

Field Prelim Paper, WINTER 2010

Measuring the Meaning of Political Concepts in Chinese Online Deliberation

Daniel Xiaodan Zhou
February 26, 2010

Committee Members
(in alphabetical order)

Mary Gallagher
Steve Jackson
Qiaozhu Mei
Paul Resnick (chair)

Acknowledgements

I would like to thank my advisor Professor Paul Resnick for his deeply insightful advice throughout this prelim study. My cordial thanks go to Professor Mary Gallagher for her comments on China studies, to Professor Steve Jackson for teaching me the deliberation theories in SI703 and SI732, to Professor Qiaozhu Mei for his help on text mining algorithms such as the topic model and random walk model. I'd like to thank my collaborator Yawen Lei for her contribution in Section 6.3. Finally I'd like to thank my fiancée Allison Riccardi for her help on grammar check and for her constant support.

Table of Contents

Acknowledgements	ii
1. Introduction.....	1
The “field”	3
2. China's Cultural Difference and its Implications.....	4
“Classical” books	4
The conception of “democracy”	8
The conception of “rights”	9
The conception of “freedom”	11
The role of government.....	12
Potential historical explanations.....	12
Online censorship and its effects.....	13
Summary and challenges.....	16
3. Online Deliberation.....	18
Deliberative democracy and public sphere theories.....	18
Critiques.....	20
Technology and online deliberation	21
Deliberation in China	23
Summary and challenges.....	25
4. Deliberation Measurement.....	27
Frameworks on measuring the quality of deliberation.....	27
Measuring specific aspects of deliberation	29
Summary and challenges.....	31
5. Measuring Meaning of Political Concepts in Online Deliberation	33
5.1. Motivation	33
Responding to theoretical challenges.....	33
Why not traditional methods.....	34
5.2. Theoretical Background.....	35
Constructing semantic meaning from text	36
Co-occurrence assumption and simplification	38
Comparison to WordNet.....	39

5.3. Computational Approaches	39
Word association	39
Concepts network.....	40
Semantic space	41
Summary and challenges.....	44
6. Future Research.....	46
6.1. Measuring the Meaning of "Democracy" in the Chinese Context.....	46
Concepts Network with Random Walk.....	47
Evaluation Framework.....	48
Empirical Study	51
Results and Discussions.....	52
6.2. Chinese Concepts of "Rights": Empirical Evidence from Political Deliberation.....	55
Research Questions	55
Research Design.....	55
Results and Discussions.....	57
6.3. Legal Culture in the State Media and on the Internet: the Salience of Law and the Construction of Legality in Two Venues of Political Communication	59
Introduction.....	59
Research question	60
Theoretical framework	61
Methodology: network analysis and qualitative content analysis	62
Preliminary results	63
7. References.....	65

1. Introduction

As China's role becomes increasingly influential in global affairs, research on China, especially on China's political reform and democratization, has attracted much attention from both practice and academia. The Internet, on the other hand, has brought a new means for Chinese political communication, and has had a profound impact in shaping people's thought by changing the way people obtain information. Thus, the study of Chinese online political deliberation is one of the key research areas to better understand China and its political challenges.

This prelim paper proposes to study the meaning of political concepts such as "democracy", "rights", "rule of law", and so on, in the context of Chinese online deliberation. The first theme of this paper is to review literature in the sociology field, including China studies and deliberation theories. I would argue that the literature demands a better understanding of political concepts in the Chinese context.

The other theme of this paper is to review literature in the technology field including computational linguistics, machine learning and text mining, regarding how to measure meanings of concepts from deliberation. I would argue that recent advances in those fields might provide potential methods to complement traditional social science methodology such as surveys, interviews and content analysis.

This paper, therefore, is a typical "SI" paper – it is guided by social theories, and applies technology to answer questions in the social sciences. Merely focusing on either social theories or technology would result in a prelim paper either in the sociology department or in the computer science department.

Basically, this paper is based on the recognition of two gaps in existing literature, as shown in Figure 1.

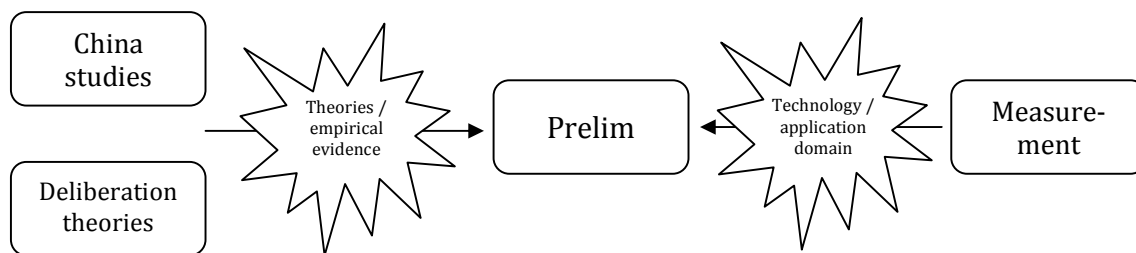


Figure 1. Gaps in literature review

The first gap is between social theories and empirical evidence. Literature in China studies and deliberation theories mostly focuses on theoretical discussion, and finds support through traditional methods such as personal observations, surveys, interviews or content analysis, and ignores the rich empirical evidence from real world Chinese online deliberation on the web. Applying computational technology could fully exploit the large volume of existing online texts and provide rich empirical support to theoretical works. The challenge here is to find and tune existing computational approaches that are both effective and intuitive to answer theoretical questions – effective in order to provide

something that traditional methods are not able to provide, intuitive in order to be more acceptable to the audience in social sciences. A comparison of computational approaches and traditional methodology will be discussed later in the paper.

The second gap is between technology and the application domain. Technology has advanced significantly in recent years. However, much technology tries to solve problems in its own domain, and limits its application to business and everyday usage. Not so much literature is about applying advanced technology to social sciences in order to answer theoretical questions. The challenge here is to find interesting research questions in the social sciences as the soil for advanced technical solutions.

This paper, then, is organized as shown in Figure 2. Section 2 and Section 3 are of the “sociology” theme, reviewing literature in China studies and online deliberation respectively. Section 4 and Section 5 are of the “technology” theme, first reviewing the

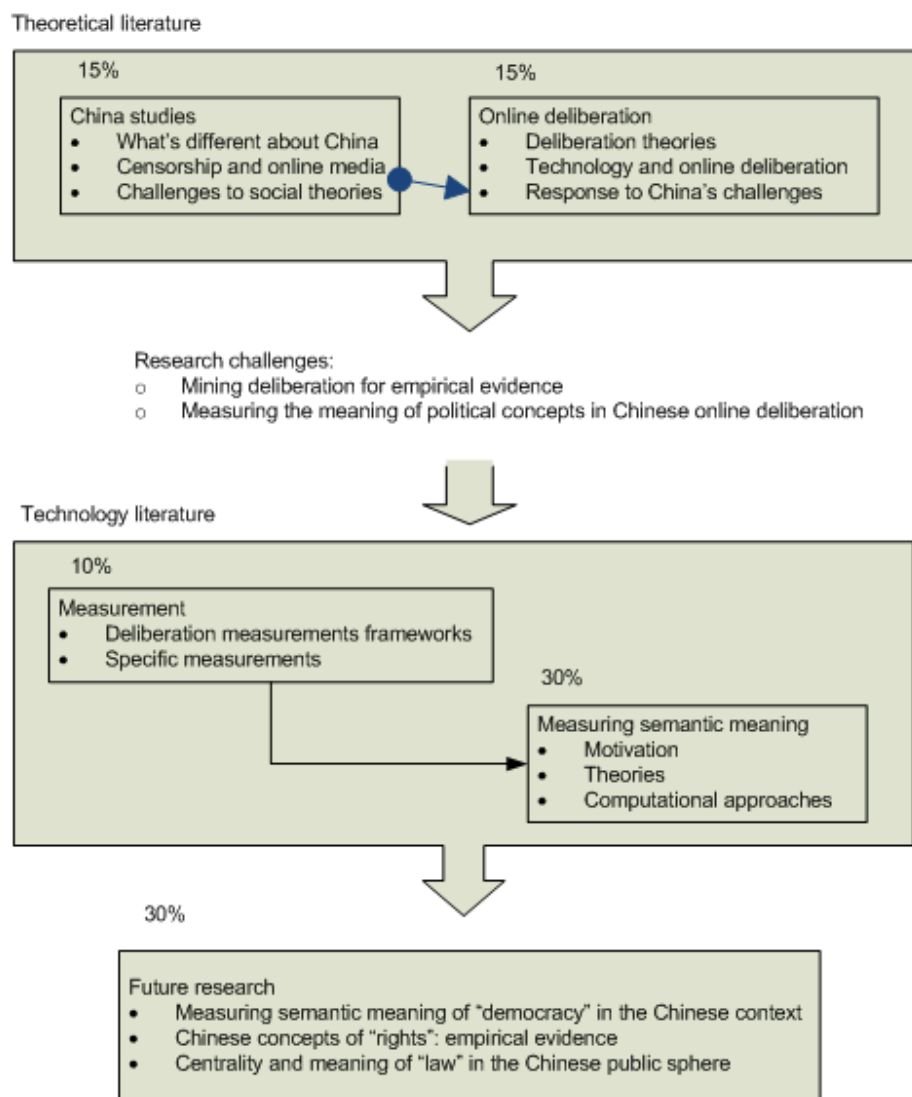


Figure 2. Outline of the paper

broad deliberation measurement and then the measurement of semantic meaning. Finally, Section 6 proposes three future research projects.

The “field”

This interdisciplinary field prelim is an intersection of China studies, online deliberation, and text analysis, as shown in Figure 3. In China studies, I focus on literature about the media and the internet with the audience of *Journal of Contemporary China*, *China Quarterly*, *Chinese Internet Research Conference*, etc. In online deliberation, I focus on literature about measuring deliberation with the audience of *Online Deliberation Conference*, *Journal of Information Technology and Politics*, *Political Communications*, etc. In text analysis, I focused on semantic network and online forums studies with the audience of *ICWSM Conference*, *WWW Conference*, etc.

To my knowledge, no existing field includes all three areas, so technically this “field prelim” has no “field” on its own. But the prelim can also affiliate with any of the three fields, and could contribute new interdisciplinary perspectives. My future research publications generated from this prelim would target an audience in any of the three fields.

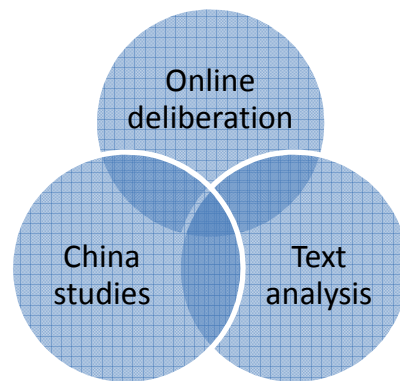


Figure 3. Interdisciplinary field

2. China's Cultural Difference and its Implications

China studies are obviously different from those without the label "China". Many scholars have reviewed the distinctive features of China studies in various fields. For example, (Walder, 2002) reviewed China studies in sociology, economics, and politics, and pointed out that theories developed in the Western context might not necessarily apply to China. (Baum, 2007) reviewed the topics and trends of political studies about China, which were quite different from political studies in the Western¹ tradition. However, those works didn't further investigate China's fundamental cultural difference.

My approach in this section is to get beyond "the international facade of verbal similarity (Nathan, 1986)", and explain the different meaning of four political concepts in the Chinese context: "democracy", "rights", "freedom" and the role of government. These concepts are widely used in Western literature and have quite a consistent meaning, but have different and inconsistent interpretation when applied in China, as I'll show later. Through examining the conceptions of those terms, I hope to give the audience a brief idea of China's cultural difference, and justify the importance of measuring the meaning of those political concepts in China's online deliberation.

In the rest of the section, I'll first review eleven books about Chinese political conceptions, each of which covers a wide range of topics and ideas. Then I'll synthesize the books with supplementary readings into four topics – the conceptions of democracy, rights, freedom and the role of government. Next, I'll offer some explanations on why the conceptions are different in China, and particularly review Chinese censorship and suggest that censorship might affect the conceptions as well. Finally, I'll summarize the challenges in the literature and argue that it is important to find empirical evidence from Chinese online deliberation to understand people's conceptions on important political ideas.

"Classical" books

Here I'll review eleven "classical" books related to the Chinese conceptions on certain political ideas. They are considered as "classical" because they are widely referenced in the field, as either recommended by Prof. Gallagher or received high number of citations in Google Scholar. I'll discuss them in the order of publishing year.

Ideology and organization in Communist China (Schurmann, 1968) argued that ideology and organization are the two pillars that hold society together; and thus the book studied communist China (before 1968) from two angles: 1) the Maoism ideology, and 2) the party-state organization. The book explained that the ideology in communist China was a combination of Marxism-Leninism as the "pure ideology", and the Thoughts of Mao as the "practical ideology" in the practice of revolution and construction in China. The book explained that the values of Chinese communist ideology diverged sharply from traditional Chinese values: the new central value was the notion of class struggle rather than the old notion of harmony, and the new ideal was the proletarianization of the Chinese people. The

¹ Prof. Gallagher pointed out that sometimes Europe has different perspectives compared to the US, and thus it was questionable to generalize "the West" as a consistent concept. Here I'll broadly use the term "Western" and American interchangeably.

communist ideology also stressed more the value of collectivity in the form of communes. “Democracy” in the communist ideology is the dialectic form of “democratic centralism”, which means the unity of democracy as mass mobilization and centralism as directing the mass. This book is also an encyclopedia of the structure and organization of communist China, covering topics such as the party-state bureaucracy, the structure of the government, and control and management system in cities and villages. This is a good book to study China in Mao’s era, but not so applicable in today’s China (China changed a lot after Deng’s reform in 1992). However, some of the key ideas in the revolutionary era such as *danwei* are still informative to China studies today (E. J. Perry, 2007).

Mao’s Revolution and the Chinese Political Culture (Solomon, 1971) studied Mao’s communist revolution and its links to traditional Chinese political culture. The book started by summarizing the Confucian-style tradition Chinese political culture, and then continued to analyze Mao’s strategy to reform the culture. For example, Mao placed the notions of struggle, confrontation, and self-assertiveness at the center of his value system rather than the Confucian harmony in order to mobilize people for actions. Mao exploited emotion to control rather than the Confucian rationality and made use of hostility rather than the Confucian benevolence in order to mobilize the masses to overcome anxiety before the old society’s authority. The book also suggested that, despite Mao’s drastic reform of the traditional culture, “these deviations have roots in the traditional political culture” and “many of the institutions of the People’s Republic continue to bear the stamp of the social traditions from which the revolution has grown”. For example, Mao’s emphasis on group life was a “line of continuity with the traditional Chinese stress on subordination of the individual to the collective”. Also, even though Mao encouraged criticism of the cadre from the masses and replaced the Confucian notion of hierarchy with the notion of peers, Mao still asserted his own authority. In short, this book bridged the gap between the Confucian political culture and the communist political culture.

Chinese Democracy (Nathan, 1986) explored the development of democratization and the meaning of democracy in contemporary China through historical studies, and compared it to the American conceptions. The cornerstone of Chinese democracy, the book argued, was built on the doctrines of Liang Qichao, who first introduced the concept of “democracy” into China together with his mentor Kang Youwei at the end of the 19th century. To Liang, democracy was but a means to empower the people to serve the state – “only when the Chinese cultivated the attitude that no one was willing to give up the smallest right would the nation stand up to its oppressors”. Based on the Confucian belief of social harmony, Liang argued that individual rights were not in conflict with the collective interest, but that “individual rights, freedom and autonomy” would lead to “group rights, freedom, and autonomy”. Liang viewed rights as “a kind of substance that could be accumulated so that the rights of individual citizens added up to the powers of the state”, which was contrary to the Western conception of “rights” as “claims against the state”. However, Liang also stressed that due to China’s backwardness, “the natural harmony of the political order could not be realized immediately”, and that the Chinese “must be trained as citizens first; until then, freedom would lead only to disorder”. Thus, Liang framed “democracy” as an instrumental concept under the framework of authoritarianism, and “laid down the rationale that would be used to justify authoritarianism, and the acceptance of

authoritarianism, throughout China's democratic era". The book also reviewed Chinese historical events and movements about democratization from the end of the 19th century to the 1980s, and particularly studied the concept of democracy in different periods under the CCP regime (which will be discussed in the next section). In addition, the book reviewed the Chinese conception of "rights" as "given by the state to individuals", and individualism as "led to personal sacrifice rather than gain". Finally, the book concluded that fundamental political reform is "futile unless the culture is reformed".

The East Asian Challenge for Human Rights (Bauer & Bell, 1999) collected 14 papers on "human rights" in the East Asian context, including theoretical debates on the legitimacy of "Asian Values", the ways of building Asian Values into a more inclusive human rights regime, and case studies of human rights in several East Asian countries. For example, the paper *Human Rights and Asian Values: A Defense of "Western" Universalism* by Jack Donnelly argued that the "Western" conception of human rights is indeed universal regardless of different cultural nuances. The paper *Human Rights and Economic Achievements* by the Nobel laureate Amartya Sen challenged the Asian Values presumption that excessive individual rights could jeopardize economic growth, and argued for the opposite. In addition, An-Na'im studied the anti-terrorist Al-Arqam case in Malaysia and showed that individual rights on religion could be legitimately suppressed for public safety; Othman studied the rights of women in a modern Islamic state and showed that Western human rights activists should strategically respect the local culture for better outcomes. All in all, this book showed a broad range of viewpoints on human rights in the East Asian context.

East Meets West (Bell, 2000) examined the "Asian Values" argument in greater detail. The book explained "Asian Values" as a term "devised by several Asian officials and their supporters for the purpose of challenging Western-style civil and political freedoms". Asian Values, the book explained, accepted a small set of crucial human rights such as "the prohibitions against slavery, genocide, murder, torture, prolonged arbitrary detention, and systematic racial discrimination", but argued differently in the grey area of debate such as "social and economic rights, minority rights and the rights of indigenous people, the rights encompassed within family law and criminal law, the freedom of speech, and the participatory rights inherent in Western-style democratic practices". It argued that Asians place special emphasis upon family and social harmony rather than individual interests, and that democratic procedures have no intrinsic value and what matters is "good government", a government that can provide economic growth, opportunities for education, social stability, and so on. It has two critiques against the US model: 1) the neglect on economic rights – "US government tends to regard freedom from the arbitrary powers of the state as most important, with a concomitant reluctance to accept economic, social, and cultural rights as human rights", and 2) the degradation of social morale – "with its excessive legalism and individualism ... contributes to various social diseases, such as high rates of drug use, collapsing families, rampant crime, growing economic inequality, and alienation from the political process". It argued "with a few exceptions, democracy has not brought good government to new developing countries" because "the governments did not establish the stability and discipline necessary for development". The goal of the book was to explain the "Asian Values" argument without too much personal judgment, and the

conclusion of the book was that different societies have different priorities, and that each specific case requires local knowledge and there's no single universal solution for all cases.

Inklings of Democracy in China (S. Ogden, 2002) is similar to (Nathan, 1986) in terms of the broad range of topics relevant to Chinese democracy. The book framed the conception of "democracy" under the context of Chinese political culture and argued that the conception of "democracy" was a "social construct", and was constantly evolving over time. The book argued that traditional Chinese political culture favors more "equality", "stability" and "social order" over individual freedom, because "freedom" has "never been in the center of Chinese political thought as it is in the Western tradition". In addition, the Chinese focus not on "rights" but on *li*, propriety or rites. After studying China's political culture and democratization history from Imperial China to the Republic and, as well as drawing conclusions from participatory observations in China, the book argued that China today has "inklings" of democracy (although it might be different from the Western conception), and that China is moving toward greater democratization and a growing pluralization despite the cultural inheritance, because of 1) the growing pluralistic interests due to economic growth, 2) participation in international affairs, and 3) to legitimize the Party's rule. Finally, the book suggested using 'human development' (measured by living standard, education, etc) as a fair and consistent standard to evaluate how much substantive freedom and democracy China has.

Political Confucianism (in Chinese) (Jiang, 2003) took the complete opposite approach and point of view. It argued that such terms as freedom, equality, democracy and human rights are merely Western not universal values, and that it is neither necessary nor desirable to import those ideas into China. On the contrary, it argued that Confucian-style authoritarian political ideas are more meritorious and applicable and should be revived in contemporary China. Furthermore, it argued that the attempt of New Confucianism to combine traditional Confucianism with liberal democracy is futile, and that Chinese political reform should revive the Confucian political practice rather than adding liberal elements. It harshly criticized Western-style individualism as "leading to immorality", and equality as "neglecting the innate human differences" and "disturbing social orders".

As an indirect support to Jiang's argument, the book *Confucian for the Modern World* (Bell & Ham, 2003) studies several cases where Confucianism could be positively applied to contemporary East Asian societies in cases such as moral education and mutual help, which are arguably better than liberal democratic practice. However, contrary to Jiang, this book argued that Confucianism and liberal democracy are not inherently incompatible, and the goal of the book was to explore the possibility of assimilating Confucianism into democracy and vice versa.

Beyond Liberal Democracy (Bell, 2006) is a book that summarizes the author Bell's ideas on human rights, democracy, and capitalism through his studies on China and other East Asian societies. The main argument was that "the main hallmark of liberal democracy – human rights, democracy, and capitalism – have been substantially modified during the course of transmission to East Asian societies that have not been shaped by liberalism to nearly the same extent."

Confucian Political Ethics (Bell, 2008b) collected 11 papers and examined Confucian perspectives on political ethics. For example, *Confucian Conceptions of Civil Society* by Richard Madsen explained “inequality” as a positive concept for social order, because the superiors, even though having unequal power over the subordinates, are also constrained by the responsibility to look after them; and thus “inequality” does not mean inferiority but rather “complementary”. Also, Madsen explained that the Confucian view on freedom was not the “freedom to choose”, but to adjust one’s inner self to be harmonious with the outside world, i.e. “creatively contextualizing those commitments which fate has assigned”. The other paper *Confucian Perspectives on Civil Society and Government* by Peter Nosco discussed the Confucian view of civil society as based on personal relationship among people, not on “voluntary association” in the Western tradition. *Territorial Boundaries and Confucianism* by Joseph Chan discussed the Confucian concept of *tian xia*, or “the world under Heaven”, as “a harmonious political order without state boundaries and governed by a sage by means of virtue, without any coercive power at all”.

Another similar book, *China’s New Confucianism* (Bell, 2008a), reviewed Confucian ideas on issues such as war, hierarchical rituals for egalitarian societies, and so on. In particular, the Confucian conception of “peace” does not simply mean the absence of violence, but a united world governed by benevolence (*ren*); and “just war” is to restore peace to *tian xia*. In short, these Confucian conceptions on “inequality”, “freedom”, “civil society”, “just war” and so on are interpreted systematically different from the Western tradition.

