Econometric Methods for Social Spillovers and Networks

University of St. Gallen, September 26th to October 3rd, 2016

Course Description

This course will provide an overview of econometric methods appropriate for the analysis of social and economic spillovers and networks. Many social and economic activities are embedded in networks. Furthermore, datasets with natural graph theoretic (i.e., network) structure are increasingly available to researchers. We will review (i) how to describe, summarize and visually present network data, (ii) formal econometric models of network formation that admit heterogeneity, strategic behavior, and/or dynamics, and, finally, (iii) how to model behaviors that occur on networks (e.g., the identification of peer group effects or social spillovers). The focus will be on the formal development of methods, but selected empirical examples will also be covered, as will methods of practical computation.

Course Logistics

Instructor: Bryan Graham, Department of Economics, University of California – Berkeley

Email: `bgraham@econ.berkeley.edu`

Time: 2:15PM to 6PM daily (in classroom 24-0130 *except* for Friday 9/30 which will be in 01-103)

Office Hours: I will announce the time and location of office hours on the first day of class. I look forward to talking with all of you!

Prerequisites: The equivalent of a first year Ph.D. level sequence in econometrics. Specifically an understanding of probability and statistical inference at the level of Casella and Berger (1990, *Statistical Inference*), linear regression analysis at the level of Goldberger (1991, *A Course in Econometrics*) and some exposure to non-linear models (e.g., maximum likelihood, M-estimation). I will also assume a basic knowledge of applied matrix algebra.

Textbook: Readings preceded by a [r] in the course outline are "required" (i.e., should be read prior to class), while those preceded by a [b] are for "background" (i.e., may be useful for students interested in additional material). Students should consider purchasing the textbooks by Jackson (2008) and Newman (2010). The Easley and Kleinberg (2010) book also contains useful material (with the added attraction of being available in draft form online for free). The survey by Goldenberg, Zheng, Fienberg and Airoldi (2009) covers much of the technical literature in statistics and machine learning. A recommended general purpose reference for microeconometrics is Wooldridge (2010).

Computation: The bulk of class will be devoted to the formal development of the material, albeit with empirical illustrations as well as ample discussions of the various practicalities of implementation. However I do intend to reserve some class time for actual practice with computation. Computational examples will be done using Python. Python is a widely used general purpose programming language with good functionality for scientific computing. I highly recommend the Anaconda distribution, which is available for download at `http://continuum.io/downloads`. Some basic tutorials on installing and using Python, with a focus on economic applications, can be found online at `http://quant-econ.net`. You may also wish to install Rodeo, which is an integrated development environment (IDE) tailored to statistics or "data science" applications. Rodeo makes working in Python look and feel similar to working in Stata or MATLAB. Rodeo is also free and available at `https://www.yhat.com/`.

Students wishing to benefit fully from the computational illustrations should install Python on their computers prior to the start of class. Unfortunately I cannot help with installation related problems. Good books for learning Python, with some coverage of statistical applications, are Guttag (2013) and McKinney (2013). The former is an excellent introduction to computer science as well as Python, the latter is heavily focused on the pandas module. Graphviz is a free graph visualization program that is also useful (`http://www.graphviz.org/`).

<u>Examination:</u> Written Exam: 60% (A single A4 sheet of paper of notes may be brought to the examination. No computational aides allowed). Class Participation: 20% (Participation in class discussion, demonstration of engagement with required readings). Referee Report: 20% (3-5 page "referee report" on a networks-related paper).

