Corporate Leaders Analytics and Network System (CLANS): Constructing and Mining Social Networks among Corporations and Business Elites in China

Yuanyuan Man[1], Shuai Wang[1], Tianyu Zhang[2], T.J. Wong[2] and Irwin King[1]

[1]Department of Computer Science and Engineering

[2]School of Accountancy

The Chinese University of Hong Kong

ICONIP, 2014

Outline

1. Introduction
2. CLANS System
3. Website Illustration

Background

- Social networks are essential for business.

[1] Brian Uzzi. "Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing". In: *American sociological review* (1999), pp. 481–505.

[2] David F Larcker, Eric C So, and Charles CY Wang. *Boardroom centrality and stock returns*. Citeseer, 2010.

[3] David Larcker et al. "Back door links between directors and executive compensation". In: *Back Door Links Between Directors and Executive Compensation (February 2005)* (2005).

Image source: http://blog.beacontechnologies.com/wp-content/uploads/2013/08/social_media_for_companies.jpg

Background

- Social networks are essential for business.

- In US, social networks among firms benefit the debt financing, firm performance, and corporate governance[1][2][3].

[1] Brian Uzzi. "Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing". In: *American sociological review* (1999), pp. 481–505.

[2] David F Larcker, Eric C So, and Charles CY Wang. *Boardroom centrality and stock returns*. Citeseer, 2010.

[3] David Larcker et al. "Back door links between directors and executive compensation". In: *Back Door Links Between Directors and Executive Compensation (February 2005)* (2005).

Image source: http://blog.beacontechnologies.com/wp-content/uploads/2013/08/social_media_for_companies.jpg

Background

- In China, people like to find acquaintances to deal with problems. They think it will save them much trouble.

Image source: http://app.qpic.cn/mblogpic/7dbf15a94dd4fcb4c924/160,http://www.hbqnb.com/news/Html/chayufanhou/meeting/meeting/2008/117/08117119251795557.html

Background

- In China, people like to find acquaintances to deal with problems. They think it will save them much trouble.

- Especially, relationship plays a crucial role in Chinese business model.

Image source: `http://app.qpic.cn/mblogpic/7dbf15a94dd4fcb4c924/160`,`http://www.hbqnb.com/news/Html/chayufanhou/meeting/meeting/2008/117/08117119251795557.html`

Background

- In China, people like to find acquaintances to deal with problems. They think it will save them much trouble.

- Especially, relationship plays a crucial role in Chinese business model.
- To the best of our knowledge, few studies focus on the social network of corporations and elites in China

Image source: http://app.qpic.cn/mblogpic/7dbf15a94dd4fcb4c924/160,http://www.hbqnb.com/news/Html/chayufanhou/meeting/meeting/2008/117/08117119251795557.html

CLANS's Stakeholders

- Investors

CLANS's Stakeholders

- Investors
  - Help them take into account and highly value the social connecting issues among Chinese firms for their investment decision.

CLANS's Stakeholders

- Investors
  - Help them take into account and highly value the social connecting issues among Chinese firms for their investment decision.
- Common businessman

CLANS's Stakeholders

- Investors
  - Help them take into account and highly value the social connecting issues among Chinese firms for their investment decision.
- Common businessman
  - Help them learn more about specific information for Chinese companies, senior executives and their social networks, for better or potential commercial activities.

CLANS's Stakeholders

- Investors
  - Help them take into account and highly value the social connecting issues among Chinese firms for their investment decision.
- Common businessman
  - Help them learn more about specific information for Chinese companies, senior executives and their social networks, for better or potential commercial activities.
- Researchers

CLANS's Stakeholders

- Investors
  - Help them take into account and highly value the social connecting issues among Chinese firms for their investment decision.
- Common businessman
  - Help them learn more about specific information for Chinese companies, senior executives and their social networks, for better or potential commercial activities.
- Researchers
  - Help them conduct deep research

Objective

- Identify and analyze social networks among corporations and business elites

Objective

- Identify and analyze social networks among corporations and business elites
- Main contributions
    - Collect data from multiple sources and do the preprocessing to make it available to use

