

Métodos analíticos

El curso consiste de dos partes que exploran aspectos distintos de la teoría y práctica del análisis de datos. Una parte del curso se ocupa de los conceptos centrales y aplicaciones de la teoría moderna de redes; la otra, en explorar técnicas para el análisis de conjuntos de datos numéricos y textuales.

Objetivos

- Familiarizar al alumno con métodos de análisis de información y redes.
- Que el alumno pueda implementar algoritmos de análisis.
- Presentar al alumno distintas aplicaciones prácticas en distintas disciplinas.

Prerrequisitos

- Álgebra lineal.
- Conocimientos básicos de procesos estocásticos.
- Conocimientos básicos de sistemas dinámicos.
- Habilidades de programación en alguno de los siguientes lenguajes Python, Scala, R o Matlab (en ese orden).

Bibliografía general

- M. E. J. Newman. Networks: An Introduction. Oxford University Press, 2010.
- C. Manning, P. Raghavan and H. Schütze: Introduction to Information Retrieval, Cambridge University Press. 2008.
- C. Manning and H. Schütze: Foundations of Statistical Natural Language Processing, The MIT Press. 1999.
- A. Rajaraman and J. Ullman, Mining of Massive Datasets, Cambridge University Press, 2013.

Calificación

La calificación final será compuesta de la siguiente manera:

- 10% Participación en clase
- 10% Presentación del problema del proyecto final.
- 10% Presentación de avances en el proyecto final.
- 10% Presentación de resultados.
- 60% Reporte final. Se evaluarán los resultados, la claridad de la exposición y la presentación del documento.

Trabajo final

El trabajo final es una exploración a profundidad de un tema relacionado con lo visto en clase. El reporte escrito deberá elaborarse en LaTeX (con el formato proporcionado) y deberá tener calidad de publicación. Los alumnos entregarán avances de sus reportes los días de presentaciones —ambos: de los problemas y de los avances— para que los instructores puedan ver los avances y dar comentarios y poder orientar al alumno acerca de la presentación y el contenido. El tema del trabajo final será escogido por los alumnos en consulta con los instructores, y puede consistir en la implementación de un algoritmo, un estudio de datos con metodologías vista en clase o la replicación y explicación a detalle de algún resultado, por ejemplo. El trabajo final debe ser autocontenido; los conceptos y datos que se usen deben ser presentados con claridad, indicando qué resultados son propios y qué resultados son de otras personas (con referencias), tal como si fuera un artículo de investigación. El trabajo final se hará en equipos de dos o tres personas (no

más de tres, no menos de dos).

Instructores

Elmer Garduño Hernández (garduno@gmail.com)

Mariano Beguerisse Díaz (m.beguerisse@imperial.ac.uk)

Plan de clases	
No. Fecha	Tema
01 01/15	Introducción: Grafos y redes, tipos de redes, características
01 01/15	Similitud entre elementos: Similitud de conjuntos, minhashing
02 01/22	Conceptos básicos 1: Grados, correlaciones, asortatividad
02 01/22	Similitud entre elementos: LSH
03 01/29	Conceptos básicos 2: Centralidad, comunicabilidad
03 01/29	Similitud entre elementos: distancias
04 02/05	Small worlds, navegación
04 02/05	Procesamiento de flujos: muestreo en flujos
05 02/12	Presentaciones: Planteamiento del problema
05 02/12	Presentaciones: Planteamiento del problema
06 02/19	Redes aleatorias 1: Erdős-Renyi
06 02/19	Procesamiento de flujos: filtrado
07 02/26	Redes aleatorias 2: Barabási-Albert, configuración
07 02/26	Procesamiento de flujos: conteo
08 03/05	Dinámica de redes 1: Caminatas aleatorias, Laplaciano
08 03/05	Sistemas de recomendación: basados en contenido
09 03/12	Dinámica de redes 2: Contagio, percolación
09 03/12	Sistemas de recomendación: filtrado colaborativo
10 03/19	Presentaciones: Avances
10 03/19	Presentaciones: Avances
11 03/26	Dinámica en redes 3: Consenso y sincronización
11 03/26	Modelos de lenguaje: introducción
12 04/02	Comunidades 1: Clustering y Modularidad
12 04/02	Modelos de lenguaje: modelos de bloom
13 04/09	Comunidades 2: Stability
13 04/09	Modelos ocultos de Márkov
04/16	DESCANSO
04/16	DESCANSO
14 04/23	Comunidades 3: Map y comunidades con traslape
14 04/23	Modelos ocultos de Márkov
15 05/30	Redes con signos y balance estructural
15 05/30	Minería de textos
16 05/07	Privacidad y privacidad en redes
16 05/07	Privacidad en bases de datos.
17 05/14	Inferencia de redes
17 05/14	Índices invertidos
18 05/21	Presentación final
18 05/21	Presentación final