Course Outline

Date	Topic	Readings
M 9/26	**Describing Networks**	[r] Graham (2015), [r] Jackson (2008, Ch. 2) [r] Newman (2010, Ch. 6 - 8), [b] Jackson et al. (forthcoming)
	Examples of networks	[r] Atalay et al. (2011), [r] De Weerdt (2004) [b] Apicella et al. (2012), [b] Matous and Todo (2016)
	Small worlds	[r] Milgam (1967)
	Centrality	[r] Bonacich (1987), [r] Calvó-Armengol et al. (2009)
	Degree distributions	[b] Mitzenmacher (2004), [b] Clauset et al. (2009)
	Triads	[b] Granovetter (1973), [b] Jackson et al. (2012)
Tu 9/27	**Nonparametrics**	[r] Hoff (2008), [b] Orbanz & Roy (2015)
	Graphons	[b] Bickel & Chen (2009), [b] Hoff (2008) [r] Chatterjee (2013), [r] Zhang et al. (2015) [r] Bhattacharya and Bickel (2015), [b] Yang et al. (2014)
	Stochastic Block Model	[b] Snijders & Nowiki (1997), [r] Daudin et al. (2008)
	EM Algorithms	[b] Gupta & Chen (2011), [b] Bickel et al. (2013)
W 9/28	**Heterogeneity**	
	β-Model, Testing	[b] Blitzstein and Diaconis (2011) [b] Chatterjee, Diaconis and Sly (2011)
	Dyadic Link Formation	[r] Graham (2014), [b] Dzemski (2014)
Th 9/29	**Strategic Network Formation**	[r] Graham (2016), [r] de Paula et al. (2015), [b] McPherson et al. (2001), [b] Leung (2015)
F 9/30	**Matching**	[r] Graham (2011, 2013), [b] Ferguson (2006) [b] Honore and Powell (1994)
M 10/3	**Peer Effects**	[r] Manski (1993)
	Network structure	[r] Bramoullé, Djebbari & Fortin (2009) [r] Calvó-Armengol, Patacchini and Zenou (2009)
	Neighborhoods	[r] Graham (forthcoming)

References

[1] Apicella, Coren L., Frank W. Marlowe, James H. Fowler & Nicholas A. Christakis. (2012). "Social networks and cooperation in hunter-gatherers," *Nature* 481 (7382): 497 - 501.

[2] Atalay, Enghin, Ali Hortaçsu, James Roberts, Chad Syverson and Lars Peter Hansen (2011). "Network structure of production," *Proceedings of the National Academy of Sciences* 108 (13): 5199 - 5202.

[3] Bhattacharya, Sharmodeep and Peter J. Bickel. (2015). "Subsampling bootstrap of count features of networks," *Annals of Statistics* 45 (6): 2384 - 2411.

[4] Bickel, Peter J. and Aiyou Chen. (2009). "A nonparametric view of network models and Newman-Girvan and other modularities," *Proceedings of the National Academy of Science* 106 (50): 21068 - 21073.

[5] Bickel, Peter, David Choi, Xiangyu Chang and Hai Zhang. (2013). "Asymptotic normality of maximum likelihood and its variational approximation for stochastic block models," *Annals of Statistics* 41 (4): 1922 - 1943.

[6] Bonacich, Phillip. (1987). "Power and centrality: a family of measures," *American Journal of Sociology* 92 (5): 1170 - 1182.

[7] Blitzstein, Joseph and Persi Diaconis. (2011). "A sequential importance sampling algorithm for generating random graphs with prescribed degrees," *Internet Mathematics* 6 (4): 489 - 522.

[8] Bramoullé, Yann, Habiba Djebbari, Bernard Fortin. (2009). "Identification of peer effects through social networks," *Journal of Econometrics* 150 (1): 41 – 55.

[9] Calvó-Armengol, Antoni. Eleonora Patacchini and Yves Zenou. (2009). "Peer effects and social networks in education," *Review of Economic Studies* 76 (4): 1239 - 1267.

[10] Chatterjee, Sourav. (2015). "Matrix estimation by universal singular value thresholding," *Annals of Statistics* 43 (1): 177 - 214.

[11] Chatterjee, Sourav, Persi Diaconis and Allan Sly. (2011). "Random graphs with a given degree sequence," *Annals of Applied Probability* 21 (4): 1400 - 1435.

[12] Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman. (2009). "Power-law distributions in empirical data," *SIAM Review* 51 (4): 661 – 703.

[13] Daudin, J-J., F. Picard and S. Robin. (2008). "A mixture model for random graphs," *Statistics and Computing* 18 (2): 193 - 183.

[14] de Paula, Aureo, Seth Richards-Shubik, and Elie T. Tamer. (2015). "Identification of preferences in network formation games," *Cemmap Working Paper* CWP29/15.

[15] De Weerdt, Joachim. (2004). "Risk-sharing and endogenous network formation, "*Insurance against Poverty*: 197 - 2016 (S. Dercon, Ed.). Oxford: Oxford University Press.

[16] Dzemski, Andreas. (2014). "An empirical model of dyadic link formation in a network with unobserved heterogeneity," *Mimeo.*

[17] Easley, David and Jon Kleinberg. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World.* Cambridge: Cambridge University Press.

[18] Ferguson, Thomas S. (2006). "U-Statistics," Lecture Note for Statistics 200C, UCLA.