To sum up, those eleven books can be placed into four categories based on their different angles on the Chinese political and cultural conception: 1) historical recounts on the development of democracy in China (Nathan, 1986; S. Ogden, 2002), 2) the Asian Values argument (Bauer & Bell, 1999; Bell, 2000, 2006), 3) Confucianism perspective (Bell, 2008a, 2008b; Bell & Ham, 2003; Jiang, 2003), and 4) Maoism/Socialism perspectives (Nathan, 1986; Schurmann, 1968; Solomon, 1971). Next, I’ll synthesize the books with supplementary readings and discuss the different conception on four political concepts.

The conception of “democracy”

The concept “democracy” is hard to define. In the liberal tradition, “democracy” is defined as “a government that has a mandate from the people, accomplished through free and fair elections, and that guarantees individual freedoms, such as religion, assembly, and speech and the right to live one’s own life without excessive government regulation.”² On the other hand, the Marxist view is fundamentally opposing to liberal democracy, believing that “the capitalist state cannot be democratic by its nature, as it represents the dictatorship of the bourgeoisie” – all “independent” media and most political parties are controlled by capitalists and one “either needs large financial resources or to be supported by the bourgeoisie to win an election”³. Thus, the Marxist interpretation of “democracy” is “proletarian democratic centralism”, where the majority of proletarian class rules over the bourgeois and capitalist class (Schurmann, 1968; Solomon, 1971). In fact, Wikipedia listed 25 different types of democracy, each interpreting “democracy” differently. To that, (Collier

² <http://www.answers.com/topic/liberal-democracy>

³ http://en.wikipedia.org/wiki/Democracy_in_Marxism

& Levitsky, 1997) proposed a framework of using adjectives in front of “democracy” to convey its precise meaning. In short, the term “democracy” does not have an objective, scientific definition agreed on by all peoples for all times (S. Ogden, 2002).” More discussion on different definitions of democracy can be found in (Christiano, 2008; Schmitter & Karl, 1991)

What then is the meaning of "democracy" in the Chinese context? Even though some argue that similar concept of democracy was found in Chinese Confucian tradition, the exact term "democracy" was never invented in China on its own (Bell, 2000). In the beginning of the 20th century when “democracy” was first introduced into China by Kang Youwei and Liang Qichao, it was interpreted not as an end in itself, but as a means to unite the people and strengthen the nation (Nathan, 1986). Sun Yat-Sen's interpretation of democracy, as described in his Three Principles of the People⁴ and inherited by Chiang Kai-Shek, was more about “tutelage democracy”— that the people were given certain rights by the state in order to serve the state, and that individual freedom is constrained in order to force people into a unity rather than “a sheet of loose sand” (Nathan, 1986) .

Under the CCP regime, the meaning of “democracy” was also constantly evolving. (Nathan, 1986) explained the difference between Mao’s interpretation (especially in Cultural Revolution) and Deng’s interpretation in Gengshen Reforms. In Mao’s era, "democracy" was promoted in propaganda as the majority proletarian class ruling over the minority capitalist and bourgeois class. It was further explained in (Schurmann, 1968; Solomon, 1971) that “democracy” only meant mass mobilization of the people led by CCP. In Deng’s era, however, the interpretation of democracy was still framed as “class struggle” as in Mao’s time, but Deng argued that there was no “bourgeoisie within the party” as suggested by Mao. Deng also accepted more liberal ideology in the “open economy” policy, but constrained democracy under four lines: socialism, dictatorship of the proletariat, Marxism-Leninism-Maoism, and party leadership.

In the past decade under the new leadership of President Hu, China achieved fast economic growth and CCP government began to prompt the notion of “harmonious society” instead of Maoist “class struggle”. The Chinese political culture is moving towards more pluralism (S. Ogden, 2002), and the meaning of “democracy” in contemporary China remains a puzzling question. Although the government dubbed China as a democratic country, it is not clear what kind of “democracy” it is. Furthermore, it is quite likely that the Chinese people interpret “democracy” differently in political deliberation. Measuring the meaning of “democracy” in Chinese online deliberation would be an interesting research topic.

The conception of “rights”

The Chinese conception of rights is different compared to the Western tradition. By studying the teachings of ancient Chinese philosophers, the doctrines in Mao’s era and the ethical discourses of contemporary political leaders, (E. J. Perry, 2008) observed that they "all emphasized economic development, and specifically poverty alleviation for the hard-pressed peasantry, as a cornerstone of their claim to political legitimacy". Perry also

⁴ http://en.wikipedia.org/wiki/Three_Principles_of_the_People

observed that Chinese protesters framed their human rights claims as requests for basic survival needs, which can be characterized as more of "rules consciousness" than "rights consciousness". In contrast, the American conception of rights is more of civil rights in terms of Locke and Jefferson – "Tyranny, not poverty, was offered as the reason for rightful rebellion. To Americans, liberty rather than livelihood is the foundation of political morality." Perry's conclusion was that the Chinese conception of rights is more of "the rights to subsistence and development" rather than civil rights in terms of Locke and Jefferson.

Perry's argument was not flawless. First, the representativeness of Perry's sample was questionable. She just selectively chose the examples in favor of her argument, while ignoring the opposite side of examples, such as the 1989 student movement, the Tibet uprising, and the Falungong movement, etc, to show that Chinese contemporary conception of human rights is consistent with Western literature. Second, Perry overlooked the strong Western influence in contemporary China. It is true that a lot of people might consider human rights as a basic survival need and economic comfort, but still there are many people, especially the growing middle class population, who are influenced considerably by the West and think of rights as civil rights. Despite the problem of lack of support, Perry's paper did provide an intriguing viewpoint to China studies – Chinese conception of rights is not necessarily consistent with the Western literature.

Perry's argument was echoed in other studies as well. For example, (S. Ogden, 2002) reviewed several surveys that might count as empirical support for Perry's work. In a 1990 poll conducted among rural and urban Chinese youth on "What do you consider the greatest happiness for yourself?", the civil rights concepts such as "freedom" were ranked much lower than socio-economic rights concepts such as "successful career", "warm family", etc. In a 1998 survey of 350 Beijing college students on the "top 12 development issues for China over the next five years", the civil rights issue "defends freedom of speech" was ranked at 10th, following economic rights issues such as "promote economic development" (1st) and "maintain economic stability" (2nd). Those surveys, however, were limited in sample size and representiveness, and were more or less outdated.

Furthermore, (Nathan, 1986) studied 11 Chinese constitutions in 4 periods of Chinese modern history – the imperial dynasty, the liberal early republic, the authoritarian Guomindang, and the socialist People's Republic – and noted that political rights in China were consistently regarded as a grant given by the state to the citizens, in order to make them contribute their energies to the needs of the nation, rather than to protect the individuals against the government. The author pointed out six differences of Chinese constitutions compared to the American constitutions on rights. First, none of the Chinese constitutions considered rights as derived from "human personhood", but granted by the state, which was the opposite in America. Second, since rights were granted, all Chinese constitutions granted different rights, whereas in America rights were defined quite consistently. Third, some rights in each Chinese constitution were presented as goals to be realized. Fourth, every Chinese constitution gave the government the power to limit rights by acts of legislation, whereas in America, rights are considered to form a limit to law. Fifth, none of the Chinese constitutions established an effective procedure for independent

review of a law's constitutionality. Sixth, none of the Chinese constitutions provided the procedure to elect the president. In short, (Nathan, 1986) suggested that the Chinese conception of rights is very different from that in the US.

The conception of "freedom"

The conception of freedom is closely related to the conception of democracy and rights. Two forms of freedom – negative freedom and positive freedom – were discussed in (Berlin, 1969). Negative freedom can be defined loosely as “the absence of obstacles, barriers or constraints” (Carter, 2008). Positive freedom, on the other hand, is “the possibility of acting – or the fact of acting – in such a way as to take control of one's life and realize one's fundamental purposes” (Carter, 2008). Intuitively, positive freedom is to say that sometimes individuals are not free or able to make the choice in their best interest without positive intervention from the outside. For example, the freedom of young children is limited considerably because they might harm themselves if allowed to act freely. Another example might be to help drug addicts quit using drugs by forcefully constraining their free actions – although the addicts wish to quit, sometimes they cannot control themselves without help from others.

The Western conception of freedom is usually in the negative sense. Many Western scholars such as (Berlin, 1969) opposed positive freedom because they argue that the logic of positive freedom runs the risk of authoritarianism – the authorities in power might abuse their power not for the good of the people, but for their selfish benefits, in the sense that “power tends to corrupt; absolute power corrupts absolutely”.

On the other hand, literature shows that the Chinese conception of freedom has a strong sense of positive freedom, which is quite different from the Western conception. First, Chinese tradition believes that people have intrinsic values that they need to develop, such as filial piety, or to live as “the way of heaven” (Bell & Ham, 2003). Second, the Chinese tradition acknowledges that people have the duality of inferior desires and higher goals that they often need outside force to help them subdue their inferior desires in order to achieve the higher goals (Bell & Ham, 2003). Together, these two reasons create the legitimacy for positive freedom. Authoritarian rules and disciplines are usually treated not as a constraint to freedom, but as liberation to access one's potential.

From studying Chinese political culture, (S. Ogden, 2002) also viewed the Chinese conception of “freedom” in the positive sense. The author hinted that in developing countries such as China, the population often acts irresponsibly, and even illegally, in order to “sustain and better their lives in a society of scarce resources”, and that the constraint on individual freedom is necessary to assure others' freedom. In addition, the author argued that the Chinese people tend to view the state as “a caretaker, with responsibility for fulfilling the paternalistic role of a patriarchal state”, and people would not prefer freedom without the guarantee of good life – food, clothing, housing, education, medical care and security. Furthermore, the author argued that individuals in the Chinese political culture don't yearn for much freedom – “individuals in an authoritarian political culture tend to be less tolerant of difference and, like their government, usually are not deeply concerned

with the right of an individual to express difference freely". Finally, the author proposed to use "human development" as a fair measure of freedom in the Chinese political culture.

(Nathan, 1986) also suggested the Chinese conception of freedom complies with positive freedom. Nathan noted that, instead of seeing the constraints on individual freedom (usually by moral order) as limiting the power of the individuals, the Confucians saw it as "calling for the fulfillment of the people's innate promptings as social beings", or "moral constraints ... are intended not to make [the ruler] less of a king but to help him be a king". This view was explained in (Bell, 2008b) as well. Nathan also noted that Liang Qichao paired freedom with discipline, defining "freedom as acting autonomously as long as one does not invade others' freedom, and discipline as the virtue of respecting others' freedoms. Thus freedom is marked by obedience to law and to the public interest." In short, the conception of "freedom" in the Chinese context is quite different.

The role of government

The conception of democracy, rights and freedom is closely related to the role of government, or, state capacity (Tilly, 2007). American political tradition emphasizes limited government, especially on the federal government. Until the recent financial crisis, Americans usually think of government as "state-as-problem" rather than "state-as-solution", to use the terminology in (Brinkerhoff, 2008). The role of government should be a referee, not a player, a servant, not a father. Distrust of the government can be found even in the US Constitution, such as allowing the citizens to have guns to prevent political tyranny. This might be too extreme, but in general, Americans do not like the government⁵.

On the other hand, the Chinese have a more paternalistic view of the role of government. As (E. J. Perry, 2008) pointed out, "Chinese statecraft since the times of Mencius has envisioned a more proactive role for government – which was expected to promote economic welfare and security." In communist China, the party-state government and village communes were also viewed as the protector of people to provide food and security (Schurmann, 1968). From the study of the urban poor, Perry hinted that people have a mental reliance on the government to provide welfare (E. J. Perry, 2008). A similar idea can be also found in (Nathan, 1986) and (S. Ogden, 2002). In addition, (Walder, 2002) discussed the local governments' active role in participating in economic growth rather than passively isolate themselves from the economy. In short, the Chinese view of the government is consistently paternalistic beginning with the Confucian China, through the communist China, and to today (E. J. Perry, 2007).

Potential historical explanations

This section attempts to offer three potential explanations from a historical perspective to answer why China might have adopted these conceptions discussed earlier.

First, to many Chinese people, a powerful central government is preferable in order to prevent civil wars, natural disasters, and foreign invasions. Historically, when China had no

⁵ According to Prof. Gallagher, this is one of the cases where the US is different from Europe and the rest of the world.

central government, there were always wars among landlords. Prosperity and peace came only when a powerful central government was in place. Also, some argued that a powerful central government is required to fight against natural disasters such as the annual flooding of the Yellow River, which is only possible when people are coordinated and motivated throughout the country. Furthermore, China in its history has suffered a lot from foreign invasions such as the Mongols. In modern history (mid-19th to early 20th century), China was forced into the Opium war with Britain and invaded by Japan during WWII – such fresh memories remind people the necessity of a strong central government. Under a strong central government, it is natural to form the conception of democracy, freedom and rights as described earlier.

Second, the cultural heritage from both Confucianism and the communist revolution place a higher value on collective interests over individual interests, which is fundamentally different from the popular Western value system based on liberalism and individualism. For more than two thousand years, Confucianism taught people the virtue of obedience to authorities (Bell, 2008a). People learned this and lived this for a long time, carrying the tradition from generation to generation. The communist revolution brought Marxism-Leninism and Maoism, which deviated considerably from traditional Confucianism. But still, collective values were higher than individual values (Schurmann, 1968). And even today when communism is greatly diluted, the communist tradition of collective interests is still reflected in the political culture today (E. J. Perry, 2007). In short, the liberal ideology, so popular in the West, has never fully grown in China.

Finally, education and propaganda also play an important role in maintaining the established political culture of Confucianism and socialism ideologies. (Y Zhao, 1998) argued that mass media in China is CCP's "mouthpiece" for propaganda, promoting official ideologies such as "harmonious society" all over the media. Also, (S. Zhao, 1998) studied education campaigns to promote nationalism and love of the state. Even though there is undoubtedly influence of liberalism and individualism from the West, the state apparatus is still powerful, e.g., (Brady, 2009) studied mass persuasion and propaganda and argued that it was effective to legitimize China's popular authoritarianism. In particular, I'll study censorship and its effects on people's ideas of political concepts in the next section.

Online censorship and its effects

The study of China's online deliberation has to take into account censorship and its effects. Despite recent media reform and commercialization, mass media in China is still tightly controlled by the CCP to maintain its "commanding heights" (Yuezhi Zhao, 1998). The Internet, especially social media⁶, is also heavily regulated. There's a large literature studying Chinese censorship and its effects. For example, (Harwit & Clark, 2001) studied three layers of Internet control in China – physical network control, network content control, and foreign influence control. (Chiao, 2009) studied Chinese censorship under the framework of legal economics. (Rayburn & Conrad, 2004) studied four tiers of censorship

⁶ I'll use "social media" as an umbrella term for all kinds of user-generated-contents such as posts in BBS/forums, blogs, or comments. It is distinguished from the "broadcasting" website model where communication is uni-directional.

using technology: gatekeeper (the fiber cable to the internet), ISPs, ICPs, and client-side. (Tan, 1999) examined the convergence of media regulation across all Chinese telecommunication regimes. (Zhang, 2006) conducted interviews with high officials in the regulation regime and summarized the “unspoken” guidelines behind censorship. (Zittrain & Edelman, 2003) studied the technologies of censorship and listed blocked websites about political dissidents, health (hunger China, AIDS China, etc.), news (BBC, CNN, etc), and so on. (M. Chase & J. C. Mulvenon, 2002) studied how the dissidents used the Internet and the counterstrategies of the government.

Next, I’ll synthesize the literature and show 1) the common censorship practice, and 2) the effects of censorship. And finally I’d like to argue that censorship might further obfuscate the meaning of political concepts in online deliberations, and suggest that future research should take into consideration the effects of censorship, as shown in Section 6.1.

To legalize online censorship, the state agencies, led by the Propaganda Department of CCP, have enacted several regulatory laws. For example, "Regulations on Internet News and Information Services" (CNNIC, 2005) identifies 11 categories of forbidden topics, such as “spreading rumors, disturbing social order, and undermining social stability”, “agitating illegal gathering, association, demonstrations, protests, and the gathering of masses to disturb social order”, and “engaging in activities in the name of illegal civil organizations”. In addition, the Propaganda Department can conduct “administrative intervention” by issuing “internal documents” to supervise special cases (Chiao, 2009), which makes the censorship policy elusive to the observers.

To enforce such regulations, the first approach was “low tech solutions for high tech problems” (M. Chase & J. Mulvenon, 2002). For example, cyber police squads, “as many as 30,000 in one estimate, are patrolling Chinese cyberspace, deleting politically incorrect content in real time, blocking websites, monitoring networking activities of citizens, and tracking down and arresting offending individuals (Zhao, 2008).”

The high tech solutions are also exploited extensively to censor outside information (Zittrain & Edelman, 2003). By using blacklists, keyword filtering and intelligent pattern recognition algorithms, the "Great Firewall" is able to block political dissenting sites, pornography sites, as well as rather innocuous social sites such as Twitter, Facebook, Youtube, MySpace, Blogger, etc, where people can contribute content freely. Also, (M. Chase & J. Mulvenon, 2002) suggested that the government actively adopted hacking attacks such as denial-of-service to bring down dissidents’ websites such as *minghuai.cn*. The recent security breach of *Google.cn* targeting to dissidents’ email is another example of the government’s “high-tech” approach⁷.

In addition to passively blocking unwanted information, the state also employs “active” censorship. Recently, the party branches in several universities started hiring students, known as the “50 cents army” (*wumaodang*), to publish articles in online discussions for 50 cents a piece in favor of the government, and sometimes to employ personal attacks against those who have different opinions (Elgan, 2009). The disruptive nature of “50 cents army”

⁷ <http://googleblog.blogspot.com/2010/01/new-approach-to-china.html>

discourages many ordinary citizens to join online discussions simply because the innocent people want to avoid getting picked on by the “army”.

Despite of state-led censorship, the government also “outsources”⁸ censorship to ISPs (Internet Service Providers) that host online discussions and blogs. “Regulation on Internet Bulletin Board Services” holds ISPs responsible for deleting unwanted contents; failing to comply will result in severe punishment (CNNIC, 2000). Lacking “safe harbor” protection, many ISPs usually employ even stricter censorship rules (MacKinnon, 2008), or simply discourage any political topics. Foreign ISPs like the local branches of Microsoft and Yahoo also employ self-censorship to avoid troubles (E. Gutmann, 2004).

Self-censorship is also extended to individuals through the state's psychological control of injecting fear into people. (Link, 2002) described it vividly as “anaconda in the chandelier”: “The Chinese government's censorial authority in recent times has resembled not so much a man-eating tiger or fire-snorting dragon as a giant anaconda coiled in an overhead chandelier. Normally the great snake doesn't move. It doesn't have to. It feels no need to be clear about its prohibitions. Its constant silent message is “You yourself decide,” after which, more often than not, everyone in its shadow makes his or her large and small adjustments—all quite ‘naturally’.”

Finally, the censorship policy is inconsistent and vague. (Lieberthal, 1992) talked about “fragmented authoritarianism” and the tension between the central government and local governments. In fact, the central government is quite tolerant of sensitive topics such as corruption and political reform, given that the discussion is under the premise of one-party system (Deutsche-Welle, 2007). Interviews with high officials in the Propaganda Department revealed that central leadership even considered pushing more press freedom to facilitate power struggles within the central committee and to add bargaining power against local governments (Zhang, 2006).

The paragraphs above reviewed the practice of online censorship in China. Other literature studied how censorship affects online communication and thus affects people. For example, (Shi, 2003) studied how the difference of ownership (state-owned vs. private-owned) and the channel (traditional channel vs. online channel) imposed different degrees of censorship, and showed that ownership didn't matter that much, and the channel was very important, i.e., the online channel had less censorship than the traditional channel regardless of ownership. (Elegant, 2007) argued that censorship created the “millennium generation” of Chinese youth who are apathetic to politics, which is detrimental to the Chinese political culture in the long run. (Thompson, 2006) argued that even under censorship, the Internet still creates an unprecedented opportunity for the youth to experience “democracy” through non-political exercises such as expressing personal ideas through blogs, and voting for super idols.

Undoubtedly, censorship affects the results of online deliberation. An interesting question is to study whether and how censorship could change people's conception on basic political

⁸ The term “outsource” was suggested by Prof. Gallagher.

terms such as “democracy” or “rights” as we discussed above. This is a gap in current literature, and will be addressed in Section 6.

Summary and challenges

Table 1. Summary of the China studies literature

	References	Democracy	Rights	Freedom	Role of gov't	Other concepts
Classical books	(Schurmann, 1968)	X		X	X	organization, Marxism, Leninism, ideology, centralism
	(Solomon, 1971)	X		X	X	luan (confusion), authority, ideology, democratic centralism
	(Nathan, 1986)	X	X	X	X	sovereignty, constitution, law, propaganda
	(Bauer & Bell, 1999)	X	X	X	X	social justice, legality, citizenship
	(Bell, 2000)	X	X	X	X	authoritarianism
	(S. Ogden, 2002)	X	X	X	X	civil society, human development
	(Jiang, 2003)		X	X	X	centralism, political Confucianism
	(Bell & Ham, 2003)	X	X		X	constitution, accountability, capitalism, law, decentralization
	(Bell, 2006)	X	X	X	X	capitalism, just war, citizenship
	(Bell, 2008b)			X	X	civil society, territory, citizenship, inequality
(Bell, 2008a)			X	X	communism, just war, hierarchical rituals, political Confucianism	
Supplementary readings	(Collier & Levitsky, 1997)	X				
	(Christiano, 2008)	X				
	(Schmitter & Karl, 1991)	X				
	(E. J. Perry, 2008)		X			
	(Berlin, 1969)			X		
	(Carter, 2008)			X		
	(Brinkerhoff, 2008)				X	
	(Walder, 2002)				X	

To sum up, this section reviewed classical books and supplementary papers about the Chinese conception on democracy, rights, freedom and the role of government. I showed that the Chinese conception places higher values on collective interests and welfare rights and is quite different from the Western conception based on liberalism and individualism. I offered three tentative explanations on what makes the Chinese conception different. And finally I studied the practice and effects of online censorship in China.

To understand the different meaning of political concepts in the Chinese context has significant implications for studies about China. For example, given that the Chinese people interpreted “rights” as economic rights, then perhaps the authoritarian regime is indeed better than liberal democracy because the social elites of the party might make better decisions for economic development. Therefore, social and political theories developed in Western context might need re-evaluation before introducing to the Chinese context. It is also important to make salient the difference between the Chinese and Western conception

on basic political terms so that scholars with different backgrounds and assumptions are able to communicate well to make academic advance.