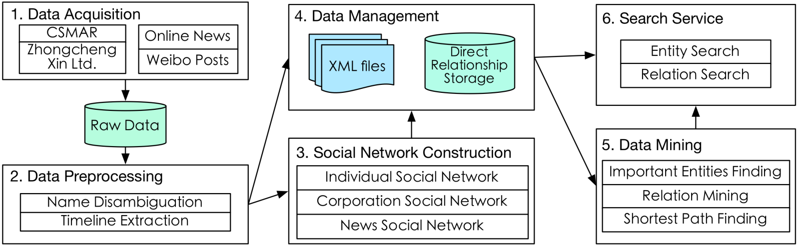
Objective

- Identify and analyze social networks among corporations and business elites
- Main contributions
  - Collect data from multiple sources and do the preprocessing to make it available to use
  - Construct a business social network and formulate similarity relations among individuals and corporations

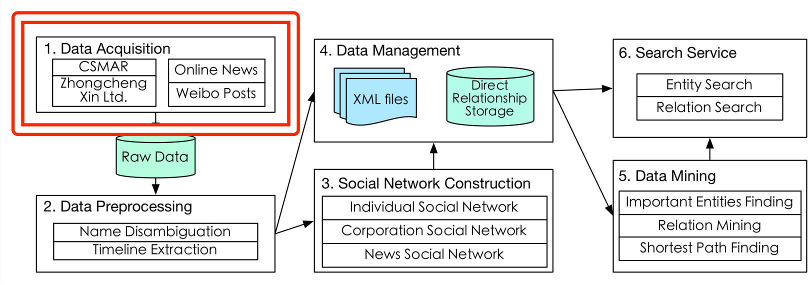
Objective

- Identify and analyze social networks among corporations and business elites
- Main contributions
  - Collect data from multiple sources and do the preprocessing to make it available to use
  - Construct a business social network and formulate similarity relations among individuals and corporations
  - Conduct data mining to discover more implicit information

System Overview



System Overview



Data

- Corporate leaders' biography

Data

- Corporate leaders' biography
  - China Securities Market and Accounting Research Database (CSMAR), contains listed firms' annual reports from year 1999 to 2012

Data

- Corporate leaders' biography
  - China Securities Market and Accounting Research Database (CSMAR), contains listed firms' annual reports from year 1999 to 2012
  - A personnel database compiled by Zhongcheng Xin Ltd.

Data

- Corporate leaders' biography
  - China Securities Market and Accounting Research Database (CSMAR), contains listed firms' annual reports from year 1999 to 2012
  - A personnel database compiled by Zhongcheng Xin Ltd.

Data

- Corporate leaders' biography
  - China Securities Market and Accounting Research Database (CSMAR), contains listed firms' annual reports from year 1999 to 2012
  - A personnel database compiled by Zhongcheng Xin Ltd.

Sources	Records	Biography format
CSMAR	399,216 (2,500 companies covered)	text
Zhongcheng Xin Ltd.	84,859 (68,289 people covered)	structured

Data

- Corporate leaders' biography
  - China Securities Market and Accounting Research Database (CSMAR), contains listed firms' annual reports from year 1999 to 2012
  - A personnel database compiled by Zhongcheng Xin Ltd.

Sources	Records	Biography format
CSMAR	399,216 (2,500 companies covered)	text
Zhongcheng Xin Ltd.	84,859 (68,289 people covered)	structured

- Senior managers and listed firms related online news and Sina Weibo posts.

Data

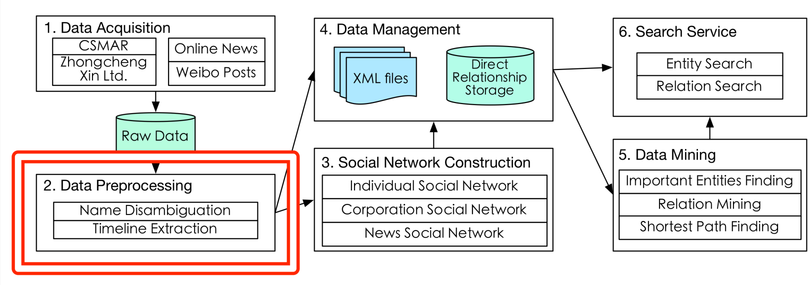
- Corporate leaders' biography
  - China Securities Market and Accounting Research Database (CSMAR), contains listed firms' annual reports from year 1999 to 2012
  - A personnel database compiled by Zhongcheng Xin Ltd.

Sources	Records	Biography format
CSMAR	399,216 (2,500 companies covered)	text
Zhongcheng Xin Ltd.	84,859 (68,289 people covered)	structured

- Senior managers and listed firms related online news and Sina Weibo posts.