[19] Goldenberg, Anna, Alice X. Zheng, Stephen E. Fienberg and Edoardo M. Airoldi. (2009). "A survey of statistical network models," *Foundations and Trends in Machine Learning* 2 (2): 129 - 233.

[20] Graham, Bryan S. (2011). "Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers," *Handbook of Social Economics* 1B: 965 - 1052 (J. Benhabib, A. Bisin, & M. Jackson, Eds.). Amsterdam: North-Holland.

[21] Graham, Bryan S. (2013). "Comparative static and computational methods for an empirical one-to-one transferable utility matching model," *Advances in Econometrics* 31: 153 - 181.

[22] Graham, Bryan S. (2014). "An econometric model of link formation with degree heterogeneity," *NBER Working Paper No. 20341.*

[23] Graham, Bryan S. (2015). "Methods of Identification in Social Networks," *Annual Review of Economics* 7.

[24] Graham, Bryan S. (2016). "Homophily and transitivity in dynamic network formation," *cemmap Working Paper CWP16/16.*

[25] Graham, Bryan S. (forthcoming). "Identifying and estimating neighborhood effects," *Journal of Economic Literature.*

[26] Granovetter, Mark S. (1973). "The strength of weak ties," *American Journal of Sociology* 78 (6): 1360 - 1380.

[27] Gupta, Maya R. and Yihua Chen. (2011). "Theory and Use of the EM Algorithm," *Foundations and Trends in Signal Processing* 4 (3): 223 - 296.

[28] Guttag, John V. (2013). *Introduction to Computation and Programming Using Python.* Cambridge, MA: MIT Press.

[29] Hoff, Peter. (2008). "Modeling homophily and stochastic equivalence in symmetric relational data," *Advances in Neural Information Processing Systems* 20: 657 - 664 (J. Platt, D. Kroller & S. Roweis, Eds.). Cambridge, MA: MIT Press.

[30] Honoré, Bo E. and James L. Powell. (1994). "Pairwise difference estimators of censored and truncated regression models," *Journal of Econometrics* 64 (1): 241 - 278.

[31] Jackson, Matthew O. (2008). *Social and Economic Networks.* Princeton: Princeton University Press.

[32] Jackson, Matthew O., Tomas Rodriguez-Barraquer and Xu Tan. (2012). "Social capital and social quilts: network patterns of favor exchange," *American Economic Review* 102 (5): 1857-97.

[33] Jackson, Matthew O., Brian W. Rogers and Yves Zenou. (forthcoming). "The economic consequences of social network structure," *Journal of Economic Literature.*

[34] Leung, Michael. (2015). "Two-step estimation of network-formation models with incomplete information," *Journal of Econometrics* 188 (1): 182 - 195.

[35] Manski, Charles F. (1993). "Identification of endogenous social effects: the reflection problem," *Review of Economic Studies* 60 (3): 531 - 542.

[36] Matous, Petr and Yasuyuki Todo. (2016). "Energy and resilience: The effects of endogenous interdependencies on trade network formation across space among major Japanese firms," *Network Science* 4 (2): 141 - 163.

[37] McKinney, Wes. (2013). *Python for Data Analysis.* Cambridge: O'Reilly Media, Inc.

[38] McPherson, Miller, Lynn Smith-Lovin and James M. Cook. (2001). "Birds of a feather: homophily in social networks," *Annual Review of Sociology* 27 (1): 415 - 444.

[39] Milgram, Stanley (1967). "The small-world problem," *Psychology Today* 1 (1): 61 - 67.

[40] Mitzenmacher, Michael (2004). "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics* 1 (2): 226 - 251.

[41] Newman, M. E. J. (2010). *Networks: An Introduction.* Oxford: Oxford University Press.

[42] Orbanz, Peter and Daniel M. Roy. (2015). "Bayesian models of graphs, arrays and other exchangeable random structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (2): 437 - 461.

[43] Snijders, Tom A. B. and Krzysztof Nowicki. (1997). "Estimation and prediction for stochastic blockmodels for graphs with latent block structures," *Journal of Classification* 14 (1): 75 - 100.

[44] Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: The MIT Press.

[45] Yang, Justin J., Qiuyi Han and Edoardo M. Airoldi. (2014). "Nonparametric estimation and testing of exchangeable graph models," *International Conference on Artificial Intelligence and Statistics (AISTATS):* 1060 - 1067.

[46] Zhang, Yuan, Elizaveta Levina and Ji Zhu. (2015). "Estimating network edge probabilities by neighborhood smoothing," *arXiv:1509.08588v2 [stat.ML].*