Analysis of Social Information Networks

A. Chaintreau, Columbia University

Spring Semester 2011

The availability of social information networks connecting users together through their online behaviors is a unique research opportunity for computer scientists. This is not only because today many computing applications and technologies learn to integrate them, but also because, for the first time, these networks describing individuals are considered primarily as *computational* objects. Hence, in addition to its sheer "complexity" and apparent lack of structure, information on a social network is remarkable for the algorithm design challenge that it poses: how to extract, exploit, and sometimes explain social properties using the minimum amount of information exchange and computation.

This course is an introduction to the algorithmic methods that have been shown to cope with the specific constraints of social information networks, and form today the "toolbox" of social computing applications.

1 Motivation and Overview

A brief history: *Social information networks*, broadly defined, characterize attributes of individuals (*e.g.,* interests, sociological profile) and the structure underlying their interactions (*e.g.,* friendship, business relationship). Traditionally it always was a latent force of sociological studies, and an important business factor in marketing, advertisement, and employment. In the 1990s, descriptive data of large information network became gradually available. A number of remarkable empirical observations, coupled with unified models to describe them across various application domains, gave birth to a new branch of natural physics. "Complexity" was put forward as their common denominator.

Social information networks are now, and for a decade, an object of *computational analysis*. This analytical method claims that at the essence of understanding these networks lies the way simple algorithms (similar to computer programs) are able to deduce from local information a global property that would otherwise require complete knowledge. In other words, whenever exhaustive enumeration is not cost-efficient or is simply infeasible (due to a lack of global data, or its overwhelming size), the answer to the "complexity" aforementioned lies in hidden structures that can be exploited by algorithms.

Where are we know? Today, it is hard *not* to see the impact of social information networks, for three reasons:

- The field and interest with these objects have grown with their sizes and scope, and their (partial) availability. For a company like Facebook, storing and making available information networks relating a population comparable to the world's two largest country is by itself an active area of computer science research. For almost any large company, large data becomes a commodity, and its ability to use them an important competitive factor.

- Whatever domains of applications where social information networks are integrated with technology, they typically change best practice. Business is conducted differently when the price or the service you receive depends on your friends. Media and entertainment are almost entirely redefined, incorporating contributions beyond the traditional content creator/consumer model. Public issues like health, environment, economics and development are likely to quickly change as algorithms and computer systems using social information networks are introduced.

- In response to these new challenges, original computational methods were designed, together with a characterization of their result and speed. It significantly extended our understanding of the "Complexity" of social information network by revealing what any algorithms can and cannot do. It also impacted in many ways how information is exploited for several applications.

Objectives: This course is an introduction to the computational analysis of social information network, with an emphasis on acquiring the set of theoretical skills that allow to mathematically justify the design of algorithmic methods. Examples where social informations are used by real systems will be given and analyzed, with some discussion on their broader impact.

During this course, students are expected to acquire (and will be evaluated on) the following set of skills:

- Relate a phenomenon to common dynamics observed and analyzed among users of a social network dealing with information; develop a critical eye towards future research topics.

- Formulate a problem dealing with social information in relation to ranking, algebraic spectral methods, and optimization.

- Propose innovative solutions using simple distributed algorithms running on top of social networks, with a mathematical justification of their convergence.

In other words, if you wish to understand from first principles how search engine and recommender systems find relevant information, how does information propagates over Facebook or Twitter, and what is the algorithmic potential of using social information, this course will help you find the answers.

The following topics, although they are related and will be mentioned at times, will *not* be covered or evaluated:

- Strategic and game theoretic aspects of social information networks.
- System principles to build and run large social networking services.
- Machine learning and Bayesian inference on social information networks.

If you wish to acquire a deep knowledge in any of these three topics (which are covered in different lectures) this course may provide you with a good complement.

2 Detailed contents

The course is made of three parts

1. Introduction: The first lecture aims at getting yourself acquainted with social information network and developing a critical eye towards research in this domain.
2. Fundamentals: In a series of 7 lectures, the course will cover a series of classical results that form the core of our current understanding of social information network. Practice will be encouraged by homework.
3. Empirical studies and advanced topics: The last 6 lectures are devoted to recent studies on online social networks (4 lectures) and some advanced topics (2 lectures). By nature this part covers less mature material, which are more prone to lead to a promising research direction. Highly interactive, each lecture will be organized with a set of recent research papers and a presentation and discussion where students will be expected to answer some challenges and open ended questions.

2.1 Introduction

2.1.1 Thursday, January 20th: Introduction

The role of Social Information Networks today.

The "small-world" phenomenon, Weak-ties and homophily, and the efficiency of decentralized search.

2.2 Fundamentals

These are organized along four "trails", corresponding to different computational aspect of social information network:

- Popularity (Trail 1): What is the popularity profiles exhibited by items and nodes in a social or information network? Can we explain and exploit them?
- Spread (Trail 2): How do opinions and behavior spread by influence between the nodes? Can we use this influence to one's advantage?
- Epidemy (Trail 3): How do information or virus propagates along the edges of a network? Can we design algorithm to leverage it?
- Structures (Trail 4): Given a data sets describing a social information network, what are the underlying structures allowing to rank nodes, to map them according to similarity, and to partition them? How can such structure be extracted from large datasets?

2.2.1 Thursday, January 27th: Trail 1

Popularity and dynamics of reinforcement.

- The ubiquitous power law.
- How do power-laws build up by reinforcement, optimization, artefacts?
- Fragility of power law, inoculation. Addressing the long tail with replication.

