

Fall 2020 – Tuesday – 12:30-3:20pm

Section: 28836D

Location: On Zoom (until further notice)

Instructor: Prof. Emilio Ferrara

Office: Zoom (until further notice)

Office Hours: Tuesday morning, before class, on request

Contact Info: emiliofe@usc.edu

Zoom link: <https://usc.zoom.us/j/4987174380>

Teaching Assistant / Grader: Emily Chen

Office: On Zoom (until further notice)

Office Hours: On request, via Zoom & Slack channel

Contact Info: echen920@usc.edu

I. Course Description

Learn how to unleash the full power and potential of Social Web data for research and business application purposes!

The Social Web pervades all aspects of our lives: we connect and share with friends, search for jobs and opportunities, rate products and write reviews, establish collaborations and projects, all by using online social platforms like Facebook, LinkedIn, Yelp and GitHub. We express our personality and creativity through social platforms for visual discovery, collection and bookmarking like Tumblr and Pinterest. We keep up-to-date, communicate and discuss news and topics of our interest on Twitter and Reddit.

In this course, we will explore the opportunities provided by the wealth of social data available from these platforms. You will learn how to acquire, process, analyze and visualize data related to social networks and media activity, users and their behaviors, trends and information spreading. This journey will bring through the lands of data mining and machine learning methods: supervised and unsupervised learning will be applied to practical problems like social link analysis, opinion mining, and building smart recommender systems. We will explore open-source tools to understand how to extract meaning from human language, use network analysis to study how human connect and discover affinities among people's interests and tastes by building interest graphs.

Proficiency in a programming language is expected. All coursework is done in Python. (This is not an introductory coding course. Please refer to DSCI classes like DSCI 510 for intro coding).

II. Student Learning Outcomes

Taking this course, you should expect to learn about:

- Networks.
 - o Statistical descriptors of networks: link analysis, centrality, and prestige.
 - o Network clustering: modularity and community detection.
 - o Dynamics of information and epidemics: threshold and information cascade models.
 - o Network biases and network manipulation: paradoxes, bots, disinformation.
 - o Network visualization algorithms: spring-like layouts, multidimensional scaling, Gephi.

- Applications of texts and documents analysis.
 - o Natural Language Processing and Part-of-speech tagging.
 - o Sentiment Analysis.
 - o Topic Modeling.

- Supervised learning: Crush course on Data Classification.
 - o Eager vs. Lazy learning: Decision Trees
 - o Ensemble methods: Random Forest
 - o Classification performance evaluation: Precision/Recall/F1, Accuracy and ROC Curves.

- Unsupervised learning: Crush course on Clustering Data.
 - o Distance and similarity measures & K-means clustering.
 - o Hierarchical Clustering and Dendrograms.
 - o Clustering performance evaluation.

All topics will be explored from an applied, practical, computational perspective. This will allow the interested student to deepen the rigorous theoretical implications of the methods in other courses offered by USC (for example, CSCI-567 Machine Learning). Students interested in ethics and privacy of data should consider CSCI-599 AI in Society. Throughout the course, we will deliver several “hands-on” sessions with live coding, data analysis, and problem-solving!

Recommended Preparation: course work or background on coding in Python is advisable; otherwise, experience with any other programming language is strongly recommended!

III. Required Readings and Supplementary Materials

Recommended textbooks (total Amazon price [new/used]: \$100/\$60)

1. Web Data Mining (2nd Ed.) —by Bing Liu (Amazon price [new/used]: \$48/\$35)
2. Mining the Social Web (2nd Ed.) —by Matthew A. Russell (Amazon price [new/used]: \$27/\$15)
3. Programming Collective Intelligence —by Toby Segaran (Amazon price [new/used]: \$25/\$10)

4. Network Science Book —by La szl  Barab si
(FREE: <http://barabasilab.neu.edu/networksciencebook/>)
5. Dive into Python —by (FREE: <http://www.diveintopython.net/>)

Some details: Book 1 will provide insights on methods and approaches studied throughout the course from a machine learning perspective; Books 2 and 3 will serve as recipe books to effectively design and make those methods work with Social Web data; Books 4 and 5 are free resources we will exploit to gather additional material on networks and Python programming. Technical, recommended (non-required) Python “cookbooks”:

- Python Data Visualization Cookbook —by Igor Milovanovi  (ebook: \$14)
- Learning IPython for Interactive Computing and Data Visualization —by Cyrille Rossant (ebook: \$10)
- Learning scikit-learn: Machine Learning in Python —by Ra l Garreta and Guillermo Moncecchi (ebook: \$10)

The readings for this class are provided in Appendix A of this syllabus.

