F19 ATCM3385 – Social Networks

Syllabus 2019-08-18 version

Instructor:
Dr. Maximilian Schich, Associate Professor in ATEC
Email: mxs135130@utdallas.edu (subject line "[ATCM3385 SocNets]") | Web: http://www.schich.info
Office hours: after class or by appointment

Time: Tuesdays & Thursdays 11:30 am – 12:45 pm | **Room:** ATC 2.101

Box folders:
Assignment upload: https://utdallas.box.com/v/F19-ATCM3385-upload (expires 2019-12-31)
Material download: https://utdallas.box.com/v/F19-ATCM3385-download (expires 2019-12-31)

Course description & learning outcomes
The ATCM 3385 *Social Networks* course introduces the theoretical perspectives and practical applications of the study of social networks with emphasis on the impact of communication technologies on the creation, maintenance, and transformation of social networks in contemporary media environments. Students will also learn to analyze and visualize networks.

Required Materials
This course will build on two required texts. *Newman's book chapter provides a gentle yet foundational introduction to "Social Networks".* The chapter is non-technical (aka no math), while the book as a whole is one of "the" introductions and references in network science. As networks are everywhere, Newman's book is likely to become a long-term friend in your bookshelf, no matter if you develop as an artist, a humanist, an engineer, or a scientist. *Brath and Jonker's book is a practical introduction to basic network analysis and visualization.* It teaches marketable skills, and as such provides a practical foundation for the semester-long project assignment. Hints to find further sources and a number of classics regarding our topic of social networks are attached at the end of this syllabus.

1. Newman, Mark. "Chapter 4. Social Networks." In ***Networks. Second Edition***. Oxford university press, 2018. pp. 47-69 [ISBN: 978-0198805090] [>10,000 citations]
2. Brath, Richard, and David Jonker. "***Graph analysis and visualization: Discovering Business Opportunity in Linked Data.***" John Wiley & Sons, 2015. [ISBN: 978-1118845844]

Schedule, assignments, and course specific policies
The full ***course schedule*** including submission reminders is detailed on page 2. The ***assignments*** are specified on pages 3 and 4. The ***course-specific policies*** on page 5 complement the standard UT Dallas policies that every student must know and follow (available at http://go.utdallas.edu/syllabus-policies). The ***appendices*** on pages 6 to 8 provide useful hints regarding additional literature, useful sources of data, and tools feeding into the semester-long research project.

Course schedule

Week	Tuesday	Theory	Thursday	Practice
1	2019-08-20	Intro / What? Why? How?	2019-08-22	How are we connected?
2	2019-08-27	Social Network Literature	2019-08-29	Data preparation (form & shape)
3	2019-09-03	Types of Social Networks	2019-09-05	Point & Click tools
4	2019-09-10	SNA: Classic Methods & Concepts	2019-09-12	Stats & Layout
5	2019-09-17	NetSci: Tails, Small Worlds, etc.	2019-09-19	Visual Attributes
6	2019-09-24	Honest Signals & Reality Mining	2019-09-26	Explore & Explain
	2019-09-30	Class social network analysis & preliminary research project slides due submit `YourNetID-a1-20190930.pdf` & `YourNetID-a2-20190930.pdf` * (due midnight)		
7	2019-10-01	Computational Social Science	2019-10-03	Preliminary project presentations
8	2019-10-08	Health, Happiness, & Friendship	2019-10-10	Python & D3.js
9	2019-10-15	Running, Dating, & Diversity	2019-10-17	Visual Analysis
10	2019-10-22	Social Images & Multimedia	2019-10-24	Images to Networks
11	2019-10-29	Modularity, Nestedness, Confusion	2019-10-31	Community Finding
12	2019-11-05	Temporal Networks	2019-11-07	Time/Flow/Dynamics
13	2019-11-12	Network Multiplicity	2019-11-14	Networks & Space(s)
14	2019-11-19	Game Theory & Misinformation	2019-11-21	Big Data & Animation
	2019-11-26	Fall break	2019-11-28	Fall break
	2019-12-02	Final project slides due submit `YourNetID-a2-20191202.pdf` * (due midnight)		
15	2019-12-03	Final project presentations	2019-12-05	Final project presentations
	2019-12-16	Final grades available in Orion		

*** Assignment submission guideline:**
Generic and short: Create the file with the filename as specified above, replace "YourNetID" with your NetID, and submit your assignment to https://utdallas.box.com/v/F19-ATCM3385-upload. – Please be aware that non-compliant filenames will not count as submitted. Particularly avoid submitting files such as "word.doc", "a1.pdf", etc. – Within the PDF files, always include all NetIDs of your group, the course number, the assignment number, the due date, and your project title on the first page. Paginate the whole file.