2.2.2 Thursday, February 3rd: Trail 2, part 1/2

Cascading and the spread of behavior in social network

- Life under the influence
- Models of cascading behaviors

2.2.3 Thursday, February 10th: Trail 2, part 2/2

Maximizing the spread of influence

- A general model and conditions
- Hardness and approximation

2.2.4 Thursday, February 17th: Trail 3

Epidemic spread and algorithms

- How do virus and information spread as a function of topology?
- Aggregating information through gossip?

2.2.5 Thursday, February 24th: Trail 4, part 1/3

Ranking between nodes

- Measure of importance.
- Spectral algorithms and their convergence.

2.2.6 Thursday, March 3rd: Trail 4, part 2/3

Assessing similarity between the nodes

- Collaborative filtering and recommendations.
- Spectral analysis of data

2.2.7 Thursday, March 10th: Trail 4, part 3/3

How to partition a network?

- Metric of community and heuristic algorithm.
- Spectral partitioning, and why it works.

2.3 Advanced topics

2.3.1 Empirical studies 1: Online social networks and their evolution

2.3.2 Empirical studies 2: Communities and their evolution

2.3.3 Empirical studies 3: User behaviors and interactions

2.3.4 Empirical studies 4: Content and social media

2.3.5 Advanced topic 1: Mobility and Space in social networks

2.3.6 Advanced topic 2: TBD

3 Logistics

Where and when? Thursday 2:10-4pm (w. 5mn pause at 3pm), room TBA

Who is teaching? Augustin Chaintreau (instructor), Zeinab Abbassi (TA).

Prerequisite? The course requires no other knowledge that simple discrete probability, linear algebra and elementary graph theory. If you would like a refresh before taking the course, you may consider review the following notions which will be introduced.

- Homogeneous Markov Chain: Chap.1-3 in P. Bremaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues* (2010) Springer.

- Graph Theory: Chap.1 in R. Diestel, *Graph Theory* (2010), Springer.
- Linear Algebra: Matrix, eigenvalues, eigenvectors.

Grading scheme: The evaluation will be based on:

- a mid-term exam which deals with the "Fundamentals" materials (30%)
- class participation (with presentation in the second half) (30%)
- a final project (research case study or topic review) (40%)

For organizational purpose, the division between presentations and projects will depend on the enrollment. It will be finalized an announce on the third course of the lectures.

Textbook, reading: There is no requisite reading before the course.

Unfortunately the topic covered in this course is not described in a textbook at the graduate level. The book[1] *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, by D. Easley and J. Kleinberg may be used as a very good introduction to this course (and other topics in the domain). Relevant parts are I and IV-VI.

4 Additional resources & Full Bibliography

Related teaching

- Michael Kearns, U. Penn,
  Networked Life, Spring 2004-2010,
  `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/`

[1] available at: `http://www.cs.cornell.edu/home/kleinber/networks-book/`

4.1 Small-world phenomenon

[1] Lada A Adamic. . . . How to search a social network. *Social Networks*, Jan 2005.

[2] M Granovetter. The strength of weak ties: A network theory revisited. *Sociological theory*, Dec 1983.

[3] Jon Kleinberg. Navigation in a small world. *Nature*, Jan 2000.

[4] Jon Kleinberg. The small-world phenomenon: an algorithm perspective. *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, May 2000.

[5] S Milgram. The small world problem. *Psychology today*, Jan 1967.

[6] Duncan Watts and S Strogatz. Collective dynamics of 'small-world'networks. *Nature*, Dec 1998.

4.2 Popularity and Reinforcement

[7] Dimitris Achlioptas, Aaron Clauset, David Kempe, and Cristopher Moore. On the bias of traceroute sampling. *J. ACM*, 56(4):1–28, Jun 2009.

[8] D Alderson, J Doyle, and W Willinger. Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*, Dec 2005.

[9] Q Chen, H Chang, R Govindan, and S Jamin. The origin of power laws in internet topologies revisited. *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 2:608–617, 2002.

[10] Byung-Gon Chun, Kamalika Chaudhuri, Hoeteck Wee, Marco Barreno, Christos Papadimitriou, and John Kubiatowicz. Selfish caching in distributed systems: a game-theoretic analysis. *PODC '04: Proceedings of the twenty-third annual ACM symposium on Principles of distributed computing*, Jul 2004.

[11] A Clauset, CR Shalizi, and MEJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[12] Edith Cohen and Scott Shenker. Replication strategies in unstructured peer-to-peer networks. *SIGCOMM '02: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, Aug 2002.

[13] Alex Fabrikant, Elias Koutsoupias, and Christos Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. *ICALP '02: Proceedings of the 29th International Colloquium on Automata, Languages and Programming*, Jul 2002.

[14] Stratis Ioannidis and Peter Marbach. On the design of hybrid peer-to-peer systems. *SIGMETRICS '08: Proceedings of the 2008 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, Jun 2008.

[15] EF Keller. Revisiting "scale-free" networks. *Bioessays*, 27(10):1060–1068, 2005.

[16] N Laoutaris, G Bestavros, and A Matta. Distributed selfish caching. *IEEE Transactions on Parallel and Distributed Systems*, Dec 2007.

[17] N Laoutaris, O Telelis, and V Zissimopoulos. Distributed selfish replication. *IEEE Transactions on Parallel and Distributed Systems*, Dec 2006.

[18] C Leng, W Terpstra, B Kemme, and W Stannat. Maintaining replicas in unstructured p2p systems. *Proceedings of the 2008 ACM CoNEXT Conference*, Dec 2008.

[19] C Leng, W Terpstra, B Kemme, and W Stannat. Slides: Maintaining replicas in unstructured p2p systems. *Proceedings of the 2008 ACM CoNEXT Conference*, Dec 2008.