Materiales de lectura

Clase 1 - 01/15

Introducción: Grafos y redes, tipos de redes, características

- Newman - Capítulos 1-5 (hojear, ver ejemplos), Capítulo 6, secciones 1-4 y 6.
- Brossard, D. & Scheufele, D. A. Social science. Science, new media, and the public. *Science*, **2013**, 339, 40-41.
- Giles, J. Computational social science: Making the links. *Nature*, **2012**, 488, 448-450.

Similitud entre elementos: Similitud de conjuntos, minhashing

- Ullman MMDS - Chapter 3

Clase 2 - 01/22

Conceptos básicos 1: Grados, pesos, correlaciones, asortatividad, caminos.

- Newman - Capítulo 6, secciones 9-12. Capítulo 8 sección 3.
- Newman, M. E. J. Assortative Mixing in Networks, *Phys. Rev. Lett.*, **2002**, 89, 208701.

Similitud entre elementos: LSH

- Ullman MMDS - Chapter 3
- Syntactic Clustering of the Web. Andrei Z. Broder, Steven C. Glassman, Mark S. Manasse, Geoffrey Zweig, <http://www.hpl.hp.com/techreports/Compaq-DEC/SRC-TN-1997-015.pdf>

Clase 3 - 01/29

Conceptos básicos 2: Centralidad, comunicabilidad.

- Newman - Capítulo 7, secciones 1-7.
- Estrada, E. & Hatano, N. Communicability in complex networks, *Phys. Rev. E*, **2008**, 77, 036111.
- Kleinberg, J. M. Authoritative sources in a hyperlinked environment. *J. ACM, ACM*, **1999**, 46, 604-632.
- Cooper, K. & Barahona, M. Role-based similarity in directed networks, *arXiv:1012.2726*, **2010**.

Similitud entre elementos: distancias

- Ullman MMDS - Chapter 3

Clase 4 - 02/05

Small worlds, navegación, clustering coefficient.

- Newman - Capítulo 8, secciones 2 y 6.
- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks, *Nature* **1998**, 393, 440-442.
- Kleinberg, J. M. Navigation in a small world, *Nature*, **2000**, 406, 845-845.
- Stanley Milgram, "The Small World Problem", *Psychology Today*, 1967, Vol. 2, 60-67.

Procesamiento de flujos: muestreo en flujos

- Ullman MMDS - Chapter 4

Clase 5 - 02/12

Presentaciones: Planteamiento del problema

Clase 6 - 02/19

Redes aleatorias 1: Erdős-Rényi

- Newman - Capítulo 12.
- Erdős, P. & Rényi, A. On the Evolution of Random Graphs. *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, **1960**, 17-61.
- Bollobás, B. Random Graphs, *Cambridge University Press*, **2001**.

Procesamiento de flujos: filtrado

- Ullman MMDS - Chapter 4

Clase 7 - 02/26

Redes aleatorias 2: Barabási-Albert, configuración y otros.

- Newman - Capítulo 13 y capítulo 14, secciones 1-3.
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks, *Science*, **1999**, 286, 509-512.
- Clauset, A.; Shalizi, C. & Newman, M. Power-Law Distributions in Empirical Data, *SIAM Review*, **2009**, 51, 661-703.

Procesamiento de flujos: conteo

- Ullman MMDS - Capítulo 4
- HyperLogLog: the analysis of a near-optimal cardinality estimation algorithm. Philippe Flajolet, Éric

Clase 8 - 03/05

Dinámica de redes 1: Caminatas aleatorias, Laplaciano

- Newman - Capítulo 6 secciones 12-14.
- Chung, F. R. Spectral graph theory *American Mathematical Society*, **1997**.
- Lovász, L. *Random walks on graphs: A survey, Combinatorics, Paul Erdős is eighty*, **1993**, 2, 1-46 .
- Callaghan, T.; Mucha, P. J. & Porter, M. A. Random walker ranking for NCAA Division IA football, *American Mathematical Monthly, Mathematical Association of America*, **2007**, 114, 761-777 .

Sistemas de recomendación: basados en contenido

- Ullman MMDS - Capítulo 9

Clase 9 - 03/12

Dinámica de redes 2: Contagio, percolación.

