

Course Form for PKU Summer School International 2018

Course Title	Computational Social Science
	计算社会科学
Teacher	Jonathan ZHU, Winson PENG
First day of classes	July 2, 2018
Last day of classes	July 12, 2018
Course Credit	2 credits
Course Description	
Objective	
<p>The course aims to introduce computational social science (CSS) to students in various data science disciplines such as computer science, mathematics, statistics, electronic engineering, bio-/medical informatics, etc., who are interested in individual human behaviors and aggregate social processes. Assuming the enrolled students to have already known how to perform basic data analysis using a variety of algorithms or tools, the course focuses on the fundamental principles (e.g., what are good or bad practices) and the research design (e.g., how to plan rigorous studies to answer causal questions). Ample examples from the existing literature will be used to help illustrate the principles and apply the research design. At the end of the course, the students are expected to be able to do the following: (1) communicating and collaborating with social scientists using their vocabulary and reasoning logic; (2) planning and implementing rigorous computational studies on human behaviors or social processes; (3) evaluating the quality of other social science studies and suggesting feasible improvements.</p>	
Pre-requisites /Target audience	
<p>Pre-requisites: Basic knowledge of statistical analysis, text mining, machine learning Target audience: Senior undergraduate students and graduate students in various disciplines of data science (computer science, mathematics, statistics, electronic engineering, bio-/medical informatics, etc.)</p>	
Proceeding of the Course	
No	
Assignments (essay or other forms)	
Readings, In-class and online discussions, and take-home exercises	
Evaluation Details	

Attendance and Discussions: 25%	
Assignments: 45%	
Exam: 30%	
Text Books and Reading Materials <ul style="list-style-type: none"> Matthew J. Salganik (2018). Bit by Bit: Social Research in the Digital Age. Princeton University Press; Earl Babbie (2011). Principles of Social Research, 11th Edition. Tsinghua University Press (or later edition after 11th from Wordsworth); Additional readings are given in the Class Schedule. 	
Academic Integrity (If necessary) <p>Students are allowed to discuss readings and assignments among classmates in and outside the class, but are discouraged to seek help from any living person outside the class. However, individual-based writing assignments must be independently completed (i.e., without any plagiarism).</p>	
CLASS SCHEDULE (Subject to adjustment)	
Session 1: Introduction to CSS and Research Design	Date: 7/2/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) The basic features of computational social science (CSS, e.g., what it is, how it has come about and why, what it aims to do, and how it works), primary purpose of social science research (descriptive, explanatory, or predictive), and most common research designs in social research (within-subjects design and between-subjects design).	
【Questions】 What is CSS? How is it related to and different from social sciences and data science and why? How to study social phenomena cross time, space, and population?	
【Readings, Websites or Video Clips】 <ol style="list-style-type: none"> Babbie Ch1, Ch3-4 Salganik Ch1, Ch3 Peng & Zhu (2013) Campbell Video1 1.1-1.7; 4.1-4.7; 5.1-5.8; 6.1-6.8 	
【Assignments for this session (if any)】 Nil	
Session 2: Experiment in Social Research	Date: 7/3/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) Principles and procedure of social science experiment (i.e., comparison between subjects with and without exposure to stimulus) as the key solution to problems and challenges in causal reasoning.	
【Questions】 Why is it difficult to determine causes of social processes and how to deal with the problems?	

【Readings, Websites or Video Clips】 1. Babbie Ch8 2. Salganik Ch4 3. Bond et al. (2012) 4. King et al. (2017) 5. Campbell Video1 7.1-7.8	
【Assignments for this session (if any)】 Exercise 1. Review and critique of the research design and the causal reasoning used in a set of published CSS studies	
Session 3: Data for CSS	Date: 7/5/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) How to collect data on human behaviors through online and offline means; technical, operational, and quality challenges involved.	
【Questions】 What are the major advantages and shortcomings of online data as compared with offline data?	
【Readings, Websites or Video Clips】 1. Babbie Ch9, Ch11 2. Salganik Ch2, Ch5 3. Liang & Zhu (2017) 4. Campbell Video2 2.1-2.10	
【Assignments for this session (if any)】 Nil	
Session 4: Sampling	Date: 7/6/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) Principles and methods of classic probability sampling; opportunities and challenges of applying probability sampling to online data.	
【Questions】 When does probability sampling work for online data and when doesn't? How to deal with the problems?	
【Readings, Websites or Video Clips】 1. Babbie Ch7 2. Zhu et al. (2011) 3. Xu & Zhu (2016) 4. Campbell Video2 3.1-3.9	
【Assignments for this session (if any)】 Nil.	
Session 5: Measurement	Date: 7/7/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) How to quantify human thoughts and behaviors from empirical data; how to link latent	