[20] Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker. Search and replication in unstructured peer-to-peer networks. *SIGMETRICS '02: Proceedings of the 2002 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, Jun 2002.

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

[22] Herbert Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, Dec 1955.

[23] W Terpstra, J Kangasharju, and C Leng. Bubblestorm: resilient, probabilistic, and exhaustive peer-to-peer search. *portal.acm.org*, Dec 2007.

[24] Saurabh Tewari and Leonard Kleinrock. Analysis of search and replication in unstructured peer-to-peer networks. *SIGMETRICS '05: Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, Jun 2005.

4.3 Spreads and influence on social networks

Empirical studies

[25] S Aral. Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Sci*, 2010.

[26] S Aral, E Brynjolfsson, and MW Van Alstyne. Productivity effects of information diffusion in networks. *Proceedings of the 28th Annual International Conference on Information Systems*, 2007.

[27] S Aral, L Muchnik, and A Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544, 2009.

[28] S Aral and D Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Proceedings of the 31th Annual International Conference on Information Systems*, 2010.

[29] D Centola. The spread of behavior in an online social network experiment. *Science*, Jan 2010.

[30] K Chaudhuri and F Chung Graham. . . . A network coloring game. *INTERNET AND NETWORK ECONOMICS*, Jan 2008.

[31] NA Christakis and JH Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370, 2007.

[32] NA Christakis and JH Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249, 2008.

[33] J. H Fowler and N. A Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *BMJ*, 337(dec04 2):a2338–a2338, Dec 2008.

[34] Mark Granovetter. Threshold models of collective behavior. *The American Journal of Sociology*, 83(6):1420–1443, May 1978.

[35] Shawndra Hill, Foster Provost, and Chris Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, 21(2):256–276, May 2006.

[36] R Iyengar, C Van den Bulte. . . , and TW Valente. Opinion leadership and social contagion in new product diffusion. *Marketing Science*, Jan 2010.

[37] S Judd and M Kearns. . . . Behavioral dynamics and influence in networked coloring and consensus. *Proceedings of the . . .* , Jan 2010.

[38] M Kearns. . . , S Judd, and J Tan. . . . Behavioral experiments on biased voting in networks. *Proceedings of the . . .* , Jan 2009.

[39] M Kearns. . . and S Suri. . . . An experimental study of the coloring problem on human subject networks. *Science*, Jan 2006.

[40] Michael Kearns and Jinsong Tan. Biased voting and the democratic primary problem. *INTERNET AND NETWORK ECONOMICS*, Jan 2008.

[41] Stephen Morris. Contagion. *The Review of Economic Studies*, 67(1):57–78, Jan 2000.

[42] D Strang and SA Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual review of sociology*, 24(1), 1998.

[43] TW Valente. Social network thresholds in the diffusion of innovations. *Social Networks*, 18(1):69–89, 1996.

Diffusion on random graphs

[44] Hamed Amini, Moez Draief, and Marc Lelarge. Marketing in a random network. *Network Control and Optimization*, Feb 2009.

[45] D Centola. The spread of behavior in an online social network experiment. *Science*, Jan 2010.

[46] D Centola, VM Eguíluz, and MW Macy. Cascade dynamics of complex propagation. *Physica A: Statistical Mechanics and its Applications*, 374(1):449–456, 2007.

[47] D Centola and M Macy. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–34, 2007.

[48] M Lelarge. Diffusion and cascading behavior in random networks. *Note*, 2010.

[49] Marc Lelarge. Diffusion of innovations on random networks: Understanding the chasm. *WINE '08: Proceedings of the 4th International Workshop on Internet and Network Economics*, Dec 2008.

[50] Marc Lelarge. Efficient control of epidemics over random networks. *SIGMETRICS '09: Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems*, Jun 2009.

[51] Duncan Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of ...*, Jan 2002.

Maximizing influence

[52] E Berger. Dynamic monopolies of constant size. *Journal of Combinatorial Theory, Series B*, 83(2):191–200, 2001.

[53] J.-C Bermond, J Bond, D Peleg, and S Perennes. The power of small coalitions in graphs. *Discrete Applied Mathematics*, 127(3), May 2003.

[54] D Kempe, Jon Kleinberg, and Eva Tardos. Influential nodes in a diffusion model for social networks. *Proceedings of 32nd International Colloquium on Automata, Languages and Programming (ICALP)*, pages 1127—1138, Jan 2005.

[55] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug 2003.

[56] J Kostka and Y Oswald.... Word of mouth: Rumor dissemination in social networks. *STRUCTURAL INFORMATION AND COMMUNICATION COMPLEXITY*, Jan 2008.

[57] Marc Lelarge. Diffusion of innovations on random networks: Understanding the chasm. *WINE '08: Proceedings of the 4th International Workshop on Internet and Network Economics*, Dec 2008.

[58] Marc Lelarge. Efficient control of epidemics over random networks. *SIGMETRICS '09: Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems*, Jun 2009.

[59] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug 2007.

[60] Dominic Meier, Yvonne Pignolet, Stefan Schmid, and Roger Wattenhofer. On the windfall of friendship: inoculation strategies on social networks. *EC '08: Proceedings of the 9th ACM conference on Electronic commerce*, Jul 2008.

[61] E Mossel and S Roch. Submodularity of influence in social networks: From local to global. *SIAM Journal on Computing*, Dec 2010.

Variants of Spread

[62] MJ Brzozowski, T Hogg, and G Szabo. Friends and foes: ideological social networking. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 817–820, 2008.

[63] K Chaudhuri and F Chung Graham.... A network coloring game. *INTERNET AND NETWORK ECONOMICS*, Jan 2008.

[64] Peter Dodds and Duncan Watts. Universal behavior in a generalized model of contagion. *Physical review letters*, 92(21):218701, May 2004.

