Sixth LACCEI International Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2008)
"Partnering to Success: Engineering, Education, Research and Development"
4-6 June 2008, Tegucigalpa, Honduras.

Rhetorical Analysis as a Means to Discover Networks of Influence

Robin Burk
United States Military Academy, West Point, NY USA, Robin.Burk@usma.edu

Kevin Huggins
United States Military Academy, West Point, NY USA, Kevin.Huggins@usma.edu

Lester Knotts
United States Military Academy, West Point, NY USA, Lester.Knotts@usma.edu

ABSTRACT

In this work we present a research program to use computational analysis of rhetorical styles and content in order to identify key influencers and the growth of networks of shared ideology, concepts or ideas. Research in social networks have focused on identifying persons of influence by the number of links that are connect them to other people. In such networks all parties are aware of their connections to adjacent people. However, such social networks are not the only way for ideas and ideological influence to propagate. The Internet and other information and communication technologies facilitate the spread of ideas and insights outward from a source without requiring that the source be aware of those who are influenced directly or indirectly. It is commonly the case, however, that key ideas, concepts or ideological assertions retain a characteristic vocabulary, syntactic formulation and other rhetorical features as they are adopted by new persons in the growing network of those influenced by them. Such characteristics can be computationally identified, thereby providing a way to discover networks and persons of influence separate from social network analysis in the usual sense. We describe an approach and research program for this purpose.

Keywords: network science, rhetorical analysis, computational linguistics, influence networks

1. INTRODUCTION

The study of rhetoric provides insight into characteristically human activities. Although rhetoricians do not agree on a single definition, a useful application of rhetoric describes the art or study by which discourse is adapted to enlighten the understanding, and to move the passion. We are more interested in enlightenment than with persuasion in the use of symbols to create new knowledge. Master rhetorician Kenneth Burke helps us to understand that "Rhetoric is rooted in an essential function of language itself, a function that is wholly realistic and continually born anew: the use of language as a symbolic means of inducing cooperation in beings that by nature respond to symbols . . . wherever there is rhetoric, there is meaning. "

Converting human knowledge into the 0s and 1s of binary code and then back into symbols understandable to humans using a variety of mechanical interfaces is a central task in computer science. Such conversions increasingly go beyond simple data capture to include presentation and analysis of online collections of news articles, web sites, interactive comments and multimedia products such as the videos on YouTube. Beyond the

sophisticated human-machine interface that makes these possible is usually a grander purpose – human interaction designed to inform and move others If the study of rhetoric is, as Charles Bazerman (1988) writes, a "study of how people use language and other symbols to realize human goals and carry out human activities . . . ultimately a practical study offering people great control over their symbolic activity" then rhetorical analysis can be expected to shed light on a key human activity whose purpose specifically includes influencing others.

Network Science is the study of network representations of physical, biological, information, cognitive and social phenomena that grow and spread as nodes of activity spawn or influence other nodes (Committee, 2005). The goal of network science as a new discipline is to gain insight into the shared principles and behaviors that underly networked phenomena of all kinds and ultimately to be able to create predictive models of these phenomena. For some phenomena such a social networks a fair amount of investigation has already taken place. Other networked phenomena such as cognitive influence are less familiar but play an increasingly important role in our highly connected contemporary world. Characterizing and modeling such networks is interesting both in its own right and insofar as it can contribute to the development of network science as an independent discipline.

We are particularly interested in studying networks of ideological or conceptual influence for several reasons. First, they are poorly understood. Second, they increasingly impact our social, economic, scientific, political, religious and other areas of life through media such as the World Wide Web and other Internet means for sharing information. Such networks have particular impact during stressful situations such as security threats (Arquila and Ronfelt, 2001), but they also function in benign milieus such as the diffusion of scientific research results or increased cross-cultural understanding. And third, a computational approach to studying networks of influence offers a challenging opportunity to apply techniques and insights from several disciplines to a complex human activity for which we currently lack rigorous models.