The key challenge here is to find more empirical evidence, which is lacking in current literature, to how the Chinese people, not the government or the textbooks, interpret such political concepts as “rights”, “freedom”, and “democracy”. One of the best sources for such empirical evidence could come from online deliberation, which is next section’s topic.

3. Online Deliberation

In this section I'll review literature grouped into four topics: 1) theories about deliberation, 2) critiques of them, 3) technology and online deliberation, and 4) deliberation in China. The goal is to review and present the theoretical foundation necessary to study deliberation in China. More importantly, I'd like to show that deliberation, particularly online deliberation, is important to democracy and Chinese political reform, and thus I'd argue that it is of theoretical importance to study Chinese online deliberation empirically by measuring the meaning of political concepts within the deliberation.

Deliberative democracy and public sphere theories

Deliberation is important and deserves much attention because theories argue that deliberation plays a central role in democracy and political life. (Gastil & Keith, 2005) described three periods of deliberative democracy theories: the progressive era (1890s-1940, expanding), 'mass society' (1940s-1960s, contracting), and 'deliberative renaissance' (1970s-, expanding). Following that historical trajectories, I'll describe some deliberation theorists and their ideas in chronological order.

In (Dewey, 1927), the author emphasized the importance of communication, and concluded that the problem of the public was "improvement of the methods and conditions of debate, discussion and persuasion". The book also proposed that the ideal society can occur once local community was reestablished as "a medium of intelligence, allowing word of mouth and face-to-face interactions to give reality to public opinion". In (Barber, 1984), the author argued that we should have "strong democracy" based on citizenship and participation through deliberation, rather than simply the "thin democracy" based on liberalism.

One of the seminal works was the "public sphere" theory by (Habermas, 1989). In this work, "public sphere" was at the core of democracy, which was defined as: "The bourgeois public sphere may be conceived above all as the sphere of private people come together as a public; they soon claimed the public sphere regulated from above against the public authorities themselves, to engage them in a debate over the general rules governing relations in the basically privatized but publicly relevant sphere of commodity exchange and social labor(Habermas, 1989)."

The public sphere theory was very influential, and (Dahlberg, 2001) summarized it to have 6 major points, as quoted below:

1. Autonomy from state and economic power. Discourse must be based on the concerns of citizens as a public rather than driven by the media of money and administrative power that facilitate the operations of the market and state.
2. Exchange and critique of criticizable moral-practical validity claims. Rational-critical discourse involves engaging in reciprocal critique of normative positions that are provided with reasons and thus are criticizable, that is, open to critique rather than dogmatically asserted.
3. Reflexivity. Participants must critically examine their cultural values, assumptions, and interests, as well as the larger social context.

4. Ideal role-taking. Participants must attempt to understand the argument from the other's perspective. This requires a commitment to an ongoing dialogue with difference in which interlocutors respectfully listen to each other.
5. Sincerity. Each participant must make a sincere effort to make known all information, including their true intentions, interests, needs, and desires, as relevant to the particular problem under consideration.
6. Discursive inclusion and equality. Every participant affected by the validity claims under consideration is equally entitled to introduce and question any assertion whatsoever. Inclusion can be limited by inequalities from outside of discourse - by formal or informal restrictions to access. It can also be limited by inequalities within discourse, where some dominate discourse and others struggle to get their voices heard.

Following the Habermas public sphere theory, Gutmann et al introduced the idea of “deliberative democracy” (A. Gutmann & Thompson, 2004). Compared to the simple “aggregative democracy” where collective decisions are made by aggregating the votes, “deliberative democracy” can be defined as “a form of government in which free and equal citizens (and their representatives), justify decisions in a process in which they give one another reasons that are mutually acceptable and generally accessible, with the aim of reaching conclusions that are binding in the present on all citizens but open to challenge in the future (A. Gutmann & Thompson, 2004).” Gutmann et al further discussed that the aims of deliberative democracy were: 1) to promote the legitimacy of collective decisions, 2) to encourage public-spirited perspectives on public issues, 3) to promote mutually respectful processes of decision-making, 4) to expose decisions to critical scrutiny, surfacing weakness, errors, and injustice.

Similarly, (Button & Ryfe, 2003) also summarized the answers to the question “why deliberate”⁹ as:

- Legitimacy – deliberation produces more legitimate decisions (improving acceptability, adherence, etc.)
- Efficacy – deliberation produces objectively better decisions (more sound, rational, based on better information, etc.) (an instrumentalist justification)
- Learning – the process of deliberation promotes important kinds of individual and collective learning (incl. preference formation and transformation)
- Emergence / public discovery and transformation – the process of deliberation builds and enforces norms and values of civic / public culture (‘calls a public into being’)

In addition, (McAfee, 2004) explained “why deliberate” using three models. The preferential based model views human beings as *homo economicus* who see the social world as a market in which they try to maximize their own preferences, thus “democracy would be ‘rule by the people’ in a way that helps them maximize, as much as possible, individuals’ preferences.” The rational proceduralist model argues that citizens are guided by a will to “come up with universalizable norms – or at least norms that are acceptable to all those affected by any given policy”; deliberation then is a way in which individuals

⁹ Quoted from SI 732 class slides, by Prof. Jackson

collectively decide whether a policy is legitimate. The integrative model views deliberation as a process through which people “grapple with the consequences of various public problems and proposals; they see politics as a difficult matter of deciding what kinds of communities they are making for themselves”.

Critiques

(Young, 1996) criticized the deliberative model from two perspectives. First, *exclusiveness* – “by restricting their concept of democratic discussion narrowly to critical argument, most theorists of deliberative democracy assume a culturally biased conception of discussion that tends to silence or devalue some people or groups”. Second, *assuming unity* – “deliberative theorists, moreover, tend inappropriately to assume that processes of discussion that aim to reach understanding must either begin with shared understandings or take a common good as their goal.” To address these problems, Young proposed *communicative democracy*, where “differences of culture, social perspective, or particularist commitment” and “greeting, rhetoric and storytelling” were all considered as political communication and incorporated in the deliberative model.

(Schudson, 1997) argued that conversation is not the “soul of democracy”, and that “institutions and norms of democracy give rise to democracy conversations rather than that the inherent democracy of conversation gives rise to politically democratic norms and institutions.” In order to make this argument, the author distinguished between the “social” model and “problem solving” model of conversation, and argued that conversation is not the panacea for the problems of democracy.

Drawing conclusions from participatory observations and interviews with ordinary Americans, (Eliasoph, 1998) explained how Americans came to political apathy due to the anomalies of public discussion. In (Mansbridge, 1999), the author argued that everyday talk is a crucial part of the full deliberative system even though it’s not always deliberative. In (Mutz, 2006; Sunstein, 2001), both authors argue that un-regulated deliberation might lead to balkanization and polarization, where people prefer to communicate with similar minds, which in turn re-enforce their prior opinions.

(Fung, 2004) summarized criticisms against the deliberative model from five perspectives:

- The strong-rationality perspective: it is not easy for people to acknowledge others’ perspectives in deliberative settings due to lack of rationality.
- The strong-egalitarian perspective: inequalities in society prevent deliberative democracy from creating fair laws and regulations.
- The social-capital perspective: weak social capital among people might thwart the performance of deliberation.
- The cultural-difference perspective: deliberative democracy can be culturally biased against non-Western, non-white, non-middle-class cultures.
- The expertise perspective: ignorance and lack of competence from ordinary people might jeopardize decision making where expertise is required.

(Button & Ryfe, 2003) also summarized the limits, pathologies, and challenges of the deliberative democracy model¹⁰:

- The deliberation industry (lobbyists, paid consultancy)
- Costs (money, time, efficiency)
- Selection, manipulation, and the management of outcomes
- Consultation without power (participation as enrollment)

(Levine, 2007) argued that “it is imperative that those engaged in civil society have the skills, experience, and numbers needed to make it a viable public force. If its members are without skills or experience, the civil society suffers”. The author argued that deliberation would probably fall apart if people didn’t know “how to discuss issues civilly and on terms of mutual respect, without recourse to violence”. To address those problems, Levine argued that the necessary condition for deliberative democracy would be successful campaigns of civic engagement and civic education.

To sum up, theories in favor of deliberation basically argue that deliberation is at the core of democracy, and that it has both intrinsic values – deliberation is people’s rights – and instrumental values – only through deliberation can people be informed and make better democratic decisions. The critiques, however, challenged the prerequisites, assumptions and practical feasibility of deliberation, and some critiques suggested changes to improve the theories and outcome of deliberation, such as “civic education” (Fung, 2004) or “communicative democracy” (Young, 1996). Nevertheless, none of the critiques challenged the core values of deliberation to democracy, and the ferment debate on the pros and cons of deliberation justifies the theoretical importance of the topic – certainly, deliberation is an important topic to democracy and political studies.

Technology and online deliberation

Theories discussed in the previous section focused on “offline” deliberation, i.e., face-to-face deliberation in social settings. In this section, I’ll discuss the impact of technology and the Internet on online deliberation, and review four clusters of literature as shown in Table 2. The goal is to show that online deliberation might be different from offline deliberation in terms of conveying the meaning of political concepts, and that better technical designs might help people better understand the meaning of the political concepts.

The first cluster of literature, studies in CMC and CSCW, demonstrated that “distance matters”. For example, (Hollan & Stornetta, 1992) showed that telecommunications missed many subtle features in face-to-face communication such as “cue variety, feedback and message personality”, and that technology should not just aim to make up for those missing features but to expand new functionalities such as to facilitate archive and multi-tasking. For another example, (Olson & Olson, 2000) summarized prior literature in the area and argued that “distance matters” regardless of technological development. The general theme of this line of literature is to study the difference between the “online version” and the “offline version” in the broad field of telecommunication and collaborative work.

¹⁰ Quoted from SI 732 class slides, by Prof. Jackson

Table 2. Clusters of literature on online deliberation

Clusters	Argument/Findings	Examples
Technology effects in CMC and CSCW	“distance matters”	(Hollan & Stornetta, 1992; Olson & Olson, 2000)
Face-face deliberation vs. online deliberation	online deliberation has both pros and cons	(Albrecht, 2006; Dahlberg, 2001; Min, 2007)
Online deliberation applications	online deliberation has many useful applications	(Gastil & Levine, 2005; Gordon & Koo, 2008; van de Donk, 2004)
Technology as control	technology can be actively designed for social benefits	(Beniger, 1986; L. Lessig, 2006)

A similar theme was found in the online deliberation literature as well. One cluster of literature compared online deliberation to face-to-face deliberation. For example, (Albrecht, 2006) studied the constituents of participants in online deliberation and showed that it represented more youngsters and less women, and that digital divide indeed exists. (Min, 2007) compared online versus face-to-face in terms of civic engagement and showed that both can increase participants’ issue knowledge, political efficacy, and willingness to participate in politics. (Dahlberg, 2001) criticized online deliberation on the increasing commoditization, lack of reflexivity and respectful listening, difficulty of verifying authenticity, digital divide, and individuals dominance. All of these studies demonstrated that online deliberation has both pros and cons compared to the traditional offline deliberation.

Another cluster of literature discussed examples of online deliberation applications and innovative technology to improve online deliberation. For example, (Gastil & Levine, 2005) described the National Issues Forums, Deliberative Polling, Citizen Juries and other digital government applications using deliberation. (Gordon & Koo, 2008) studied deliberation in the virtual world *Second Life* and argued that it enabled “previously disempowered individuals” to “form politically powerful groups”. (van de Donk, 2004) explored the role of online deliberation in cyber-protests and argued that the online setting might “change the political game in favor of resource-poor players”. The *Opinion Space*¹¹ visualized individuals’ viewpoints in a map and showed semantic distance from other viewpoints.

The first three clusters of literature mainly studied the effects of technology in the “passive” sense; the fourth and final cluster of literature introduced the idea of “technology as control” and argued that technology could be actively designed to achieve social benefits. (Beniger, 1986) argued that the technology revolution in the beginning of the century was in fact a “control revolution”, where control was shifted to technology. For example, since the location of railroad has direct influence to how farms sell produce and extract profits, the railroad companies which possess the technology of constructing railroad would then have power and control over the farms. Similarly, (Lawrence Lessig, 1999) argued that “Code is law; architecture is control; software is power”. To illustrate his argument, Lessig introduced the concept of the four modalities – law, architecture, norms and market – and how they serve as both constraints and protectors to regulate certain behaviors, as shown

¹¹ <http://opinion.berkeley.edu>

in Figure 4. One example was that, although the law doesn't suppress pornography, social norms would suppress it. "While modalities of constraint can be used as swords against the individual (powers), modalities of protection can be used as shields (rights)" (L. Lessig, 2006). Lessig used the example of "free expression" and showed how technology as control could promote free expression: "The law therefore is an imperfect protection. But on top of this list of protectors of speech in cyberspace is (once again) architecture. Relative anonymity, decentralized distribution, multiple points of access, no necessary tie to geography, no simple system to identify content, tools of encryption—all of these features and consequences of Internet protocol make it difficult to control speech in cyberspace. The architecture of cyberspace is the real protector of speech there; it is the real 'First Amendment in cyberspace', and this First Amendment is no local ordinance."

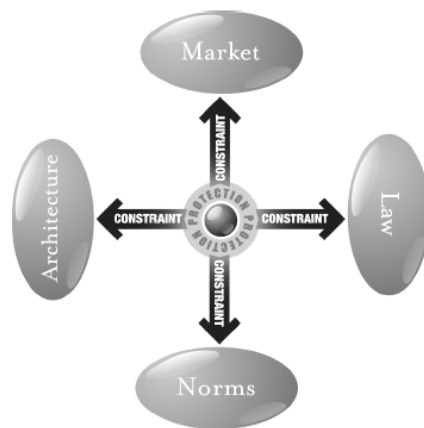


Figure 4. Four modalities (Lessig, 1999)

To sum up, these four clusters of literature suggested that online deliberation could be different from offline deliberation in terms of conveying the meaning of political concepts, and this should be considered in future research design. Also, the idea of "technology as control" suggested that technology should play an active role to promote social benefits in online deliberation. One advantage of computational approaches to measure the meaning of political concepts, compared to traditional methods, was that they could be adopted to improve people's common understanding of political concepts in online deliberation. An example will be discussed in Section 6.1.

Deliberation in China

As (Nathan, 1986) pointed out, "together with competitive elections, an independent press [also implied free deliberation] was one of two reforms that the more radical among the activists considered essential to the democratization of Chinese politics". This section reviews literature about deliberation in China studies. The goal is to point out that measuring the meaning of political concepts is important to study deliberation in China.

The first notable study of civil society, public sphere and deliberation in the Chinese context was the 1993 Rowe-Wakeman-Huang debate in the journal *Modern China*. By studying the city Hankow in late imperial Qing Dynasty, (Rowe, 1993) argued that the necessary elements of civil society and the public sphere such as capitalism, proprietorship,

the notion of natural rights and social contract, were already found at that time, and suggested that the theories of civil society and the public sphere that originated in the West were also applicable to China. On the contrary, (Wakeman Jr, 1993) criticized Rowe's "mechanical application" of Habermas's model on to the Chinese historical experience, and stressed the role of the state which was missing in Rowe's argument. On the other hand, (Huang, 1993) proposed the idea of "the third realm" between the state/society dichotomy central to the Rowe-Wakeman debate, and argued that China had no such conflict as "societal autonomy against the state", but rather the reality was "state-society relation in the third realm".

Another notable work was a collection of papers in the book *The Search for Deliberative Democracy in China* (Leib & He, 2006), with the goal to "develop forms of deliberative democracy more appropriate for the Chinese context" instead of "of looking to the US as the standard". For example, the paper *Human Nature, Communication, and Culture: Rethinking Democratic Deliberation in China and the West* by Shawn Rosenberg argued that "inequality" between seniors and juniors, superiors and subordinates, is important to Chinese deliberation in order to reflect social order and promote social harmony; this is in contrast to the Anglo-American assumption of equality and autonomy in deliberation. The paper *Western Theories of Deliberative Democracy and the Chinese Practice of Complex Deliberative Governance* by Baogang He argued that deliberation doesn't necessarily require the assumption of "liberal democracy", and proposed the idea of "authoritarian deliberation", where deliberation plays an important consultancy role to the authoritarian government. Baogang He's other article *Participatory and Deliberative Institutions in China* argued that deliberation could release the tension between the state and the people and thus strengthen the state's legitimacy. He explained three forms of deliberation in authoritarian China – consultative meeting, citizen evaluation meeting and village representative assembly – and explained two positive outcomes: enhancement of local governance and empowerment of citizens. The paper also argued that emotion could enhance deliberation rather than distort it. *The Native Resources of Deliberative Politics in China* by Shengyong Chen argued that deliberation in China should be smoother because of the concepts of "harmony" and "people first". The book finally presented several papers of empirical studies, such as "democratic talk in all sincerity", "deliberative poll", and a case study of reforming Peking University with deliberative democracy.

The literature reviewed above mainly addressed theoretical studies about deliberation in China. The other line of literature focused on online deliberation and the Internet public sphere in China, which is discussed below.

Guobin Yang is one of the pioneers in the study of Chinese Internet deliberation. Through survey data and case studies, (Guobin Yang, 2003) suggested that the Internet has three usages: 1) engaging people in public debate and problem articulation, and serving as supervisory role, 2) forming online communities, 3) facilitating mobilization and protest. Through participatory observation on two online forums, (G. Yang, 2003a) studied the co-evolution of the Internet and civil society in China: "The internet facilitates civil society activities by offering new possibilities for citizen participation; civil society facilitates the development of the Internet by providing the necessary social basis". (G. Yang, 2003b)

studied the transnational public sphere dubbed as “Chinese culture sphere” using an ethnography method. (G. Yang & Calhoun, 2007) studied how deliberation affected policy making on the dam construction project on the Nu River. (Guobin Yang, 2009) systematically studies online activism in China.

There are other studies in this area too. For example, through case studies, both (Bates, 2008) and (Wang, 2007) argued that under strict state regulation of mass media, the Internet would be the most powerful form of public sphere in China, and would lead to the rise of democracy. (Y. Zheng & Wu, 2005) studied the Internet as 1) communication tool, 2) public sphere, and 3) means for collective action.

However, all the literature reviewed so far focused only on deliberation itself, without relating deliberation studies to a broader research agenda in China studies. As an exception, (Leib & He, 2006) studied the relation of deliberation and democratization in China, but their empirical evidence came mostly from specific case studies. (Lei, 2009) studied how the Internet increased Chinese netizens’ “rights consciousness” through survey data, but it also raised the question of finding more empirical evidence. All in all, I’d argue that measuring the meaning of political concepts from Chinese online deliberation would be an important contribution to fill the gap between deliberation studies and other research agenda. An example is given in Section 6.3.

Summary and challenges

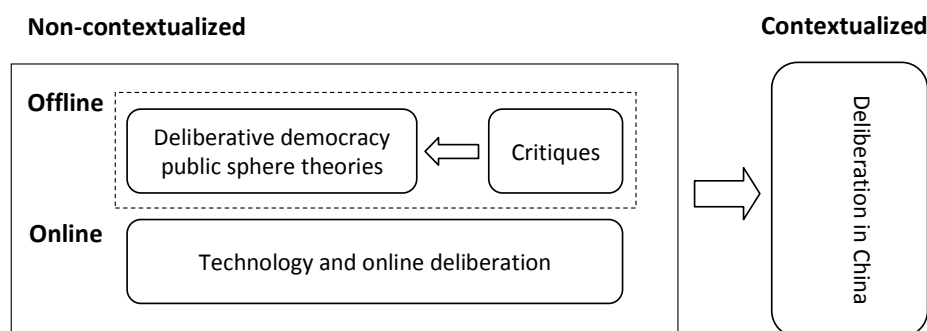


Figure 5. Literature about deliberation

To sum up, this section reviewed literature in four topics: 1) theories about deliberation, 2) critiques on them, 3) technology and online deliberation, and 4) deliberation in China. The relationship of the four topics is illustrated in Figure 5. The first two topics justified that deliberation is indeed important to democracy and should be at the core of political studies, despite the legitimate critiques on the deliberation model. Then the discussion of technology pointed out that technology and the Internet setting would affect deliberation, and thus studies on online deliberation are important as well. Finally I reviewed deliberation studies particularly targeting to the Chinese context. I pointed out that future research should address the gap between the study of deliberation and other research agendas in China studies. The challenges of this section can be summarized as follows:

- Theoretical debate requires empirical evidence (the gap in the deliberation theories literature).
- Online deliberation is different from face-to-face deliberation, and this needs to be addressed in future research.
- Better technical design is needed to improve online deliberation, particularly in helping people achieve common understanding of basic political concepts
- Censorship would affect online deliberation, and this needs to be addressed in future research.
- The study of online deliberation in China can be related to other research agenda, in particular, to the study of measuring the meaning of political concepts.

In this prelim study, I'm particularly interested in finding empirical evidence to understand the meaning of political concepts in Chinese online deliberation, and I believe that it will be a contribution to the field. In the next two sections, I will first discuss literature about measuring deliberation in general in Section 4, and then about measuring meaning in Section 5.

4. Deliberation Measurement

The previous two sections reviewed literature in China studies and online deliberation, where I made two arguments:

1. Political concepts in the Chinese context have different meanings, and finding empirical evidence of their meanings in online deliberation would contribute to the field.
2. Studies about deliberation in China need to take into account other research agendas, and discovering the underlying meaning of basic political ideas such as “democracy”, “rights” and “rule of law” would be interesting, but was missing in current literature.

In short, the two theoretical sections raised the question of measuring the meaning of political concepts in Chinese online deliberation, and the key is to find empirical evidence to fill the gap in theoretical literature.

In this section, however, I'll survey literature in the broad area of deliberation measurement. My goals are: 1) to show that these measurements are important for empirical study of deliberation in China, and 2) to argue that, in order to apply these measurements to the study of Chinese online deliberation, it is necessary, or at least beneficial, to take into account the meaning measurement of various political concepts, which was not systematically studied in the literature.

In the rest of this section, I'll first discuss the frameworks of measuring the general quality of deliberation, and then proceed to discuss measurements of six specific aspects of deliberation (measuring the meaning of concepts is deferred to Section 5). The measurement problem can be divided into two smaller questions: 1) what to measure, and 2) how to measure. I'll address these questions in the discussion.

Frameworks on measuring the quality of deliberation

One cluster of literature supports the usage of the Habermas public sphere theory (Habermas, 1989) as theoretical background. This line of work first addressed the question of “what to measure” by identifying the ideal criteria of deliberation, and then discussed operationalization of “how to measure”. The common theme was to measure the general quality of deliberation against the ideal criteria.