[65] N Immorlica, J Kleinberg, M Mahdian, and T Wexler. The role of compatibility in the diffusion of technologies through social networks. *Proceedings of the 8th ACM conference on Electronic commerce*, pages 75–83, 2007.

[66] S Judd and M Kearns.... Behavioral dynamics and influence in networked coloring and consensus. *Proceedings of the ...*, Jan 2010.

[67] M Kearns..., S Judd, and J Tan.... Behavioral experiments on biased voting in networks. *Proceedings of the ...*, Jan 2009.

[68] M Kearns... and S Suri.... An experimental study of the coloring problem on human subject networks. *Science*, Jan 2006.

[69] Michael Kearns and Jinsong Tan. Biased voting and the democratic primary problem. *INTERNET AND NETWORK ECONOMICS*, Jan 2008.

[70] Jure Leskovec, D Huttenlocher, and J Kleinberg. Governance in social media: A case study of the wikipedia promotion process. *Arxiv preprint arXiv:1004.3547*, 2010.

[71] Erez Lieberman, Christoph Hauert, and Martin A Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):309–312, Jan 2005.

4.4 Epidemy

[72] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE/ACM Transactions on Networking (TON*, 14(SI), Jun 2006.

[73] M Cao, DA Spielman, and AS Morse. A lower bound on convergence of a distributed network consensus algorithm. *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 2356–2361, 2006.

[74] Cha, H Haddadi, F Benevenuto, and K Gummadi. Measuring user influence in twitter: The million follower fallacy. *Proceedings of the 4th ...*, Dec 2010.

[75] F Chierichetti, S Lattanzi, and A Panconesi. Rumor spreading in social networks. *Proceedings of the 36th Internatilonal Collogquium on ...*, Dec 2009.

[76] F Chierichetti, S Lattanzi, A Panconesi, and S Universita di Roma. Rumour spreading and graph conductance. *Proceedings of SODA*, 2010.

[77] A Ganesh, L Massoulié, and D Towsley. The effect of network topology on the spread of epidemics. *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, 2:1455–1466, 2005.

[78] Z Haas, J Halpern, and L Li. Gossip-based ad hoc routing. *IEEE/ACM Transactions on Networking (TON)*, Dec 2006.

[79] S Hedetniemi and A Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, Dec 1988.

[80] R Karp, C Schindelhauer, S Shenker, and B Vocking. Randomized rumor spreading. *FOCS '00: Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, Nov 2000.

[81] F Liljeros, C Edling, L Amaral, and H Stanley. The web of human sexual contacts. *Nature*, Dec 2001.

[82] Laurent Massoulie and Moez Draief. Networks and epidemics. *Books in preparation*, pages 1–76, May 2010.

[83] Boris Oreshkin, Mark Coates, and Michael Rabbat. Optimization and analysis of distributed averaging with short node memory. *IEEE Transactions on Signal Processing*, 58(5), May 2010.

[84] B Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, Dec 1987.

[85] Michael Rabbat. On spatial gossip algorithms for average consensus. *SSP '07: Proceedings of the 2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, Aug 2007.

[86] Michael Rabbat, Jarvis Haupt, Aarti Singh, and Robert Nowak. Decentralized compression and predistribution via randomized gossiping. *IPSN '06: Proceedings of the 5th international conference on Information processing in sensor networks*, Apr 2006.

[87] S Sanghavi, B Hajek, and Laurent Massoulié. Gossiping with multiple messages. *Information Theory*, Dec 2007.

[88] Devavrat Shah. Gossip algorithms. *Foundations and Trends® in Networking*, 3(1), Jan 2009.

[89] Devavrat Shah and Tauhid Zaman. Detecting sources of computer viruses in networks: theory and experiment. *SIGMETRICS '10: Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, Jun 2010.

[90] Dan-Cristian Tomozei and Laurent Massoulié. Distributed user profiling via spectral methods. *SIGMETRICS '10: Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, Jun 2010.

[91] Deniz Üstebay, Boris Oreshkin, Mark Coates, and Michael Rabbat. Greedy gossip with eavesdropping. *IEEE Transactions on Signal Processing*, 58(7), Jul 2010.

4.5 Extracting Structures from Data

Nodes ranking

[92] Dimitris Achlioptas. . . , A Fiat, A Karlin, and Frank McSherry. Web search via hub synthesis. *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 500 – 509, 2001.

[93] Alon Altman and Moshe Tennenholtz. Ranking systems: the pagerank axioms. *EC '05: Proceedings of the 6th ACM conference on Electronic commerce*, Jun 2005.

[94] M Bianchini and M Gori. . . . Inside pagerank. *ACM Transactions on Internet Technology*, 5(1):92—128, Jan 2005.

[95] Allan Borodin, Gareth Roberts, Jeffrey Rosenthal, and Panayiotis Tsaparas. Finding authorities and hubs from link structures on the world wide web. *WWW '01: Proceedings of the 10th international conference on World Wide Web*, Apr 2001.

[96] Allan Borodin, Gareth Roberts, Jeffrey Rosenthal, and Panayiotis Tsaparas. Link analysis ranking: algorithms, theory, and experiments. *Transactions on Internet Technology (TOIT*, 5(1), Feb 2005.

[97] David Kempe and Frank McSherry. A decentralized algorithm for spectral analysis. *Journal of Computer and System Sciences*, 74(1), Feb 2008.

[98] Jon Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM*, 46(5), Sep 1999.

[99] Frank McSherry. A uniform approach to accelerated pagerank computation. *WWW '05: Proceedings of the 14th international conference on World Wide Web*, May 2005.

