By Jure Leskovec

STANFORD UNIVERSITY

# KDD 2011 Tutorial

# Social Media Analytics

## Tracking, Modeling and Predicting the Flow of Information through Networks



## Tutorial information

Online social media represent a fundamental shift of how information is being produced, transferred and consumed. User generated content in the form of blog posts, comments, and tweets establishes a connection between the producers and the consumers of information.

Tracking the pulse of the social media outlets, enables companies to gain feedback and insight in how to improve and market products better. For consumers, the abundance of information and opinions from diverse sources helps them tap into the wisdom of crowds, to aid in making more informed decisions.

The tutorial investigates techniques for social media modeling, analytics and optimization:

- How do we collect massive amounts of social media data and what techniques can be used for correcting for the effects and biases arising from incomplete and missing data?
- What methods can be used to extract and track the flow of interesting pieces of information that spread and diffuse among the users? How can we identify the subset of content that is discussing not only a specific entity, but higher level concepts? %that are topically relevant?
- Having identified the subset of relevant content, how do we identify the most authoritative or influential authors? How do we quantify the influence of users on the adoption and spread of different topics? How do we maximize the overall influence?
- How do we tease apart emerging topics of discussion from the constant chatter in the blogosphere and other social media? How do we extract and model the temporal patterns by which information grows and fades over time?
- How do we predict popularity of memes and other pieces of information that spread through the social media networks?
- The information spreads via implicit networks. How do we identify and infer such networks of influence and diffusion? How do we discover implicit links between users?
- How does sentiment flow through networks and how does polarization occur?
- How do we overcome the information overload and provide users with rich and coherent experience?
- How to deal with unreliable and often conflicting information? What notions of trust are appropriate?

Social Media data comes in many forms: blogs (Blogger, LiveJournal), micro-blogs (Twitter, FMyLife), social networking (Facebook, LinkedIn), wikis (Wikipedia, Wetpaint), social bookmarking (Delicious, CiteULike), social news (Digg, Mixx), reviews (ePinions, Yelp), and multimedia sharing (Flickr, Youtube). Tutorial will investigate methods and case studies for analyzing such data and extracting actionable analytics.

## Tutorial outline

- **Part 1: Information flow in social media (slides)**
  - Collecting social media data
  - Extracting and tracking the flow of relevant information
  - Predicting and modeling the flow of information
  - Identifying networks of information flow
  - Influence maximization in networks
- **Part 2: Rich user interactions (slides)**
  - Predicting and recommending links in network
  - Modeling tie strenght
  - Modeling trust and distrust, frieds and foes
  - How users evaluate one another and the social media content

## Tutorial Video

(click to view the video)

## Tutorial slides

**Tutorial slides are available:**

- Part 1: Flow of information through networks
- Part 2: Rich user interactions

## Location

Tutorial will be held at ACM SIGKDD Conference in San Diego, CA on Sunday August 21 2011.

**Tutorial will start at 8:30 go till 12:00 with a 30 min break at 10:00. Room: Manchester H.**

## Who should attend

Social media data arises in so many different areas of data mining and predictive analytics so the tutorial should be of theoretical and practical interest to a large part of the world-wide-web and data mining community.

The tutorial will not require prior knowledge beyond the basic concepts covered in introductory machine learning and algorithms classes.

## Presenter

Jure Leskovec is an assistant professor of Computer Science at Stanford University. His research focuses on the analysis and modeling of large real-world social and information networks as the study of phenomena across the social, technological, and natural worlds. Problems he investigates are motivated by large scale data, the Web and Social Media. Jure received his PhD in Machine Learning from Carnegie Mellon University in 2008 and spent a year at Cornell University. His work received six best paper awards, won the ACM KDD cup and topped the Battle of the Sensor Networks competition.