Sources	People	Company
Online news	16,374,279	1,126,299
Sina Weibo	19,445,929	2,367,619

System Overview



- Problems

Data Preprocessing

Name Disambiguation - Problem One

- Problems
  - Extract individual entities from CSMAR

- Problems
  - Extract individual entities from CSMAR
  - Map people in Zhongcheng Xin Ltd. to CSMAR

Data Preprocessing

Name Disambiguation - Problem One

- Problems
  - Extract individual entities from CSMAR
  - Map people in Zhongcheng Xin Ltd. to CSMAR
- Solutions
  - If two records share a high similarity of cognizable features (like name, age, gender, company and birthplace) over a defined threshold, we consider them as the same person.

- Problems
  - Extract individual entities from CSMAR
  - Map people in Zhongcheng Xin Ltd. to CSMAR
- Solutions
  - If two records share a high similarity of cognizable features (like name, age, gender, company and birthplace) over a defined threshold, we consider them as the same person.
- Result
  - Identify 87,458 individual entities in CSMAR and find the common 46,130 people in two datasets

Data Preprocessing

Name Disambiguation - Problem Two

- Problem
  - a popular name in our database matches multiple people from Internet sources

- Problem
  - a popular name in our database matches multiple people from Internet sources
- Solutions
  - Two-round labeling method for Weibo posts

Data Preprocessing

Name Disambiguation - Problem Two

- Problem
  - a popular name in our database matches multiple people from Internet sources
- Solutions
  - Two-round labeling method for Weibo posts
    - First round, use related keywords(like company name, positions and college) related to certain people to do the filtering. Use the result as training set to do the next round.

Data Preprocessing

Name Disambiguation - Problem Two

- Problem
  - a popular name in our database matches multiple people from Internet sources
- Solutions
  - Two-round labeling method for Weibo posts
    - First round, use related keywords(like company name, positions and college) related to certain people to do the filtering. Use the result as training set to do the next round.
    - Second round, transform the text into vector expression based on tf-idf, and calculate the cosine similarity, set a threshold and get the final result.

Data Preprocessing

Name Disambiguation - Problem Two

- Problem
  - a popular name in our database matches multiple people from Internet sources
- Solutions
  - Two-round labeling method for Weibo posts
    - First round, use related keywords(like company name, positions and college) related to certain people to do the filtering. Use the result as training set to do the next round.
    - Second round, transform the text into vector expression based on tf-idf, and calculate the cosine similarity, set a threshold and get the final result.
  - Three-round labeling method for online news

Data Preprocessing

- Problem
  - a popular name in our database matches multiple people from Internet sources
- Solutions
  - Two-round labeling method for Weibo posts
    - First round, use related keywords(like company name, positions and college) related to certain people to do the filtering. Use the result as training set to do the next round.
    - Second round, transform the text into vector expression based on tf-idf, and calculate the cosine similarity, set a threshold and get the final result.
  - Three-round labeling method for online news
    - The first two rounds are the same.
    - Third round, apply the Latent Semantic Allocation to the labeled result from the second round, map all the documents to vectors in the lower dimensional latent semantic space, calculate document similarities, set the threshold and label the rest.

Data Preprocessing

Name Disambiguation - Problem Two

- Problem
  - a popular name in our database matches multiple people from Internet sources
- Solutions
  - Two-round labeling method for Weibo posts
    - First round, use related keywords(like company name, positions and college) related to certain people to do the filtering. Use the result as training set to do the next round.
    - Second round, transform the text into vector expression based on tf-idf, and calculate the cosine similarity, set a threshold and get the final result.
  - Three-round labeling method for online news
    - The first two rounds are the same.
    - Third round, apply the Latent Semantic Allocation to the labeled result from the second round, map all the documents to vectors in the lower dimensional latent semantic space, calculate document similarities, set the threshold and label the rest.
- Result
  - Select random sample of 1000 posts and news, Weibo posts is 98% and online news is 86% by precision rate.

- Problem
  - for people in CSMAR, who are not covered in Zhongcheng Xin Ltd. data, we analyze personal unstructured profiles and extract structured timeline information

Data Preprocessing

- Problem
  - for people in CSMAR, who are not covered in Zhongcheng Xin Ltd. data, we analyze personal unstructured profiles and extract structured timeline information
- Solutions
  - For education timeline, due to its simple expression, we employ rule-learning algorithm.