[100] Frank McSherry and Marc Najork. Computing information retrieval performance measures efficiently in the presence of tied scores. *ECIR'08: Proceedings of the IR research, 30th European conference on Advances in information retrieval*, Mar 2008.

Similarity

[101] Dimitris Achlioptas.... Random matrices in data analysis. *PKDD '04: Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Sep 2004.

[102] Dimitris Achlioptas..., A Fiat, A Karlin, and Frank McSherry. Web search via hub synthesis. *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 500 – 509, 2001.

[103] Dimitris Achlioptas... and Frank McSherry. Fast computation of low rank matrix approximations. *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing*, Jul 2001.

[104] Dimitris Achlioptas... and Frank McSherry. Fast computation of low-rank matrix approximations. *Journal of the ACM (JACM*, 54(2), Apr 2007.

[105] Reid Andersen, Christian Borgs, Jennifer Chayes, Uriel Feige, Abraham Flaxman, Adam Kalai, Vahab Mirrokni, and Moshe Tennenholtz. Trust-based recommendation systems: an axiomatic approach. *WWW '08: Proceeding of the 17th international conference on World Wide Web*, Apr 2008.

[106] Yossi Azar, Amos Fiat, Anna Karlin, Frank McSherry, and Jared Saia. Spectral analysis of data. *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing*, Jul 2001.

[107] Ravi Boppana. Eigenvalues and graph bisection: An average-case analysis. *Proceedings of the 28th Annual Symposium on Foundations of Computer Science*, Oct 1987.

[108] Soumen Chakrabarti, Byron Dom, S Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins, David Gibson, and Jon Kleinberg. Mining the web's link structure. *Computer*, 32(8), Aug 1999.

[109] Soumen Chakrabarti, Byron Dom, Prabhakar Raghavan, Sridhar Rajagopalan, David Gibson, and Jon Kleinberg. Automatic resource compilation by analyzing hyperlink structure and associated text. *WWW7: Proceedings of the seventh international conference on World Wide Web 7*, Apr 1998.

[110] M Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18(6), Dec 1989.

[111] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Recommendation systems: A probabilistic analysis. *FOCS '98: Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, Nov 1998.

[112] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Recommendation systems: a probabilistic analysis. *Journal of Computer and System Sciences*, 63(1), Aug 2001.

[113] Christos Papadimitriou, Prabhakar Raghavan, Hisao Tamaki, and Santosh Vempala. Latent semantic indexing: a probabilistic analysis. *Journal of Computer and System Sciences*, 61(2), Oct 2000.

Partitioning

[114] Noga Alon. Eigenvalues and expanders. *Combinatorica*, Jan 1986.

[115] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM*, 56(2), Apr 2009.

[116] J Audibert. . . . Minimax policies for adversarial and stochastic bandits. *Proceedings of the 22nd Annual Conference on . . .* , Jan 2009.

[117] Sebastien Bubeck. Jeux de bandits et fondations du clustering. pages 1–247, May 2010.

[118] Sébastien Bubeck and Ulrike Luxburg. Nearest neighbor clustering: A baseline method for consistent clustering with arbitrary objective functions. *The Journal of Machine Learning Research*, 10, Dec 2009.

[119] F Chung and S.-T Yau. Eigenvalues, flows and separators of graphs. *STOC '97: Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, May 1997.

[120] Fan Chung and S.-T Yau. A near optimal algorithm for edge separators (preliminary version). *STOC '94: Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, May 1994.

[121] FRK Chung. Spectral graph theory: Chapter 1. *American Mathematical Society*, 3:8, 1997.

[122] FRK Chung. Spectral graph theory: Chapter 2. *American Mathematical Society*, 3:8, 1997.

[123] FRK Chung. Spectral graph theory: Chapter 3. *American Mathematical Society*, 3:8, 1997.

[124] FRK Chung. Spectral graph theory: Chapter 4. *American Mathematical Society*, 3:8, 1997.

[125] Miroslav Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25:619—633, Jan 1975.

[126] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM (JACM*, 51(3), May 2004.

[127] U Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, Jan 2007.

[128] Frank McSherry. Spectral partitioning of random graphs. *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 529 – 537, 2001.

[129] A Ng and M Jordan. . . . On spectral clustering: Analysis and an algorithm. . . . *Systems 14: Proceeding of the 2001 . . .* , Jan 2001.

[130] D Spielman. . . . Spectral partitioning works: Planar graphs and finite element meshes. *Linear Algebra and its Applications*, Jan 2007.

[131] Daniel A Spielman. Spectral graph theory and its applications. *Tutorial at FOCS 2007*, pages 1–75, Oct 2007.

4.6 Empirical studies

Graph of online social networks and their evolution

[132] L Backstrom. . . . Supervised random walks: Predicting and recommending links in social networks. *Arxiv preprint arXiv:1011.4071*, Jan 2010.

[133] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: membership, growth, and evolution. *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug 2006.

[134] TY Berger-Wolf and J Saia. A framework for analysis of dynamic social networks. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 523–528, 2006.

[135] Ronald Burt. Structural holes versus network closure as social capital. *Chapter in Social Capital: Theory and Research*, pages 1–30, Aug 2000.

[136] Ronald Burt. Structural holes and good ideas. *The American Journal of Sociology*, 110(2):349–399, Sep 2004.

[137] F Chierichetti, R Kumar, and S Lattanzi. On compressing social networks. *Proceedings of the 15th ACM SIGKDD international . . .* , Dec 2009.

[138] A Clauset, C Moore, and M Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, Dec 2008.

[139] Dennis Fetterly, Mark Manasse, Marc Najork, and Janet Wiener. A large-scale study of the evolution of web pages. *WWW '03: Proceedings of the 12th international conference on World Wide Web*, May 2003.

