Web Science in the SONIC Research Group

Willem Pieterson
Northwestern University
SONIC Research Group
2240 Campus Drive, FSB 1.459
Evanston IL, 60208
wpieterson@northwestern.edu

Noshir Contractor
Northwestern University
SONIC Research Group
2240 Campus Drive, FSB 2.133
Evanston IL, 60208
nosh@northwestern.edu

ABSTRACT
In this poster, we present an overview of the Web Science research conducted in the SONIC research group at Northwestern University. The angle taken by SONIC is to look at the Web from a network perspective. The research is theoretically rooted in a Multi-Theoretical, Multi-Level framework and in the empirical research new methods and tools are being applied to Web data analysis. Goals of the research are 1) to understand the Web, 2) to enable the Web and 3) re-conceptualize the Web.

Categories and Subject Descriptors
A.0 [General]: Web Science and Networks. Social Network Analysis, Hypergraphs, Tools and methods. Virtual worlds.

General Terms
Management, Experimentation, Theory.

Keywords
Social Network Analysis, Network Theory, Web Science

1. INTRODUCTION
The Web is the largest human information construct ever created and its impact on society is enormous. If we want to ensure that the Web benefits the human race as a whole, we first need to our best to understand it. The Web is a core component of the Internet, the large underlying network linking computers all over the world. Understanding the structures and characteristics of these networks may aid in getting that better understanding of the Web. The SONIC research group at Northwestern University applies network science to the Web and tries to aid in understanding the Web and on top of that enable the web to ensure its (social) benefits. Finally, the group explores new approaches based on network theories to study, understand and enable the Web. In this poster we present an overview of the Web related research conducted at SONIC from a network perspective. Hereby we focus on the **theoretical** dimension of the research in which we apply a Multi-Theory, Multi-Level (MTML)[1] approach to Web Science. A second focus is on the **empirical** part of the research in which we develop new methods and tools to test MTML using Web-enriched data

2. THE SONIC RESEARCH GROUP
The Science of Networks in Communities (SONIC) research group advances social network theories, methods, and tools to better understand and meet the needs of diverse communities. SONIC develops cutting-edge techniques to study and improve social and knowledge networks in different communities. Using a comprehensive methodology, SONIC researchers model, assess, validate, and enable the emergence of social and knowledge networks. SONIC helps communities discover their existing communication and knowledge networks, diagnose their network's health by measuring its robustness, connectedness, bottlenecks, etc., and design the social incentives and technical infrastructure needed for networks to function at optimal effectiveness. The four contexts in which SONIC operates are:

- Science and Engineering Teams
- Business contexts
- Virtual worlds
- Social Issues

2.1 Theoretical line of Research
Different social science theories make differential predictions about networks. Some of these mechanisms are unique, some are complementary, some are duplicative and some compete [1]. None of the theories, however, have the potential to provide definitive, exhaustive explanations of network phenomena. Therefore, in order to understand networks, we have to combine different theories, and different levels (e.g. person, triad, group) on which these theories may apply. This Multi-Theory, Multi-Level (MTML) framework identifies such network properties as mutuality and density and shows how these properties correspond to theoretical mechanisms in social theories [1, p.21].

2.2 Empirical line of Research
Given the dynamic characteristic and ever increasing size of the Web, The empirical work in SONIC focuses on three activities:

1. *Getting data*
   - Building and developing tools to automate data collection (e.g. from Wikipedia[4])
2. *Manipulating data*
   - Combining, collating and querying data from large, diverse sources
3. *Analyzing data*
   - Statistical techniques to discern emergence of structural signatures in networks
   - Building tools to optimize team assembly and performance
   - Algorithms optimized for large scale network analysis

Examples of tools developed in SONIC are Cyber-infrastructure for Inquiring Knowledge Networks on the Web (C-IKNOW)[2] a powerful web application for social network analysis investigation and the C-IKNOW Semantic Recommender[2][2] (see below).

3. THE RESEARCH
3.1 Understand the Web
First focal point of the research is to help increasing the understanding of the Web and the connection between peoples' behavior online and offline. We do this through three main activities: 1) comparing online and offline behavior, for example the economics of online communities [3], 2) modeling online and

offline behavior and 3) making predictions regarding offline and online behavior. These predictions vary from predicting the establishment of online friendship relationships [4] to online collaborations in teams and co-authorships of article on for example WikiPedia [5].

3.2 Enable the Web

The prime focus of SONIC in Enabling the Web is to make predictions and do recommendations regarding individuals and teams. Based on the notion that many recommendations involve the creation or change of social ties, we can use MTML to distinguish between different types of recommendations. Table 1 below gives an example of a number of heuristics based on MTML and how these can be applied in collaboration and citation settings.

