GENERAL RESOURCES SOCIAL NETWORKS

To accompany the class COMS–4995 Introduction to Social Network, Fall 2016, Columbia University

Related teaching (by chronological order of first course)

- Michael Kearns, U. Penn, Networked Life, Spring 2004–2012, http://www.cis.upenn.edu/ mkearns/teaching/NetworkedLife/
- Jon Kleinberg, Cornell U., The Structure of Information Networks, Fall 2007, http://www.stanford.edu/class/cs224w/
- Dan Spielman, Yale U., Graphs and Networks, Fall 2007–2010, http://www.cs.yale.edu/homes/spielman/462/
- David Kempe, USC, Structure and Dynamics of Networked Information, Spring 2008, http://www-bcf.usc.edu/~dkempe/CS59908/index.html
- Jure Leskovec, Stanford, Social and Information Network Analysis, Fall 2009–2010, http://www.stanford.edu/class/cs224w/
- K Daron Acemoglu, Asu Ozdaglar, Networks, Fall 2009, http://stellar.mit.edu/S/course/6/fa09/6.207J/
- Adam Wierman, Caltech, The ideas behind the web, Winter 2010–2012, http://courses.cms.caltech.edu/cs144/
- Mung Chiang, Princeton, Networks: Friends, Money, and Bytes, Fall 2011, http://www.network20q.com/

N.B.: You may want to subscribe (for free) to the available online version of these classes (e.g., Mung Chiang, Michael Kearns) that are available on http://www.coursera.org

Textbooks

Unfortunately, although there are now a few books on the topics (and many of them very well written), none of them contains entirely the contents that is required to cover at the graduate level a computer science class oriented towards algorithmic aspect of social network. Here are some of them that could be useful for your own study:

The most readable and relevant books for this class are:

- Networks, Crows and Markets, by D. Easley and J. Kleinberg, Cambridge (2010), electronic copy available for free
 - http://www.cs.cornell.edu/home/kleinber/networks-book/
- Epidemics and Rumours in Complex Networks, by M. Draief and L. Massoulié, Cambridge (2009),
- Complex Social Networks, by F. Vega Redondo, Cambridge (2007),

Other books, primarily on economic aspect of social networks:

- Social and Economic Networks, by Matthew O. Jackson, Princeton (2008),
- Connections: An Introduction to the Economics of Networks, by S. Goyal, Princeton (2009),

Books on a complex-system approach to reproduce properties of networks in physics, biology and other disciplines:

• Networks, An Introduction, by M. Newman, Oxford (2010),

Books on mathematical analysis of some random dynamics on graphs (notso-relevant)

- Random Graph Dynamics, by R. Durett, Cambridge (2007), (primarily establishing properties of connectivity and diameter for various random graph models)
- Probability on Graphs, by G. Grimmett, Cambridge (2010), (random trees, percolation, contact process, random clusters, Ising model, voter model)

The 10 papers that will make you a social expert:

1. S.Milgram, "The small world problem," Psychology today, 1967.

- 2. M. Granovetter, "The strength of weak ties: A network theory revisited," Sociological theory, vol. 1, pp. 201–233, 1983.
- 3. M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a Feather: Homophily in Social Networks," Annual review of sociology, vol. 27, pp. 415–444, Jan. 2001.
- M. O. Lorenz, "Methods of measuring the concentration of wealth," Publications of the American Statistical Association, vol. 9, no. 70, pp. 209–219, 1905. + H. Simon, "On a Class of Skew Distribution Functions," Biometrika, vol. 42, no. 3, pp. 425–440, 1955.
- 5. R. I. M. Dunbar, "Coevolution of Neocortical Size, Group-Size and Language in Humans," Behav Brain Sci, vol. 16, no. 4, pp. 681–694, 1993.
- 6. D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory.," Psychological Review, vol. 63, no. 5, pp. 277–293, 1956.
- 7. M. Granovetter, "Threshold Models of Collective Behavior," The American Journal of Sociology, vol. 83, no. 6, pp. 1420–1443, May 1978.
- B. Ryan and N. C. Gross, "The diffusion of hybrid seed corn in two Iowa communities," Rural sociology, vol. 8, no. 1, pp. 15–24, 1943. + S. Asch, "Opinions and social pressure," Scientific American, 1955.
- 9. R. S. Burt, Structural Holes: The Social Structure of Competition. Harvard University Press, 1992.
- 10. F. Galton, "Vox Populi," Nature, vol. 75, no. 1949, pp. 450–451, Mar. 1907.