[140] S Garg, T Gupta, N Carlsson, and A Mahanti. Evolution of an online social aggregation network: An empirical study. *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 315–321, 2009.

[141] Jon Kleinberg, Siddharth Suri, Éva Tardos, and Tom Wexler. Strategic network formation with structural holes. *SIGecom Exchanges*, 7(3), Nov 2008.

[142] Jon Kleinberg, Siddharth Suri, Éva Tardos, and Tom Wexler. Strategic network formation with structural holes. *EC '08: Proceedings of the 9th ACM conference on Electronic commerce*, Jul 2008.

[143] G Kossinets and Duncan Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88, 2006.

[144] R Kumar, J Novak, and A Tomkins. Structure and evolution of online social networks. *Link Mining: Models, Algorithms, and Applications*, pages 337–357, 2010.

[145] Jérôme Kunegis and Andreas Lommatzsch. Learning spectral graph transformations for link prediction. *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, Jun 2009.

[146] S Lattanzi and D Sivakumar. Affiliation networks. *Proceedings of the 41st annual ACM symposium on . . .* , Dec 2009.

[147] Jure Leskovec, Lars Backstrom, R Kumar, and A Tomkins. Microscopic evolution of social networks. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 462–470, 2008.

[148] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. *WWW '10: Proceedings of the 19th international conference on World wide web*, Apr 2010.

[149] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. *KDD '05: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, Aug 2005.

[150] David Liben-Nowell and Jon Kleinberg. The linkprediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019—1031, Jan 2007.

[151] A Mislove, HS Koppula, KP Gummadi, P Druschel, and B Bhattacharjee. Growth of the flickr social network. *Proceedings of the first workshop on Online social networks*, pages 25–30, 2008.

[152] Alan Mislove, M Marcon, and K Gummadi. Measurement and analysis of online social networks. *Proceedings of the . . .* , Dec 2007.

[153] G Palla, A Barabási, and T Vicsek. Quantifying social group evolution. *Nature*, Dec 2007.

[154] S Redner. Networks: Teasing out the missing links. *Nature*, Dec 2008.

[155] HH Song, TW Cho, V Dave, Y Zhang, and L Qiu. Scalable proximity estimation and link prediction in online social networks. *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 322–335, 2009.

[156] Panagiotis Symeonidis, Eleftherios Tiakas, and Yannis Manolopoulos. Transitive node similarity for link prediction in social networks with positive and negative links. *RecSys '10: Proceedings of the fourth ACM conference on Recommender systems*, Sep 2010.

[157] C Tantipathananandh and T Berger-Wolf. Constant-factor approximation algorithms for identifying dynamic communities. *Proceedings of the 15th ACM SIGKDD international ...*, Dec 2009.

[158] C Tantipathananandh, T Berger-Wolf, and David Kempe. A framework for community identification in dynamic social networks. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 717–726, 2007.

[159] M Torkjazi, R Rejaie, and W Willinger. Hot today, gone tomorrow: On the migration of myspace users. *Proceedings of the 2nd ACM workshop on Online social networks*, pages 43–48, 2009.

[160] E Zheleva, H Sharara, and L Getoor. Co-evolution of social and affiliation networks. *Proceedings of the 15th ACM ...*, Dec 2009.

Community

[161] Aaron Clauset, M Newman, and Cristopher Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70(6):066111, Dec 2004.

[162] GW Flake, S Lawrence, and CL Giles. Efficient identification of web communities. *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, page 160, 2000.

[163] GW Flake, S Lawrence, CL Giles, and FM Coetzee. Self-organization and identification of web communities. *Computer*, 35(3):66–70, 2002.

[164] S Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[165] S Fortunato and M Barthélemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36, 2007.

[166] M Girvan and MEJ Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821, 2002.

[167] R Kumar, P Raghavan, S Rajagopalan, and A Tomkins. Trawling the web for emerging cyber-communities. *Computer Networks*, 31(11-16):1481–1493, 1999.

[168] Jure Leskovec, Kevin Lang, Anirban Dasgupta, and Michael Mahoney. Statistical properties of community structure in large social and information networks. *WWW '08: Proceeding of the 17th international conference on World Wide Web*, Apr 2008.

[169] Jure Leskovec, Kevin Lang, and Michael Mahoney. Empirical comparison of algorithms for network community detection. *WWW '10: Proceedings of the 19th international conference on World wide web*, Apr 2010.

[170] MEJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577, 2006.

[171] MEJ Newman and M Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):26113, 2004.

[172] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, Jun 2005.

[173] Y Zhang, J Wang, Y Wang, and L Zhou. Parallel community detection on large networks with propinquity dynamics. *Proceedings of the 15th ACM SIGKDD international ...*, Dec 2009.

User Behaviors

[174] Eytan Bakshy, Brian Karrer, and Lada A Adamic. . . . Social influence and the diffusion of user-created content. *EC '09: Proceedings of the tenth ACM conference on Electronic commerce*, Jul 2009.

[175] AL Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[176] Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. Characterizing user behavior in online social networks. *IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, Nov 2009.

[177] M Cha, Alan Mislove, and K Gummadi. A measurement-driven analysis of information propagation in the flickr social network. *Proceedings of the 18th . . .* , Dec 2009.

[178] J Chan and E Daly. . . . Decomposing discussion forums and boards using user roles. *AAAI Conference on Weblogs and Social . . .* , Jan 2010.

[179] J Jiang, C Wilson, X Wang, P Huang, W Sha, Y Dai, and BY Zhao. Understanding latent interactions in online social networks. *Proc. of the ACM SIGCOMM Internet Measurement Conference*, 2010.

