a node measures its average distance to all other nodes. It can be a method of finding nodes that could distribute information very efficiently through a network. Nodes having a highest closeness score get the shortest distances to all of other nodes. Therefore, it is important to find such important nodes to find out who can spread the information quickly and efficiently. The top 10 nodes are presented in the following [Table 7](#). All tools are providing different rankings although UCINET and NetMiner agreed on top 7 rankings.

Table 7. Top 10 nodes by out closeness centrality.

R		UCINET 6		NetMiner 4		
	Elite	Closeness	Elite	Closeness	Elite	Closeness
1	Xi Jinping	3.74312E-06	Li Keqiang	0.180791	Li Keqiang	0.1670
2	Li Keqiang	3.7431E-06	Xi Jinping	0.180596	Xi Jinping	0.1666
3	Sun Chunlan	3.7427E-06	Wang Qishan	0.180446	Wang Qishan	0.1581
4	Hu Chunhua	3.7427E-06	Wang Yang	0.180172	Wang Yang	0.1521
5	Han Zheng	3.7427E-06	Zhang Dejiang	0.180073	Zhang Dejiang	0.1482
6	Xu Qiliang	3.74263E-06	Liu Yunshan	0.180053	Liu Yunshan	0.1477
7	Wei Fenghe	3.74256E-06	Wen Jiabao	0.180015	Wen Jiabao	0.1473
8	Wang Qishan	3.74243E-06	Sun Chunlan	0.179884	Yu Zhengsheng	0.1455
9	Fan Changlong	3.74239E-06	Zhang Gaoli	0.179842	Sun Chunlan	0.1445
10	Wang Yang	3.74212E-06	Liu Yandong	0.179817	Zhang Gaoli	0.1434

3.7.4 Eigenvector Centrality

In graph theory, eigenvector centrality can be a way of measuring the impact of a node within a network. A high eigenvector value implies that a node is linked to a lot of nodes who themselves have high values. The eigenvector centrality has been computed on 3 different tools. In analysis of eigenvector centrality, we have found that the top 3 persons are *Li Keqiang*, *Xi Jinping*, and *Wang Yang* upon which 3 tools (i.e., R, UCINET and NetMiner) show consensus. However, there is some difference in centrality ranks between R, UCINET versus NetMiner. Besides this, other ranks of top 10 eigenvector centrality nodes are presented in [Table 8](#).

Table 8. Top 10 nodes by out closeness centrality.

R		UCINET 6		NetMiner 4		
	Elite	Eigenvector	Elite	Eigenvector	Elite	Eigenvector
1	Li Keqiang	1.000	Li Keqiang	0.2188	Li Keqiang	0.2186
2	Wang Yang	0.867	Xi Jinping	0.2143	Xi Jinping	0.2171
3	Xi Jinping	0.867	Wang Yang	0.2121	Wang Yang	0.2119
4	Wang Qishan	0.768	Zhao Leji	0.1761	Zhao Leji	0.1761
5	Liu Yandong	0.760	Liu Yandong	0.1746	Liu Yandong	0.1741
6	Zhang Gaoli	0.750	Xu Qiliang	0.1738	Xu Qiliang	0.1738
7	Zhang Dejiang	0.738	Zhang Gaoli	0.1724	Zhang Gaoli	0.1719
8	Liu Yunshan	0.707	Han Zheng	0.1715	Hu Chunhua	0.1714
9	Yu Zhengsheng	0.675	Hu Chunhua	0.1715	Han Zheng	0.1714
10	Wang Huning	0.673	Sun Chunlan	0.1715	Sun Chunlan	0.1714

4. Conclusion

SNA is a technique that is gaining momentum globally. It enables users to gather vital information such as the connections that exist between individuals, the people or groups that affect public opinion, and the parties that influence important decisions. For this reason, SNA is often introduced across many media. The main purpose of this study is to highlight aspects of data definition, collection and data quality maintenance, in constructing databases for the evaluation of the interconnection paths of social networking among the political and socioeconomic power elites in China. It is difficult for users to form a clear image of the elites and their relationships by using fragmentary and distributed information data. This study addresses this challenge by establishing a database that collates people's profiles. The information pertaining to the power elites of China was collected, and based on the data, their network of relationships was identified and visualized, which enables the user to intuitively understand the relationships that exist between the relevant individuals.

We also focused our research on applying three SNA tools in combination with the collected data to compare centrality rankings. We found that all three tools are providing slightly different results for the top 10 nodes. This is because there are various libraries in tools. Therefore to get accurate centralities comparison study using different SNA tools, it is necessary to find the accurate result upon which the majority of tools reach a consensus.

We intend to expand access of this database to the general public in the future, with additional and improved applications based on an actor's sentimental or interest measurement approach and calculate each elite's sentiment (e.g. like, dislike or recommend) towards on the other elite. We will consider three factors such as elite sentiment similarity, interpersonal sentimental influence, and elite's reputation similarity into our social network analysis system to develop an accurate centrality prediction using sentimental evaluation. The results of this study are significant academically, economically and diplomatically, in understanding the relationship dynamics of China's power elites. We believe that further research in this area is necessary, and intend to apply state of the art technologies such as deep learning in advancing the findings of the present study.