Descriptions and timelines are subject to change at the discretion of the professor.

Assignments

The two assignments below are (almost literally) adopted from another outstanding textbook. It is equally general as the required book by Newman and is another likely long-term friend in your bookshelf:

3. Barabási, Albert-László with Márton Pósfai. "**Network Science.**" Cambridge University Press, 2016. pp. viii-ix. http://networksciencebook.com/ (Preface)

Assignment a1: Class social network analysis

As a warmup to network analysis, you will analyze an anonymized version of our class social network, prepared together at the begin of the semester. The assignment is to analyze the class social network properties using the network analysis and visualization tools you have acquired until midterm. The assignment allows you to explore a relatively small network that you are invested in and understand. The assignment offers a preparation for the more extensive network analysis you will perform in your final research project. The homework is assigned after hands-on practice on software in class, so you will be already familiar with the online tools available for network analysis.

Assignment a2: Semester-long Research Project

The final project is the most rewarding part of the class, offering the students the opportunity to combine and utilize all the knowledge they acquired. Students are asked to select a network of interest to them, map it out and analyze it. Some procedural details enrich this assignment: [a] The project is carried out in pairs. If the class composition allows, the students are asked to form professionally heterogenous pairs: students from different pathways are asked to work together, like a student who can code with a student who can draw. This forces the students to collaborate outside their expertise level and comfort zone, a common ingredient of interdisciplinary research. The instructor does not do the pairing, but students are encouraged to find their partners. [b] A few weeks into the course one class is devoted to preliminary project presentations. Each group is asked to offer a five-minute presentation with no more than five slides, offering a preview of the dataset they selected. Students are advised to collect their own data - simply downloading a single dataset already prepared for network analysis is not acceptable. Indeed, one of the goals of the project is to experience the choices and compromises one must make in network mapping. Manual mapping is allowed, such as looking up the interaction of characters in a novel or a historical text. Digital mapping is encouraged, like scraping data from a website or a database that is not explicitly organized as a network map, but the students must reinterpret and clean the data to make it amenable for network analysis. For example, one can systematically scrape data from Wikipedia to identify relationships between writers, scientists or concepts. Alternatively, you can compare at least two existing network datasets, which puts more emphasis on analysis and visualization. [c] It is important to always emphasize that the purpose of the final project is to test a student's ability to analyze a network. Consequently, students must stay focused on exploring the network aspect of the data, and avoid being carried away by other tempting questions their dataset poses that would take them away from this goal. [d] The course ends with the final project presentations.

Preliminary research project presentation & submission

Each group has 5 minutes to present their preliminary project. Time limit is strictly enforced. Present 5 slides in no more than five minutes: ***[1]*** On the first slide, give the *presentation date*, your *project title*, the *names & NetIDs* of all group members; ***[2]*** Introduce your network, discussing its nodes and links; ***[3]*** Tell us how you will collect the data and estimate the size of the network (N, L), i.e. number of nodes and links. Make sure that N > 100. ***[4]*** Tell us what questions you are planning to ask. We understand that they may change as you advance with your project and the class. ***[5]*** Tell us why you care about your network.

(final deliverable see next page)

Final research project presentation & submission

Each group has 10 minutes to present their final project. Time limit is strictly enforced. ***[i]*** On the first slide, give the *presentation date*, your *project title*, the *names & NetIDs* of all group members; ***[ii]*** Tell us about your data and the data collection method. Show an entry of the data source to offer a sense of where you started from. ***[iii]*** Analyze the network: Measure N, L, and their time dependence if you have a time dependent network; degree distribution, average path length, clustering coefficient C(k), component sizes, the weight distribution P(w) if you have a weighted network. Visualize the network with sense-making layouts as a node-link diagram, as a matrix, etc.; enrich the nodes and links with metadata and measurements; visualize communities; measure and visualize other properties whichever is appropriate for your project. ***[iv]*** It is not sufficient to simply measure and visualize things - you need to discuss the insights you gained, always asking [a] What was your expectation? [b] What is the proper random reference? [c] How do the results compare to your expectation? [d] What did you learn from each quantity?

There is no need to write a report - submit the presentation as a pdf file (see schedule on page 2 for details).