2. PREVIOUS WORK

2.1 RHETORIC

Using symbols and mutual understanding of symbols is a peculiarly human characteristic. To begin with no system at all—neither oral nor written, advancing to the technology of pencil and paper to record marks that held meaning for persons who were not present for the original discourse or event required a complex interconnectivity of humans in agreement about morphemes (the smallest linguistic unit with meaning) and graphemes (letters and letter combinations that make up morphemes) and phonemes (the smallest unit in language capable of conveying a distinct meaning). In our acquistion of language as children we take on these complex systems with amazing ease but it is typically not until we are adults that we have mastered the subtleties of our mother tongue. The complexity of any natural (human) language is due in part to the size of a language's vocabulary and the possible syntactic arrangements of words into sentences. However, to an even greater degree the complexity of natural language is a function of features such as polysemy (two or more meaning for the same word), homphony (two words that sound alike but are written in different ways and have different meanings) and metaphorical usages (both frozen and active metaphor). This complexity is resolved during human communication in a variety of ways: by restating or summarizing the meaning of an utterance or text and often through interactive dialogue. Morover, understanding and interpretation of texts (in particular) is often spread through teaching, editorial writings and scholarly commentaries. The result is a growing network of shared understanding, a sophisticated network of human interaction whose intricacy is compounded when those interactions are mediated by computer-based systems. In such cases the human interpretation of language is augmented by the need to translate speech or text into machine symbols and back into human terms.

The rhetorical structure and use of human language has long been the topic of study and analysis (Ehninger, 1975). The ancient Greeks and Romans were concerned with the grammatical structure of language and ways in which a speaker might best get his point across and influence the audience, often for political purpose. Those

who were merely fluent in the techniques of communication without concern for serious and socially valuable purpose were, however, dismissed as mere sophists by those who like Socrates, viewed conversation as above all an opportunity for us to become wiser. By the eighteenth century Western students of rhetoric had shifted their focus away from speaker techniques to an understanding of the psychology and interests of the audience. In keeping with the Enlightenment they encouraged speakers and writers to use data (rather than logical syllogisms), plain and dignified language and clear grammar to draw the audience along towards a conclusion. By the early twentieth century rhetoric, as with other disciplines, was especially focused on the characteristic use of language within social groups, whether ethnic, economic or otherwise identified.

Today we find that the Internet has dis-intermediated communication of all kinds. This has brought about both opportunities and challenges. On the one hand it is now inexpensive and easy for news, scientific results, literature and opinion pieces to be promulgated far beyond one's own country. Machine translation is rapidly improving, making texts from other languages accessible to a wide audience around the globe. Individuals can make their voices heard online immediately without relying on a newspaper editor to publish (or not) a letter in response to some article written last week. On the other hand, over recent years we have also seen the potential for dis-intermediated communication to enable the spread of terrorist ideology, to undercut the economic basis of professional journalism and to impact the economic viability of businesses and the political viability of candidates for public office around the world.

2.2 COMPUTATIONAL APPROACHES TO SOCIAL NETWORK DISCOVERY

Social network analysis has been a well-developed discipline since at least the publication of Wasserman and Faust's *Social Network Analysis: Methods and Applications* in 1994. That work discusses network topologies and isomorphisms, sociometric formalisms, power relationships and data collection for (manual) network analysis. More recent work has extended this sort of analysis to such domains as email traffic, weblog and web page references and tracing the citation linkages between scientists in the scholarly literature for a given discipline of study. In general this work, like most social network analysis, seeks to identify relations between persons of interest through the fact of communications between them – but not necessarily through the content of those communications.

2.3 COMPUTATIONAL APPROACHES TO RHETORIC

Every computational approach to a natural phenomenon begins by extracting features of that phenomenon into a model that can be represented, ultimately, in the binary 1s and 0s of computer circuits and storage devices. Computational models for natural language have been a key computer science topic since the 1950s when four great scientists - John McCarthy (then at Dartmouth), Marvin Minsky (then at Harvard University), Nathaniel Rochester of IBM and Claude Shannon (then at MIT) – submitted their famous "PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE" to the Rockafeller Foundation.

> We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to **make machines use language**, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer. [Dartmouth 1956]

Alas, artificial intelligence in general and natural language processing (NLP) in particular proved to be somewhat more difficult than these great men anticipated. Initial attempts included symbolic reasoning approaches to both syntax and semantics (meaning), resulting in generative grammars; semantic networks such as Princeton University's WordNet and University of California at Berkeley's Framenet and more recent work in formal ontologies. During the 1990s many computational linguists adopted statistical approaches to computational

linguistics (Manning and Schuetze, 1999. However, purely statistical and machine learning approaches have also fallen short of desired goals for NLP and more recent research tends to combine multiple approaches into hybrid techniques (Jurafsky and Martin, 2000 and Mitkov, 2005).