- Newman - Capítulos 16 y 17.
- Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *N Engl J Med*, **2007**, 357, 370-379.
- May, R. M. & Lloyd, A. L. Infection dynamics on scale-free networks *Physical Review E*, **2001**, 64, 066112.
- Preciado, V. M.; Zargham, M.; Enyioha, C.; Jadbabaie, A. & Pappas, G. Optimal Vaccine Allocation to Control Epidemic Outbreaks in Arbitrary Networks, *arXiv:1303.3984*, **2013**.

Sistemas de recomendación: filtrado colaborativo

- Ullman MMDS - Capítulo 9

Clase 10 - 03/19

Presentaciones: Avances

Clase 11 - 03/26

Dinámica en redes 3: Consenso y sincronización.

- Newman - Capítulo 18.
- Strogatz, S. H. Exploring complex networks. *Nature*, **2001**, 410, 268-276.
- Barahona, M. & Pecora, L. M. Synchronization in Small-World Systems, *Phys. Rev. Lett.* **2002**, 89, 054101.

- Preciado, V. M. & Jadbabaie, A. Moment-based analysis of synchronization in small-world networks of oscillators *Proceedings of the 48th IEEE Conference on Decision and Control*, **2009**, 1690-1695.
- Olfati-Saber, R.; Fax, J. & Murray, R. Consensus and Cooperation in Networked Multi-Agent Systems *Proceedings of the IEEE*, **2007**, 95, 215-233.

Modelos de lenguaje: introducción

- Manning SNLP - Capítulo 6

Clase 12 - 04/02

Comunidades 1: Clustering y Modularidad.

- Newman - Capítulo 11.
- Porter, M. A.; Onnela, J.-P. & Mucha, P. J. Communities in Networks. *Notices of the American Mathematical Society*, **2009**, 56, 1082.
- Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, **2008**, P10008.
- Fortunato, S. Community detection in graphs. *Physics Reports*, **2010**, 486, 75 - 174.

Modelos de lenguaje: modelos de bloom

- Smoothed Bloom filter language models: Tera-Scale LMs on the Cheap. David Talbot and Miles Osborne, <http://acl.ldc.upenn.edu/D/D07/D07-1049.pdf>

Clase 13 - 04/09

Comunidades 2: Problemas de resolución de modularidad. Markov stability.

- Lancichinetti, A. & Fortunato, S. Limits of modularity maximization in community detection. *Physical Review E, APS*, **2011**, 84, 066122.
- Delvenne, J.-C.; Schaub, M. T.; Yaliraki, S. N. & Barahona, M. Mukherjee, A.; Choudhury, M.; Peruani, F.; Ganguly, N. & Mitra, B. (Eds.) The Stability of a Graph Partition: A Dynamics-Based Framework for Community Detection, *Dynamics On and Of Complex Networks, Volume 2, Springer New York*, **2013**, 221-242 .
- Delvenne, J.-C.; Yaliraki, S. & Barahona, M. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences*, **2010**, 107, 12755-12760.
- Lambiotte, R.; Delvenne, J. & Barahona, M. Laplacian Dynamics and Multiscale Modular Structure in Networks. *arXiv:0812.1770*, **2008**

Modelos ocultos de Markov

- Manning SNLP - Capítulo 9

04/16 DESCANSO

Clase 14 - 04/23

Comunidades 3: Map y comunidades con traslape

- Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, **2008**, *105*, 1118-1123 .
- Schaub, M. T.; Lambiotte, R. & Barahona, M. Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation. *Phys. Rev. E, American Physical Society*, **2012**, *86*, 026112
- Evans, T. S. & Lambiotte, R. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E, American Physical Society*, **2009**, *80*, 016105.
- Lancichinetti, A.; Fortunato, S. & Kertész, J. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics, IOP Publishing*, **2009**, *11*, 033015

Modelos ocultos de Markov

- Manning SNLP - Capítulo 9

Clase 15 - 05/30

Redes con signos y balance estructural

- Newman - Capítulo 7 sección 11.
- S.A. Marvel, S.H. Strogatz and J.M. Kleinberg. Energy landscape of social balance. *Physical Review Letters* 103, 198701 (2009).
- S.A. Marvel, J. Kleinberg, R.D. Kleinberg and S.H. Strogatz. Continuous-time model of structural balance. *Proceedings of the National Academy of Sciences* 108, 1771-1776 (2011).

Minería de textos

- Manning IR - Capítulo 2

Clase 16 - 05/14

Inferencia de redes

- Marbach, D.; Prill, R. J.; Schaffter, T.; Mattiussi, C.; Floreano, D. & Stolovitzky, G. Revealing strengths and weaknesses of methods for gene network inference, *Proceedings of the National Academy of Sciences*, **2010**, *107*, 6286-6291.
- De Smet, R. & Marchal, K. Advantages and limitations of current network inference methods. *Nature Reviews Microbiology, Nature Publishing Group*, **2010**, *8*, 717-729
- Tenenbaum, J. B.; De Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction *Science*, **2000**, *290*, 2319-2323.