Data Preprocessing

Timeline Extraction

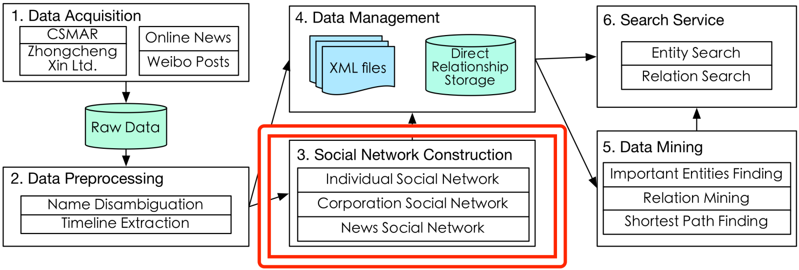
- Problem
  - for people in CSMAR, who are not covered in Zhongcheng Xin Ltd. data, we analyze personal unstructured profiles and extract structured timeline information
- Solutions
  - For education timeline, due to its simple expression, we employ rule-learning algorithm.
  - For work timeline, due to expression's diversity and complexity, we employ HMM model.

Data Preprocessing

Timeline Extraction

- Problem
  - for people in CSMAR, who are not covered in Zhongcheng Xin Ltd. data, we analyze personal unstructured profiles and extract structured timeline information
- Solutions
  - For education timeline, due to its simple expression, we employ rule-learning algorithm.
  - For work timeline, due to expression's diversity and complexity, we employ HMM model.
- Result
  - 95.1% precision rate for education timeline, 83.1% for working timeline.

System Overview



- We define alumni relationship as the closeness of the relationship between two alumni

- We define alumni relationship as the closeness of the relationship between two alumni
  - Four criteria: major, degree, time of enrollment, intersection school time

- We define alumni relationship as the closeness of the relationship between two alumni
  - Four criteria: major, degree, time of enrollment, intersection school time
  - Deduce 13 types of relationships

- We define alumni relationship as the closeness of the relationship between two alumni
  - Four criteria: major, degree, time of enrollment, intersection school time
  - Deduce 13 types of relationships
    - The closest relationship means that two people are classmates (same major, same degree and same time of enrollment), weight is 0.9

- We define alumni relationship as the closeness of the relationship between two alumni
  - Four criteria: major, degree, time of enrollment, intersection school time
  - Deduce 13 types of relationships
    - The closest relationship means that two people are classmates (same major, same degree and same time of enrollment), weight is 0.9
    - The weight of farthest relationship is 0.1 (with different major, different degree and no intersection school time)

Social Network Construction
Colleague Social Network

- Let position rank (PS) denoted as a representation of job level by integer ranging from 0 to 9.

- Let position rank (PS) denoted as a representation of job level by integer ranging from 0 to 9.
  - The higher position rank has a larger value

- Let position rank (PS) denoted as a representation of job level by integer ranging from 0 to 9.
  - The higher position rank has a larger value
  - The PS of the board chairman is 9

- Let position rank (PS) denoted as a representation of job level by integer ranging from 0 to 9.
  - The higher position rank has a larger value
  - The PS of the board chairman is 9
  - The PS of the CEO is 8

- Let position rank (PS) denoted as a representation of job level by integer ranging from 0 to 9.
    - The higher position rank has a larger value
    - The PS of the board chairman is 9
    - The PS of the CEO is 8
    - The PS of the independent director is 1

- Let position rank (PS) denoted as a representation of job level by integer ranging from 0 to 9.
    - The higher position rank has a larger value
    - The PS of the board chairman is 9
    - The PS of the CEO is 8
    - The PS of the independent director is 1
- Let value relation between two colleagues denoted as the average position rank of the two people.

- Let position rank (PS) denoted as a representation of job level by integer ranging from 0 to 9.
    - The higher position rank has a larger value
    - The PS of the board chairman is 9
    - The PS of the CEO is 8
    - The PS of the independent director is 1
- Let value relation between two colleagues denoted as the average position rank of the two people.
- Let close relation between two colleagues denoted as the intersection years that they work together.