Programming tools for social network analysis:

- NetworkX (http://networkx.lanl.gov) is a software package for Python that contains several functions for network analysis. It also contains documentation, including Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, "Exploring network structure, dynamics, and function using NetworkX", in Proceedings of the 7th Python in Science Conference, pp. 11–15 (2008), and was recently used in the book Social Network Analysis for Startups: Finding connections on the social web, M. Tsvetovat, A. Kouznetsov, O'Reilly (2011).
- NodeXL (http://nodexl.codeplex.com) is a template that can be used within Excel 2007 and 2010, it was recently covered in a help book D. Hansen, B. Shneiderman, M. A. Smith, Analyzing Social Media Networks with NodeXL: Insights from a Connected World, Morgan Kauffman
- SNAP (http://snap.stanford.edu) is a network analysis and mining library written in C++, it also contains some help/tutorial and slides from an introductory lectures are at:

http://www.stanford.edu/class/cs224w/tutorial/file/SNAP_slides.pdf
To boost the cool factors of your graphs, you may also want to check visualizations tools (and cool examples) like Graphviz (http://www.graphviz.org) or Data Driven Documents (http://d3js.org) (ex-Protovis, http://mbostock.github.com/protovis/)

Where to find data sets about social networks

• Twitter makes about 5% of their public tweets available through the gardenhose API. Check it out at:

https://dev.twitter.com/docs/streaming-api/methods

• Max Planck Institute has made data from IMC 2007 paper, WOSN 2008 papers, WWW 2009 paper, and WOSN 2009 paper, as well as Alan Mislove's PhD Thesis publicly available. Details at:

http://socialnetworks.mpi-sws.org/

• KONECT (The Klobenz Network Collection) contains various sort of networks from online and offline cases, they are available at the following URL

http://konect.uni-koblenz.de/

• Stanford Large Network Dataset Collection makes several data sets (not limited to social network) available at the following URL

http://snap.stanford.edu/data/index.html

• MLData is another repository website

http://mldata.org/

• The following is pasted from J. Kleinberg's webpage

There are a number of interesting network datasets available on the Web; they form a valuable resource for trying out algorithms and models across a range of settings.

Collaboration and citation networks: For the 2003 KDD Cup competition, Johannes Gehrke, Paul Ginsparg, and I provided a dataset based on the arXivpre-print database, which allows one to study the networks of co-authorships and citations among a large community of physicists. Here is the KDD Cup dataset and a paper describing the competition in more detail.

http://www.cs.cornell.edu/projects/kddcup/

http://www.cs.cornell.edu/home/kleinber/chrono.html

Internet topology: The network structure of the Internet can be studied at several levels of resolution. Here is a dataset at the autonomous system (AS) level.

http://topology.eecs.umich.edu/data.html

Web subgraphs: There are many such datasets available for download. One set is maintained by Panayiotis Tsaparas; the experiments that used this data are described in his Ph.D. thesis, and in other papers linked from his home page.

http://www.cs.toronto.edu/~tsap/experiments/download/download.html

P. Tsaparas, Link Analysis Ranking, Ph.D. Thesis, Department of Computer Science, University of Toronto, 2004.

Semantic networks: Free association datasets for words have been collected by cognitive scientists; these are constructed by compiling the free responses of test subjects when presented with cue words. (For example, a test subject presented with the cue word ice' might react with the word cold,' cream,' or water.')

http://w3.usf.edu/FreeAssociation/

University of South Florida Free Association Norms.

• The following is pasted from Max Planck Institute website:

Data from our IMC 2007 paper, our WOSN 2008 papers, our WWW 2009 paper, our WOSN 2009 paper, and Alan Mislove's PhD Thesis is publicly available by emailing Alan Mislove at amislove (at) mpi-sws (dot) org. Each of the data sets has been anonymized to protect the privacy of the social network users.

Alan Mislove, Massilmiliano Marcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee. Measurement and Analysis of Online Social Networks. In Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC'07), San Diego, CA, October 2007.

http://www.mpi-sws.mpg.de/~amislove/publications/SocialNetworks-IMC.pdf

Meeyoung Cha, Alan Mislove, Ben Adams, Krishna P. Gummadi. Characterizing Social Cascades in Flickr. In Proceedings of the 1st Workshop on Online Social Networks (WOSN'08), Seattle, August 2008.

http://www.mpi-sws.mpg.de/~amislove/publications/Cascades-WOSN.pdf

Meeyoung Cha, Alan Mislove, and Krishna P. Gummadi. A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. InProceedings of the 18th Annual World Wide Web Conference (WWW'09), Madrid, Spain, April 2009.

http://www.mpi-sws.org/~amislove/publications/Flickr-WWW.pdf

Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. Characterizing User Behavior in Online Social Networks. InProceedings of Usenix/ACM SIGCOMM Internet Measurement Conference (IMC), Chicago, Illinois, November 2009.

http://an.kaist.ac.kr/~mycha/docs/imc126-benevenuto.pdf

• The following is pasted from J. Leskovec's course resource section:

Snap network datasets http://snap.stanford.edu/data

Yahoo! Webscope Catalog of datasets http://www.stanford.edu/class/cs345a/YahooData.pdf

DBLP http://dblp.uni-trier.de/xml/

KDD Cup Dataset http://www.cs.cornell.edu/projects/kddcup/datasets.html

Internet Topology AS Graphs http://topology.eecs.umich.edu/data.html

Wikipedia http://users.on.net/~henry/home/wikipedia.htm or http://wiki.dbpedia.org/Downloads33

Movie Ratings: http://www.imdb.com/interfaces or http://www.grouplens.org/taxonomy/term/14

Trust network datasets http://www.trustlet.org/wiki/Trust_network_datasets

Mark Newman's pointers: http://www-personal.umich.edu/~mejn/netdata/