One notable work is (Janssen & Kies, 2004) that summarized seven previous works (Dahlberg, 2002; Graham & Witschge, 2003; Jankowski & Van Os, 2004; Jensen, 2003; Schneider, 1997; Steenbergen et al., 2003; Wilhelm, 1999). It proposed to measure six ideal criteria of deliberation based on (Dahlberg, 2001): 1) autonomy from state and economic power; 2) exchange and critique of criticizable moral-practical validity claims; 3) reflexivity; 4) ideal role-taking, 5) sincerity, 6) discursive inclusion and equality. It mainly used three methods: 1) content analysis with human coders, 2) discourse analysis (qualitative meta-reading and case studies), 3) surveys and interviews. The summary of (Janssen et al, 2004) is listed in Figure 5.

1. **Themmatization and reasoned critique of validity claims**
 - a. Reciprocity
 - i. Content analysis with a 'reply-count'
 - ii. Visualization of 'web of reciprocity'
 - b. Justification
 - i. Content analysis based on argumentation theory
 - ii. Survey > Questions on how participants evaluate the 'rational character' of the debate: did they think other people used good arguments in defending their positions? Questions on information gathering behaviour: did participant use information to 'back up' claims? What kind of information?
2. **Reflexivity**
 - a. Content analysis based on argumentation theory (with an evaluation of the content of the argumentation)
 - b. Meta reading of 'reflexivity': when does it become the subject of conversation
 - c. Survey > Questions on participant's own reflexivity: where they ever convinced by other people's arguments? Did they have to change their initial opinions? Questions on the reflexivity of other participants: did your arguments ever convince anyone? How 'reflexive' was the conversation in general: did positions shift or not?
3. **Ideal role taking**
 - a. Respectful listening
 - i. Metareading of 'respect'
 - b. Ongoing
 - i. The 'one-timer effect' as a general statistic
 - ii. ??
4. **Sincerity**
 - a. Meta reading of 'sincerity'
 - b. Survey > Questions on the perceived sincerity of other participants: did you ever think participants were being insincere? Questions on personal sincerity: have you ever been insincere?
5. **Inclusion and discursive equality**
 - a. Inclusion
 - i. Internet penetration/usage rates of country
 - ii. Survey > Demographics (compare with country usage rates)
 - b. Discursive equality
 - i. '%participants-%contributions' statistic
 - ii. Meta reading of 'equality'
 - iii. Survey > Questions concerning behaviour of other participants: did they treat you as an equal? Questions concerning personal behaviour: have you treated others equal?
6. **Autonomy of state and economic power**
 - a. Case study report
 - b. Meta reading on 'autonomy'

Figure 6. Summary of deliberation measurement framework by (Janssen et al, 2004)¹²

Another later work (Stromer-Galley, 2007) proposed to measure six other ideal criteria in political deliberation: 1) reasoned opinion expression, 2) references to external sources when articulating opinions, 3) expressions of disagreement and hence exposure to diverse perspectives, 4) equal levels of participation during the deliberation, 5) coherence with regard to the structure and topic of deliberation, and 6) engagement among participants. This work also relies on content analysis and human coders as the operationalization approach. The emphasis of this work is to develop a valid and reliable coding scheme for human coders.

¹² The “??” in 3.b.ii is indeed in the paper. I assume it means survey and participatory observation according the text of the paper.

There are three common problems of this line of literature on deliberation measurement. First, their goal was too broad and ambitious and thus lacked practical value. They tried to define a few ideal criteria and then test the quality of deliberation according to those criteria. But how did they justify the validity of the selected criteria? How could they tell, say, whether “disagreement” should be included in “reflexivity” but not “openness”?

Second, their operationalization approaches were all ad-hoc to some extent, and it was hard to evaluate their validity. Also, content analysis based on human coding was not scalable to large text corpuses, and thus was limited in practice. They failed to consider computational feasibility for automated measurement.

Finally, all those deliberation measurements were based on Habermas public sphere theory, which assumed shared understanding or a common good as their goal, as criticized by (Young, 1996) and others. That required, at least partially, common understanding of political concepts such as “democracy”. That assumption might be satisfied in the Western context, but certainly not in the Chinese context. Therefore, measuring the meaning of political concepts should also be integrated in such frameworks, especially when they are applied in the Chinese context.

Measuring specific aspects of deliberation

The deliberation measurement frameworks discussed above try to measure the general quality of deliberation. The other line of work only measures some specific aspects of deliberation without trying to establish the overall ideal criteria for deliberation. Thus it can't measure the overall quality of deliberation, but it might have more practical value. I'll very briefly review six measurements.

Measuring politeness, rudeness and hostility. This cluster of literature studied how to measure politeness, rudeness and hostility in deliberation. For example, (Spertus, 1997) recognized several heuristics that suggesting hostility, and then applied machine learning techniques to detect hostile messages. (Abbasi, 2007) applied a probabilistic disambiguation technique to measure the usage of violence and hatred in extremist groups. (Burke & Kraut, 2008) measured politeness using human coders guided by the politeness theory by (Brown & Levinson, 1987), and showed that politeness is positively related to desired outcomes in online communities such as increased reply rates.

Measuring sentiment. This type of work applied sentiment analysis techniques and measured positive or negative sentiment in text messages. (Pang et al., 2002; Turney, 2002) were among the earliest to propose how to measure positive or negative messages using machine learning techniques such as SVM. (Bautin et al., 2008) extended sentiment analysis technique to nine different languages, and showed that their “entity sentiment scores” were consistent across nine languages. In particular, (van Atteveldt et al., 2008) studied the positive and negative relations between political figures and actions in the Dutch newspaper.

Measuring political position. This type of work tried to assess party affiliation and political position on specific issues from text analysis. For example, (Laver et al., 2003; Martin & Vanberg, 2008) proposed a simple algorithm to classify articles into conservative or liberal

- the algorithm first associated a conservative-liberal “wordscore” to each word from labeled data, and then aggregated the “wordscore” to infer the label for each new article. This algorithm might be too simple for the technical mind, but its advantage was its intuitiveness for political scientists. For another example, (Gamon et al., 2008) used a combination of text analysis, link analysis and a technique called “emotional charge” that automatically classified political blogs into conservative, liberal, and independent.

Measuring polarization and diversity. (Adamic & Glance, 2005) applied social network analysis and studied political blogs before the presidential election in 2004, and found that liberal blogs tended to link to liberal blogs, and conservatives to conservative blogs. On the contrary, (Kelly et al., 2005) found that different camps of people indeed intensively interacted with each other in USENET. (Wu & Huberman, 2008) used simple statistics and studied reviews on IMDB and Amazon, and argued that they found no group polarization. (Munson, Zhou, & Resnick, 2009) proposed three measurements of diversity: inclusion, alienation, and proportional representation, and found that the proposed “sidelines” algorithm indeed improved diversity in news aggregators.

Measuring buzzwords and topics. This type of work measured what people were talking about. For example, (Leskovec et al., 2009) discussed tracking “memes”, popular quotes in news, and found the “heart-beat” curve of news cycles. (Qiaozhu Mei & Zhai, 2005) generated themes from blogs using the CPLSA method, and studied the themes evolution using HMM method.

Measuring opinion change. Different from other measurements that directly measure certain aspects of online deliberation, this type of work measured the indirect outcome of deliberation. For example, (Price & Cappella, 2002) designed a controlled group experiment and found that online deliberation could result in participant opinion change, which was measured by a self reported survey.

This brief survey of measuring deliberation in six aspects is not intended to be comprehensive; rather, it is to show some examples of deliberation measurement in different directions. Measuring the meaning of political concepts as they appear in deliberation is yet another direction, but it was somewhat neglected in current literature relevant to China studies.

I’d also like to argue that such measurements, when applied to Chinese online deliberation, would become more informative if we also interpret the meaning behind various political concepts. For example, imagine that a study showed that Chinese people have positive sentiment to “democracy”, then it would be better to learn whether the sentiment is towards liberal democracy or the “democracy” promoted by the government.

Summary and challenges

Table 3. Summary of deliberation measurement

	References	What to measure	How to measure?	Require measure meaning?
Deliberation measurement frameworks	(Schneider, 1997)	equality, diversity, reciprocity and quality	simple statistics	Yes
	(Jensen, 2003)	form, dialogue, openness, tone, argumentation, reciprocity	simple statistics, content analysis, surveys	Yes
	(Jankowski & Van Os, 2004)	reciprocity, reflexivity, ideal role taking	N/A	Yes
	(Wilhelm, 1999)	topography, topicality, inclusiveness, design and deliberation	N/A	Yes
	(Steenbergen, Bächtiger, Spörndli, & Steiner, 2003)	participation, level of justification, content of justification, respect, and constructive politics (DQI)	content analysis, human coding	Yes
	(Graham & Witschge, 2003)	process of understanding (consisting of rational-critical debate, reciprocity, reflexivity and empathy), sincerity, equality and freedom	content analysis, human coding	Yes
	(Janssen & Kies, 2004)	reciprocity, justification, reflexivity, ideal role taking (respectful listening, ongoing), sincerity, inclusion and discursive equality, autonomy.	content analysis, human coding, discourse analysis (meta-reading), surveys, interviews	Yes
	(Stromer-Galley, 2007)	reasoned opinion expression, sourcing, disagreement, equality, topic, engagement	content analysis, human coding	Yes
Specific measurement	(Abbasi, 2007; Burke & Kraut, 2008; Spertus, 1997)	politeness, rudeness, hostility	sentiment analysis, machine learning	Opt.
	(Bautin, Vijayarenu, & Skiena, 2008; Godbole, Srinivasaiah, & Skiena, 2007; Pang, Lee, & Vaithyanathan, 2002; Turney, 2002; van Atteveldt, Kleinnijenhuis, Ruigrok, & Schlobach, 2008)	sentiment	sentiment analysis, machine learning	Opt.
	(Gamon et al., 2008; Laver, Benoit, & Garry, 2003; Martin & Vanberg, 2008)	political position	machine learning	Opt.
	(Adamic & Glance, 2005; Kelly, Fisher, & Smith, 2005; Wu & Huberman, 2008)	polarization, diversity	social network analysis	Opt.
	(Leskovec, Backstrom, & Kleinberg, 2009; Qiaozhu Mei & Zhai, 2005)	buzzwords, topics	Meme-tracking, text mining	Opt.
	(Price & Cappella, 2002)	opinion change	controlled group experiment, panel design	Opt.

Table 3 summarizes the literature relevant to deliberation measurement. In order to apply these measurements to the study of Chinese online deliberation, it is necessary, or at least

beneficial, to take into account the meaning measurement of various political concepts. Therefore, it is important to study how to measure the meaning of political concepts in Chinese online deliberation.

This survey of literature is not intended to be comprehensive; rather, my goal here is to sample existing works and identify the most appropriate direction to measure online deliberation in the Chinese context. The direction I choose for this prelim is to measure a specific aspect of online deliberation – semantic meaning of basic political concepts. All measurements discussed in this section are also open for my future research, but are not the focus of this prelim. In the next section, I'll review literature about measuring semantic meaning.

5. Measuring Meaning of Political Concepts in Online Deliberation

Compared to other measurements such as politeness or sentiments that can be represented as numeric scores or nominal labels, the “meaning” measurement can have multiple representations in different situations. I’ll briefly introduce my proposed representations here, and then discuss them in detail in the rest of the Section 5 and 6.

First, the meaning of a political concept can be measured by a vector of semantic similarity to other concepts. For example, the meaning of “democracy” can be represented as $SV_{\text{democracy}} = [0.4 \text{ election}, 0.5 \text{ human rights}, 0.1 \text{ class}, -0.5 \text{ one party system}]$ in the liberal sense, and $[0.1 \text{ election}, 0 \text{ human rights}, 0.4 \text{ class}, 0.5 \text{ one party system}]$ in the socialist sense. I’d like to call this representation *semantic vector* (SV), and if we normalize it so that $\|SV\| = 1$, I’ll call it *semantic vector – normalized* (SV-N). To transform the vector into a random variable distribution, where all dimensions are positive and add up to 1), I’ll call it *semantic distribution* (SD); a commonly used approach was discussed in (Coccaro & Jurafsky, 1998). Both semantic vector and semantic distribution are used in Section 6.1, 6.2 and 6.3

Second, we can categorize the elements in the semantic distribution and collapse them into a few categories. For example, if we categorize “election” and “human rights” as LIBERAL, and “class” and “one party system” as SOCIALISM, then the meaning of “democracy” can be represented as $[0.9 \text{ LIBERAL}, 0.1 \text{ SOCIALISM}]$ and $[0.1 \text{ LIBERAL}, 0.9 \text{ SOCIALISM}]$ respectively. I’d like to call it *semantic distribution – categorized* (SD-C) An example of SD-C is given in Section 6.2.

Third, we can measure the difference of two semantic distributions using the KL divergence score. I’ll call it *semantic distribution - divergence* (SD-D). We can then learn the intra-corpus consistency as measured by SD-D for the same corpus over a period of time. An example is given in Section 6.2. Or we can measure the inter-corpus divergence using SD-D for different corpuses of different deliberation settings. Examples are found in Section 6.1 and Section 6.2

In the rest of the section, Section 5.1 discusses the motivation. Section 5.2 discusses theories from computational linguistics and cognitive psychology that support the computational approaches. Section 5.3 discusses the operationalization of three computational approaches.

5.1. Motivation

Responding to theoretical challenges

Political concepts usually have rich meanings, and the meanings are different in different contexts. As discussed in Section 2, political concepts such as “democracy”, “rights”, “freedom” and so on all have a multitude of meanings. Empirical evidence to help understand the precise meaning of those political concepts in the Chinese context has significant value, I’d argue, for theoretical research in political science, legal studies, and social sciences in general.

First, such evidence can respond to the challenges in China studies discussed in Section 2. For example, critics argued that the “Asian values” argument was mainly from government propaganda, which was not legitimate to represent the people (Bauer & Bell, 1999). Understanding the meaning of important political concepts in the Chinese online deliberation will help scholars understand the people’s opinions of “Asian values”. For another example, literature argue that China’s economic development might lead to democracy (Lipset, 1959), or slow down democratization (Gallagher, 2002). Learning people’s conception on “democracy” could shed light on that debate by showing people’s mental readiness for liberal democracy.

Second, this study can respond to the challenges in online deliberation discussed in Section 3. For example, communication theories argue that a common ground is necessary for good communication results (Clark, 1996); understanding the Chinese people’s conception of political terms can shed light on whether they have the common ground to carry out effective deliberation. Also, we can study how different technological designs in online deliberation could change people’s conception, which will be discussed in Section 6.1 and 6.2.

Finally, this research can facilitate scholarly, as well as political, communication between different groups of people, particularly between China and the US. Learning the mental model of what is intended by “democracy”, “human rights” or “state capacity” from the other side could avoid misunderstanding. Furthermore, by making salient the latent semantic meaning of concepts (e.g. displaying them in a block in relevant discussions), it will have tremendous civic education value to the people.

Why not traditional methods

“Traditional methods” here refer to the popular methods used in social sciences, such as content discourse analysis, qualitative meta-reading, case studies, surveys and interviews. These are also the main methods used in Chinese online deliberation studies. In contrast, this prelim proposes to use computational approaches to measure semantic meaning. One might ask why not just use the established traditional methods?

The main advantage of computational approaches is that, in addition to measurement, the same technique can be applied directly to improve online deliberation, in the sense of “technology as control”. For example, the semantic distribution for “democracy” can be shown to the users during deliberation and help them understand the meaning of the concept, and perhaps help the users navigate to other diverse discussion threads related to the concept. In contrast, traditional methods can only fulfill the observing functionality.

In terms of measurement, computational approaches have many advantages as well. The first is that they can automatically process large corpuses of textual data without excessive human labor. Second, compared to surveys and interviews, computational approaches can take advantage of the richness of existing web data, saving the efforts of collecting data. Third, the outcome can be objective, reducing biases introduced by human coders. Finally, rather than sampling a small portion of data, they can directly process the population data and thus avoiding sampling confound.

However, the first disadvantage of computational approaches is that it loses large amounts of interesting information that is only accessible through qualitative study. Second, it might not be as intuitive as the traditional methods, and might have resistance in the audience. Third, there are still unsolved computational problems, and thus we might not obtain all the answers we want.

The best way, then, is to combine computational approaches together with traditional methods. Computational approaches can provide the “big picture” of the data from massive text processing, and then we can use traditional methods such as content analysis and surveys/interviews to provide in-depth, qualitative study. I’ll give an example of combining traditional methods with computation approaches in Section 6.3.

5.2. Theoretical Background

This section briefly summarizes theories about texts and meanings in computational linguistics (Biber, Conrad, & Reppen, 2004; Bolshakov & Gelbukh, 2004; Callow, 1998; K. M. Carley, 1997; Cheng, Fleischmann, Wang, & Oard, 2008; Orpin, 2005), artificial intelligence (Fellbaum, 1998; Sowa & Borgida, 1991), cognitive psychology (Griffiths, Steyvers, & Tenenbaum, 2007), and content analysis (Holsti, 1969; Krippendorff, 2004; Popping, 2000; Roberts, 1997).

The exact definitions of “texts”, “words”, “concepts” and “meanings” are discussed in the areas of linguistics (Callow, 1998), cognitive psychology (Griffiths et al., 2007), and philosophy (C. K. Ogden, Richards, Malinowski, & Crookshank, 1949), which are out of the scope of this paper. Generally speaking, I’ll define “texts” and “words” as tangible, symbolic representation of languages, “concepts” as mental image of “words” (in this paper, I’ll use the term “words” and “concepts” interchangeably unless special notice), “meanings” as structures in the human brain which people experience as ideas and thoughts (Bolshakov & Gelbukh, 2004).

Apparently, words have different meanings in different contexts, and the same meaning might be represented in different words. The many-many relationship between text and meaning is shown in Figure 7. Figure 8 shows the two level representations, where text is at the surface level and meaning is at the deep level. Figure 9 shows language functions like encoder/decoder in the communication channel to transfer meaning through text.

The online setting might be different, because we might not be able to infer people’s mental model from the text they submitted (e.g., people copy/paste text that doesn’t represent their mental model). However, for the rest of the discussion, I’ll assume the deliberation texts to be analyzed match the mental model of the authors and recipients of the texts. That is, it is valid to infer what people think and conceptualize from what they write and talk.

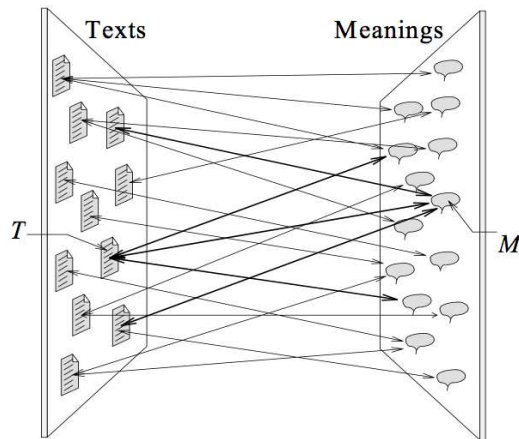


Figure 7. Text-meaning many-many relations (Bolshakov et al, 2004)

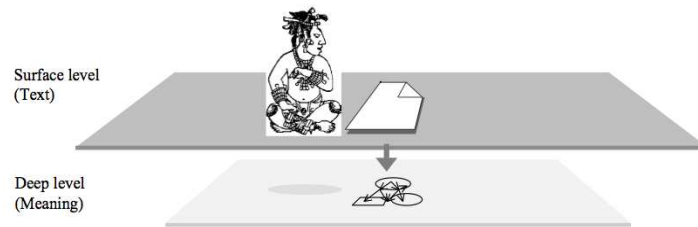


Figure 8. Two levels of representation (Bolshakov et al, 2004)

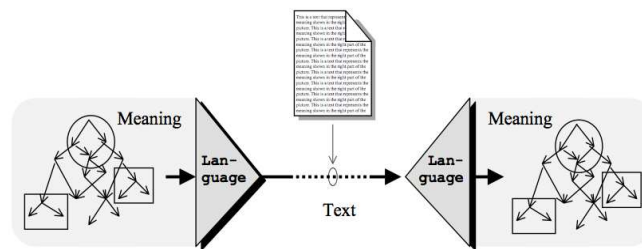


Figure 9. Language function as encoder/decoder (Bolshakov et al, 2004)

Constructing semantic meaning from text

How do concrete words represent abstract meanings? One line of theory is that words have “innate” meanings, and can be “deduced” from certain rules (Callow, 1998; Stubbs, 2006). An example is the “semes” theory, which is discussed in the next paragraph. The other line of theory is that the meanings of words can be empirically “induced” or “constructed” from corpus of text (Callow, 1998; Stubbs, 2006). Following the “inductive” line, there are three other types of theoretical work, “metaphors”, “topics” and “relations”, which are discussed below respectively.

Semes. The first type of theory attempts to find “a universal representation via decomposition and even atomization of the meaning of several semantic components”(Bolshakov & Gelbukh, 2004). The atomic elements of meaning are usually called semantic primitives, or *semes*. The meaning of a word can then be represented by semes. To use the example in (Bolshakov & Gelbukh, 2004), “matar” in Spanish means “to kill”, and it can be decomposed into MATAR(x)->CAUSAR(CESAR(VIVIR(x))), where CAUSAR (to cause), CESAR (to stop) and VIVIR (life) are all elementary semes. The problem with this theory is that there is no universal agreement on what is the elementary set of semes. Besides, it is hard to design computational approaches to infer semes from texts.

Metaphors. The second type of theory basically argues that semantic meaning of a word is embedded in the metaphors:

Metaphor is for most people a device of the poetic imagination and the rhetorical flourish— a matter of extraordinary rather than ordinary language. Moreover, metaphor is typically viewed as characteristic of language alone, a matter of words rather than thought or action. For this reason, most people think they can get along perfectly well without metaphor. We have found, on the contrary, that metaphor is pervasive in everyday life, not just in language but in thought and action. Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature.(Lakoff & Johnson, 1980)

For example, the metaphor "argument is war" can be found in everyday language, e.g., “your claims are *indefensible*”, “he *attacked* every weak point in my argument”, “his criticisms were *right on target*”, “I’ve never *won* an argument with him. You disagree? Okay *shoot!*”, “If you use that *strategy*, he’ll *wipe you out*. He *shot down* all of my arguments”. Then, the meaning of “argument” can be explained by “war”. (Neuman & Nave, 2009) discussed the operationalization of measuring meaning based on the metaphor theory through human coding, but it didn’t propose a computational procedure.

