Online Political alignment and Information Diffusion

Bruno Gonçalves

Research Associate
Center for Complex Networks and Systems Research
School of Informatics and Computing
Indiana University, Bloomington, IN

Recent years have witnessed a veritable explosion of new online services and digital devices. As these new tools gradually permeate every aspect of our lives, the digital footprints left behind by our use of them become increasingly detailed: When we use a cell phone, where we are, who we are talking to and for how long, are automatically recorded. When we use a search engine or read an online news source, what we are reading or searching for is logged along with personally identifiable information. When we share pictures of ourselves and our friends in the latest dominant online social network, who our friends are, how often we interact with them and whether or not we see them in real life is also stored. Such digital footprints provide a unique view of human behavior (1) on a scale that was unimaginable just a decade or two ago. Instead of relying on small scale surveys of a few hundreds or thousands of volunteers, studies can now be made using the information of millions of unsuspecting individuals who dont even know they are participating in a scientific or market study. This unprecedented amount of data has attracted the attention of researchers in areas as diverse as physics or computer science and has opened the flood gates of relevant sociological questions that are now within reach of a detailed computational analysis.

An interdisciplinary area by excellence, this Computational approach to the study of Human Behavior and Social Science has important technological, commercial and scientific applications: Models of online traffic can help predict and prepare for spikes in demand; Understanding the way buyers choose one product over another can help a company suggest the right products; Clearly defining and quantifying typical human behavior based on observations of tens of millions individuals would provide validation for sociological and behavioral economical theories alike.

Of particular interest in this area is the study of how information, ideas and behaviors spread from person to person through the social network in which we are all embedded. Emerging online platforms provide a unique opportunity to study social behavior and human interaction. An analysis of the way individuals utilize microblogging services like Twitter to talk and interact with each other can provide valuable insight onto the way people interact in other, less constrict, contexts.

As a first step in this direction, I helped build a dedicated data mining infrastructure to obtain a real-time view of political discourse and interactions on Twitter in preparation for the 2010 midterm elections. This data and respective analysis are are kept updated on the `http://truthy.indiana.edu` website and has led us to identify the fundamental characteristics of a meme that allowed for the training of a machine learning algorithm capable of detecting meme injection attempts (2). Separating the wheat from the chaff, distinguishing between real patterns are those created artificially, is the first step towards characterizing the true behavior of society.

Building on these results we were able to identify the fundamental characteristics of political discourse associated with both side of the aisle that we were able to use to automatically determine whether a given user is Left or Right leaning based solely on their behavior and choice of topics (3). Using the geographical information contained within geolocated tweets or self dis-

closed through the users Twitter profile, one might be able to observe spatio-temporal patterns in the way information diffuses. Do specific types of information or ideas tend to originate in a given geographical area and then diffuse to the rest of the country? Which person or area is the most influential? How effectively does a person become infused in his or hers new environment upon moving from one area of the country to another?

Topological properties of the information diffusion network provide us with a dynamical view of the underlying social network. The structure of this network can provide us with insight about the way in which its actors self organize into groups. For example, we have been able to show, using the network of Twitter users that have retweeted messages with political content, we were able to identify the clusters of Democrat and Republic supporters (4). Undoubtedly, much more information about who we are and how we behave is inscribed in the network that we build with our interactions.

The key for further developments in this area is to collect and analyze data that can serve as a proxy for the analysis of Human Behavior and Interactions. The *Truthy* platform and many others are continuously collecting valuable data on how information spreads and people interact with each other. As we have seen many times before in the history of science, the availability of new data eventually brings forth a theoretical tipping point where questions that could never have been answered before come within reach. We are quickly approaching such a tipping point in Social Science in general and in the area of information diffusion in particular. Large, unique datasets are becoming increasingly available. Computational tools for their analysis are under continuous development. From the convergence of these two trends new theoretical frameworks are sure to emerge with increased explicative and predictive power in the fields of sociology, epidemiology and information diffusion.

Answering the type of questions that this workshop intends to address will help us gain foundational insights on individual and social behavior and help unlock the doors to understanding the functional mechanisms that underlie the functioning of modern society, the emergence of social norms, or the way in which individual interactions can impact societal processes.

References

[1] B. Gonçalves and J. J. Ramasco. Human dynamics revealed through web analytics. *Phys. Rev. E*, 78:026123, 2008.

[2] J. Ratkiewicz, M. Meiss, M. Conover, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer. Truthy : Mapping the spread of astroturf in microblog streams. In *WWW '11 Proceedings of the 20th international conference companion on World wide web*, page 249, 2011.

[3] M. Conover, J. Ratkiewicz, B Gonçalves, J. Haff, A. Flammini, and F. Menczer. Predicting the political alignment of twitter users. *Under Review*, 2011.

[4] M. Conover, J. Ratkiewicz, M. Francisco, B Gonçalves, A. Flammini, and F. Menczer. Political polarization on twitter. In *To Appear in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.