- Let position rank (PS) denoted as a representation of job level by integer ranging from 0 to 9.
  - The higher position rank has a larger value
  - The PS of the board chairman is 9
  - The PS of the CEO is 8
  - The PS of the independent director is 1
- Let value relation between two colleagues denoted as the average position rank of the two people.
- Let close relation between two colleagues denoted as the intersection years that they work together.
- Let colleague relationship denoted as a combination of value relation and close relation.

- The colleague weight between person p_i and p_j is defined as

$$\omega_{p_i,p_j} = \sum_{t \in L(p_i,p_j)} \frac{PS_{t,p_i} + PS_{t,p_j}}{2} , \tag{1}$$

where $L(p_i, p_j)$ denotes a collection of the intersection years that person p_i and p_j used to work with each other, and PS_{t,p_i} denotes the position rank of person p_i in the year t. At the end, all the weights are normalized, which is also applied in the following weight calculation.

- We define the individual social network as an undirected graph $G(V, E)$.

- We define the individual social network as an undirected graph $G(V, E)$.
- In $G(V, E)$, every edge (relationship) has weighted value, which is defined as

$$W_{i,j} = \alpha\omega_{i,j}^{al} + \beta\omega_{i,j}^{co} \ , \tag{2}$$

- We define the individual social network as an undirected graph $G(V, E)$.
- In $G(V, E)$, every edge (relationship) has weighted value, which is defined as

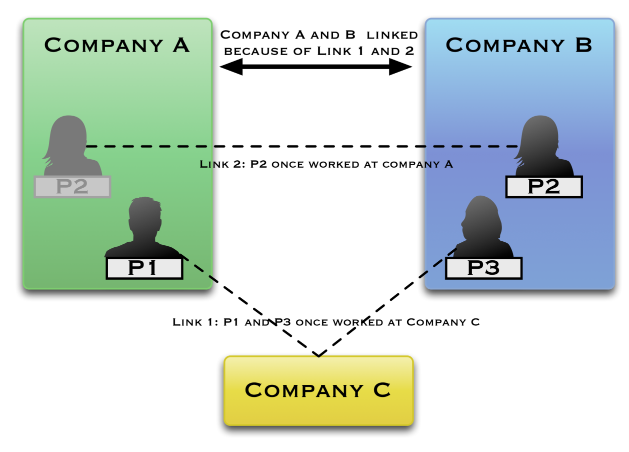
$$W_{i,j} = \alpha\omega^{al}_{i,j} + \beta\omega^{co}_{i,j} \ , \tag{2}$$

where $\omega^{al}_{i,j}$ is a weight for alumni relationship, $\omega^{co}_{i,j}$ for colleague relationship; α and β denotes the corresponding percentage.

Definition

We define the corporation social network as an directed graph $\hat{G}(\hat{V},\hat{E})$. In $\hat{G}(\hat{V},\hat{E})$, every vertex (corporation) has feature set $P_i = \{p_i^1, p_i^2, \cdots, p_i^n\}$ and every direct edge (relationship) has weighted value $W_{i,j} = (\omega_{i,j}^{gp}, \omega_{i,j}^{nk})$. n is the size of the set (total number of staffs); $\omega_{i,j}^{gp}$ is a weight for group membership, $\omega_{i,j}^{nk}$ for network relationship.

Example



Company A
Company A and B linked because of Link 1 and 2
Company B
P2
Link 2: P2 once worked at company A
P2
P1
P3
Link 1: P1 and P3 once worked at Company C
Company C

- $\omega^{gp}_{i,j}$, $\omega^{nk}_{i,j}$ are defined as follows:

$$\omega^{gp}_{i,j} = \sum_{p^k_i \in P_i \cap P_j} PS_{p^k_i} * \omega^{gp}_{p^k_i} , \tag{3}$$

$$\omega^{nk}_{i,j} = \sum_{(p^k_i, p^r_j) \in L_2(P_i, P_j)} PS_{p^k_i} * \omega^{nk}_{p^k_i, p^r_j} . \tag{4}$$

  $PS_{p^k_i}$ denotes the position rank of person p^k_i in corporation i; $\omega^{gp}_{p^k_i}$ is a weight for p^k_i connecting P_i with P_j; $L_2(P_i, P_j)$ denotes a collection of connections between $(P_i - P_i \cap P_j)$ and $(P_j - P_i \cap P_j)$; $\omega^{nk}_{p^k_i, p^r_j}$ denotes a weight between p^k_i and p^r_j calculated in the previous equation.
- Corporation weight from corporation i to j is defined as $W_{i,j} = \alpha\omega^{gp}_{i,j} + \beta\omega^{nk}_{i,j}$

- Help investors identify how people and firms are connected with each other in news

- Help investors identify how people and firms are connected with each other in news
- We construct the news social network for individual and corporation based on Weibo posts and online news.

