Econometric Analysis of Network Data

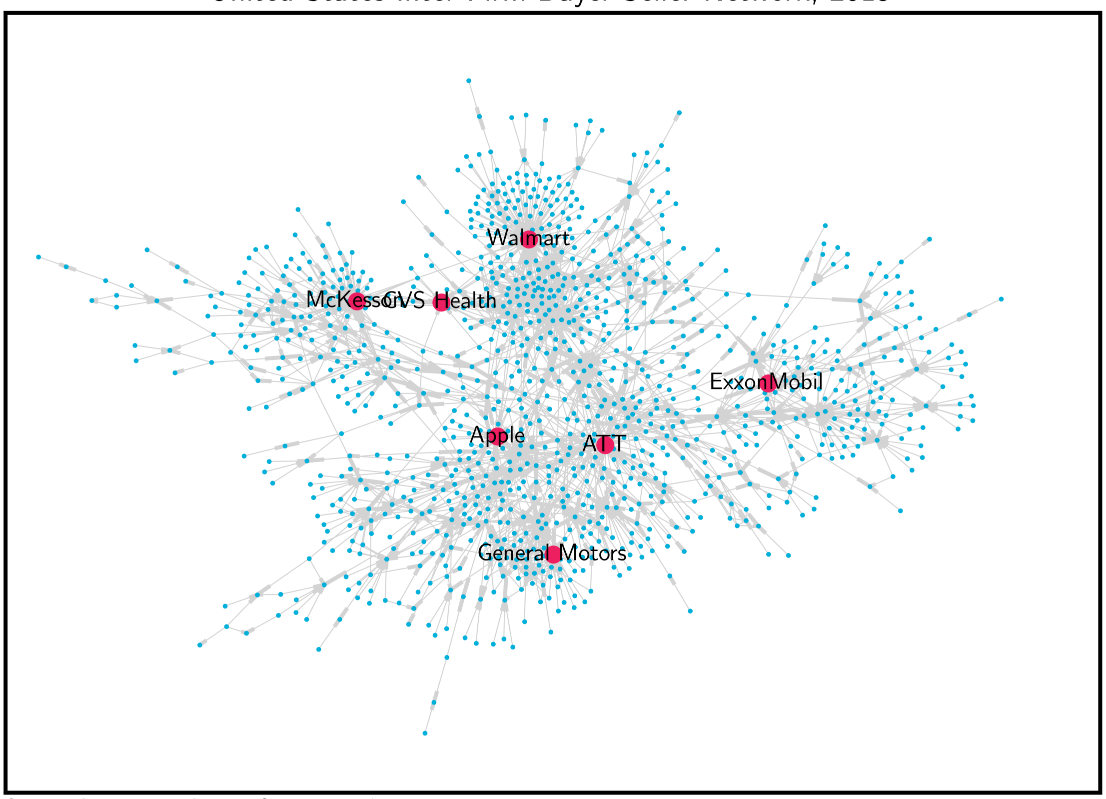
Singapore Management University

May 29th to June 1st, 2017

Course Description

This course will provide an overview of econometric methods appropriate for the analysis of social and economic 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 (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.

United States Inter-Firm Buyer-Seller Network, 2015



Source: Compustat - Capital IQ and author's calculations.
Raw data available at https://wrds-web.wharton.upenn.edu/wrds/ (Accessed January 2017)

Course Logistics

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

Email: `bgraham@econ.berkeley.edu`

Time: To be determined.

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 ideally 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 (2006) and Newman (2010), but doing so is not necessary. The survey by Goldenberg et al. (2009) covers much of the technical literature in statistics and machine learning, but is now somewhat dated.

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/`.

Good books for learning Python, with some coverage of statistical applications, are:

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

2. VanderPlas, Jake. (2016). *Python Data Science Handbook: Essential Tools for Working with Data.* Cambridge: O'Reilly Media, Inc.

The former is an excellent introduction to computer science as well as Python, the latter covers the core Data Science packages available in Python.

The code I will provide will execute properly in Python 2.7. Nevertheless I recommend installing Python 3.5. This is the latest Python release. Graphviz is a free graph visualization program that is also useful (`http://www.graphviz.org/`).