[180] Gueorgi Kossinets, Jon Kleinberg, and Duncan Watts. The structure of information pathways in a social communication network. *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug 2008.

[181] Ravi Kumar, Mohammad Mahdian, and Mary McGlohon. . . . Dynamics of conversations. *KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, Jul 2010.

[182] Kristina Lerman and Aram Galstyan. Analysis of social voting patterns on digg. *WOSP '08: Proceedings of the first workshop on Online social networks*, Aug 2008.

[183] R Malmgren, Jake Hofman, Luis Amaral, and Duncan Watts. Characterizing individual communication patterns. *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, Jun 2009.

[184] R Malmgren and D Stouffer. A poissonian explanation for heavy tails in e-mail communication. *Proceedings of the . . .* , Dec 2008.

[185] Fabian Schneider, Anja Feldmann, Balachander Krishnamurthy, and Walter Willinger. Understanding online social network usage from a network perspective. *IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, Nov 2009.

[186] M Valafar, R Rejaie, and W Willinger. Beyond friendship graphs: a study of user interactions in flickr. *Proceedings of the 2nd ACM workshop on Online social networks*, pages 25–30, 2009.

[187] B Viswanath, Alan Mislove, and M Cha. On the evolution of user interaction in facebook. *Proceedings of the 2nd . . .* , Dec 2009.

[188] Christo Wilson, Bryce Boe, Alessandra Sala, Krishna Puttaswamy, and Ben Zhao. User interactions in social networks and their implications. *EuroSys '09: Proceedings of the 4th ACM European conference on Computer systems*, Apr 2009.

Content and Social Media (Blogs, Twitter)

[189] E Adar and L Adamic. Tracking information epidemics in blogspace. *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, pages 207 – 214, 2005.

[190] N Agarwal, H Liu, L Tang, and PS Yu. Identifying the influential bloggers in a community. *Proceedings of the international conference on Web search and web data mining*, pages 207–218, 2008.

[191] J Aizen, D Huttenlocher, J Kleinberg, and A Novak. Traffic-based feedback on the web. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5254, 2004.

[192] Eytan Bakshy, Brian Karrer, and Lada A Adamic.... Social influence and the diffusion of user-created content. *EC '09: Proceedings of the tenth ACM conference on Electronic commerce*, Jul 2009.

[193] Cha, H Haddadi, F Benevenuto, and K Gummadi. Measuring user influence in twitter: The million follower fallacy. *Proceedings of the 4th ...*, Dec 2010.

[194] M Cha, Alan Mislove, and K Gummadi. A measurement-driven analysis of information propagation in the flickr social network. *Proceedings of the 18th ...*, Dec 2009.

[195] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, Oct 2007.

[196] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on Networking (TON*, 17(5), Oct 2009.

[197] Meeyoung Cha, Alan Mislove, Ben Adams, and Krishna Gummadi. Characterizing social cascades in flickr. *WOSP '08: Proceedings of the first workshop on Online social networks*, Aug 2008.

[198] Abhinandan Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: Scalable online colaborative filtering. 2007.

[199] P Domingos and M Richardson. Mining the network value of customers. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66, 2001.

[200] K El-Arini, G Veda, D Shahaf, and C Guestrin. Turning down the noise in the blogosphere. *Proceedings of the 15th ACM SIGKDD international ...*, Dec 2009.

[201] Sharad Goel, Andrei Broder, Evgeniy Gabrilovich, and Bo Pang. Anatomy of the long tail: ordinary people with extraordinary tastes. *WSDM '10: Proceedings of the third ACM international conference on Web search and data mining*, Feb 2010.

[202] M Götz, Jure Leskovec, and Mary McGlohon.... Modeling blog dynamics. *AAAI Conference on ...*, Jan 2009.

[203] L Guo, E Tan, S Chen, X Zhang, and Y Zhao. Analyzing patterns of user content generation in online social networks. *Proceedings of the 15th ACM SIGKDD international ...*, Dec 2009.

[204] BA Huberman, DM Romero, and F Wu. Social networks that matter: Twitter under the microscope. *First Monday*, 14(1):8, 2009.

[205] R Kumar, J Novak, P Raghavan, and A Tomkins. On the bursty evolution of blogspace. *World Wide Web*, 8(2):159–178, 2005.

[206] H Kwak, C Lee, H Park, and S Moon. What is twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web*, pages 591–600, 2010.

[207] Lerman. Social information processing in news aggregation. *Internet Computing, IEEE*, 11(6):16 – 28, 2007.

[208] Jure Leskovec, Lada A Adamic..., and B Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, Dec 2007.

[209] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. *Proceedings of the 15th ACM SIGKDD international ...*, Dec 2009.

[210] Jure Leskovec, M McGlohon, C Faloutsos, N Glance, and M Hurst. *Cascading Behavior in Large Blog Graphs Patterns and a Model*. 2007.

[211] Jure Leskovec and A Singh. . . . Patterns of influence in a recommendation network. *Advances in Knowledge Discovery . . .* , Jan 2006.

[212] D Liben-Nowell and J Kleinberg. Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences*, 105(12):4633, 2008.

[213] J.-P Onnela and F Reed-Tsochas. Spontaneous emergence of social influence in online systems. *Proceedings of the National Academy of Sciences*, 107(43):18375–18380, Oct 2010.

[214] M Richardson and P Domingos. Mining knowledge-sharing sites for viral marketing. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, page 70, 2002.

[215] T Sakaki, M Okazaki, and Y Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. *Proceedings of the 19th international conference on World wide web*, pages 851–860, 2010.

[216] D Shahaf and C Guestrin. Connecting the dots between news articles. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 623–632, 2010.