Please refer to the Website for links to the articles:

<http://www.emilio.ferrara.name/data-science-for-communication-social-networks/>

IV. Description and Assessment of Assignments & Assignment Submission Policy

Attendance policy

- Students are strongly encouraged to attend live sessions via Zoom. This is the best way to engage with the course and keep up to date with class activities and assignments. However, there will be no penalty for failing to attend live sessions, and students who miss live sessions will be able to keep up with the class by reviewing class recordings and engaging through asynchronous class activities and assignments. Note that university guidelines dictate that faculty should only maintain normal attendance, participation, and assessment expectations for students when the class time falls within reasonable learning hours in the student’s time zone, defined as 7:00am to 10:00pm in the student’s time zone.

Participation

- Participation is part of your grade. The current modality of teaching includes synchronous and asynchronous lessons, assignments, exercises, and evaluation processes, and each affords different types of participation. In a synchronous zoom environment, when proper, participation is synchronous. This may include talking, when appropriate, during a zoom meeting, or using the chat function to generate or participate in discussions. Blackboard offers different types of participation, including participating in forums and completing asynchronous assignments.
- The distant nature of these learning environments may prompt us to forget that we are a community and that each of us and our ideas deserve respect. For this reason, it is imperative that you remember to respect the opinion of others, regardless of how much you disagree.

- Class participation and engagement are essential ingredients for success in your academic career, therefore during class please set aside or turn off cell phones. Please avoid other distractions during the lectures.

Zoom etiquette

- Although you are not obligated to turn your camera on, we highly recommend it. Please wear appropriate clothing. Please keep your microphone off during zoom class, except when you're asked to unmute for discussion or questions. Please use appropriate backgrounds.

Other policies

- Students will be expected to do all readings and assignments, and to attend all meetings unless excused, via email, at least 24 hours prior. Assignments shall be submitted through Blackboard unless otherwise communicated by the instructor.
- The following misconducts will automatically result in a zero weight for that component of the grade: (1) failing to attend class on the day of your presentation; (2) failing to attend meetings of your group's Hackathon and/or final presentation; (3) failing to submit your final paper by the expected date. Extenuating circumstances will normally include only serious emergencies or illnesses documented with a doctor's note.

Detailed assignments description

Assignments

Reaction debriefs: Each student will prepare a "synthesis and reaction" debrief in response to the weekly readings. This will be a brief note, aimed at summarizing in one paragraph the gist of the paper, and provide comments or inputs for discussions, including questions, critiques, and/or theoretical and methodological concerns or ideas. These will be used to guide the discussion session of each class. (Reaction debriefs are not graded)

Readings & discussion

During each lecture (starting lecture 2), one student will hold a 15m presentation on one of the daily reading of choice and will help moderate a discussion session about it. The list of readings is available at the end of the syllabus (see Appendix A).

Mid-Term Hackathon

The mid-term exam is in the form of a collaborative hackathon project. The goal is to develop crucial abilities such as:

- Intellectual development: leveraging expertise and multidisciplinary backgrounds, sharing ideas and knowledge.
- Teamwork skills: effective brainstorming, communication and presentation, and group problem-solving.
- Project management skills: ability to set goals, map progress, prototyping-delivery, and matching deadlines.

If possible, we suggest that participants form groups of 2 members with the goal of solving a single problem. Students are encouraged to form groups with members from different academic background when possible. Each group will propose or receive a different problem.

We will propose several problems of interest for the course, as well as receive your explicit solicitations, that should be agreed upon with the Instructor during the first 4 weeks, in the form of a short one-page proposal clearly stating:

- What is the problem?
- Why it is deemed relevant.
- How the group plans to solve the problem.
- Bibliographic references to at least one relevant related paper.

All project proposals will be subject to our approval. Groups will be assigned an approved project, either selected among those proposed by the Instructor, or by the group itself. Each group will receive a 30m slot for the presentation of their results, in which each member of the group is expected to discuss at least one critical task of the project. The grading of the projects will be in part based on crowd-sourced ratings attributed by other fellow students and submitted in anonymous form at the end of each presentation day.

Final Paper

A serious final paper will be expected. The manuscript will be at least 3,000 words (excluding references) and no more than 4,000 (excluding references) and will include appropriate figures and tables, and unlimited number of references. The work should cover the following points:

- Statement of the problem & Why the problem is important.
- How the problem was faced —including a description of methodology and dataset(s).
- Discussion of results, findings, and limitations of the study.
- Related literature & Final remarks/conclusions.