concepts and observed variables; how to control for measurement noises.	
【Questions】 How to evaluate and improve the quality (i.e., validity and reliability) of quantified data?	
【Readings, Websites or Video Clips】 1. Babbie Ch5-6 2. Lazer et al. (2014) 3. Newman (2017)	
【Assignments for this session (if any)】 Exercise 2. Review and critique of the data collection, sampling, and measurement used in a set of published CSS studies.	
Session 6: Multivariate Analysis	Date: 7/9/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) Principles, procedure, and exemplar applications of multivariate analysis (e.g., multiple regressions) to describe, explain, and predict causal processes with multiple causes.	
【Questions】 How to control for confounding effects, identify inflated or suppressed effects, and mediated/moderated effects?	
【Readings, Websites or Video Clips】 1. Babbie Ch14-16 2. To be added	
【Assignments for this session (if any)】 Nil.	
Session 7: Multilevel and Temporal Analysis	Date: 7/10/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) Multilevel analysis: principles, procedure, and exemplar applications of multilevel analysis (e.g., mixed effects model and network analysis) to identify socially structured effects on individuals and emergent effects from individuals on social structure. Temporal analysis: principles, procedure, and exemplar applications of temporal analysis (e.g., time series analysis, cohort analysis, survival analysis, sequence analysis) to model dynamic characteristics of individual human behaviors and aggregated social processes.	
【Questions】 Multilevel analysis: How to quantify individual variability within global regularities? How to determine societal and collective effects on individuals? How to determine individual contributions to the society? Temporal analysis: How to separate and quantify different dynamic processes (e.g., auto-regression, periodicities, long-term trends, short-term shocks, individual habits, social constraints, etc.) underlying social data?	
【Readings, Websites or Video Clips】 1. Akciland & Zhu (2015) 2. Peng & Zhu (2012) 3. Wimmer & Lewis (2010) 4. Murdock et al. (2017)	

5. Robinson & Deng (2015) 6. Xu et al. (2013) 7. Zhu et al. (2018)	
【Assignments for this session (if any)】 Nil.	
Session 8: Research Ethics and Student Presentation	Date: 7/11/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) Ethical concerns for traditional social science research; new challenges for computational social science; causes of and possible solutions to threats to privacy, confidentiality, security, copyrights, and other issues. Students present the results of assignment 3.	
【Questions】 When does computational social science research become harmful to individuals and the society and how to mitigate the negative consequences?	
【Readings, Websites or Video Clips】 1. Babbie Ch3 2. Salganik, Ch6-7 3. Campbell (2) Video 6.1-6.9	
【Assignments for this session (if any)】 Exercise 3: Analysis and interpretation of online data with a complex structure (i.e., multivariate, multilevel, and time-stamped)	
Sources of Textbooks, Videos, and Additional Readings	
Textbooks/Videos: <ul style="list-style-type: none"> Babbie = Earl Babbie (2007), <i>The practice of social research</i>, 11th ed. Tsinghua University Press. Salganik = Matthew J. Salganik (2018). <i>Bit by bit: Social research in the digital age</i>. Princeton University Press. Campbell = Cameron Campbell: <i>Social science approaches to the study of Chinese society</i>, part 1 and part 2, Coursera.com 	
Additional Readings: <ul style="list-style-type: none"> Ackland, R., & Zhu, J. J. H. (2015). Social network analysis. In P. Halfpenny & R. Procter (Eds.), <i>Innovations in digital research methods</i> (pp. 221-244). Sage Publications. Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. <i>Nature</i>, 489, 295. King, G., Schneer, B., & White, A. (2017). How the news media activate public expression and influence national agendas. <i>Science</i>, 358(6364), 776. Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. <i>Science</i>, 343(6176), 1203-1205. Liang, H., & Zhu, J. J. H. (2017). Big data, collection of (social media, harvesting). In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), <i>International Handbook of</i> 	

Communication Methods, Wiley & Sons.

- Murdock, J., Allen, C., & DeDeo, S. (2017). Exploration and exploitation of Victorian science in Darwin's reading notebooks. *Cognition*, 159, 117-126.
- Newman, M. (2017). Measurement errors in network data. *arXiv preprint arXiv:1703.07376*.
- Peng, T. Q., Zhang, L., Zhong, Z. J., & Zhu, J. J. H. (2013). Mapping the landscape of Internet studies: Text mining of social science journal articles, 2000-2009. *New Media & Society*, 15(5), 644-664.
- Peng, T. Q., & Zhu, J. J. H. (2012). Where you publish matters most: A multilevel analysis of factors affecting citations of internet studies. *Journal of the American Society for Information Science and Technology*, 63(9), 1789-1803.
- Robinson, W. N., & Deng, T. (2015). *Data Mining Behavioral Transitions in Open Source Repositories*. Paper presented at the 2015 48th Hawaii International Conference on System Sciences.
- Wimmer, A., & Lewis, K. (2010). Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *American Journal of Sociology*, 116(2), 583-642.
- Xu, P., Wu, Y., Wei, E., Peng, T.-Q., Liu, S., Zhu, J. J. H., & Qu, H. (2013). Visual Analysis of Topic Competition on Social Media. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2012-2021.
- Xu, X. K., & Zhu, J. J. H. (2016). Flexible sampling large-scale social networks by self-adjustable random walk. *Physica A*, 463, 356-365.
- Zhu, J. J. H., Mo, Q., Wang, F., & Lu, H. (2011). A random digit search (RDS) method for sampling of blogs and other web content. *Social Science Computer Review*, 29(3), 327-339.
- Zhu, J. J. H., Chen, H. X., Peng, T. Q., Liu, X. F., & Dai, H. X. (2018). How to measure sessions of mobile device use? Quantification, Evaluation, and Applications. *Mobile Media & Communication*.