## References

Adamic, L.A. & Glance, N.
*The political blogosphere and the 2004 U.S. election: divided they blog*
LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery. 2005, pp. 36-43

Adar, E. & Adamic, L.A.
*Tracking Information Epidemics in Blogspace*
Web Intelligence. 2005, pp. 207-214

Adar, E., Zhang, L., Adamic, L.A. & Lukose, R.M.
*Implicit Structure and the Dynamics of Blogspace*
Workshop on the Weblogging Ecosystem. 2004

Agichtein, E., Castillo, C., Donato, D., Gionis, A. & Mishne, G.
*Finding high quality content in social media, with an application to community-based question answering*
WSDM '08: ACM International Conference on Web Search and Data Minig. 2008, pp. 183-194

De Choudhury, M., Lin, Y.-R., Sundaram, H., Candan, K.S., Xie, L. & Kelliher, A.
*How Does the Data Sampling Strategy Impact the Discovery of Information Diffusion in Social Media?*
ICWSM '10: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. 2010

Fisher, D., Smith, M. & Welser, H.T.
*You Are Who You Talk To: Detecting Roles in Usenet Newsgroups*
HICSS '06: Proceedings of the 39th Annual Hawaii International Conference on System Sciences. 2006, Vol. 3, pp. 59b

Gilbert, E. & Karahalios, K.
*Predicting tie strength with social media*
CHI '09: Proceedings of the 27th international conference on Human factors in computing systems. 2009, pp. 211-220

Glance, N., Hurst, M., Nigam, K., Siegler, M., Stockton, R. & Tomokiyo, T.
*Deriving marketing intelligence from online discussion*
KDD '05: Proceeding of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining. 2005, pp. 419-428

Goetz, M., Leskovec, J., Mcglohon, M. & Faloutsos, C.
*Modeling blog dynamics*
International Conference on Weblogs and Social Media. 2009

Gomez-Rodriguez, M., Leskovec, J. & Krause, A.
*Inferring Networks of Diffusion and Influence*
KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. 2010

Gruhl, D., Guha, R., Kumar, R., Novak, J. & Tomkins, A.
*The predictive power of online chatter*
KDD '05: Proceeding of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining. 2005, pp. 78-87

Gruhl, D., Guha, R., Liben-Nowell, D. & Tomkins, A.
*Information Diffusion Through Blogspace*
WWW '04: Proceedings of the 13th international conference on World Wide Web. 2004, pp. 491-501

Guha, R., Kumar, R., Raghavan, P. & Tomkins, A.
*Propagation of trust and distrust*
WWW '04: Proceedings of the 13th international conference on World Wide Web. 2004, pp. 403-412

Kempe, D., Kleinberg, J.M. & Tardos,
*Maximizing the spread of influence through a social network*
KDD '03: Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining. 2003, pp. 137-146

Kumar, R., Novak, J., Raghavan, P. & Tomkins, A.
*On the bursty evolution of blogspace*
WWW '02: Proceedings of the 11th international conference on World Wide Web. 2003, pp. 568-576

Kwak, H., Lee, C., Park, H. & Moon, S.
*What is Twitter, a Social Network or a News Media?*
WWW'10: Proceedings of the 19th International World Wide Web Conference. 2010

Leskovec, J., Backstrom, L. & Kleinberg, J.
*Meme-tracking and the dynamics of the news cycle*
KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009, pp. 497-506

Leskovec, J., Huttenlocher, D. & Kleinberg, J.
*Predicting Positive and Negative Links in Online Social Networks*
WWW '10: Proceedings of the 19th International Conference on World Wide Web. 2010

Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. & Glance, N.
*Cost-effective Outbreak Detection in Networks*
KDD '07: Proceeding of the 13th ACM SIGKDD international conference on Knowledge discovery in data mining. 2007

Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N. & Hurst, M.
*Cascading behavior in large blog graphs*
SDM '07: Proceedings of the SIAM Conference on Data Mining. 2007

Leskovec, J., Singh, A. & Kleinberg, J.M.
*Patterns of Influence in a Recommendation Network*
PAKDD '06: Proceedings of the 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2006, pp. 380-389

Myers, S. & Leskovec, J.
*On the Convexity of Latent Social Network Inference*
NIPS '10: Advances in Neural Information Processing Systems. 2010

Sadikov, S., Medina, M., Leskovec, J. & Garcia-Molina, H.
*Correcting for Missing Data in Information Cascades*
WSDM '11: ACM International Conference on Web Search and Data Minig. 2011

Watts, D.J. & Dodds, P.S.
*Influentials, Networks, and Public Opinion Formation*
Journal of Consumer Research, 2007, Vol. 34(4), pp. 441-458

Yang, J. & Leskovec, J.
*Patterns of Temporal Variation in Online Media*
WSDM '11: ACM International Conference on Web Search and Data Minig. 2011

Yang, J. & Leskovec, J.
*Modeling Information Diffusion in Implicit Networks*
ICDM '10: IEEE International Conference On Data Mining. 2010