(continued on next page)

Course-specific policies

Assignment submission:
To get full credit, please follow the submission guidelines on the schedule page of this syllabus. Please be aware that you will lose credit, if you don't follow the submission file-name convention as given in the schedule. Following the file-name convention is required to disambiguate and allocate credit as all uploads go to the same Box folder. Submissions with file names such as "word.pdf", without due date, without assignment specification, and without your NetID in the file name and on the first page of the PDF file will not count as submitted for grading.

Relative Shares Contributing to Your Grade:
30% Participation & Attendance + 10% Project presentation + 60% Final project = TOTAL 100%

Assignment specific grading criteria:
[1] Use of network tools (completeness / correctness);
[2] Ability to extract information insights from your data using the network tools.
[3] Overall quality of the project/presentation.

General grading criteria:
All work should conform to professional and ethical standards, so "proofread" and edit work that you submit in this class for clarity, mechanics, and style issues. This applies to graphics/visuals or any other chosen genre of expression, much like text. Professionalism also means that you use appropriate source citations wherever and whenever necessary. You should not submit any work for this course that you developed for another course without written permission from the instructor(s) of both courses. While you may explore topics across courses, the work you submit for this course should be substantially different from the work that you submit in any other course. All individual grades are scored out of 100 points. Because the grades are weighted, a simple average will not determine your grade.

Grade Rubric:
To make a C or less, simply do not follow the assignment instructions, refuse to collaborate in groupwork, disregard filename and file format conventions, and/or turn your project in after the deadline. To make a B, follow the assignment instructions, avoid "spelling and grammar errors" in writing and graphics, follow filename and file format conventions, turn it in on time, and include academic references where necessary. To make an A, aim for excellence in terms of presentation and content, including references and graphical layout.

Participation & Attendance:
More than just attendance, this grade reflects how you share your ideas, participate in classwork, engage your classmates, and behave with respect toward them. Your comments and insights contribute to the class' success, so you must attend class prepared to discuss material as a public, interactive process. Everyone benefits when you engage alternative perspectives, challenge interpretations, and invite constructive arguments as long as you do so respectfully. This grade explicitly includes civility and professionalism in all course communication and behavior, such as contributing to conversations, respecting others' opinions, working together in a spirit of cooperation, and actively listening to those who are speaking. Some of the ways you can demonstrate your skills in this area include (but certainly are not limited to): Keeping the class in the foreground of your attention; Showing respect to your peers and to the instructor in your listening and communicating behaviors; Participating actively in class rather than simply waiting to be called on; Adding value with your contributions to discussion, such as connecting disparate ideas, bringing topical information to the table, and asking insightful questions; Taking responsibility for the consequences of your choices and actions; Demonstrating a strong work ethic by engaging all ungraded work (such as completing assigned readings, minor homework, and in-class exercises) with a mindfulness and timeliness to reflect a professional approach to the class.

Late Work:
If a personal situation arises during the semester that may affect your classroom performance, please talk to me sooner rather than later. In other words, be proactive. If you wait until the end of the semester, I cannot help you. I can work with you more easily if you speak to me when the situation arises. I can't help you if I don't know you need help. You can have make-up privileges for university-specified circumstances, including religious holy days and university-sponsored activities. If you must miss class or deadlines for such reasons, you must make arrangements with me in advance.

Digital Devices:
You may use laptops, tablets, cell phones, and other digital devices so long as you use them responsibly and respectfully and particularly if you use them to enhance the class-experience. If your digital device disturbs other students or interferes with your ability to participate meaningfully in class activities, you may be asked to remove the distraction and/or leave class, thus losing credit for any of the day's activities. Please silence device notification settings before class begins and refrain from accepting calls in class. However, if I see you are texting or emailing with a friend, watching a movie, or playing video poker, I will not be happy and may ask you to leave from that day's class.

APPENDIX 1: Hints to find additional literature (optional yet highly recommended)

Choosing and working on your project assignment, you may get inspiration from existing literature, which can be hard to find. A classic approach is mining relevant and more novel literature via "forward bibliography", starting from "highly cited" older literature. Here's an example: Search for "Wasserman Faust 1994" at scholar.google.com, click on "Cited by ...", check "Search within citing articles", and then search for a chosen topic within the search box. Searching for "Twitter" within articles citing Wasserman & Faust returns the article by Huberman at al. below, which is spot on. You could repeat the whole procedure, until you arrive in the present. It may make sense to write down interesting terms you learn on the way, or to think about alternative terms, such as "mass communication" replacing "news media" in your searches. Try using such found terms and alternative terms to restrict and broaden your search. Starting from a given publication, the procedure of "forward bibliography" allows you to explore the "river-delta" of follow-up research; simply looking into the bibliography within a given publication takes you in the other direction, to the equally branched out sources of research in the past. Exploring literature via citations in this way, you are already navigating a network that is deeply intertwined with social influence between scholars. Here are some highly-cited classics related to "Social Networks", which provide a good starting point for "forward bibliography":