Course Outline

Date	Topic	Readings
Lec 1	**Describing**	[r] de Paula (2016); Graham (2015); Jackson (2008, Ch. 2)
(~2 hours)	**Networks**	[b] Goldenberg et al. (2009)
	Examples of networks	[b] Atalay et al. (2011); Mizuno et al. (2014)
		[b] Apicella et al. (2012); Acemoglu et al. (2012)
	Small worlds	[b] Milgram (1967)
	Degree distributions	[b] Mitzenmacher (2004)
	Homophily	[b] McPherson et al. (2001)
	Triads	[b] Granovetter (1973); Jackson et al. (2012)
Lec 2	**Shock Propogation**	[r] Carvalho (2014)
(~2 hours)	**& Centrality**	[r] Bonacich (1987)
		[b] Acemoglu et al. (2012)
		[b] Acemoglu et al. (2016)
Lec 3	**Nonparametrics:**	[r] Bickel & Chen (2009); Diaconis & Janson (2008)
(~2 hours)	**Graphons**	[b] Orbanz & Roy (2015)
	Estimation	[r] Chatterjee (2015); Zhang et al. (2015)
	Stochastic Block Model	[b] Daudin et al. (2008)
Lec 4	**Nonparametrics:**	[r] Holland & Leinhardt (1976)
(~2 hours)	**Network Moments**	[b] Nowicki (1991)
	Theory	[r] Bickel et al. (2011)
	Computation	[r] Bhattacharya & Bickel (2015)
		[r] Blitzstein & Diaconis (2011); Chatterjee et al. (2011)
Lec 5	**Models of Network**	
(~2 hours)	**Formation**	
	Dyadic Link Formation	[r] Graham (2014)
	Dynamic Models	[r] Graham (2016)
	Strategic models	[r] de Paula et al. (2015)
Lec 6	**Peer Effects**	[r] Manski (1993)
(~2 hours)		[r] Bramoullé et al. (2009)
		[b] Blume et al. (2015); Serpa & Krishnan (2017)

References

Acemoglu, D., Akcigit, U., & Kerr, W. (2016). Networks and the macroeconomy: an empirical exploration. *NBER Macroeconomics Annual*, 31(1), 273 – 335.

Acemoglu, D., Carvalho, V., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5), 1977 – 2016.

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers,. *Nature*, 481(7382), 497 – 501.

Atalay, E., Hortaçsu, A., Roberts, J., & Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13), 5199 – 5202.

Bhattacharya, S. & Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *Annals of Statistics*, 43(6), 2384 – 2411.

Bickel, P. J. & Chen, A. (2009). A nonparametric view of network models and newman-girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50), 21068 – 21073.

Bickel, P. J., Chen, A., & Levina, E. (2011). The method of moments and degree distributions for network models. *Annals of Statistics*, 39(5), 2280 – 2301.

Blitzstein, J. & Diaconis, P. (2011). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*, 6(4), 489 – 522.

Blume, L. E., Brock, W. A., Durlauf, S. N., & Jayaraman, R. (2015). Linear social interaction models. *Journal of Political Economy*, 123(2), 444 – 496.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92, 1170–1182.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1), 41 – 55.

Carvalho, V. (2014). From micro to macro via production networks. *Journal of Economic Perspectives*, 28(4), 23–48.

Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *Annals of Statistics*, 43(1), 177 – 214.

Chatterjee, S., Diaconis, P., & Sly, A. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability*, 21(4), 1400 – 1435.

Daudin, J.-J., F., P., & Robin, S. (2008). A mixture model for random graphs,". *Statistics and Computing*, 18(2), 183 – 193.

de Paula, Á. (2016). *Econometrics of network Models*. Technical Report CeMMAP Working Paper CWP06/16, CeMMAP.

de Paula, Á., Richards-Shubik, S., & Tamer, E. (2015). *Identification of preferences in network formation games*. CeMMAP Working Paper CWP29/15, CeMMAP.

Diaconis, P. & Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica*, 28(1), 33 – 61.

Goldenberg, A., Zheng, A., Fienberg, S. E., & Airoldi, E. M. (2009). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2), 129–333.

Graham, B. S. (2014). *An econometric model of network formation with degree heterogeneity*. NBER Working Paper 20341, National Bureau of Economic Research.

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7(1), 465–485.

Graham, B. S. (2016). *Homophily and transitivity in dynamic network formation*. NBER Working Paper 22186, National Bureau of Economic Research.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360 – 1380.

Holland, P. W. & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7, 1 – 45.

Jackson, M. (2006). The economics of social networks. In R. Blundell, W. Newey, & T. Persson (Eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*. Cambridge: Cambridge University Press.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton: Princeton University Press.

Jackson, M. O., Rodriguez-Barraquer, T., & Tan, X. (2012). Social capital and social quilts: network patterns of favor exchange. *American Economic Review*, 102(5), 1857–1897.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem,. *Review of Economic Studies*, 60(3), 531 – 542.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27(1), 415 – 444.

Milgram, S. (1967). The small-world problem. *Psychology Today*, 1(1), 61 – 67.

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2), 226 – 251.

Mizuno, T., Souma, W., & Watanabe, T. (2014). The structure and evolution of buyer-supplier networks. *Plos One*, 9(7), e100712.

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

Nowicki, K. (1991). Asymptotic distributions in random graphs with applications to social networks. *Statistica Neerlandica*, 45(3), 295 – 325.

Orbanz, P. & Roy, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable randoms structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2), 437 – 461.

Serpa, J. C. & Krishnan, H. (2017). The impact of supply chains on firm-level productivity. *Management Science*.

Zhang, Y., Levina, E., & Zhu, J. (2015). Estimating network edge probabilities by neighborhood smoothing. *arXiv:1509.08588v2 [stat.ML]*.