Topics. The third type of theory focuses on disambiguating meanings for a word in different topics (Blei et al., 2003; Griffiths et al., 2007; Hofmann, 1999). For example, the word “bank”, when referenced in financial topics, means the financial institute; but when it’s referenced in river-related topics, then it means the slope alongside the river. This theory assumes that texts are associated with a limited number of predefined meanings, and the task is to disambiguate the meaning of a word according to its topic.

Relations. The last type of theoretical work argues that the semantic meaning of a word is defined by its relations to other words in the same unit of analysis such as document or n-word window (K. Carley, 1986; K. Carley & Palmquist, 1992; Krippendorff, 2004). For example, from the snippet “dogs are animals that can bark”, we can define the meaning of “dog” as its IS-KIND relation with “animal” and CAN relation with “bark”. This is extensively used in the knowledge representation literature such as (Sowa & Borgida, 1991). Another example is to use the simple co-occurrence relationship between words. For example, suppose the word "democracy" co-occurs frequently with "human rights", "election" and "freedom", which follows the liberal tradition, then its meaning is different from the Marxist definition of "democracy" which co-occurs frequently with "classes", "state" and "capitalism".

Table 4. Theories of constructing meaning from text

Ideas	Arguments	References
Semes	Meaning can be “deduced” from logic using the semantic primitive “semes”	(Bolshakov & Gelbukh, 2004)
Metaphors	Meaning can be constructed from metaphors	(Bolshakov & Gelbukh, 2004; Lakoff & Johnson, 1980)
Topics	Meaning is disambiguated according to the topics	(Blei et al., 2003; Griffiths et al., 2007; Hofmann, 1999)
Relations	Meaning can be defined by a concept’s relation to other concepts; the relations could either have semantic info, or simply the co-occurrence	(K. Carley, 1986; K. Carley & Palmquist, 1992; Krippendorff, 2004; Sowa & Borgida, 1991)

Co-occurrence assumption and simplification

The “relations” theory is widely used in many applications, especially in the semantic networks literature (Lehmann, 1992; Sowa & Borgida, 1991). This prelim paper also uses the “relations” theory as the theoretical base, but with some additional assumptions and simplification.

Here I assume that the semantic meaning of a word is simply defined by its mere co-occurrence relations with other words, regardless of the semantic nature of the relationship. That is, if two words co-occur in the same unit of analysis such as document, paragraph or the 50-word window, then they are considered to be related, and the semantic information in their relation is ignored.

This simple co-occurrence assumption has some problems. By ignoring the semantic information in word relations, we can’t learn if co-occurrence is due to true semantic relation, or just due to statistical co-incidence. For example, the words “cat” and “dog” co-occur frequently, but apparently they don’t share the same meaning. Some argue that this assumption tends to find word relevance but not semantic meaning (Griffiths et al., 2007). In addition, I would argue that this assumption tends to find the extension of a concept, not the intension. Despite those shortcomings, the simplification was still proved worthy in many practices in content analysis (K. Carley & Palmquist, 1992), because we can learn a lot from co-occurrences in massive text processing. Without this simplification, computational approaches would be hard, if possible at all. More discussion on the assumptions can be found in (K. Carley & Palmquist, 1992; Popping, 2000).

Furthermore, the co-occurrence simplification is also based on the key assumption that the ‘bag-of-words’ or vector space model of documents will preserve most of the relevant information in many cases. More discussion of this “bag-of-words” assumption can be found at (Manning et al., 2008).

Comparison to WordNet

The most similar work was perhaps WordNet (Fellbaum, 1998). The goal of WordNet was also to construct the semantic meaning of words from their relations to other words. Here I'll discuss the difference of this work to WordNet.

First, WordNet was built on top of the general purpose text corpus, called the Brown corpus (Fellbaum, 1998). It represented everyday usage of English, and its goal was to define the meaning of words in a general purpose dictionary for language use, not for use in specific political context. On the other hand, this paper proposes to apply computational approaches to different Chinese online deliberation text corpuses; the goal is to find the meaning of a few selected political concepts in the context of political deliberation.

Second, in WordNet, the semantic relations among words were manually coded by lexicographers as PART-WHOLE, KIND-OF, IS-A, and so on. This labor-intensive labeling is impossible for automatic computation. On the other hand, this paper adopts the simplified assumption and only considers the simple co-occurrence relation, which permits computational approaches.

5.3. Computational Approaches

Based on the theories and assumptions discussed above, here I'll review three classes of computational approaches that construct semantic meaning from a text corpus.

Word association

The first class of approaches is the simple statistical count of word co-occurrences. One example is Key Word In Context (KWIC) that simply counts keywords in the context, and shows a keyword and the other words co-occurring near to it for qualitative analysis (Adolphs, 2006). Another approach, mutual information (MI) (Church & Hanks, 1990), is discussed in details here. Mutual information is defined as:

$$PMI(V_1, V_2) = \log_2 \frac{P(V_1, V_2)}{P(V_1)P(V_2)}$$

In the formula, V_1 and V_2 are two words; $P(V_1, V_2)$ is the probability that V_1 and V_2 co-occur in the corpus; $P(V_1)$ is the probability V_1 occurs; $P(V_2)$ is the probability V_2 occurs.

If there's a genuine association between V_1 and V_2 , then $P(V_1, V_2)$ would be greater than $P(V_1)P(V_2)$, thus $MI(V_1, V_2) > 0$; if there's no interesting information, then $MI(V_1, V_2) \approx 0$; if V_1 and V_2 follow a complementary distribution, then $MI(V_1, V_2) < 0$. The semantic vector of a word V_i can be defined as $SV(V_i) = [PMI(V_i, V_1), PMI(V_i, V_2), \dots]$.

In fact, the formula above is also known as point-wise mutual information (PMI), and the "full" mutual information is defined in the following formula, which means to add up all four cases of whether or not V_1 and V_2 co-occur (Manning et al., 2008).

$$MI(V_1, V_2) = \sum_{e_t \in (0,1)} \sum_{e_t \in (0,1)} \log_2 \frac{P(V_1 = e_t, V_2 = e_t)}{P(V_1 = e_t)P(V_2 = e_t)}$$

Mutual information has been shown useful for lexicography applications and identifying semantic classes (Church & Hanks, 1990). However, its shortcoming is that the semantic vector is computed pair-wise between words, and thus implicitly assumes that word associations are independent from one another, which might not be true in reality. As a result, we can only get the pair-wise relationship among words, not any global properties of the model such as the centrality of certain words (compared to the network approach).

Concepts network

The basic idea is to construct a network on top of the underlying text corpus, where the network nodes are the words and the network edges are the relations between words. This approach is also called “semantic network”, “concepts map” or “map analysis”, and was widely used for extracting meanings from text (K. Carley & Palmquist, 1992). For example (K. Carley, 1986) studied a small group of students at MIT and their notion of writing changes over a semester of technical writing class. (Steyvers & Tenenbaum, 2005) studied different ways of constructing the network and showed that they were scale-free networks. I use the term “concepts network” to emphasize the co-occurrence simplification which is different from the semantic network literature. An example of concepts network is shown in Figure 9.

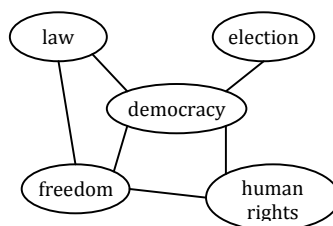


Figure 10. A simple concepts network

(K. Carley & Palmquist, 1992) formulized four steps to construct a network from underlying text. The first step was “identifying concepts” as the network nodes. Considerations include level of analysis – what constitutes a concept, irrelevant information, predefined or interactive concept choice, level of generalization (whether to collapse certain words), creation of translation rules (synonyms, etc), level of implication for concepts, existence or frequency, and number of concepts (K. Carley, 1993). Here, each “concept” was a cluster of words meaning the same thing. This was identified by human coders using domain knowledge¹³.

The second step was “defining relationships” as the network edges. This step requires defining 1) strength, 2) sign, 3) directionality, and 4) meaning to the edges. “Strength” denotes how directly two concepts are related. “Sign” denotes whether the two concepts are positively related or negatively related. “Directionality” denotes whether the relationship is from node A to node B, vice versa, or both. “Meaning” denotes the semantic

¹³ Prof. Mei suggested using the micro-clustering methodology proposed in (Lin & Pantel, 2002) to collapse words into concept. However, in the scope of this paper, I’ll just take the human coders.

information such as KINDOF or IS-A. According to our simplification, I'll only consider "strength", which is calculated by the frequency of co-occurrence.

The third step was "extracting statement". Statement here is defined as two words and their relationship. This step is to retrieve the important relations that are of interests to the research question, and neglect the rest. And the fourth step was "displaying and analyzing maps", where the researchers can visualize and analyze the network.

The semantic vector of the word V_i is just defined as $SV(V_i) = [E(V_i, V_1), E(V_i, V_2), \dots]$, where $E(V_i, V_j) \in [0, \infty)$ is the strength of the edge between V_i and V_j , measured by the frequency of co-occurrence of V_i and V_j . The semantic distribution of V_i can be defined as:

$$E(V_i) = \sum_{j \in N} (E(V_i, V_j))$$

$$SD(V_i) = \left[\frac{E(V_i, V_1)}{E(V_i)}, \frac{E(V_i, V_2)}{E(V_i)}, \dots \right]$$

The problem with the concepts network approach was that it doesn't take into account iterative definitions of latent semantics. For example, "democracy" could be universally interpreted as "election" and "majority wins" in any text corpus. However, in the liberal corpus, "election" could be interpreted as direct election, and "majority wins" as wins by the majority of people; but in the authoritarian context, "election" would be interpreted as election within the government, and "majority wins" as the majority of those who are in power. In such a case, we won't be able to capture the different subtle meaning of "democracy" without iteratively looking at the indirectly associated concepts and the latent semantics. To address this problem, I propose a new approach "concepts network with random walk", which will be discussed in Section 6.1.

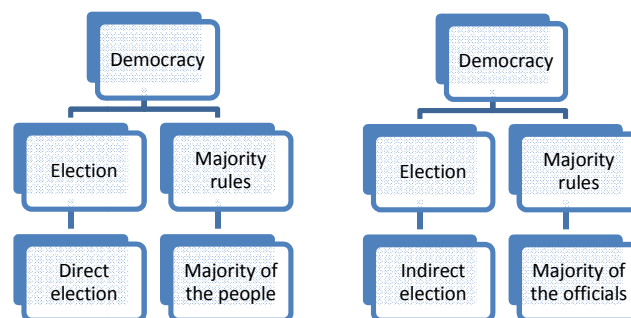


Figure 11. Iterative definition

Semantic space

The basic idea is to represent words/concepts as vectors in the vector space (it is different from semantic vector (SV) discussed earlier in that the vectors here are from the word-document matrix, whereas SV are from the word-word matrix). And the relationship among concepts would then be reduced to measuring the distance or angles between the

vectors. Some of the approaches includes latent semantic analysis (LSA) (Deerwester et al., 1990), probabilistic latent semantic analysis (PLSA) (Hofmann, 1999), latent dirichlet allocation (LDA) (Blei et al., 2003), context probabilistic latent semantic analysis (CPLSA) (Qiaozhu Mei & Zhai, 2006), etc. I'll discuss LSA first.

The first step of LSA was to construct a word-document matrix M , where each cell value $M_{ij}=f(V_i, D_j)$ which is the frequency of word V_i in document D_j . One optimization is to set $M_{ij}=\log(f(V_i, D_j))$ in order to scale the high frequency words. Another optimization is to set $M_{ij}=f(V_i, D_j)/\text{entropy}(V_i)$ in order to underplay the common words. The next step is to perform the singular vector decomposition (SVD) on M :

$$M = UDV'$$

Then we have the words matrix UD , with rows as the vectors for the words. An optional step is to perform a dimension reduction to remove the not-so-important dimensions in D that have low diagonal values.

Finally, to calculate the relations among words, we just compute the cosine similarity between the word vectors in UD using the formula below. For a word V_i , its semantic vector would be $SV(V_i) = [R(V_i, V_1), R(V_i, V_2), \dots]$, where $R(V_i, V_j)$ is:

$$R(V_i, V_j) = \cos(UD_i, UD_j) = \frac{UD_i \cdot UD_j}{|UD_i||UD_j|}$$

The PLSA model (Hofmann, 1999) assumes there are n latent topics ($z_1, z_2, \dots, z_i, \dots, z_n$) in the corpus (where n is given), and each document and each word has probability conditional on z . The intuition of PLSA is shown in Figure 12 as the "generative model" – given a document d in probability $P(d)$, we can generate a topic z using $P(z|d)$, and then generate a word w using $P(w|z)$; we've already known $P(d,w)$ as the probability of word w in document d , and the goal is to find the distribution of z_i using maximum likelihood estimation so that this process is most likely.

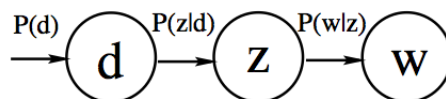


Figure 12. The "generative model" of PLSA, (Hofmann, 1999)

Formally, we have:

$$P(d, w) = P(d)P(w|d) = P(d) \sum_{z \in Z} P(w|z)P(z|d)$$

Assuming d and w are independent conditioned on z , and applying Bayesian theorem, we have:

$$P(d, w) = \sum_{z \in Z} P(d)P(z|d)P(w|z) = \sum_{z \in Z} P(z)P(d|z)P(w|z)$$

We've already known $P(d,w)$ from the word-document matrix, and then we'll apply the standard Expectation Maximization approach (EM) (Dempster, Laird, & Rubin, 1977) approach to estimate $P(z)$, $P(d|z)$ and $P(w|z)$. The EM approach is a form of maximum likelihood estimation, and is described in (Hofmann, 1999), copied below. By applying Bayesian theorem, we can also get $P(z|w)$ and $P(z|d)$ based on the estimation of $P(z)$, $P(d|z)$ and $P(w|z)$. We can then use $P(z|w)$ as the vector and calculate cosine distance between words as the semantic vector.

$$P(z|d, w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z' \in \mathcal{Z}} P(z')P(d|z')P(w|z')}$$

$$P(w|z) \propto \sum_{d \in \mathcal{D}} n(d, w)P(z|d, w),$$

$$P(d|z) \propto \sum_{w \in \mathcal{W}} n(d, w)P(z|d, w),$$

$$P(z) \propto \sum_{d \in \mathcal{D}} \sum_{w \in \mathcal{W}} n(d, w)P(z|d, w).$$

PLSA is different from LSA in several ways. First, PLSA is based on solid probability theories, while LSA uses SVD that does not have solid theories on the application domain. Second, the results of PLSA are statistical distributions, while SVD are vectors that might be negative, and hard to interpret. Lastly, PLSA is proven to generate better results than LSA in many domains such as information retrieval. However, LSA can generate the exact results while PLSA is estimation; besides, LSA is more intuitive than PLSA and might be more readily accepted in the social sciences community.

The last approach I will introduce here is the Latent Dirichlet Allocation (LDA) (Blei et al., 2003). This model is shown to be a generalized version of PLSA that uses Dirichlet distribution for the latent topics, and PLSA is equivalent to LDA under a uniform Dirichlet prior distribution (Girolami & Kaban, 2003). For that reason, I'm not going to discuss LDA in detail here.

The advantages of LDA over PLSA is that: 1) it fixes the overfitting problem of PLSA, that the parameters of PLSA grow linearly as document grows; 2) it can be applied to unobserved data, i.e., you can first apply it on the training set, and then apply it on the testing set, while PLSA is unsupervised and each time you have unobserved data you have to calculate the whole model again. However, I'd argue that PLSA already suffices the purpose of measuring semantic distribution.

LSA, PLSA and LDA are the three classical models in this area. There are at least 20+ papers in the KDD/CIKM community that improve the three models. I won't review them here because this prelim is not to design advanced algorithms in this area.

Although the semantic space approach was proved to be a good alternative in many areas such as IR, it is not as intuitive as the concepts network approach and it can't be visualized as a graph. Thus, it might have limited usage in the social sciences. Also, none of the LSA/PLSA/LDA approach goes beyond calculating point-wise similarity between words, although the word vectors themselves are iteratively constructed. Finally, all of the

methods assume the coarse level of unit of analysis, i.e. document or paragraph, and can't use n-word window.

Summary and challenges

Table 5 summarizes the features of the three classes of semantic representation model on a text corpus. Computationally, the three classes can be easily transformed into one another. For example, using the pair-wise relation between words in MI or LSA, we can construct the concepts network; given a network, it is easy to map it to a vector space and then apply LSA technique. The distinction of the classes is mainly conceptual rather than instrumental.

Figure 13 shows the map of literature reviewed in this section. First I reviewed four theories of measuring semantic meaning, “semes”, “metaphors”, “topics”, and “relations”. And then to the “relations” theory, I made the “co-occurrence” simplification. And finally, based on the simplification, I reviewed three computational approaches.

Table 5. Comparison between three computational approaches

	Word Association	Concepts Network	Semantic Space
Algorithms	KWIC, PMI, MI, etc	Varied	LSA, PLSA, LDA, etc.
References	(Church & Hanks, 1990; Manning, Raghavan, & Schtze, 2008)	(K. Carley, 1986, 1993; K. Carley & Palmquist, 1992; Steyvers & Tenenbaum, 2005)	(Blei, Ng, & Jordan, 2003; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Hofmann, 1999)
Pair-wise words relations (assume independence)	Yes	Yes/No*	Yes/No+
Global properties (centrality, etc)	No	Yes	No
Visualization	No	Yes	No
Iterative definition	No	No <Gap>	Yes
Unit of analysis (support n-word window)	Yes	Yes	No
Parameters optimization	Ad-hoc <Gap>	Ad-hoc <Gap>	Ad-hoc <Gap>

*. “Yes” without random walk, “No” with random walk

+. Even though the word relations are calculated pair-wise (the reason for “Yes”), the vector of each word is calculated without the assumption for independence (the reason for “No”)

The first gap in the literature is that the concepts network approach doesn't consider the iterative definition problem, as discussed earlier. I propose to use the Random Walk approach to address it, which will be discussed in Section 6.1.

The second gap, also the major problem of all three approaches – “word association”, “concepts network” and “semantic space” – is that they all require arbitrary coding choices. For example, researchers have to make ad-hoc decisions on questions such as how to decide the n-word window size, which words to be discarded, what is the optimal number

of hidden topics in the PLSA model, and so on. Different settings might result in quite different models, e.g., (De Choudhury, Hofman, Mason, & Watts, 2010) argued that the arbitrary choice of cut-off threshold in a network would result in very different predictions results. Naturally, the research context would provide cues to make the coding choices, but it is not always obvious to make the best choice in all cases. To address this problem, I'll propose an evaluation framework that has some objective and systematic guidelines to make the arbitrary parameter optimization choices, which will be discussed in Section 6.1.

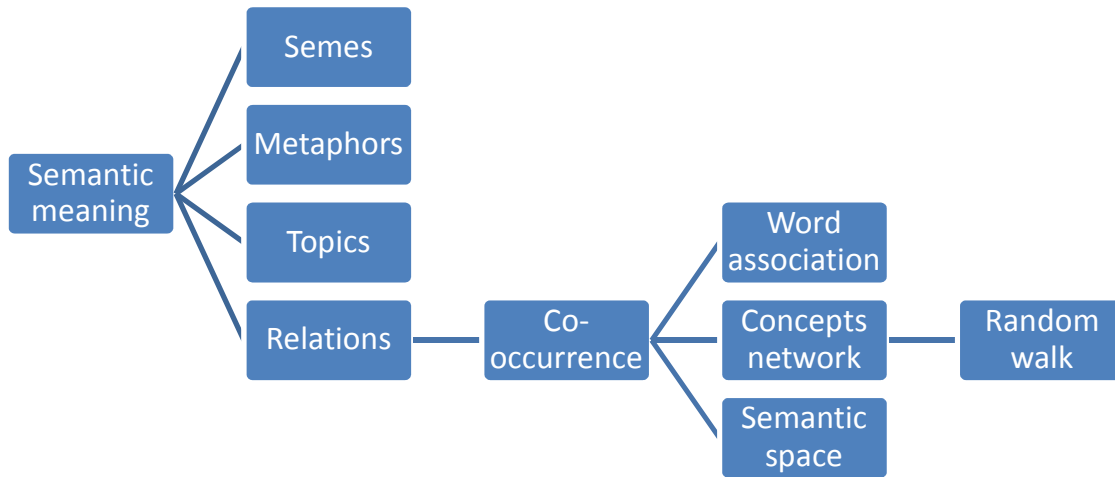


Figure 13. Structure of literature about measuring meaning

6. Future Research

Future research to fill the gap in the literature can be both technical and substantive. The technical contributions are 1) to propose an evaluation framework to optimize constructing semantic representation model from text corpus, and 2) to propose the algorithm of constructing concepts network with random walk. This is discussed in Section 6.1 with the targeted audience of ICWSM conference.

In the substantive area, future research can measure the semantic meaning of such political concepts as “democracy”, “human rights”, “rule of law”, and “government” from Chinese online deliberation. Examples are given in Section 6.2 and 6.3 with the targeted audience of *Perspectives on Political Science* journal and *International China Internet Research* conference.

6.1. *Measuring the Meaning of "Democracy" in the Chinese Context*¹⁴

This paper applies text analysis techniques and tries to measure different meanings of "democracy" in 4 different contexts: 1) the Chinese official propaganda, 2) heavily-regulated Chinese online political discussion, 3) lightly-regulated Chinese online political discussion, and 4) US political blogs¹⁵.

The paper's contribution is twofold. The technical contribution is to propose an evaluation framework to find optimized ways of constructing a semantic representation model from a text corpus. In addition, the paper proposes a novel way of constructing concepts network with random walk. We try to answer these questions:

- Do different semantic models with different parameters lead to different results?
- Is the proposed evaluation framework valid?
- Do complex models generate better results than simple models?
- Does the concepts network with random walk approach generate better results?

The substantive contribution is to use computation techniques to provide empirical evidence for social science research. Since the paper is mainly for technical readers, we prefer to be less rigorous from a social science perspective, and the goal is to present interesting findings even though the results are preliminary. In particular, we try to answer these substantive questions:

- What is the meaning of “democracy” in Chinese deliberation?
- Is the meaning different in different contexts? To what extent?
- What are the possible reasons for potential democratic reform in China?
- How, if ever, will democratization happen in China, and what kind?

¹⁴ This section is based on my SI 755 final paper. Most data in the “Empirical Study” and “Results and Discussions” sections are fake in order to illustrate the point. The tone of this part is as if I have done the research. I plan to submit it to ICWSM 2011.

¹⁵ Comparing to the US might be more interesting to the ICWSM readers, even though it might introduce the “cultural difference” confound.