4. Boyd, Danah M., and Kate Crawford. "**Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon.**" *Information, communication & society* 15, no. 5 (2012): 662-679. [>2900 citations]
5. Papacharissi, Zizi, ed. "***A networked self: Identity, community, and culture on social network sites.***" Routledge, 2010. [698 citations]
6. Jackson, Matthew O. "***Social and economic networks.***" Princeton university press, 2010. [>4900 citations]
7. Huberman, Bernardo, Daniel M. Romero, and Fang Wu. "**Social networks that matter: Twitter under the microscope.**" *First Monday* 14, no. 1 (2009). [>1500 citations]
8. Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis et al. "**Computational social science.**" *Science* 323, no. 5915 (2009): 721-723. [>2600 citations]
9. Boyd, Danah M., and Nicole B. Ellison. "**Social network sites: Definition, history, and scholarship.**" *Journal of computer-mediated Communication* 13, no. 1 (2007): 210-230. [>16800 citations]
10. Eagle, Nathan, and Alex Sandy Pentland. "**Reality mining: sensing complex social systems.**" *Personal and ubiquitous computing* 10, no. 4 (2006): 255-268. [>2800 citations]
11. Freeman, Linton. "**The development of social network analysis.**" *A Study in the Sociology of Science* 1 (2004). [>2700 citations]
12. Wasserman, Stanley, and Katherine Faust. "***Social network analysis: Methods and applications.***" Vol. 8. Cambridge university press, 1994. [>33,000 citations]
13. Dunbar, Robin I.M. "**Coevolution of neocortical size, group size and language in humans.**" *Behavioral and brain sciences* 16, no. 4 (1993): 681-694. [>2500 citations]
14. Granovetter, Mark S. "**The strength of weak ties.**" In *Social networks*, pp. 347-367. Academic Press, 1977. [52,500 citations]
15. Milgram, Stanley. "**The small world problem.**" *Psychology today* 2, no. 1 (1967): 60-67. [>8700 citations]

As a caveat, consider that not all relevant literature is highly cited, due to the fact that "the most frequent pattern in social spreading is extinction". Also consider that relevant literature may fail to contain the term "social network". This is due to the fact that some say "tomäto" when others say "tomāto". Inversely the same word may have different meanings in different fields. This makes it hard yet often highly worthwhile to "excavate" obscure but relevant literature, both preceding our current conventions of language or just living in a different "filter bubble". See for example:

16. Oettinger, A. G. "**Information resources: knowledge and power in the 21st century.**" *Science* (New York, NY) 209, no. 4452 (1980): 191-198. (84 citations)
17. Moreno, Jacob L., and Helen H. Jennings. "**Statistics of social configurations.**" *Sociometry* (1938): 342-374. (298 citations)

Finally, if you are looking for further relevant terminology and ideas, the "wisdom of the crowd" may be valuable, as evident in Wikipedia under "Social networking service", "Social network", and "Social network analysis". So, check it out. But always consider that "social influence can undermine the wisdom of crowd effect".

APPENDIX 2: Useful sources of data

In recent years, a great variety of social network data have become available openly online, ready for reuse and to learn. For your semester-long research project, I highly recommend you take a look at the repositories below, where a large number and variety of datasets can be found, from the very small to the seriously big. The data sets are often associated with the research literature for which the data have been collected originally. Many of the data sets are more or less convenient to download & process, encouraging further experimentation and analysis, including novel visualizations.

Here is a number of dedicated collections of (social) network data sets:

Colorado Index of Complex Networks (ICON) – https://icon.colorado.edu/

Koblenz Network Collection KONECT – http://konect.uni-koblenz.de/

Network Repository – http://networkrepository.com/

Stanford Large Network Dataset Collection (SNAP) – https://snap.stanford.edu/data/index.html

UCI Network Data Repository – http://networkdata.ics.uci.edu/index.html

(Social) network data is also found in more general repositories:

Nature Scientific Data – https://www.nature.com/sdata/

Figshare – https://figshare.com (search for "social network dataset")

Kaggle – https://www.kaggle.com/datasets?search=social+network

"The 70 Online Databases that define our planet" – https://www.technologyreview.com/s/421886/