Research in NLP and computational linguistics is not confined to studying the use of English. Work on Romance languages (including Spanish) is maturing rapidly, with standardized evaluation events such as ROMANSEVAL, co-sponsored by the Association for Computational Linguistics and EURALEX.

Given the complexity of NLP in general and the plethora of techniques applied to sub-problems, how can we computationally capture rhetorical style and content in particular? One approach has been crafted by Boris Katz, Ozlem Uzuner and others at the Computer Science & Artificial Intelligence Laboratory at MIT. They empirically evaluate texts of various sorts for similarity of *expression* and *content* using a variety of syntactical, lexical and machine learning techniques. (Uzuner et al, 2004); Uzuner et al., 2005; Uzuner and Katz, 2005a and 2005b). This approach specifically does not require that one construct and validate a formal model for rhetorical style; instead, it infers similarity of rhetorical style and content between a given text and others. In this way it sidesteps the larger problems of full natural language understanding while providing a useful tool for identifying influence as evidenced by rhetorical style and content in emails, speeches, documents, scholarly articles and other texts.

3. USING RHETORICAL ANALYSIS TO IDENTIFY THE INFLUENTIAL PERSON IN A NETWORK

Traditional approaches to influence in social networks adopt a power / social relationship model (Marsden, 1993). More recent work on computational discovery of social networks emphasize discovery and potentially disruption of network growth and structure (Carley, 2003)

We believe that in addition to networks of explicit social relationships there is a growing phenomenon of electronically-mediated cognitive or ideological networks of ideas and influence which may go undetected or under-detected by existing social network research approaches.

As an example, social networking Internet sites such as Facebook and MySpace explicitly link member pages. The influence of a member can be measured, in social network terms, by the number of people who link to it as 'friends' or by the number of people to whom the member links. Similarly, a weblogger whose entries are explicitly linked by other bloggers is part of a visible social network.

However, if that weblogger's characteristic phrases, rhetorical style and themes are adopted by others without explicit hyperlinks to any given entry on his site, he has become an influencer in a cognitive network which relies on ideas rather than explicit social connectedness. An anecdotal example of such influencer is Glenn Reynolds, a law professor who blogs at http://www.Instapundit.com . Shortly after the attacks of 9/11/2001 he referred to the passengers of the airline flight who forced their hijackers to crash the plane into a field in Pennsylvania (rather than into the White House or Congress) as "a pack, not a herd". This phrase has become popular as a shorthand for the idea of citizens responding to emergencies with cooperative initiative rather than passively relying on official action. Few who use the phrase today cite Reynolds as the source; nonetheless, he has become one of the most influential writers of an individual weblog, attracting nearly 80 million visits a year to his site.

We believe that networks of influence exist in many domains besides public commentary on weblogs or the use of the Internet for ideological recruitment by terror groups. In the following section of the paper we outline a research program which will use rhetorical analysis to discover key influencers in the domain of computing and engineering science.

4. RESEARCH PROGRAM

Researchers who study the activities of scholarship and scholarly influence have typically applied social network metrics such as papers co-authored and papers cited (White and McCain, 1998) despite known problems with this approach. The growth of the Web and of open access to scientific literature has prompted calls for more

sophisticated application of citation and other bibliometric approaches (Cronin, 2001) but these continue to depend on explicit linkages to identify scholarly influence.

We propose to initiate an alternative research program that uses rhetorical analysis to identify networks of scholarly influence without depending on citation links. Such a program might begin by selecting highly cited articles and authors in computer science whose work appears in online document libraries such as CiteSeer, the ACM Portal and the IEEE online journal library. Using these as reference documents, we will search the journals in which they were published, along with peer journals on related topics, identify subsequent papers whose rhetorical and content similarity suggest they have been influenced by the selected authors despite the lack of explicit citation. As a control, we will also calculate rhetorical and content similarity between the reference documents and articles that explicitly cite them.