Social Network Construction

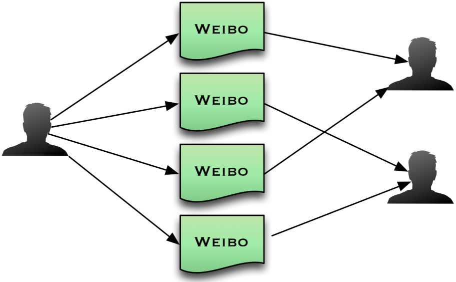
News Social Network

- Help investors identify how people and firms are connected with each other in news
- We construct the news social network for individual and corporation based on Weibo posts and online news.
  - If two individuals or corporations are mentioned in the same posts or news, we consider they have a link.

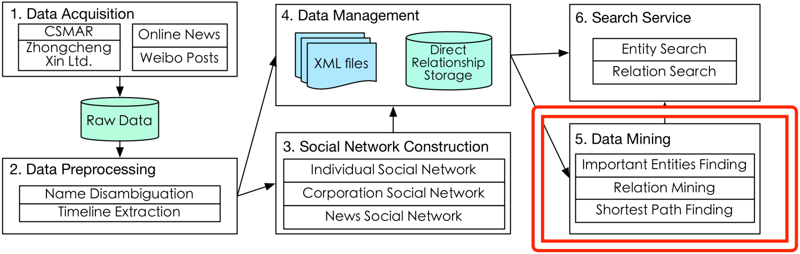
Social Network Construction

News Social Network

- Help investors identify how people and firms are connected with each other in news
- We construct the news social network for individual and corporation based on Weibo posts and online news.
  - If two individuals or corporations are mentioned in the same posts or news, we consider they have a link.



System Overview



- Aim to find important individuals and corporations

- Aim to find important individuals and corporations
- For individual social network
  - A method takes into consideration of both personal and network information

- Aim to find important individuals and corporations
- For individual social network
  - A method takes into consideration of both personal and network information
  - The basic idea is that an important person knows someone a, then a is also important.

- Aim to find important individuals and corporations
- For individual social network
  - A method takes into consideration of both personal and network information
  - The basic idea is that an important person knows someone a, then a is also important.
  - Steps: firstly, we assign every individual with initial score according to position rank; secondly, we distribute the score according to the weight of the out-link edge; third, the algorithm will stop if the change is less than a threshold.

- Aim to find important individuals and corporations
- For individual social network
  - A method takes into consideration of both personal and network information
  - The basic idea is that an important person knows someone a, then a is also important.
  - Steps: firstly, we assign every individual with initial score according to position rank; secondly, we distribute the score according to the weight of the out-link edge; third, the algorithm will stop if the change is less than a threshold.
- For corporation social network
  - Use PageRank algorithm

Data Mining

- Relation Mining
  - Aim to find out important people's link between two corporations' link

Data Mining

- Relation Mining
  - Aim to find out important people's link between two corporations' link
- Shortest path finding
  - People-to-people
    - Direct connection: schoolmate, family, friend or colleague
    - Indirect connection between them through closest connected intermediate nodes

Data Mining

- Relation Mining
  - Aim to find out important people's link between two corporations' link
- Shortest path finding
  - People-to-people
    - Direct connection: schoolmate, family, friend or colleague
    - Indirect connection between them through closest connected intermediate nodes
  - People-to-company
    - Direct connection: employment relationship
    - Indirect connection: find out the possible link to the people who worked in the company

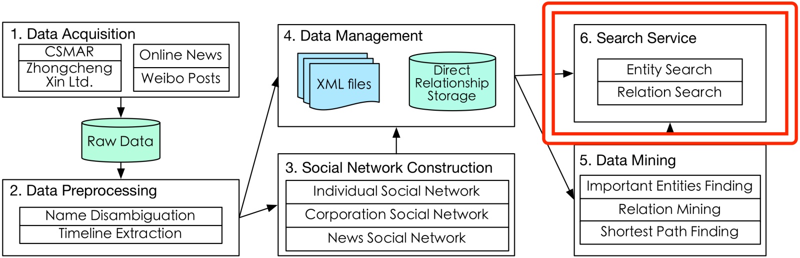
Data Mining

- Relation Mining
  - Aim to find out important people's link between two corporations' link
- Shortest path finding
  - People-to-people
    - Direct connection: schoolmate, family, friend or colleague
    - Indirect connection between them through closest connected intermediate nodes
  - People-to-company
    - Direct connection: employment relationship
    - Indirect connection: find out the possible link to the people who worked in the company
  - Company-to-company
    - Direct connection: cooperative relationship
    - Indirect connection