Colloquial Term	MTML Theory	Network of Relations	Importance (centrality measure)	Node Attributes
Most Qualified	Self Interest	Agent-artifact relations (e.g., citation)	In-degree within the network of concern	Measures of qualification (e.g., # of publications)
Friend-of-a-Friend	Balance	Agent-agent relations (e.g., collaborations)	Distance (e.g., # of geodesics)	
Birds of a Feather	Homophily			Measures of similarity (e.g., gender)
Social Exchange	Social Exchange	Agent-agent relations		Measures of tie strength
Follow the Crowd	Contagion	Agent-agent relations (e.g., collaborations, citations)	In-degree within the network of concern	

Table 1. MTML concepts and colloquial terms for recommendations

The C-IKNOW Semantic Recommender is specifically developed to make recommendations based on Semantic Web data. It uses the Resource Description Framework (RDF) as representation for instance data and is able to use and integrate various data types. In the future different heuristics will be combined to improve the recommendation engine and p*/ERGM models will be included. Finally structural matching using the taxonomy of data (e.g. specific) topics will be used to identify recommendations.

3.3 Re-Conceptualize the Web

With the increasing complexity and size of the Web comes the need to develop new ways to model and represent the Web. In many cases the Web contains so many data points (e.g. people in an online game) on various different levels (e.g roles in games) that it becomes virtually impossible to present these data using common formats (e.g. tables and graphs). To solve this problem, SONIC researchers try to develop new ways to look to the Web, by adopting Network Science principles. One key example is the use of Hypergraphs [6] to represent data. Hypergraphs are generalizations of graphs in which edges (relations) can connect any number of (types of) nodes (e.g. different types of people).

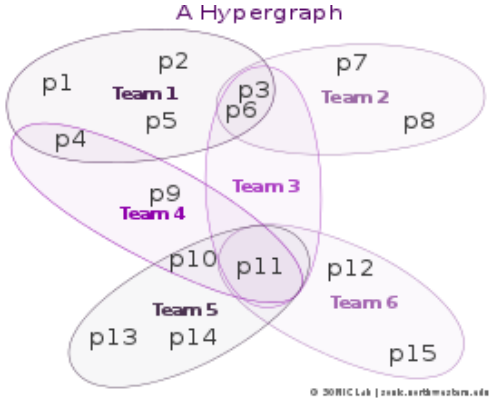

Figure 2. Example of a Hypergraph

Hypergraphs offer a richer way to generate insights about the web as an ecosystem of individuals, teams, and communities. Work in progress on Hypergraphs involves 1) developing metrics to analyze teams as Hypergraphs and 2) developing techniques to assess the emergence and outcomes of teams and ecosystems of teams.

4. CONCLUSIONS

The complexities and size of the Web call for new theories and methods to understand and enable the Web. Using network science in general and Social Network Analysis theories and tools in particular, we have a powerful and unique perspective on the Web. The perspective takes a very *relationship* oriented focus on the Web and in that it distinguishes itself from traditional Web Science approaches. Using MTML and the strong focus on theory as well as empirical research, SONIC aids in the understanding and enabling of the Web.

5. RESEARCHERS INVOLVED

The following researchers in the SONIC group are involved in Web Science related research projects: Prof. Dr. Noshir Contractor, Dr. Yun Huang, Dr. Hugh Devlin, Brooke Foucault, Brian Keegan, Mengxiao Zhu, Ryan Whalen, Alina Lungeanu, Sophia Sullivan, Liwen Ouyang & Zhe Zhang. The following developers and research technologists are involved: York Yao, Jinling Li, Joe Gilborne, Anup Sawant, Curie Chang, Anthony Vashevko & Nicholas Bennett.

6. REFERENCES

[1] Monge, P. & Contractor, N. 2003. *Theories of Communication Networks*. Oxford University Press

[2] Fazel-Zarandi, M., Devlin, H. J., Huang, Y. and Contractor, N. Expert recommendation based on social drivers, social network analysis, and semantic data representation. In Proceedings of the Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (Chicago, Il, 2011). ACM, 2011.

[3] Castronova, T., Williams, D., Shen, C., Ratan, R., Xiong, L., Huang, Y., & Keegan, B. (2009). As real as real? Macroeconomic behavior in a large-scale virtual world. New Media & Society, 11(5).

[4] Foucault Welles, B. & Contractor, N. 2011. From Strangers to Friends Online: A Mixed-Methods Investigation of Friendship Formation on the Internet. Paper presented at the National Communication Association 97th Annual Convention. New Orleans, LA, USA. Top 4 Paper Selection in Human Communication Technology Division, Manuscript Under Review.

[5] Keegan, B. Gergle, D. & Contractor, N. (2012). Do Editors or Articles Drive Collaboration? Multilevel Statistical Analysis of Wikipedia Coauthorship. 2012 AMC Conference on CSCW, Seattle, WA, USA

[6] Contractor, N., DeChurch, L., Carson, J., & Sullivan, S. 2012. On the "Hyper" Edge: A Multilevel Framework for Using Hypergraphs to Understand Collectives. Proceedings from Sunbelt XXXII. Crowne Plaza Hotel Redondo Beach.

Notes

[1] see http://ciknow.northwestern.edu/

[2] see http://ciknow1.northwestern.edu/semanticrecommender