Concepts Network with Random Walk

Of the three computational approaches (word association, concepts network, and semantic space), the concepts network approach has many advantages over the other two approaches: 1) it is intuitive and easy for visualization, and 2) it can reveal global properties such as network centrality. However, one disadvantage was that none of the existing network models discussed how to handle the iterative definition problem, i.e., latent semantic problem. For example, the concept “democracy” can mean the same “election and majority rules” in both American and Chinese contexts, but “election” and “majority rules” mean different things respectively, as discussed in Section 5.3

We extend the concepts network approach by adding an extra step of “random walk”, as described in (Haveliwala, Kamvar, & Jeh, 2003; Lovasz, 1993; Manning et al., 2008; Page, Brin, Motwani, & Winograd, 1998). Formally, we have a network $G = \langle V, E \rangle$. V_i is defined as a vector $[v_{i1}, v_{i2}, \dots, v_{in}]$. We normalize V_i so that $\sum v_i = 1$, which is to make V_i a random variable for the Markov chain. Then we can represent G as an adjacency matrix $P = |V| \times |V|$. Let E be another Markov chain matrix for teleportation. Then we have the random walk algorithm as:

$$A = [\alpha P + (1 - \alpha)E]^T$$
$$r = Ar$$

After the iteration converges, the vector r would have a score for each node. When E follows the uniform distribution, that is, teleportation to any random node, the algorithm is just the PageRank algorithm (Page et al., 1998), and r is a vector of PageRank scores. When E is configured so that teleportation always goes back to the same node V_i , then the algorithm is called personalized PageRank in (Haveliwala et al., 2003; Page et al., 1998), or Random Walk with Restart in (Konstas, Stathopoulos, & Jose, 2009; Lovasz, 1993).

Here, I propose to set teleportation E to always jump back to the node V_i (political concept) that we are interested in, and calculate the personalized PageRank score r_i . And r_i just happens to be the semantic distribution of V_i . We think the extra step of random walk will take into account latent semantic information, and thus make the model more accurate.

We won't discuss more sophisticated random walk approaches such as hitting time and absorbing random walk discussed in (Q. Mei, Zhou, & Church, 2008). The reason is because this paper is not an advanced algorithm paper, and our intention is to use a relatively popular algorithm to attract more social science readers. More advanced techniques will be explored in future research.

In the related work (Mihalcea, Tarau, & Figa, 2004), the authors proposed to apply PageRank-style algorithm to a semantic network, but it was different compared to the approach here. First, their network was built on top of WordNet in order to give a PageRank score for each word sense, not on top of the target text corpus they studied. Their goal was to achieve disambiguation based on predefined word sense, not for constructing meaning for text corpuses. Second, they applied PageRank but not

personalized PageRank, and their approach would achieve optimal results only when coupled together with the Lesk algorithm.

Evaluation Framework

The previous sections listed several semantic representation models, and each model has different coding choices, and each coding choice might lead to completely different results, as shown in (De Choudhury et al., 2010). This section, then, is to propose a framework to evaluate different models and optimize the parameters for the purpose of measuring meaning.

We formalize the evaluation problem as the following: We have a text corpus \mathbb{C} , which consists of documents D_1, D_2, \dots, D_n . Each D_i can be represented by a bag of words $W_i = (V_1, V_2, \dots, V_m)$. We denote $\mathbb{V} = \cup W_i$ to be the set of all words for \mathbb{C} , and \mathbb{V}_{key} to be the subset words we are interested to find their semantic vectors. We can then construct a semantic representation model $G(\mathbb{C}) = g(\mathbb{V}', \mathbb{E})$, where $\mathbb{V}' \subseteq \mathbb{V}$, and \mathbb{E} denotes the relations among \mathbb{V}' . More specifically, \mathbb{E} defines V_i as a vector of $(v_{i1}, v_{i2}, \dots, v_{i|\mathbb{V}'|})$, where v_{ij} is the strength between V_i and V_j . Our goal is to evaluate all the candidate models G_i , and find the optimal model G' with the optimal parameters to represent \mathbb{C} .

Stability, Sensitivity and Elasticity

Naturally, the best evaluation metrics for G' should be model accuracy, i.e., G' should accurately represent \mathbb{C} . However, it is impossible, by the nature of the problem, to have the “golden standard” to measure accuracy, because our goal here is to constructively measure the meaning of complex concepts, and we assume humans have subjective interpretation of these concepts. Thus, we can only design second-best metrics that provide some objective insights of the quality of model G . We suggest that the optimal G' should have high stability, high sensitivity, and high elasticity.

Stability means that if two text corpuses \mathbb{C}_1 and \mathbb{C}_2 are similar, then $G(\mathbb{C}_1)$ and $G(\mathbb{C}_2)$ should also be similar. That is, G' should *not* fluctuate too much if the underlying text corpus only has some small variations.

Sensitivity means that if two text corpuses \mathbb{C}_1 and \mathbb{C}_2 are not similar, then $G(\mathbb{C}_1)$ and $G(\mathbb{C}_2)$ should be quite different. That is, G' should be sensitive enough to capture significant semantic variance of the underlying text.

Not surprisingly, stability and sensitivity is a pair of tradeoff. If G is too stable, then it will ignore lots of information and won't be able to capture all the interesting semantic relations. An extreme example is to use constant semantic distributions for all concepts in \mathbb{V}_{key} for any corpus, ignoring any variation – in this case, stability would be the highest, but it's not useful either. On the other hand, if G is too sensitive, then the model is overfitting and it will respond to small variance and become too erratic – it is not useful either. Therefore, we define **elasticity** as the quotient of stability and sensitivity, which has a high score only when G is stable for similar text and sensitive for dissimilar text.

We use Kullback–Leibler divergence as the basis for measuring stability, sensitivity and elasticity for $G(\mathbf{C})$. Specifically, suppose \mathbf{C}_a and \mathbf{C}_b are both subsets of \mathbf{C} , and we have the semantic distribution of a concept V_i so that $\sum v_{ij}=1$. Then:

$$\text{Divergence}(G, \mathbf{C}_a, \mathbf{C}_b) = \sum_{V \in V_{\text{key}}} \text{KL}(V_{\mathbf{C}_a} | V_{\mathbf{C}_b})$$

$$\text{KL}(V_{\mathbf{C}_a} | V_{\mathbf{C}_b}) = \sum_{v_j} V_{\mathbf{C}_a}(v_j) \log \frac{V_{\mathbf{C}_a}(v_j)}{V_{\mathbf{C}_b}(v_j)}$$

Repeated Sampling

At time $t=0$, we randomly select 10% of the documents from corpus \mathbf{C} , denoted as $\mathbf{C}_{10\%}$, and divide them randomly into two subsets \mathbf{C}_a and \mathbf{C}_b . We assume \mathbf{C}_a and \mathbf{C}_b are relatively similar because they are from the same corpus and because of their randomness. From the same $\mathbf{C}_{10\%}$, we use an unsupervised classification algorithm such as SVD and divide the 10% of the documents again into another two subsets, \mathbf{C}'_a and \mathbf{C}'_b , so that \mathbf{C}'_a and \mathbf{C}'_b are most different (according to the highest dimension in SVD). Then we have the following score at time $t=0$:

$$\text{Stability}_0 = \text{Divergence}(G, \mathbf{C}_a, \mathbf{C}_b)$$

$$\text{Sensitivity}_0 = \text{Divergence}(G, \mathbf{C}'_a, \mathbf{C}'_b)$$

$$\text{Elasticity}_0 = \text{Sensitivity}_0 / \text{Stability}_0$$

We repeat this process 100 times, and then we have:

$$\text{Stability}(G, \mathbf{C}) = \frac{\sum_{t \in [0, 100)} \text{Stability}_t}{100}$$

$$\text{Sensitivity}(G, \mathbf{C}) = \frac{\sum_{t \in [0, 100)} \text{Sensitivity}_t}{100}$$

$$\text{Elasticity}(G, \mathbf{C}) = \frac{\sum_{t \in [0, 100)} \text{Elasticity}_t}{100}$$

We apply this process to all possible semantic representation models, and the optimal model $G'(\mathbf{C})$ should have low numeric stability score (which means high stability), high sensitivity score, and high elasticity score.

Note that the optimal model under this process might have dependency on the choice of the unsupervised classification algorithm. To fix this, one might employ human coders to classify the sampled documents into two dissimilar subsets for each of the 100 rounds. Or, one has to carefully choose a classification algorithm that's least likely to introduce dependency to the alternative models G .

Alternative Repeated Sampling Process

From the corpus \mathbb{C} , we randomly select 10% of the documents and divide them randomly into two subsets \mathbb{C}_a and \mathbb{C}_b . Then we use Language Style Match (LSM) (Gonzales, Hancock, & Pennebaker, 2009) to calculate $LSM(\mathbb{C}_a, \mathbb{C}_b)$ using non-substantive words. $LSM(\mathbb{C}_a, \mathbb{C}_b)$ is small if \mathbb{C}_a and \mathbb{C}_b are similar, and vice versa.

Then we repeat the process 100 times, and get $LSM_t(\mathbb{C}_a, \mathbb{C}_b)$, $t \in [0, 100]$. We used the 50 pairs of $(\mathbb{C}_a, \mathbb{C}_b)$ that have lower LSM scores to calculate the stability score, and the other 50 pairs of $(\mathbb{C}_a, \mathbb{C}_b)$ that have higher LSM scores to calculate sensitivity scores; the elasticity score would be the sensitivity score divided by the stability score.

The key difference of this alternative repeated sampling process, compared to the previous one, is that this alternative has an ordinal measurement, the similarity score, for the 100 $(\mathbb{C}_a, \mathbb{C}_b)$ pairs. The major problem of this alternative process is that we are not confident of the validity of LSM as a metrics for text corpus similarity. But we don't have other metrics that would be superior either: cosine similarity might be biased in favor of semantic space models. LSM is desirable because it only counts non-substantive words that are isolated from semantic meaning models.

Discussions

First, one might ask why we need this evaluation framework. It is needed only when different models generate different semantic representations as the outcome. If that is the case, how do we choose which model to use? If different models generate the same results, which I think is not likely, then the evaluation framework is not useful because we can use any model and still get the same results.

Then, the next question is whether the metrics of stability, sensitivity and elasticity are able to differentiate different models, given that different models generate different results. It is likely that none of the models has a sensitivity score higher than the stability score, so that the elasticity scores are not distinguishable for all alternative models; then none of the models can stand out from the evaluation framework. If this is the case, then I'd suggest the preference goes to high stability, because high stability captures the most important relations in the text corpus.

The final question concerns the validity of stability, sensitivity, and elasticity as a measure of model quality. Ultimately, model accuracy is the most important measure we care about. It's likely that due to some systematic bias, high stability, sensitivity, and elasticity could mean low accuracy. To avoid this, I'd suggest a sanity check after selecting the best model from the evaluation framework by reading a sample of the text qualitatively. We have to find a qualitative explanation for the semantic distribution we obtain. Otherwise we have to change the selected model. Given that we are not able to measure "accuracy" due to the lack of "golden standard", I'd argue that stability, sensitivity, and elasticity are still reasonable metrics.

Empirical Study

Our goal is to study the semantic meaning of “democracy” in the Chinese context, and we want to do compare Chinese social media and official media, and use US social media as comparison baseline¹⁶. We did the empirical study on 4 text corpuses:

- People's Daily (abbr. Daily): This is the official newspaper by the Chinese Communist Party (CCP), representing the conceptualization of the Chinese state.
- RisingChina forum (abbr. Rising): This is a very popular online political discussion forum sponsored by People's Daily, where the Prime Minister of China frequents. It is highly regulated, officially endorsed, and represents the mental model of Chinese political elites.
- Tianya forum (abbr. Tianya): This is one of the most popular generic-purpose Chinese forums. I'll only use the News board for political discussion purpose. It is less regulated than RisingChina, and represents average Chinese netizens.
- US political blogs (abbr. US): This represents the conceptualization of politically active Americans.

The following table listed some descriptive statistics of the 4 corpus:

Table 6. Statistics of text corpuses

	Daily	Rising	Tianya	US
Time period	2001 - 2009	2003 -2009	2003 - 2009	2007 - 2009
Language	Chinese	Chinese	Chinese	English
Nature	Official media	Social media w/ heavy regulation	Social media w/ light regulation	Social media
Text size	4.7G	7.8G	3.2G	10.1G
documents	160K	100K	96K	1M
docs/day	25	30	45	800
comments/ doc	0	20	25	10
words/doc	500	300	100	200
Data loss ratio ¹⁷	1.3%	3%	2.4%	N/A

We have 4 alternative semantic representation models: mutual information (MI), latent semantic analysis (LSA), concepts network (CN), and concepts network with random walk (CNRW). For each model, we also need to make some arbitrary coding choices. Particularly, we need to choose whether we'll use all substantive words (nouns, verbs, adjectives), or only important nouns (nouns with high $tf*idf$); also, whether two words were considered as co-occurrence in the same document, or in the 50-words window. Thus for each model, we have 4 variations, as coded below:

¹⁶ Using the US example as baseline might introduce the cultural difference confound. However, this might be more interesting to technical readers than using Taiwan/HongKong examples. We prefer interestingness than rigorous to this audience.

¹⁷ Due to network connection failure or other unexpected problems

	Words	All	Important nouns
Co-occur			
Document		a	c
50-words window		b	d

Then we have $4 \times 4 = 16$ alternative semantic representation models for each of the 4 text corpus. Our first step was to select the best model for each corpus using the evaluation framework proposed earlier. We define \mathbb{V}_{key} to have only one word “democracy”. That is, we optimize the model to be most accurate only for the word “democracy”. Our second step was to apply the optimized model for each corpus and interpret the results.

The Chinese language is different from English in that Chinese needs a “words analyzer” to group characters into words. We used ICTCLAS¹⁸ library to do that.

Results and Discussions

Optimal model. Evaluation results for the 4 datasets using the 16 alternative models are listed in Table 7. In each cell, the three scores are stability, sensitivity and elasticity. It can be shown that sensitivity scores are usually higher than the stability scores. We highlighted the lowest stability score, highest sensitivity score and highest elasticity score.

For People's Daily, the optimal alternative was MI-d because stability was higher (lower score), sensitivity was higher, and elasticity was higher. For RisingChina, it was LSA-c for the same reason. For Tianya, the 16 alternatives are not that different, so we select CNRW-b because it has other nice features (such as visualization). For the US, we used CNRW-a for the same reasons. The selected alternatives were highlighted in the table.

Surprisingly, the method MI, despite its simplicity, was the best alternative for People's Daily. That means, for some text corpuses, MI could still be better than other more complex methods. In addition, for all the 4 datasets, CNRW was constantly better than CN in terms of lower stability score, higher sensitivity score, and higher elasticity score. That means our proposed idea of concepts network with random walk is indeed useful.

Table 7. Models comparison (stability, sensitivity, and elasticity)

	MI				LSA				CN				CNRW				
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	
Daily	0.23	0.23	0.23	0.21*	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
	0.34	0.34	0.34	0.35	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
	1.12	1.12	1.12	1.13	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12
Rising	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12
Tianya	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12
US	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12

* t-test, $p < 0.05$.

¹⁸ <http://www.ictclas.org/>

Substantive results. Then, we constructed the semantic representation for all the datasets using the optimal model. Figure 14 shows the meanings of "democracy" in People's Daily, followed by Table 8 listing the top-5 related words from the 4 datasets.

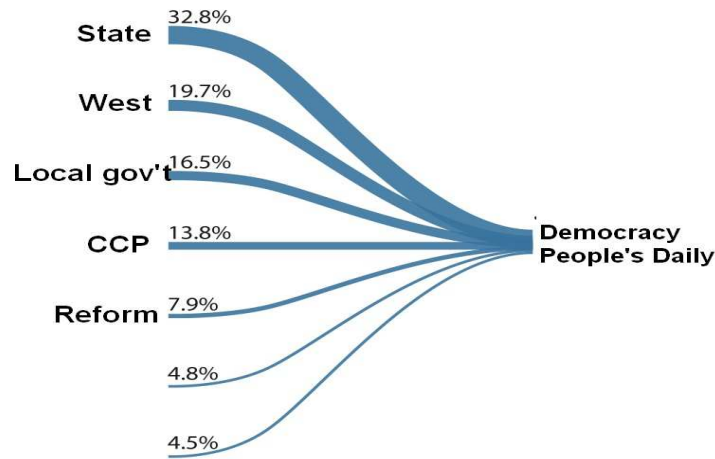


Figure 14. Meaning of democracy

Table 8. Meaning of democracy

Daily	state, West, CCP, local government, reform
Rising	West, local government, Human rights, Corruption, Election
Tianya	Human rights, Corruption, Election, Freedom, Equality
US	Freedom, Human rights, Competition, campaign

Interpretation. In order to interpret the results qualitatively, we randomly read some texts from the corpuses mentioning the word "democracy". (In each finding, we listed in parenthesis the meaning words and their rankings in relevancy to "democracy".)

In People's Daily, it was emphasized that China was already a democratic state, but different from the Western-style democracy (state 1st, West 2nd). The potential democratic reform would have to be in a top-down fashion, led by CCP (3rd) and implemented by the local government (4th).

In RisingChina, Western democracy was also seen as different from Chinese democracy (West 1st), but still, this online discussion was influenced more by the Western notion of "human rights" (3rd) and "election" (5th). Also, the reason for democracy was because it was considered as one of the most important weapons against corruption (4th).

In Tianya, "democracy" was already interpreted quite similar to the West notion (human rights 1st, election 3rd, freedom 4th, equality 5th), as shown in the US political blogs.

Meaning difference. In order to compare the different interpretations of “democracy” in the 4 text corpuses, we also ranked the top 20 meaning words of “democracy” for each corpus, and computed Spearman ranking coefficient pair-wise, as shown in Table 9. A higher score means more similarity between the two corpuses in terms of the interpreting “democracy”.

Table 9. Ranking difference

	Daily	Rising	Tianya
Rising	0.5		
Tianya	0.2	0.6	
US	0.05	0.55	0.7

We can see from the table that the difference between RisingChina and People’s Daily is smaller than that between Tianya and People’s Daily, which is consistent with common sense that regulated social media is influenced by official media. However, the difference between the two social media, RisingChina and Tianya, is smaller than their difference from People’s Daily. That means social media, even though regulated, are still different from official media. Furthermore, both of social media are closer to the US political blogs than to the official media, which means Chinese people’s conception of democracy is perhaps closer to the US than to the official propaganda. Finally, we can see that the conception of People’s Daily is almost irrelevant to the US conception.

Post study. At the end of the study, we applied all the 16 alternative models to the 4 text corpuses. The results from People’s Daily were shown in Table 10. It looks quite obvious that different semantic representation models would construct different meanings of “democracy”. Even though in many cases the meaning words were the same, the order was quite different. Therefore, it is indeed important to find the optimal model for better results. And we think the evaluation framework would lead to the optimal model.

Table 10. Results from all models

MI	a	state, West, CCP, local gov’t, reform
	b	West, local gov’t, economics dev., CCP, state
	c	CCP, state, local gov’t, people, socialist
	d	reform, CCP, election, state, West
LSA	a	West, local gov’t, economics dev., CCP, state
	b	reform, CCP, election, state, West
	c	state, West, CCP, local gov’t, reform
	d	CCP, state, local gov’t, people, socialist
CN	a	CCP, state, local gov’t, people, socialist
	b	West, local gov’t, economics dev., CCP, state
	c	state, West, CCP, local gov’t, reform
	d	reform, CCP, election, state, West
WV	a	state, West, CCP, local gov’t, reform
	b	reform, CCP, election, state, West

○	CCP, state, local gov't, people, socialist
Ⓜ	West, local gov't, economics dev., CCP, state

6.2. Chinese Concepts of “Rights”: Empirical Evidence from Political Deliberation¹⁹

This research is similar to the previous one in methodology, but different in substantive topics. Two reasons make it a separate research proposal: First, the “democracy” paper has a strong emphasis on technological contribution, and its targeted audience is the technical community. This paper focuses more on the substantive topic, and has more theoretical and qualitative research; also, the targeted audience of this paper is the community of human rights scholars and political scientists. Second, “human rights” is a quite sensitive topic in China, and this paper analyzes text corpuses from dissidents groups, which runs the risk of getting banned in China. Separating this “rights” paper at least prevents the “democracy” paper from getting into political conflict.

Research Questions

In previous sections, I reviewed literature, particularly (E. J. Perry, 2008), and showed that the Chinese conception of rights is more about welfare/economic rights, while the Anglo-American conception is more about civil rights. I also pointed out the gap in current literature of lacking empirical evidence. This paper attempts to find empirical evidence to test Perry’s argument. Specifically, I have three research questions.

RQ1 (SES difference): Do different social groups in China – including 1) workers class, 2) middle class, 3) students/intellectuals, 4) government officials, and 5) political dissidents – have different conceptions of “rights”?

RQ2 (cultural difference): Does empirical evidence support that the Chinese conception of rights is more about welfare/economic rights or more about civil rights, compared to the Western conception?

RQ3 (temporal difference): Is there a recognizable temporal pattern that shows the civil rights interpretation of “rights” becomes stronger over time?

Research Design

I will use the methodology discussed in Section 6.1 to measure the semantic meaning of “rights” in different text corpuses. In order to make the methodology more intuitive for the social science audience, I’ll just use the concepts network approach with random walk (given that we confirm that this approach is better than the other approaches mentioned in

¹⁹ To illustrate my idea, I make fake data in the “Results and Discussions” section. I plan to work with Prof. Gallagher and Ya-Wen Lei and submit a paper to *Perspectives on Politics*, where (E. J. Perry, 2008) was published.

Section 6.1). For example, the conception of “rights” could be explained by “subsistence”, “job” and “economic development” because they co-occur frequently in the text corpus. The same approach will apply to all text corpuses in order to avoid confusion.

To tell whether the conception of “rights” is explained more as welfare/economic rights or civil rights, I’ll use the categorized semantic distribution (SD-C) as the semantic representation. First, I’ll code related concepts to “rights” as *W* (stands for welfare rights, such as “food”, “subsistence”, “livelihood”, “housing”, “economics”, etc), *C* (stands for civil rights, such as “justice”, “morale”, “liberty”, “legitimate”, etc) or *N* (stands for neutral or not recognizable). The coding scheme will be decided according to (E. J. Perry, 2008) as well as to the definition in “Civil and political human rights”²⁰ and “Economic, social and cultural human rights”²¹ in Wikipedia. Two coders will be trained and code the words appeared in the text corpuses as *W*, *C* or *N*.