The final paper should be ideally based on the student's mid-term hackathon project. Text with other group members cannot be shared, figures/tables can be shared when appropriate with proper credit attribution. Grading will be based on soundness (both quality and quantity of original work). Groups of 2 students will be allowed to turn in a single joint-authored manuscript, in the format of a submission for an appropriate peer-reviewed journal or conference. Each author must contribute sufficient material to justify his/her "equal contribution" in the work. Both authors will receive the same grade for such manuscript.

V. Grading

a. Breakdown of Grade

Assignment	Points	% of Grade
Participation	15	15
Midterm exam	35	35
Final exam	50	50
TOTAL		100%

b. Grading Scale

94 to 100%: A	80% to 83%: B-	67% to 69%: D+
90% to 93%: A-	77% to 79%: C+	64% to 66%: D
87% to 89%: B+	74% to 76%: C	60% to 63%: D-
84% to 86%: B	70% to 73%: C-	0% to 59%: F

c. Grading Standards

Letter Grade	Description
A	Excellent; demonstrates extraordinarily high achievement; comprehensive knowledge and understanding of subject matter; all expectations met and exceeded.
B	Good; moderately broad knowledge and understanding of subject matter; explicitly or implicitly demonstrates good, if not thorough understanding; only minor substantive shortcomings.
C	Satisfactory/Fair; reasonable knowledge and understanding of subject matter; most expectations are met; despite any shortcomings, demonstrates basic level of understanding.
D	Marginal; minimal knowledge and understanding of subject matter; more than one significant shortcoming; deficiencies indicate only the most rudimentary level of understanding.
F	Failing; unacceptably low level of knowledge and understanding of subject matter; deficiencies indicate lack of understanding.

d. Grading Timeline

Grading Timeframe and Missing or Inaccurate Score Inquiries/Disputes

- For effective learning, students should receive timely feedback on assignments and exams. Therefore, every attempt will be made to grade assignments/exams and post grades within two weeks. Scores for all assignments and exams are regularly updated on Blackboard. You are responsible for notifying the Instructor **within one (1) week** of a score posting if you think a score is missing or inaccurate. Moreover, you only have this period of time to contest a score on an assignment/exam. If you fail to inquire/notify us of any discrepancy, missing score, or contest a score within one week of the date the score is posted, no further changes will be made.

VI. Course Schedule: A Weekly Breakdown

Important note to students: Be advised that this syllabus is subject to change - and probably will change - based on the progress of the class, events, and/or guest speaker availability, where relevant. Students should consult the Registration Calendar for dates regarding add/drop deadlines, fees, grading options, etc.

	Topics/Daily Activities & Readings	Deliverable/Due Dates
Week 1 Dates: 8/24-8/28	Introduction of the course & Planning Part 1— Networks <ul style="list-style-type: none"> Crash introduction to Networks—Statistical descriptors of networks 	(by 8/31/2020) Join Slack Channel Request Twitter API
Week 2 Dates: 8/31-9/4	Networks (continued) <ul style="list-style-type: none"> Network clustering. Modularity and community detection. Readings: Papers [27], [31], [22] and [4] Recommended Chapters: NSB:1 and NSB:2; NBS:9 and WDM:7.5	(by 9/7/2020) Forming a team and drafting a team project proposal
Week 3 Dates: 9/7-9/11	Networks (continued): <ul style="list-style-type: none"> Dynamics of information and epidemics spreading. Readings: Papers [26], [5], [6], [14] Recommended Chapters: NSB:10.1–10.3[pp.11–29] Hands-on session: mining Twitter. Readings: Papers [12] Recommended Chapters: MtSW:1[pp.5-26] Documentation: Twitter API (https://dev.twitter.com/) 	[Labor Day: Monday, September 7]