Data Mining

- Relation Mining
  - Aim to find out important people's link between two corporations' link
- Shortest path finding
  - People-to-people
    - Direct connection: schoolmate, family, friend or colleague
    - Indirect connection between them through closest connected intermediate nodes
  - People-to-company
    - Direct connection: employment relationship
    - Indirect connection: find out the possible link to the people who worked in the company
  - Company-to-company
    - Direct connection: cooperative relationship
    - Indirect connection

System Overview



Search Service

- Entity Search
  - Given any keyword, system returns a list of ranked people and companies.

Search Service

- Entity Search
  - Given any keyword, system returns a list of ranked people and companies.
- Relation Search
  - Given any two keywords, the system returns shortest path between them and the corresponding intermediate nodes and link information.

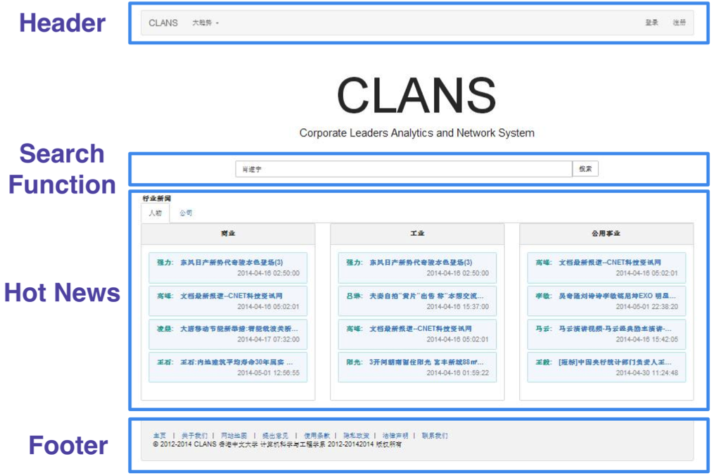
1 Introduction

2 CLANS System

3 Website Illustration

Website Illustration

Homepage



Website Illustration

Person Page

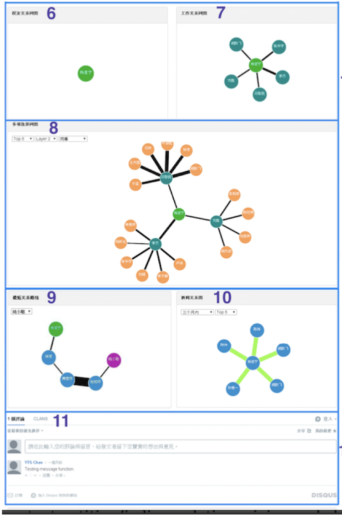
1. Basic Information

2. Timeline

3. Recent News

4. Positive News

5. Negative News



6. Alumni Social Network
7. Colleague Social Network

8. Integrated Individual Social Network

9. Shortest Path with Another Person
10. News Social Network

11. Comments

Website Illustration

Company Page

1. Basic Information

2. Timeline

3. Recent News

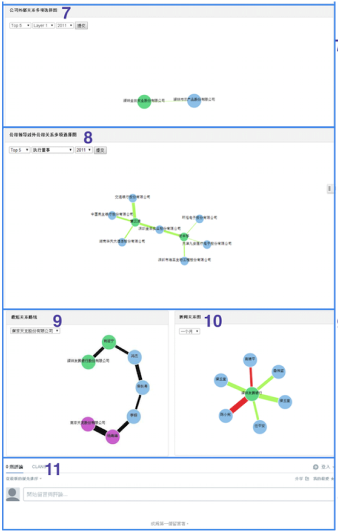
4. Positive News

5. Negative News

6. Stock Value Trend

Website Illustration

Company Page



7. Corporation Social Network

8. Corporate Leaders Social Network

9. Shortest Path with Another Firm

10. News Social Network

11. Comments

Q&A