[217] Gabor Szabo and Bernardo Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53(8), Aug 2010.

4.7 Miscellaneous

Spatial Social Processes

[218] Lars Backstrom, Jon Kleinberg, Ravi Kumar, and Jasmine Novak. Spatial variation in search engine queries. *WWW '08: Proceeding of the 17th international conference on World Wide Web*, Apr 2008.

[219] Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. *WWW '10: Proceedings of the 19th international conference on World wide web*, Apr 2010.

[220] D Brockmann, L Hufnagel, and T Geisel. The scaling laws of human travel. *Nature*, Dec 2006.

[221] J Candia, M González, and P Wang. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: . . .* , Dec 2008.

[222] G Chowell, J Hyman, S Eubank, and C Castillo-Chavez. Scaling laws for the movement of people between locations in a large city. *Physical Review E*, Dec 2003.

[223] D. J Crandall, Lars Backstrom, D Cosley, S Suri, D Huttenlocher, and Jon Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, pages 1–6, Dec 2010.

[224] DJ Crandall, Lars Backstrom, D Huttenlocher, and Jon Kleinberg. Mapping the world's photos. *Proceedings of the 18th international conference on World wide web*, pages 761–770, 2009.

[225] S Eubank, V Kumar, and M Marathe. Structural and algorithmic aspects of massive social networks. *SODA '04 Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, Dec 2004.

[226] M González, C Hidalgo, and A Barabási. Understanding individual human mobility patterns. *Nature*, Dec 2008.

[227] Marta González, Pedro Lind, and Hans Herrmann. System of mobile agents to model social networks. *Physical review letters*, 96(8):088702, Mar 2006.

[228] JP Onnela, J Saramäki, J Hyvönen, G Szabó, D Lazer, K Kaski, J Kertész, and AL Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332, 2007.

[229] C Song, Z Qu, N Blumm, and A.-L Barabasi. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, Feb 2010.

[230] I Trestian, S Ranjan, and A Kuzmanovic. Measuring serendipity: connecting people, locations and interests in a mobile 3g .... *measurement conference*, Dec 2009.

[231] G Viswanathan, S Buldyrev, S Havlin, and M Da Luz. Optimizing the success of random searches. *Nature*, Dec 1999.

[232] P Wang, M Gonzalez, C Hidalgo, and A Barabasi. Understanding the spreading patterns of mobile phone viruses. *Science*, Dec 2009.

[233] P Wang and M. C Gonzalez. Understanding spatial connectivity of individuals with non-uniform population density. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1901):3321–3329, Aug 2009.

[234] P Wang, M. C Gonzalez, C. A Hidalgo, and A.-L Barabasi. Understanding the spreading patterns of mobile phone viruses. *Science (Supplementary Information)*, 324(5930):1071–1076, May 2009.

Link Classification

[235] M Charikar, V Guruswami, and A Wirth. Clustering with qualitative information. *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 524 – 533, 2003.

[236] P Doreian and A Mrvar. Partitioning signed social networks. *Social Networks*, 31(1):1–11, 2009.

[237] R Guha, R Kumar, P Raghavan, and A Tomkins. Propagation of trust and distrust. *Proceedings of the 13th international conference on World Wide Web*, pages 403–412, 2004.

[238] J Kleinberg and Éva Tardos. Approximation algorithms for classification problems with pairwise relationships: metric labeling and markov random fields. *Foundations of Computer Science, 1999. 40th Annual Symposium on*, pages 14 – 23, 1999.

[239] J Kunegis, A Lommatzsch, and C Bauckhage. The slashdot zoo: Mining a social network with negative edges. *Proceedings of the 18th international conference on World wide web*, pages 741–750, 2009.

[240] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. *WWW '10: Proceedings of the 19th international conference on World wide web*, Apr 2010.

[241] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, Apr 2010.

[242] SA Marvel, SH Strogatz, and JM Kleinberg. Energy landscape of social balance. *Physical review letters*, 103(19):198701, 2009.

[243] Panagiotis Symeonidis, Eleftherios Tiakas, and Yannis Manolopoulos. Transitive node similarity for link prediction in social networks with positive and negative links. *RecSys '10: Proceedings of the fourth ACM conference on Recommender systems*, Sep 2010.

[244] M Szell, R Lambiotte, and S Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636, 2010.

Link Inference and Prediction

[245] S Aral, L Muchnik, and A Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544, 2009.

[246] M De Choudhury, WA Mason, JM Hofman, and Duncan Watts. Inferring relevant social networks from interpersonal communication. *Proceedings of the 19th international conference on World wide web*, pages 301–310, 2010.

[247] D. J Crandall, Lars Backstrom, D Cosley, S Suri, D Huttenlocher, and Jon Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, pages 1–6, Dec 2010.

[248] T La Fond and J Neville. Randomization tests for distinguishing social influence and homophily effects. *Proceedings of the 19th international conference on World wide web*, pages 601–610, 2010.

[249] R Iyengar, C Van den Bulte. . . , and TW Valente. Opinion leadership and social contagion in new product diffusion. *Marketing Science*, Jan 2010.

[250] Seth A Myers and Jure Leskovec. On the convexity of latent social network inference. *NIPS*, 2010.

[251] J Neville, B Gallagher, and T Eliassi-Rad. Evaluating statistical tests for within-network classifiers of relational data. *2009 Ninth IEEE International Conference on Data Mining*, pages 397–406, 2009.

[252] S Wuchty. What is a social tie? *Proceedings of the National Academy of Sciences*, 106(36):15099, 2009.

[253] R Xiang, J Neville, and M Rogati. Modeling relationship strength in online social networks. *Proceedings of the 19th international conference on World wide web*, pages 981–990, 2010.