References

- [1] S. Milgram, "The small world problem," *Psychol. Today*, vol. 2, no. 1, pp. 60–67, 1967. [Article \(CrossRef Link\)](#)
- [2] Wikipedia, "Six degrees of separation," *Wikipedia*. [Online] Available: https://en.wikipedia.org/wiki/Six_degrees_of_separation (accessed Nov. 18, 2020).
- [3] E. E. Lawrence and R. Latha, "Analysis of six degrees of separation in Facebook using Ant colony optimization," in *Proc. of 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*, pp. 1-5, 2015. [Article \(CrossRef Link\)](#)
- [4] E. Elmacioglu and D. Lee, "On six degrees of separation in DBLP-DB and more," *ACM SIGMOD Rec.*, vol. 34, no. 2, pp. 33–40, 2005. [Article \(CrossRef Link\)](#)
- [5] P. Reynolds, J. Wiener, J. Mogul, M. Shah, and A. Vahdat, "The oracle of bacon," *Retrieved Jan.*, vol. 10, p. 2000, 1999.
- [6] W. E. Hautz, G. Krummrey, A. Exadaktylos, and S. C. Hautz, "Six degrees of separation: the small world of medical education," *Med. Educ.*, vol. 50, no. 12, pp. 1274–1279, 2016. [Article \(CrossRef Link\)](#)
- [7] J. Hua, M. L. Huang, and G. Wang, "Graph layout performance comparisons of force-directed algorithms," *Int. J. Performability Eng.*, vol. 14, no. 1, pp. 67-76, 2018. [Article \(CrossRef Link\)](#)

- [8] Thomson Reuters, “Connected China - Fathom Information Design,” 2013. [Online] Available: <http://china.fathom.info/> (accessed Nov. 18, 2020).
- [9] H. J. Lee and K. Park, “Comparison of inductive and deductive data collection methods for the analysis of Chinese power elite social network,” *J. Chinese Stud.*, vol. 88, p. 181-214, 2019. [Article \(CrossRef Link\)](#)
- [10] H. Kwak and J. An, “Two tales of the world: Comparison of widely used world news datasets GDELT and EventRegistry,” in *Proc. of the International AAAI Conference on Web and Social Media*, vol. 10, no. 1, 2016.
- [11] O. Sevastaki, “Determinants of news coverage and tone of environmental events,” Tilburg University, 2018.
- [12] J. Joo, “China’s Fifth Generation Political Elite-Analysis of Actors and Structural Characteristics,” *Natl. Strateg.*, vol. 11, no. 3, pp. 149–174, 2011.
- [13] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, “Network analysis in the social sciences,” *Science (80-.)*, vol. 323, no. 5916, pp. 892–895, 2009. [Article \(CrossRef Link\)](#)
- [14] F. B. Keller, “Networks of Power: Using Social Network Analysis to understand who will rule and who is really in charge in the Chinese Communist Party,” *Lap. kajian, Draft Juli*, 2015.
- [15] F. B. Keller, “Moving beyond factions: using social network analysis to uncover patronage networks among Chinese elites,” *J. East Asian Stud.*, vol. 16, no. 1, pp. 17–41, 2016. [Article \(CrossRef Link\)](#)
- [14] D. He, “Traditional Interpersonal Ethics Variation and Power Rent-seeking and Its Correction,” *Chuanshan J.*, no. 2, pp. 214–216, 2009.
- [15] “Korea Data Agency,” “Data Quality Diagnosis Procedures and Techniques,” 2009.
- [16] D. He, “Traditional Interpersonal Ethics Variation and Power Rent-seeking and Its Correction,” *Chuanshan J.*, no. 2, pp. 214–216, 2009.
- [17] Korea Data Agency, “Data Quality Diagnosis Procedures and Techniques,” 2009.
- [18] J. S and Adler, “Introduction to Network Analysis with R,”. [Online] Available: <https://www.jessesadler.com/post/network-analysis-with-r/>.
- [19] M. Zhang, “Social network analysis: History, concepts, and research,” in *Handbook of social network technologies and applications*, Springer, 2010, pp. 3–21. [Article \(CrossRef Link\)](#)



HeeJeong Lee has a BSc. (Computer Science) from POSTECH, Korea, a MSc. (Computer Science) from the University of Edinburgh, UK, a MPhil. (Technology Policy) from Cambridge University, UK and a PhD. (Information Technology) from Monash University, Australia. She won a gold award at the 23rd International Invention, Innovation and Technology Exhibition. Her research interests include Social Network Analysis and methods using Artificial Intelligence in education. She was with Korea Telecom R&D for 10 years; and has filed 11 patents on telecommunications applications. Her move from industry to academia was motivated by a love of learning. Currently, she is an assistant professor at Pierson college, PyeongTaek University.



In Kim has a B.A. degree in Chinese, a MA in Asian Area Studies and a PhD in International Relations from HanKuk University of Foreign Studies, Korea. He was a visiting researcher of Beijing University, China. He is an author of many books including Chinese tradition and modern China. He researches, lectures and consults in Modern Chinese politics, Chinese foreign policy and one belt one road. Currently he is a director of institute of Korea & Silkroad Studies and a research professor at Pierson College, Pyeongtaek University, Korea.