Week 4 Dates: 9/14-9/18	<p>Networks (continued):</p> <ul style="list-style-type: none"> • Networks and manipulation: bots, disinformation, emotional contagion <p>Readings: [30], [39] [18], and [20]</p>	
Week 5 Dates: 9/21-9/25	<p>Networks (continued):</p> <ul style="list-style-type: none"> • Network visualization algorithms. Readings: Papers [1] and [28] Recommended Chapters: PCI:12[pp.300–302(MDS)] • Hands-on session: tutorial on Gephi. Readings: Papers [3] Recommended Chapters: NSB:10.4–10.7[pp.30–58] Documentation: Gephi Wiki https://wiki.gephi.org/index.php/Main_Page 	
Week 6 Dates: 9/28-10/2	<p>Networks (continued):</p> <p>Guest speaker: Prof. Kristina Lerman</p> <ul style="list-style-type: none"> • Bias in networks: friendship paradoxes & perception bias – network structures bias perception. • Readings: Papers [33], [23], [24] <p>Ask Me Anything (AMA) session with guest</p>	
Week 7 Dates: 10/5-10/9	<p>Part 1—Text and Documents</p> <ul style="list-style-type: none"> • Crash intro to Natural Language Processing: Part-of-Speech Tagging. Readings: Papers [16] Recommended Chapters: WDM:6.5 and MtSW:5.3–5.5[pp.190–222] • Hands-on session: Tutorial on NLP 	
Week 8 Dates: 10/12-10/16	<p>Text and Documents (continued)</p> <ul style="list-style-type: none"> • Sentiment Analysis Readings: Papers [15] and [29] Recommended Chapters: MtSW:4[pp.135–180] • Topic Modeling Readings: Papers [2] and [10] Recommended Chapters: WDM:6.7 	[Fall Recess: Thursday, October 15 and Friday, October 16]
Week 9 Dates: 10/19-10/23	Mid-term Hackathon week	Midterm Hackathon presentations
Week 10 Dates: 10/26-10/30	No Classes	

Week 11 Dates: 11/2-11/6	Part 3—Supervised Learning <ul style="list-style-type: none"> Crash intro to Supervised learning. Readings: Papers [17], [19], and [11] Recommended Chapters: WDM:3.1 Eager vs. Lazy learning—Decision Trees. Readings: Papers [21] Recommended Chapters: WDM:3.2 and WDM:3.9 	
Week 12 Dates: 11/9-11/13	Supervised Learning (continued) <ul style="list-style-type: none"> Ensemble methods & Classification performance evaluation. Readings: Papers [9] and [7] Recommended Chapters: WDM:3.3 and WDM:3.10 	
Week 13 Dates: 11/16-11/20	Part 4—Unsupervised Learning <ul style="list-style-type: none"> Crash introduction to Unsupervised learning—Distance measures & K-means clustering. Readings: Papers [38] and [37] Recommended Chapters: WDM:4.1–4.3[pp.133–147] 	
Week 14 Dates: 11/23-11/27	Unsupervised Learning (continued) <ul style="list-style-type: none"> Hierarchical clustering & Dendrograms. Readings: Papers [25], [32], [34] Recommended Chapters: WDM:4.3–4.5[pp.147–155] 	[Thanksgiving Recess: Wednesday, November 25 to Sunday, November 29]
Week 15 Dates: 11/30-12/4	Unsupervised Learning (continued) <ul style="list-style-type: none"> Clustering performance evaluation Readings: Papers [35], [36] Recommended Chapters: WDM:4.6–4.10[pp.155–165] 	Students' Course evaluations
STUDY DAYS Dates: 12/5-12/8		Projects' presentations
FINAL EXAM PERIOD Dates: 12/9-12/16		Final paper's submission

VII. Policies and Procedures

Additional Policies

Synchronous session recording notice

- Live class sessions will be recorded and made available to students through Blackboard (including transcriptions). Please remember that USC policy prohibits sharing of any synchronous and asynchronous course content outside of the learning environment. As a student, you are responsible for the appropriate use and handling of these recordings under existing SCampus policies regarding class notes (<https://policy.usc.edu/scampus-part-c/>). These rules will be strictly enforced, and violations will be met with the appropriate disciplinary sanction.

Going back to Campus

- Although we are starting the semester with online instruction only, conditions may improve. In such case, courses listed as hybrid will give opportunity to students to attend class in person. This will happen only by following the strictest health guidelines and safety protocols. These are listed in the Trojans Return page. Please take the time to read this ahead so that you are prepared in case it is possible to return to in-person instruction.

Communication

Students are encouraged to contact the instructor outside of class, to arrange for office hours and group meetings. The preferred method will be via the official Slack channel of the course. Such requests are typically answered within 48 hours. Email requests are typically answered within 72 hours.

Statement on Academic Conduct and Support Systems

Academic Integrity Policy:

The School of Communication maintains a commitment to the highest standards of ethical conduct and academic excellence. Any student found responsible for plagiarism, fabrication, cheating on examinations, or purchasing papers or other assignments will be reported to the Office of Student Judicial Affairs and Community Standards and may be dismissed from the School of Communication. There are no exceptions to the school's policy.

Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

In addition, it is assumed that the work you submit for this course is work you have produced entirely by yourself, and has not been previously produced by you for submission in another course, without approval of the instructor.

Emergency Preparedness/Course Continuity in a Crisis

In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies. See the university's site on **Campus Safety and Emergency Preparedness**.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call

studenthealth.usc.edu/counseling

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call

suicidepreventionlifeline.org

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7

on call studenthealth.usc.edu/sexual-assault

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) - (213) 740-5086 | Title IX – (213) 821-8298

equity.usc.edu, titleix.usc.edu

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298

usc-advocate.symplcity.com/care_report

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity | Title IX for appropriate investigation, supportive measures, and response.

The Office of Disability Services and Programs - (213) 740-0776

dsp.usc.edu

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Campus Support and Intervention - (213) 821-4710

campussupport.usc.edu

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101

diversity.usc.edu

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

dps.usc.edu, emergency.usc.edu

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu

Non-emergency assistance or information.

Annenberg Student Success Fund

<https://annenbergsuccessfund.usc.edu/current-students/resources/additional-funding-resources>

The Annenberg Student Success Fund is a donor-funded financial aid account available to USC Annenberg undergraduate and graduate students for non-tuition expenses related to extra- and co-curricular programs and opportunities.

Appendix A: Reading list

- [1] S. Aral and D. Walker. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341, 2012.
- [2] D. M. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.
- [3] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.
- [4] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca. Network analysis in the social sciences. *Science*, 323(5916):892–895, 2009.
- [5] D. Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010.
- [6] D. Centola. An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272, 2011.
- [7] A. Cho. Ourselves and our interactions: the ultimate physics problem? *Science*, 325(5939):406, 2009.
- [8] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, 107(52):22436–22441, 2010.
- [9] V. Dhar. Data science and prediction. *Communications of the ACM*, 56(12):64–73, 2013.
- [10] P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *Science*, 301(5634):827–829, 2003.
- [11] P. Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.
- [12] W. Fan and M. D. Gordon. The power of social media analytics. *Communications of the ACM*, 57(6):74–81, 2014.
- [13] H. Garcia-Molina, G. Koutrika, and A. Parameswaran. Information seeking: convergence of search, recommendations, and advertising. *Communications of the ACM*, 54(11):121–130, 2011.
- [14] Lorenz-Spreen, P., Mønsted, B. M., Hövel, P., & Lehmann, S. (2019). Accelerating dynamics of collective attention. *Nature communications*, 10(1), 1759.

- [15] S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.
- [16] DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society*, 2(2), 2053951715602908.
- [17] N. Jones. Computer science: The learning machines. *Nature*, 505(7482):146, 2014.
- [18] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- [19] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.
- [20] A. D. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, page 201320040, 2014.
- [21] D. Lazer, R. Kennedy, G. King, and A. Vespignani. Big data. the parable of google flu: traps in big data analysis. *Science*, 343(6176):1203, 2014.
- [22] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al. Life in the network: the coming age of computational social science. *Science*, 323(5915):721, 2009.
- [23] Lee, E., Karimi, F., Wagner, C., Jo, H. H., Strohmaier, M., & Galesic, M. (2019). Homophily and minority-group size explain perception biases in social networks. *Nature human behaviour*, 3(10), 1078-1087.
- [24] Kooti, F., Hodas, N. O., & Lerman, K. (2014, May). Network weirdness: Exploring the origins of network paradoxes. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- [25] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11623–11628, 2005.
- [26] P. T. Metaxas and E. Mustafaraj. Social media and the elections. *Science*, 338(6106):472–473, 2012.
- [27] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

- [28] L. Muchnik, S. Aral, and S. J. Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.
- [29] Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115(49), 12435-12440.
- [30] Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96-104.
- [31] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
- [32] M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.
- [33] Feld, S. L. (1991). Why your friends have more friends than you do. *American Journal of Sociology*, 96(6), 1464-1477.
- [34] M. Schich, C. Song, Y.-Y. Ahn, A. Mirsky, M. Martino, A.-L. Barabási, and D. Helbing. A network framework of cultural history. *Science*, 345(6196):558–562, 2014.
- [35] C. Staff. Recommendation algorithms, online privacy, and more. *Communications of the ACM*, 52(5):10–11, 2009.
- [36] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53(8):80–88, 2010.
- [37] A. Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, 2012.
- [38] A. Vespignani. Predicting the behavior of techno-social systems. *Science*, 325(5939):425, 2009.
- [39] Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday*, 21(11-7).
- [40] Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of broadcasting & electronic media*, 57(1), 34-52.