Then, once I have the conception of “human rights” as explained by a list of related words coded as *W*, *C* and *N*, I’ll collapse them into a (*W*, *C*, *N*) score for “rights” in that corpus as the categorized semantic distribution. For example, the conception of “rights” in the workers class corpus could be explained by (60% *W*, 30% *C* and 10% *N*), that is, “rights” in that corpus can be explained by 60% welfare/economics rights, 30% civil rights, and 10% neutral or not recognizable.

To answer RQ1, I’ll study five text corpuses, and compare the different categorized semantic distribution of “rights” in the (*W*, *C*, *N*) score. The five corpuses are:

- Interview transcripts from (M. E. Gallagher, 2006), which represent the workers class.
- Online discussion from Tianya Forums, which represents the middle class.
- Online discussion from SMTH BBS, which represents the students/intellectuals class.
- Recent articles from People’s Daily, which represent the government officials.
- Recent articles from Epoch Times, Falungong’s official media, which represent the political dissidents.

To answer RQ2, I’ll compare those five Chinese text corpuses to the US counterparts, and compare their (*W*, *C*, *N*) scores to see if they are different. The five US text corpuses are:

- Interview transcripts from labor unions of the Detroit Auto industry, which represents the workers class.
- Political blogs collected by Digg, which represents the middle class.
- Blogs from University of Michigan (<http://mblog.lib.umich.edu>), which represents students/intellectuals.
- Recent articles from WhiteHouse.org, which represents the government officials.
- Interview transcripts from protesters, which represents the dissidents.

²⁰ http://en.wikipedia.org/wiki/Civil_and_political_rights

²¹ http://en.wikipedia.org/wiki/Economic,_social_and_cultural_rights

To answer RQ3, I'll analyze the five Chinese corpuses in a monthly basis. For each month in each corpus, I'll calculate the conception of "rights" as represented by (W, C, N) score. Then, I'll calculate C/W rate – higher C/W rate means "rights" is explained more by civil rights. I'll plot the monthly C/W rate for each corpus on a timeline and see if the civil rights interpretation grows stronger.

Results and Discussions

The results for the five Chinese corpuses are listed in Table 11. We can see that the workers class and the government officials all interpret "rights" mostly in the sense of welfare/economic rights; on the other hand, political dissidents, the middle class and students/intellectuals have a stronger emphasis interpreting "rights" as civil rights. Thus, for RQ1, we can argue that Perry made a biased observation mostly for the workers class and the government officials, and ignored the middle class, political dissidents and intellectuals.

Table 11. "Human rights" as explained in five Chinese text corpuses

Corpus	Top 5 explanatory concepts of "human rights"	(W, C, N)
Working class	Food, housing, sustenance, livelihood, family	90%, 0%, 10%
Middle class	Justice, democracy, livelihood, family, law	40%, 50%, 10%
Students/Intellectuals	Democracy, family, political reform, liberty, job	30%, 50%, 20%
Government officials	Economic growth, job, law, sustenance, housing	80%, 10%, 10%
Dissidents	Democracy, election, injustice, liberty, legitimacy	10%, 80%, 10%

Table 12. "Rights" as explained in five US text corpuses

Corpus	Top 5 explanatory concepts to "human rights"	(W, C, N)
Working class	Law, justice, family, job, insurance	50%, 40%, 10%
Middle class	Justice, democracy, livelihood, family, law	30%, 60%, 10%
Students/Intellectuals	Job, fairness, law, justice, family	20%, 60%, 20%
Government officials	Justice, democracy, livelihood, family, law	50%, 40%, 10%
Dissidents	Democracy, election, injustice, liberty, legitimacy	10%, 80%, 10%

Table 13. Conception of "rights": comparison between China and the US

Corpus	China (W, C, N)	US (W, C, N)
Working class	90%, 0%, 10%	50%, 40%, 10%
Middle class	40%, 50%, 10%	30%, 60%, 10%
Students/Intellectuals	30%, 50%, 20%	20%, 60%, 20%
Government officials	80%, 10%, 10%	50%, 40%, 10%
Dissidents	10%, 80%, 10%	10%, 80%, 10%

The results for the five US corpuses are listed in Table 12. The (W,C,N) scores comparison between the Chinese corpuses and the US corpuses is listed in Table 13. First, we can see that both the workers class and the government officials in the US case interpret "rights" more in the sense of welfare/economic rights compared to the other US groups, but they

still have a strong conception of civil rights compared to the Chinese case. Second, we can see that the US case constantly interpret more civil rights in all five groups. Therefore, for RQ2, we can argue that even though Perry might have ignored middle class and political dissidents in her argument, her conclusion is still valid in that the Chinese people, no matter what groups and classes they belong to, all interpret “rights” in a more welfare/economic sense than the US case.

Finally, after breaking down the Chinese corpuses into a monthly basis, we can plot the C/W rate as shown in Figure 15. The charts show that for political dissidents, the C/W rate remains constant. However, for the workers class, C/W rate grows gradually, which mean that the workers class interpret “rights” more in the sense of civil rights over time. One explanation might be that the economy growth and modernization makes people who are not informed (such as the workers class) have more civil rights “consciousness”. Thus, for RQ3, one reasonable prediction is that, even thought Perry was right that the Chinese conception of “rights” is currently phrased as welfare/economic rights, it is likely to move towards the sense of civil rights.

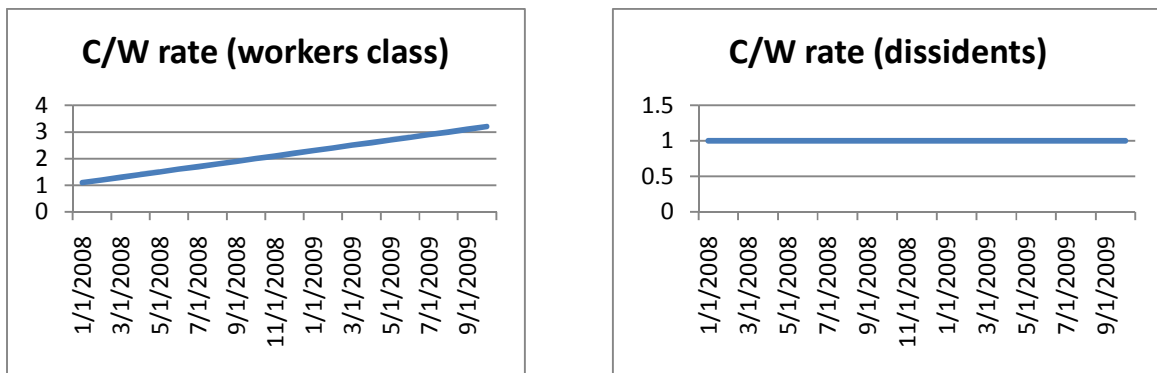


Figure 15. C/W rate on a monthly bases (Chinese corpus)

6.3. *Legal Culture in the State Media and on the Internet: the Salience of Law and the Construction of Legality in Two Venues of Political Communication*²²

This paper is different from the previous two papers in two ways. First, this paper explores the interaction of two processes – establishing rule of law and connecting people to the Internet – and tries to show that they are mutually constituted, which fills the gap between the two separate literatures. Second, this paper exploits network analysis techniques to study the centrality of “rule of law”, as well as in-depth qualitative content analysis to uncover the deeper meaning of “rule of law” from online discussion.

Introduction

We found a puzzling disencounter between two bodies of literature and between narratives in them, in which important themes and concepts have met but remained strangers. One is concerning China’s “rule of law” project; the other is regarding the rise of Chinese netizens. We suggest that bridging the two lines of literature by studying legal culture in the state media and on the Internet can help address unsettled disagreements in the two bodies of literature and deepen our understanding of two of the most important aspects of China’s modernization in our time.

Let’s proceed from China’s “rule of law” project. China is said to have initiated its legal reform around 1979 with the state’s promise and attempts to conceptualize and articulate notions of law as an objective set of rules and standards to protect rights (Lubman, 1999). It is agreed that the Chinese state adopted “rule by law” in the instrumental sense of producing rules and relevant institutions needed for a modern market economy (Goldman, 2005, p.179; Keller, 1994; Lee, 2002; Lubman, 1999; Orts, 2001). One consequence that has been often raised about this project is the expansion of legal or rights consciousness. It is suggested that people use law or rights as a weapon to strive for their interests (M. Gallagher, 2006; Goldman, 2005, p.213; Lee, 2002; O’Brien & Li, 2006). Leading figures in this literature are lawyers (Liu, 2006b; E Michelson, 2007b), judges (Liu, 2006a), law firms (Liu, 2008), parties of disputes who fight for their material interests (M. Gallagher, 2006; Lee, 2002; E Michelson, 2007a), and villagers (O’Brien & Li, 2006, p.25). Places of law encompass factories (Lee, 2002), legal aid centers (M. Gallagher, 2006), court rooms (Liebman, 2007; Liu, 2006a; E. Michelson, 2007), legal markets (Liu, 2008), and villages (O’Brien & Li, 2006). Literature collectively tells a story about the increasing importance of law in the construction and resolution of grievances. It should be noted, however, that scholars have divergent views on what types of consciousness can be inferred from these facts, how widely people have these types of consciousness, where that consciousness came from, and the consequences of this development (E. Perry, 2007, 2008).

Narratives of the rise of Chinese netizens usually begin in the early 1990s. Similar to the narratives of “rule of law,” the plots always include the explanations of why the Chinese

²² This is a collaboration research with Ya-Wen Lei from the Sociology Dept., who wrote the theoretical part of this paper. Technically this is not considered as my own work to be included in this prelim paper, but to show an example of the work derived from my prelim. The paper proposal is submitted to the 8th Chinese Internet Research Conference.

state promoted the Internet. It is suggested that the ideology of political elites—neo-technonationalism—played an important role. With a scientific mind-set, viewing information technology infrastructure as one essential part of the state building project, party elites believed the Internet could be an engine for economic and technological growth. Therefore, they decided to capitalize on the technological and commercial benefits minimizing the potential negative consequences through devices of control, such as the imposition of law and regulation (Tai, 2006, p.129; Yongnian Zheng, 2007, pp.27-8; Zhou, 2006, pp.137-8). What follow these accounts are usually themes regarding how the Chinese state and netizens have been negotiating power (Tai, 2006, ch 3; Yongnian Zheng, 2007, ch 4; Zhou, 2006, ch 7). Then episodes about how the Internet has empowered Chinese netizens as well as contributed to the emergence of public opinions, public sphere and activism are told (Tai, 2006, ch 5 & 7; Yongnian Zheng, 2007, ch 5 & 6) despite the disagreement over how “democratic” and “civil” this development is. Whereas netizens are praised for their contributions to the public, they are often criticized as being uncritical, irrational, violent, and nationalistic.

Juxtaposing the two bodies of literature reveals limitations in the literature. First, although “rule of law” is an important context in which China’s virtual space and netizens are situated, netizens and virtual world are not regarded as subject and place pertaining to law in the law and society literature. Moreover, students of Chinese Internet research tend to view law as merely an instrument for governmental control, neglecting whether law can play other roles and whether the Internet has facilitated a kind of legal culture that has gone beyond the “rule of law” script written by the state. Second, legal and rights consciousness in the Chinese context is studied in a very limited way. On the one hand, although literature indicates the existence of Chinese people who mobilize language of law and rights to fight for their private interests, we have no clue about if they have the same consciousness towards issues with public characteristics. On the other hand, literature also indicates the emergence of Chinese people who talk about public issues in the cyberspace and produce public opinions, but we don’t know through what kind of language and concepts they talk about public issues as well as form public opinions. In particular, we have no idea about whether law is a salient category in their communication, what kinds of norm (law or other norms) Chinese netizens use to evaluate right and wrong, and where these norms come from.

We argue that this gap in the literature constrains our ability to address the debates in both lines of literature. By examining cultural structures of political communication, the locations of law and other norms in these structures, the way in which law and other norms are neglected, questioned, or taken for granted, we can provide some evidence to address the debate over the political implications of the rise of Chinese netizens and the disagreements over types and sources of legal or right consciousness in the law and society literature.

Research question

To bridge this gap, we aim to understand legal culture on the Internet with reference to that promoted by the state media. We operationalize legal culture as the salience of law and the construction of legality in political communication. Following Patricia Ewick and Susan

Silbey, we refer “legality” to the meanings, sources of authority, and cultural practices that are commonly recognized as legal, regardless of who employs them or for what ends (Ewick & Silbey, 1998). We propose to address whether, how, and why the legal culture on the Internet diverges from or converges with the official version advanced by the state media.

Theoretical framework

We derive competing hypotheses from previous literature to answer the “whether” and “how” questions. On the one hand, literature on law and society leads us to conjecture that because of the consequences of the state-led rule of law project the centrality of law and the construction of legality should be similar in the two venues. According to O’Brien and Li, practices of claiming citizenship rights to demand for inclusion and policy implementation generally don’t press for wider civil and political rights nor question the legitimacy of existing laws and policies (O’Brien & Li, 2006, pp.117, 122). Therefore,

HYPOTHESIS 1: The centrality of law and the construction of legality should be similar in the two venues of political communication.

On the other hand, although they are embedded in the “rule of law project,” since that netizens are more likely to experience law’s double-edgedness in constraining and enabling their lives and that netizens are likely to encounter norms beyond the local context, we suspect that netizens’ legal and rights consciousness would be more nuanced and complex. Specifically, laws could more likely be questioned even though netizens may still recognize the necessity of law to protect individuals and maintain social order. Therefore, legality could be constructed in significantly different ways in two venues of political communication. Accordingly, we postulate:

HYPOTHESIS 2: The centrality of law should be similar in the two venues of political communication, whereas the construction of legality should be different. Specifically, law could be with more straightforward and uncontested meanings in political communication conducted by the state media. In contrast, law’s meanings are more contested and sources of authority are more diverse in the online communication.

If the analogy between netizens and mobs is true, then:

HYPOTHESIS 3: The centrality of law and the construction of legality should be different in the two venues of political communication. Specifically, law is more peripheral in the online communication.

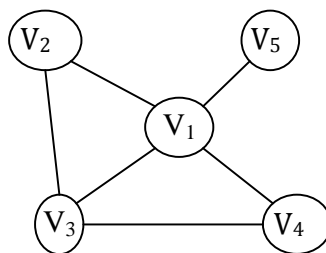
We propose a two-layered analysis, studying legal culture at both an abstract (global) level and a concrete (local) level. The purpose of a global approach is to detect the cultural symbolic structures and see how law is situated within these structures. At this level, a cultural sociology approach that draws on Saussurean linguistic tradition is an appropriate way of studying communication as it can be used to uncover the cultural structures or organizing principles in flows of communication (Alexander, 2006). We conceptualize both codes about law and other symbolic codes as the basic materials for conducting political communication. By examining how symbolic codes are configured in two venues of

political communication, we can describe and compare the centrality of codes about law and the meanings of law parsimoniously. At a local level, we study how general symbolic codes are translated into situationally specific evaluations and descriptions by examining how reasoning devices (attribution of cause, consequence, and appeal to principles) are employed in the naming-blaming-claiming process when public issues are discussed (Felstiner, Abel, & Sarat, 1980; Gamson & Lasch, 1983). Drawing on this framework, we are able to uncover according to what kind of normative standards (such as positive laws, morality, etc.) people justify their stances in the process of conducting political communication and trace where these normative principles come from.

Methodology: network analysis and qualitative content analysis

We created two text corpuses from People’s Daily and Tianya Forum. We situate our empirical study in the context of two concrete events in 2007-2008 chosen from the “Top 10 Online Stories in Year 2008”²³. Plural events allow variation in event’s characteristics. Since two events occurred around the same time, we are able to avoid confounding effects in the temporal dimension.

A quantitative method based on network analysis is able to provide a global view of legal culture in both venues and to conduct a formal mathematical test. A network is a set of “nodes” connected by “edges” (Newman, 2003). We use $G<V,E>$ to represent a network, where $V=(v_1, v_2, \dots v_n)$ denotes the set of nodes, and $E=(e_1, e_2, \dots e_n)$ denotes the set of edges, as shown in the figure below. We can also assign a value to an edge to specify the strength between the two nodes it connects. Then we can represent a node v_i as a vector $(w_{i1}, w_{i2}, \dots w_{in})$, where w_{ij} is the strength between v_i and v_j . We represent the network mathematically as a matrix M , consisting of v_i as the row vectors. A node in our study is a culture code. If codes co-occur, we assign an edge between them. In other word, we conceptualize semantic meaning of a word as its co-occurrence relations with other words in the same context.



To our purpose here, we are interested in two techniques: 1) measuring the **network centrality** of key concepts, and 2) measuring the **semantic meaning** of key concepts.

To have a contextual understanding of legal culture, we also conduct computer-assisted qualitative content analysis, in which each post in the online discussion forums and each piece of news are the unit of analysis. We read all of the articles in the news yet randomly sample 10% of the threads and read all of the posts in the selected thread. To make sure the objectivity of coding, we have two coders and test for intercoder reliability.

²³ <http://www.bjyouth.gov.cn/wlrld/243520.shtml>

Preliminary results

Figure 16 below is the concepts network constructed from People’s Daily articles related to the “Sanlu poisonous milk powder event”. The node colored red stands for the word “law”, and its related nodes are colored pink. The top words that co-occurred with "law" are society (社会), system (体系), legal documents (草案), food (食品), issue (问题), legal statute (法规), economy (经济), socialism (社会主义). It shows that the official conception of “law” is more from the state’s standpoint (society, economy, socialism, etc), not so much from people’s standpoint (rights, life, family, children, etc.)

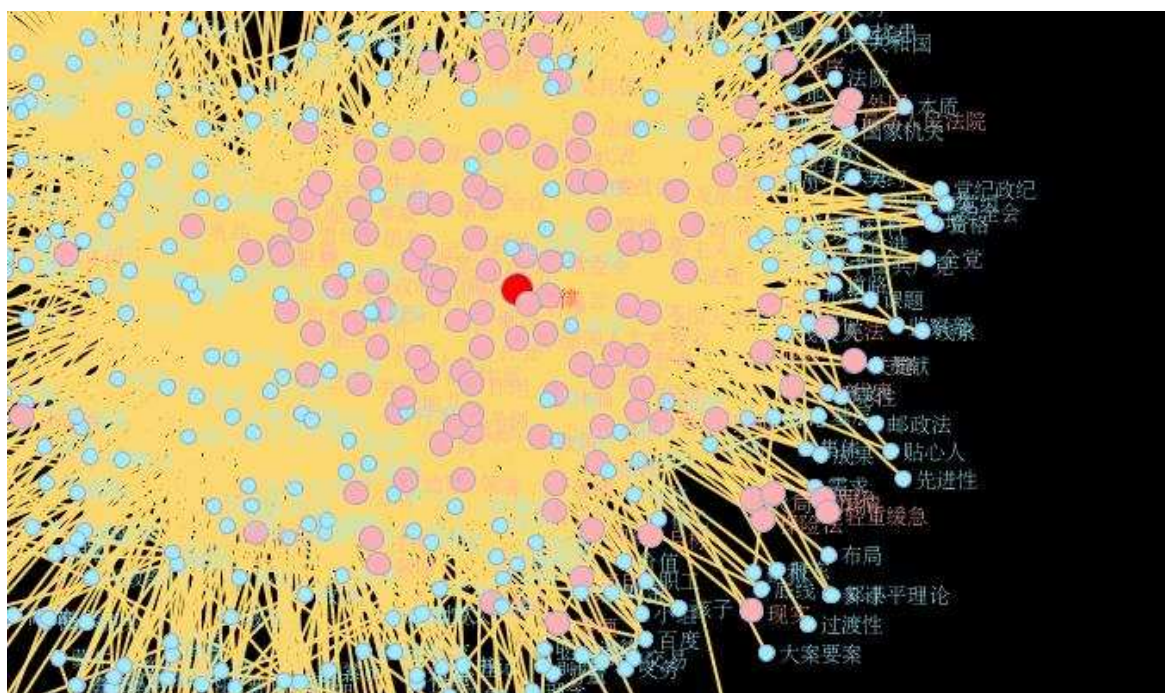


Figure 16. Local network of "law", from Sanlu event in People's Daily

Figure 17 shows the top 20 words ranked by "betweenness" centrality from the Sanlu event concepts network (synonyms have not been collapsed). Two law related words — “law” (法律) and “legality” (法) are ranked 11th and 18th respectively. Words related to government organizations, corporations, and market are highly ranked. It shows that the notion of “law” is indeed at the center of the event.

Rank	Node (Eng)	Node (Chi)	Closeness	Betweenness
1	Limited Co.	"有限公司"	0.502472415	0.247347228
2	Company	"企业"	0.579681591	0.18107263
3	Food	"食品"	0.552291523	0.118789219
4	Milk	"奶"	0.528545854	0.096760453
5	Society	"社会"	0.509291683	0.059593699
6	Issue	"问题"	0.528154338	0.057960875
7	Responsibility	"责任"	0.53090719	0.051595677
8	Industry	"业"	0.478208153	0.047938481
9	Milk powder	"奶粉"	0.523500996	0.045934642

10	Conference	"会议"	0.476928667	0.03690779
11	Law	"法律"	0.475022223	0.033837572
12	People	"人民"	0.504963425	0.031900466
13	China	"中国"	0.495144692	0.030128574
14	Quality	"质量"	0.520064447	0.02658889
15	Economy	"经济"	0.475973536	0.024846544
16	Legal cases	"案件"	0.432650702	0.020702605
17	Government	"政府"	0.482088138	0.020580483
18	Legality	"法"	0.458526274	0.01824379
19	Market	"市场"	0.470632579	0.018128681
20	The Party	"党"	0.429782011	0.016436264

Figure 17. Network centrality ranking, from Sanlu event in People's Daily

These preliminary results are very limited. More interesting results will be revealed once we compare the conceptual network using text from Tianya forum, and compare the two selected events.

Our qualitative analysis will also help explain these quantitative results. In fact, the quantitative results from the network analysis show us the "global" picture: without it, the qualitative interpretation might not be generalizable. On the other hand, the qualitative interpretation can give us a richer and deeper "local" understanding of what was indeed going on in the text. Both the "global" quantitative approach and the "local" qualitative approach are necessary for this research.