- Beguerisse-Díaz, M.; Vangelov, B. & Barahona, M. Finding role communities in directed networks using Role-Based Similarity, Markov Stability and the Relaxed Minimum Spanning Tree. *IEEE GlobalSIP Conference*, **2013**.

Índices invertidos

- Manning IR - Capítulo 2

Clase 17 - 05/07

Privacidad y privacidad en redes:

- Dwork, C. Differential privacy: A survey of results. *Theory and Applications of Models of Computation*, Springer, **2008**, 1-19.
- Backstrom, L.; Dwork, C. & Kleinberg, J. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. *Proceedings of the 16th international conference on World Wide Web*, **2007**, 181-190

Privacidad en base de datos

-

19 05/21

Presentación final

Artículos con ideas para proyectos

(Ver también material del curso sus referencias, buscar en Google Scholar, PubMed, Web of Science, etc)

Lambiotte, R. & Rosvall, M. Ranking and clustering of nodes in networks with smart teleportation. *Phys. Rev. E, American Physical Society*, **2012**, 85, 056107 .

Sohn, Y.; Choi, M.-K.; Ahn, Y.-Y.; Lee, J. & Jeong, J. Topological cluster analysis reveals the systemic organization of the *Caenorhabditis elegans* connectome. *PLoS Comput Biol*, **2011**, 7, e1001139

Schaub, M. T.; Lehmann, J.; Yaliraki, S. N. & Barahona, M. Structure of complex networks: Quantifying edge-to-edge relations by failure-induced flow redistribution. *arXiv:1303.6241*, **2013**

Beguerisse-Díaz, M.; Porter, M. A. & Onnela, J.-P. Competition for popularity in bipartite networks. *Chaos*, **2010**, 20, 043101

Davies, T. P.; Fry, H. M.; Wilson, A. G. & Bishop, S. R. A mathematical model of the London riots and their policing. **2013**, 3, 1303.

Meilă, M. Comparing clusterings---an information based distance. *Journal of Multivariate Analysis*, **2007**, 98, 873 - 895

Varshney, L. R.; Chen, B. L.; Paniagua, E.; Hall, D. H. & Chklovskii, D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput Biol*, **2011**, 7, e1001066.

Stouffer, D. B.; Sales-Pardo, M.; Sirer, M. I. & Bascompte, J. Evolutionary conservation of species' roles in food webs. *Science*, **2012**, 335, 1489-1492

Kivelä, Mikko; Arenas, Alexandre; Barthelemy, Marc; Gleeson, James P.; Moreno, Yamir; and Porter, Mason A. [2013], Multilayer Networks, arXiv:1309.7233.

Dirk Brockmann, Dirk Helbing. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science* 13 December 2013: 342 (6164), 1337-1342.

Gleeson, James P.; Cellai, Davide; Onnela, Jukka-Pekka; Porter, Mason A.; and Reed-Tsochas, Felix [2013], A Simple Generative Model of Collective Online Behaviour, arXiv:1305.7440.

Matthew J. Salganik and Duncan J. Watts. Web-based experiments for the study of collective social dynamics in cultural markets. *Topics in Cognitive Science*, 1:439-468. 2009.

Feld, S. L. Why your friends have more friends than you do. *American Journal of Sociology*, JSTOR, **1991**, 1464-1477.

J. Ugander, B. Karrer, L. Backstrom and C. Marlow, "The anatomy of the Facebook social graph." 2011 arXiv:1111.4503.

Ulanowicz, Robert E. "Quantitative methods for ecological network analysis." *Computational Biology and Chemistry* 28, no. 5 (2004): 321-339.

Saavedra, Serguei, Felix Reed-Tsochas, and Brian Uzzi. "A simple model of bipartite cooperation for ecological and organizational networks." *Nature* 457, no. 7228 (2008): 463-466.

Alon, Uri. "Biological networks: the tinkerer as an engineer." *Science* 301, no. 5641 (2003): 1866-1867.

Repositorios de datos:

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

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

<http://netwiki.amath.unc.edu/SharedData/SharedData>

<http://www.weizmann.ac.il/mcb/UriAlon/download/collection-complex-networks>

<http://vlado.fmf.uni-lj.si/pub/networks/data/>

<http://data.sioc-project.org/>

<http://thedata.harvard.edu/dvn/>