5. CONCLUSION

The research program we propose is ambitious but promising. The proliferation of online scholarly documents offers the promise of rapid diffusion of new discoveries and insights, but such knowledge diffusion faces the obstacle posed by sheer volume of available material. Insights into nature and growth of networks of scholarly influence can suggest ways to facilitate the adoption of new scientific research and technologies across existing national and other boundaries.

Although journal publications form perhaps the most tractable target for research into networks of influence, sociologists, political scientists and others may also find that such research offers a model for investigating the spread of cognitive and ideological influences which, no longer confined to expensive printed materials and slow dissemination, are literally growing at the speed of light through Internet connections and into the private browsers of millions of people around the world. Many of these are benign, if transformative. Others are less so. But all are an increasingly powerful part of our modern world. We look forward to gaining some insight into them and, along the way, contributing to the emerging broader discipline of network science *per se*.

REFERENCES

Arquilla, J., and Ronfeldt, D. (2001). Networks and Netwars: The Future of Terror, Crime and Militancy. Santa Monica, Calif.: RAND.

Bazerman, Charles. (1988) “Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science.” Madison U of Wisconsin P, 1988.

Carley, K. (2003). “Dynamic Network Analysis”. Summary of the NRC workshop on Social Network Modeling and Analysis, Ron Breiger and Kathleen M. Carley (Eds.), National Research Council. USA

Cronin, B. (2001). “Bibliometrics and Beyond: some thoughts on web based citation analysis”. Journal of Information Science, Vol. 27, No. 1.

Committee on Network Science for Future Army Applications, National Research Council. (2005) “Network Science.” The National Academies Press. Washington, D.C. USA.

Dartmouth College (1956). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Accessed from http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html .

Ehninger, D. (1975). “A Colloquy: A Synoptic View of Western Systems of Rhetoric”. Quarterly Journal of Speech, 61, 4, 448-453, Dec 75

Jurafsky, D. and Martin, J. (2000) Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, New York, NY, USA.

Manning, C. and Schuetze, H. (1999). Foundations of Statistical Natural Language Processing. The MIT Press. Cambridge, MA, USA.

Marsden, P. (1993). "Network Studies of Social Influence". Sociological Methods & Research, Vol. 22, No. 1. Sage Publications.

Mitkov, R. (ed). (2005) Oxford Handbook of Computational Linguistics. Oxford University Press, USA.

Rheingold, H. (2002). "Smart Mobs: The Next Social Revolution." MA Basic Books. Cambridge, Massachusetts, USA.

Ronfeldt, D. 2005. A long look ahead: NGOs, networks, and future social evolution. In Environmentalism and the Technologies of Tomorrow, R. Olson and D. Rejeski, eds. Washington, D.C.: Island Press.

Uzuner, O.. Davis, R. and Katz, B. (2004) "Using Empirical Methods for Evaluating Expression and Content Similarity". Proceedings of the 37th Hawaii International Coinference on System Sciences.

Uzuner, O., Katz, B. and Nahnsen, T. (2005) "Using Syntactic Information to Identify Plagiarism". Proceedings of the 2nd Workshop on Building Educational Applications Using NLP. Association for Computational Linguistics, Ann Arbor, MI, USA.

Uzuner, O. and Katz, B. (2005a). "Capturing Expression Using Linguistic Information". Proceedings of the 20th National Conference on Artificial Intelligence. American Association for Artificial Intelligence, USA.

Uzuner, O. and Katz, B. (2005b) "A Comparative Study of Language Models for Book and Author Recognition". Proceedings of the Second International Joint Conference on NLP, Jeju Island, Korea, October 11-13, 2005. Springer, Berlin Germany.

Wasserman, S. and Faust, K. (1994) Social Network Analysis: Methods and Applications. Structural Analysis and the Social Sciences series, Cambridge University Press. Cambridge, U.K.

Watts, D.J. 2003. Six Degrees: The Science of a Connected Age. New York, N.Y.: W.W. Norton.

White, H. and McCain, K. (1998) "Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972-1995". Journal of the American Society for Information Science, Vol. 49, No. 4.

Authorization and Disclaimer

Authors authorize LACCEI to publish the paper in the conference proceedings. Neither LACCEI nor the editors are responsible either for the content or for the implications of what is expressed in the paper.