7. References

- Abbasi, A. (2007). Affect intensity analysis of dark web forums. *2007 IEEE Intelligence and Security Informatics*, 282-288.
- Adamic, L. A., & Glance, N. (2005). *The political blogosphere and the 2004 US election: divided they blog*. Paper presented at the Conference on Knowledge Discovery in Data, Proceedings of the 3rd international workshop on Link discovery.
- Adolphs, S. (2006). *Introducing electronic text analysis*: Routledge.
- Albrecht, S. (2006). Whose voice is heard in online deliberation?: A study of participation and representation in political debates on the internet. *Information, Communication & Society*, 9(1), 62-82.
- Alexander, J. C. (2006). *The civil sphere*: Oxford University Press, USA.
- Barber, B. R. (1984). *Strong democracy : participatory politics for a new age*. Berkeley: University of California Press.
- Bates, B. (2008). Online Public Sphere and Democracy in China.
- Bauer, J. R., & Bell, D. A. (1999). *The East Asian challenge for human rights*: Cambridge Univ Pr.
- Baum, R. (2007). Study of Chinese Politics in the United States. In *China Watching: Perspectives from Europe, Japan, and the United States*.
- Bautin, M., Vijayarenu, L., & Skiena, S. (2008). *International sentiment analysis for news and blogs*.
- Bell, D. A. (2000). *East meets West: human rights and democracy in East Asia*: Princeton Univ Pr.
- Bell, D. A. (2006). *Beyond liberal democracy: political thinking for an East Asian context*: Princeton Univ Pr.
- Bell, D. A. (2008a). *China's new Confucianism: politics and everyday life in a changing society*: Princeton University Press.
- Bell, D. A. (2008b). *Confucian political ethics*: Princeton Univ Pr.
- Bell, D. A., & Ham, C. (2003). *Confucianism for the modern world*: Cambridge Univ Pr.
- Beniger, J. R. (1986). *The control revolution: Technological and economic origins of the information society*: Harvard Univ Pr.
- Berlin, I. (1969). Two Concepts of Liberty. In I. Berlin (Ed.), *Four Essays on Liberty*. London: Oxford University Press.
- Biber, D., Conrad, S., & Reppen, R. (2004). *Corpus linguistics: Investigating language structure and use*: University Press.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993-1022.
- Bolshakov, I. A., & Gelbukh, A. (2004). *COMPUTATIONAL LINGUISTICS: Models, Resources, Applications*.
- Brady, A. M. (2009). Mass Persuasion as a Means of Legitimation and China's Popular Authoritarianism. *American Behavioral Scientist*, 53(3), 434.
- Brinkerhoff, D. (2008). The State and International Development Management. *Public Administration Review*, 68(6).

- Brown, P., & Levinson, S. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Burke, M., & Kraut, R. (2008). *Mind your Ps and Qs: the impact of politeness and rudeness in online communities*. Paper presented at the Proceedings of the ACM 2008 conference on Computer supported cooperative work.
- Button, M., & Ryfe, D. (2003). What Can We Learn from the Practice of Deliberative Democracy? In *The Deliberative Democracy Handbook*.
- Callow, K. (1998). *Man and message: A guide to meaning-based text analysis*: Lanham, MD/Dallas: University Press of America and Summer Institute of Linguistics. xvi.
- Carley, K. (1986). An approach for relating social structure to cognitive structure. *Journal of Mathematical Sociology*, 12(2), 137-189.
- Carley, K. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological methodology*, 23, 75-126.
- Carley, K., & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social Forces*, 70(3), 601-636.
- Carley, K. M. (1997). Network text analysis: The network position of concepts. In *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 79-100).
- Carter, I. (2008). Positive and Negative Liberty. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
- Chase, M., & Mulvenon, J. (2002). *You've got dissent!: Chinese dissident use of the Internet and Beijing's counter-strategies*: Rand Corp.
- Cheng, A. S., Fleischmann, K. R., Wang, P., & Oard, D. W. (2008). *Advancing social science research by applying computational linguistics*. Paper presented at the Proceedings of the 2008 Annual Meeting of the American Society for Information Science and Technology.
- Chiao, B. (2009). *The Great Firewall of China: The Law and Economics of Internet Censorship or 中国防火长城--互联网审查的法律经济学 (in Chinese)*: China Economic Publishing House (中国经济出版社).
- Christiano, T. (2008). Democracy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1), 22-29.
- Clark, H. H. (1996). *Using language*. New York: Cambridge University Press.
- CNNIC. (2000). 互联网电子公告服务管理规定 Regulation on Internet Bulletin Board Services. from <http://cnnic.cn/html/Dir/2000/10/08/0653.htm>
- CNNIC. (2005). 互联网新闻信息服务管理规定 Regulations on Internet News and Information Services. from <http://cnnic.cn/html/Dir/2005/09/27/3184.htm>
- Coccaro, N., & Jurafsky, D. (1998). *Towards better integration of semantic predictors in statistical language modeling*. Paper presented at the Fifth International Conference on Spoken Language Processing.
- Collier, D., & Levitsky, S. (1997). Democracy with adjectives: Conceptual innovation in comparative research. *World Politics*, 430-451.
- Dahlberg, L. (2001). Computer-Mediated Communication and The Public Sphere: A Critical Analysis. *Journal of Computer-Mediated Communication*, 7(1).

- Dahlberg, L. (2002). *Net-Public Sphere Research: Beyond the 'First Phase'*. Paper presented at the Euricom Colloquium: Electronic Networks and Democracy.
- De Choudhury, M., Hofman, J., Mason, W., & Watts, D. J. (2010). *Inferring "Relevant" Network Structures from Social Communication*. Paper presented at the International Conference on the World Wide Web, 2010.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1-38.
- Deutsche-Welle. (2007). 十七大前的主义之争 Ideology Competition in China Before the 17th State Congression. from <http://www.dw-world.de/dw/article/0,2144,2516886,00.html>
- Dewey, J. (1927). *The public and its problems*: New York, H. Holt and Company [c1927].
- Elegant, S. (2007). China's Me Generation. *Time Magazine*.
- Elgan, M. (2009). How China's '50 Cent Army' Could Wreck Web 2.0. *Datamation*, from <http://itmanagement.earthweb.com/columns/article.php/3795091/How%20China%20s%2050%20Cent%20Army%20Could%20Wreck%20Web%202.0.htm>
- Eliasoph, N. (1998). *Avoiding Politics: How Americans Produce Apathy in Everyday Life*: Cambridge University Press.
- Ewick, P., & Silbey, S. (1998). *The common place of law: Stories from everyday life*: University of Chicago Press.
- Fellbaum, C. (1998). *WordNet: An electronic lexical database*: MIT press Cambridge, MA.
- Felstiner, W., Abel, R., & Sarat, A. (1980). Emergence and Transformation of Disputes: Naming, Blaming, Claiming..., *The. Law & Soc'y Rev.*, 15, 631.
- Fung, A. (2004). *Empowered Participation: Reinventing Urban Democracy*: Princeton University Press.
- Gallagher, M. (2006). Mobilizing the law in china:" informed disenchantment" and the development of legal consciousness. *Law and Society Review*, 40(4), 783.
- Gallagher, M. E. (2002). " Reform and Openness": Why China's Economic Reforms Have Delayed Democracy. *World Politics*, 54(3), 338-372.
- Gallagher, M. E. (2006). Mobilizing the law in china:" informed disenchantment" and the development of legal consciousness. *Law and Society Review*, 40(4), 783.
- Gamon, M., Basu, S., Belenko, D., Fisher, D., Hurst, M., & Konig, A. C. (2008). *BLEWS: Using Blogs to Provide Context for News Articles*. Paper presented at the Proceedings of the 2008 International Conference on Weblogs and Social Media (ICWSM).
- Gamson, W., & Lasch, K. (Eds.). (1983). *The Political Culture of Social Welfare Policy*: New York, NY: Academic Press.
- Gastil, J., & Keith, W. (2005). A Nation That (Sometimes) Likes to Talk: A Brief History of Public Deliberation in the United States. In *The Deliberative Democracy Handbook*.
- Gastil, J., & Levine, P. (2005). *The deliberative democracy handbook: Strategies for effective civic engagement in the 21st century*: San Francisco: Jossey-Bass.

- Girolami, M., & Kaban, A. (2003). *On an equivalence between PLSI and LDA*. Paper presented at the Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval.
- Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). *Large-scale sentiment analysis for news and blogs*. Paper presented at the ICWSM'07.
- Goldman, M. (2005). *From comrade to citizen: The struggle for political rights in China*: Harvard Univ Pr.
- Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2009). Language Style Matching as a Predictor of Social Dynamics in Small Groups. *Communication Research*.
- Gordon, E., & Koo, G. (2008). Placeworlds: Using Virtual Worlds to Foster Civic Engagement. *Space and Culture*, 11(3), 204.
- Graham, T., & Witschge, T. (2003). In search of online deliberation: Towards a new method for examining the quality of online discussions. *Communications*, 28(2), 173-204.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114(2), 211-244.
- Gutmann, A., & Thompson, D. (2004). *Why Deliberative Democracy?*: Princeton University Press.
- Gutmann, E. (2004). *Losing the New China: A Story of American Commerce, Desire and Betrayal*: Encounter Books.
- Habermas, J. (1989). *The Structural Transformation of the Public Sphere: An Inquiry Into a Category of Bourgeois Society*: MIT Press.
- Harwit, E., & Clark, D. (2001). Shaping the internet in China. Evolution of political control over network infrastructure and content. *Asian Survey*, 41(3), 377-408.
- Haveliwala, T., Kamvar, S., & Jeh, G. (2003). An analytical comparison of approaches to personalizing pagerank. *Preprint, June*.
- Hofmann, T. (1999). *Probabilistic latent semantic analysis*. Paper presented at the Uncertainty in Artificial Intelligence '09.
- Hollan, J., & Stornetta, S. (1992). *Beyond Being There*. Paper presented at the ACM SIGCHI'92.
- Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*: Longman Higher Education.
- Huang, P. C. C. (1993). "Public Sphere"/"Civil Society" in China?: The Third Realm between State and Society. *Modern China*, 216-240.
- Jankowski, N. W., & Van Os, R. (2004). Internet-based political discourse: a case study of electronic democracy in Hoogeveen. *Democracy Online. The Prospects for Political Renewal through the Internet*, 181-193.
- Janssen, D., & Kies, R. (2004). *Online forums and deliberative democracy: Hypotheses, variables and methodologies*. Paper presented at the Prepared for the Conference on Empirical Approaches to Deliberative Politics, European University Institute, Florence (May 22-23, 2004) <http://edc.unige.ch/publications/e-workingpapers/onlineforums.pdf> [Last accessed February 28, 2005].
- Jensen, J. L. (2003). Public Spheres on the Internet: Anarchic or Government-Sponsored-A Comparison. *Scandinavian Political Studies*, 26(4), 349-374.
- Jiang, Q. (2003). *Political Confucianism*: Beijing: The Jointly Press.
- Keller, P. (1994). Sources of Order in Chinese Law. *The American Journal of Comparative Law*, 711-759.

- Kelly, J., Fisher, D., & Smith, M. (2005). Debate, Division, and Diversity: Political Discourse Networks in USENET Newsgroups. *Online Deliberation*.
- Konstas, I., Stathopoulos, V., & Jose, J. M. (2009). *On social networks and collaborative recommendation*. Paper presented at the Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*: Sage Publications, Inc.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*: Chicago London.
- Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(02), 311-331.
- Lee, C. (2002). From the specter of Mao to the spirit of the law: Labor insurgency in China. *Theory and Society*, 31(2), 189-228.
- Lehmann, F. (1992). *Semantic Networks in Artificial Intelligence*.
- Lei, Y.-W. (2009). Chinese Netizens and China's Democratization: The Political Consequences of the Rise of the Chinese Netizens.
- Leib, E. J., & He, B. (2006). *The search for deliberative democracy in China*: Palgrave Macmillan.
- Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). *Meme-tracking and the Dynamics of the News Cycle*. Paper presented at the Proc. 15th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining.
- Lessig, L. (1999). *Code: and other laws of cyberspace*.
- Lessig, L. (2006). *Code: Version 2.0*: Basic Books, Inc. New York, NY, USA.
- Levine, P. (2007). *The Future of Democracy: Developing the Next Generation of American Citizens*: Tufts University Press.
- Lieberthal, K. G. (1992). Introduction: the "fragmented authoritarianism" model and its limitations. In *Bureaucracy, Politics, and Decision-making in Post-Mao China, Studies on China* (Vol. 14).
- Liebman, B. L. (2007). China's Courts: Restricted Reform. *Columbia Journal of Asian Law*, 21, 1.
- Lin, D., & Pantel, P. (2002). *Concept discovery from text*.
- Link, P. (2002). China: The Anaconda in the Chandelier. *NEW YORK REVIEW OF BOOKS*, 49, 67-70.
- Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *The American Political Science Review*, 53(1), 69-105.
- Liu, S. (2006a). Beyond Global Convergence: Conflicts of Legitimacy in a Chinese Lower Court. *Law & Soc. Inquiry*, 31, 75-1035.
- Liu, S. (2006b). Client influence and the contingency of professionalism: The work of elite corporate lawyers in China. *LAW AND SOCIETY REVIEW*, 40(4), 751.
- Liu, S. (2008). Globalization as Boundary-Blurring: International and Local Law Firms in China's Corporate Law Market. *Law & Society Review*, 42(4), 771-804.
- Lovasz, L. (1993). Random walks on graphs: A survey. *Combinatorics*, 2(1), 1-46.
- Lubman, S. (1999). Bird in a Cage: Chinese Law Reform After Twenty Years. *Nw. J. Int'l L. & Bus.*, 20, 383.
- MacKinnon, R. (2008). Studying Chinese blog censorship. from <http://rconversation.blogs.com/rconversation/2008/11/studying-chines.html>
- Manning, C. D., Raghavan, P., & Schtze, H. (2008). *Introduction to information retrieval*: Cambridge University Press New York, NY, USA.

- Mansbridge, J. (1999). Everyday Talk in the Deliberative System. *Deliberative Politics: Essays on Democracy and Disagreement*, 211–242.
- Martin, L. W., & Vanberg, G. (2008). A robust transformation procedure for interpreting political text. *Political Analysis*, 16(1), 93.
- McAfee, N. (2004). Three Models of Democratic Deliberation. *Journal of Speculative Philosophy*, 18(1), 44-59.
- Mei, Q., & Zhai, C. (2005). *Discovering evolutionary theme patterns from text: an exploration of temporal text mining*. Paper presented at the Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining.
- Mei, Q., & Zhai, C. (2006). *A mixture model for contextual text mining*. Paper presented at the Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining.
- Mei, Q., Zhou, D., & Church, K. (2008). Query suggestion using hitting time.
- Michelson, E. (2007a). Climbing the Dispute Pagoda: Grievances and Appeals to the Official Justice System in Rural China. *American Sociological Review*, 72(3), 459-485.
- Michelson, E. (2007). Climbing the Dispute Pagoda: Grievances and Appeals to the Official Justice System in Rural China (Vol. 72, pp. 459-485): American Sociological Association.
- Michelson, E. (2007b). Lawyers, Political Embeddedness, and Institutional Continuity in China's Transition from Socialism 1. *American Journal of Sociology*, 113(2), 352-414.
- Mihalcea, R., Tarau, P., & Figa, E. (2004). *PageRank on semantic networks, with application to word sense disambiguation*. Paper presented at the Proceedings of the 20th International Conference on Computational Linguistics (COLING-04).
- Min, S. J. (2007). Online vs. face-to-face deliberation: Effects on civic engagement. *JOURNAL OF COMPUTER MEDIATED COMMUNICATION*, 12(4), 1369.
- Munson, S., Zhou, D. X., & Resnick, P. (2009). *Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators*. Paper presented at the ICWSM'09.
- Mutz, D. C. (2006). *Hearing the Other Side: Deliberative Versus Participatory Democracy*: Cambridge University Press.
- Nathan, A. J. (1986). *Chinese democracy*: Univ of California Pr.
- Neuman, Y., & Nave, O. (2009). Metaphor-based meaning excavation. *Information Sciences*.
- Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, vol.45(no.2), 167-256.
- O'Brien, K., & Li, L. (2006). *Rightful resistance in rural China*: Cambridge Univ Pr.
- Ogden, C. K., Richards, I. A., Malinowski, B., & Crookshank, F. G. (1949). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*: Routledge & Kegan Paul.
- Ogden, S. (2002). *Inklings of democracy in China*: Harvard Univ Pr.
- Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction*, Vol. 15, 139-179.
- Orpin, D. (2005). Corpus linguistics and critical discourse analysis: Examining the ideology of sleaze. *International Journal of Corpus Linguistics*, 10(1), 37-61.
- Orts, E. (2001). Rule of Law in China. *Vand. J. Transnatl. L.*, 34, 43.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). *The PageRank citation ranking: Bringing order to the Web*.

- Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up?: sentiment classification using machine learning techniques*. Paper presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10.
- Perry, E. (2007). Studying Chinese Politics: Farewell to Revolution? *CHINA JOURNAL-CANBERRA*, 57, 1.
- Perry, E. (2008). Chinese Conceptions of "Rights": From Mencius to Mao—and Now. *Perspectives on Politics*, 6(01), 37-50.
- Perry, E. J. (2007). Studying Chinese Politics: Farewell to Revolution? *CHINA JOURNAL-CANBERRA*, 57, 1.
- Perry, E. J. (2008). Chinese Conceptions of "Rights": From Mencius to Mao-and Now. *Perspectives on Politics*, 6(01), 37-50.
- Popping, R. (2000). *Computer-assisted text analysis*: Sage Pubns Ltd.
- Price, V., & Cappella, J. N. (2002). Online Deliberation and its Influence: The Electronic Dialogue Project in Campaign 2000. *IT and Society*, 1(1), 303-329.
- Rayburn, J., & Conrad, C. (2004). China's Internet structure: Problems and control measures. *International Journal of Management*, 21(4), 471-480.
- Roberts, C. W. (1997). *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts*: Lawrence Erlbaum Associates.
- Rowe, W. T. (1993). The problem of "civil society" in late imperial China. *Modern China*, 139-157.
- Schmitter, P. C., & Karl, T. L. (1991). What democracy is... and is not. *Journal of Democracy*, 2(3), 75-88.
- Schneider, S. M. (1997). *Expanding the Public Sphere through Computer-Mediated Communication: Political Discussion about Abortion in*. Citeseer.
- Schudson, M. (1997). Why Conversation Is Not the Soul of Democracy. *Critical Studies in Mass Communication*, 14(4), 297-309.
- Schurmann, F. (1968). Ideology and organization in Communist China.
- Shi, L. (2003). *The Policy Deviation in China's New Media: A Comparative Content Analysis*. Paper presented at the Annual meeting of the International Communication Association.
- Solomon, R. H. (1971). *Mao's revolution and the Chinese political culture*: Univ of Michigan Center for chinese.
- Sowa, J. F., & Borgida, A. (1991). *Principles of Semantic Networks: Explorations in the representation of knowledge*: Morgan Kaufmann San Mateo, CA.
- Spertus, E. (1997). *Smokey: Automatic recognition of hostile messages*. Paper presented at the PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE.
- Steenbergen, M. R., Bächtiger, A., Spörndli, M., & Steiner, J. (2003). Measuring political deliberation: a discourse quality index. *Comparative European Politics*, 1, 21-48.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1), 41-78.
- Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of Public Deliberation*, 3(1), 12.
- Stubbs, M. (2006). Inferring Meaning: Text, Technology and Questions of Induction. In A. Mehler (Ed.), *Aspects of automatic text analysis*: Springer Verlag.
- Sunstein, C. R. (2001). *Republic. com*: Princeton University Press.

- Tai, Z. (2006). *The Internet in China: Cyberspace and Civil Society*: Routledge.
- Tan, Z. (1999). Regulating China's Internet: Convergence toward a coherent regulatory regime. *Telecommunications Policy*, 23(3-4), 261-276.
- Thompson, C. (2006, April 23). Google's China Problem (and China's Google Problem). *New York Times*.
- Tilly, C. (2007). *Democracy*.
- Turney, P. D. (2002). *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. Paper presented at the Proceedings of the 40th annual meeting of the Association for Computational Linguistics.
- van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., & Schlobach, S. (2008). Good news or bad news? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations. *Journal of Information Technology & Politics*, 5(1), 73-94.
- van de Donk, W. (2004). *Cyberprotest: New Media, Citizens and Social Movements*: Routledge.
- Wakeman Jr, F. (1993). The civil society and public sphere debate: Western reflections on Chinese political culture. *Modern China*, 108-138.
- Walder, A. G. (2002). The Transformation of Contemporary China Studies, 1977-2002. *Produced by The Berkeley Electronic Press, UC/IAS Edited Volumes, 3*.
- Wang, X. (2007). *Online Public Spheres: How Internet Discussion Forums Promote Political Participation in China*. Paper presented at the the annual meeting of the NCA 93rd Annual Convention. from http://www.allacademic.com/meta/p190487_index.html
- Wilhelm, A. G. (1999). Virtual sounding boards: how deliberative is online political. *Digital democracy: Discourse and decision making in the information age*, 153.
- Wu, F., & Huberman, B. A. (2008). Public Discourse in the Web Does Not Exhibit Group Polarization.
- Yang, G. (2003a). The co-evolution of the Internet and civil society in China. *Asian Survey*, 43(3), 405-422.
- Yang, G. (2003). The Internet and Civil Society in China: a preliminary assessment. *Journal of Contemporary China*, 12(36), 453-475.
- Yang, G. (2003b). The Internet and the rise of a transnational Chinese cultural sphere. *Media, Culture & Society*, 25(4), 469.
- Yang, G. (2009). *The Power of the Internet in China: Citizen Activism Online*.
- Yang, G., & Calhoun, C. (2007). Media, civil society, and the rise of a green public sphere in China. *China Information*, 21(2), 211.
- Young, I. M. (1996). Communication and the Other: Beyond Deliberative Democracy. In S. Benhabib (Ed.), *Democracy and Difference: Contesting the Boundaries of the Political* (pp. 120-135).
- Zhang, L. L. (2006). Behind the 'Great Firewall': Decoding China's Internet Media Policies from the Inside. *Convergence*, 12(3), 271.
- Zhao, S. (1998). A state-led nationalism: The patriotic education campaign in post-Tiananmen China. *Communist and Post-Communist Studies*, 31(3), 287-302.
- Zhao, Y. (1998). *Media, market, and democracy in China*: University of Illinois Press Urbana.
- Zhao, Y. (1998). *Media, market, and democracy in China: Between the party line and the bottom line*: Univ of Illinois Pr.
- Zhao, Y. (2008). *Communication in China: Political Economy, Power, and Conflict*: Rowman & Littlefield Publishers, Inc.

- Zheng, Y. (2007). *Technological empowerment: the Internet, state, and society in China*: Stanford Univ Pr.
- Zheng, Y., & Wu, G. (2005). Information technology, public space, and collective action in China. *Comparative Political Studies*, 38(5), 507.
- Zhou, Y. (2006). *Historicizing online politics: Telegraphy, the Internet, and political participation in China*: Stanford University Press.
- Zittrain, J., & Edelman, B. G. (2003). Internet filtering in China. *IEEE Internet Computing*.