Here are two Interesting personal collections:

Network data on Mark Newman's personal website – http://www-personal.umich.edu/~mejn/netdata/

Awesome Network Analysis datasets – https://github.com/briatte/awesome-network-analysis#datasets

Increasingly open data is also directly attached in the online supporting material of scientific journal papers. In other cases, data analyzed in a paper is made available in public repositories, such as Figshare above. In yet other cases, data is published in more obscure places. See for example:

Facebook 100 – https://archive.org/details/oxford-2005-facebook-matrix

Even though ever more data is readily available, it is of course still worthwhile to collect your own data. And of course, sampling a small social network qualitatively in person may be just as interesting as a large-scale crawl of social media service with billions of users.

APPENDIX 3: Access to tools

Network analysis and visualization has gone mainstream across many fields in the last two decades. As a consequence, an increasing number of more or less convenient tools have become available. The point here is not to choose the right one and only. All the tools listed below have their strengths and weaknesses. This list is obviously also exhaustive, but only provides a number of outstanding examples. We will focus on **must see & try** examples in bold.

Obtaining and processing data for analysis:

- **Microsoft Excel** – https://www.utdallas.edu/oit/o365/ [both over- & underestimated]
- OpenRefine – http://openrefine.org/ ["A free, open source, powerful tool for working with messy data"]
- Data Science at the Command Line – https://www.datascienceatthecommandline.com/ [to get and process data]

Graph databases:

- **Wikidata** – https://www.wikidata.org/wiki/Wikidata:Tools [not only data, but also a toolkit]
- neo4j – https://neo4j.com/developer/get-started/ [think *big*, think *network* not *table*, think *walk* not *join*]

Point & click node-link diagrams & basic network analysis:

- **Cytoscape** – https://cytoscape.org/ [Think: The Photoshop for networks; open source; great UX]
- **Gephi** – https://gephi.org/ [Think: The Corel Draw for networks ☺ ; better layouts; more complicated UX]
- NodeXL – https://www.smrfoundation.org/nodexl/ [network analysis within Excel]
- UCINET – https://sites.google.com/site/ucinetsoftware/home [a classic in social network analysis]
- Pajek – http://vlado.fmf.uni-lj.si/pub/networks/pajek/ [another classic in social network analysis]
- Visione – https://visone.info/ [yet another classic in social network analysis]
- LINKURIOUS – https://linkurio.us/ [a recent proprietary solution]
- ORA-LITE & ORA-PRO – http://www.casos.cs.cmu.edu/projects/ora/ & http://netanomics.com/ [open & proprietary]
- Nodegoat – https://nodegoat.net/ [a specialized data management tool for historical (social) network analysis]
- Palladio – http://hdlab.stanford.edu/palladio/ [Visualize complex historical data with ease]

Point & click matrix visualization:

- **Bertifier** – https://aviz.fr/bertifier [will change your life]
- **Microsoft Excel** – https://www.utdallas.edu/oit/o365/ [underestimated & neglected by developers for this use]
- Google FACETS DIVE – https://pair-code.github.io/facets/ [could be hacked to do social networks; think how?]
- WebMeV (Multiple Experiment Viewer) – http://mev.tm4.org [it never hurts to steal a tool from systems biology]

Using code for network analysis & visualization:

- **Jupyter notebooks & Python, R, etc.** – https://jupyter.org/ & https://github.com/networkx/notebooks
- **Observable notebooks & D3.js** – https://observablehq.com/ [D3.js; search: "social network" for examples]
- Vega-Lite: A Grammar of Interactive Graphics – https://vega.github.io/vega-lite/ [revolutionary for standard plots]
- Python magic words: scipy, numpy, pandas, scikit-learn, statsmodels, graph-tool, igraph, networkX, nerworkit, snap, *matplotlib*, *seaborn*, *bokeh*, *altair*. [There is a similar ecocosmos for R, Matlab, Mathematica, and other alternatives].

Network visualization showcases for inspiration:

- **Visualcomplexity.com** – http://www.visualcomplexity.com/vc/index.cfm?domain=Social%20Networks [the Louvre]
- IBM Watson News Explorer – http://news-explorer.mybluemix.net/ [real-time network mapping]
- Twitter Company by Santiago Ortiz – http://moebio.com/newk/twitter/ [still inspiring]

Finally, there is the eternal toolkit that will never go out of fashion:

- **Pencil, paper, your own thought, in conversation with peers!**
- Linear algebra ☺