Rank-Based Models of Network Structure and the Discovery of Content

Adam Douglas Henry and Paweł Pralat

1 Division of Public Administration, West Virginia University, Morgantown, WV 26506-6322, USA
2 Department of Mathematics, West Virginia University, Morgantown, WV 26506-6310, USA

Abstract. Research on self-organizing networks, especially in the context of the Web graph, holds great promise to understand the complexity that underlies many social systems. We argue that models of social network structure should begin to consider how structure arises from the “content” of networks, a term we use to describe attributes of network actors that are independent of their structural position, such as skill, intelligence, or wealth. We propose a rank model of how content (operationalized as attribute rank relative to other individuals) may change amongst agents over time within a stochastic system. We then propose a model of network self-organization based on this rank model. Finally, we demonstrate how one may make inferences about the content of networks when attributes are unobserved, but network structures are readily measured. This approach holds promise to enhance our study of social interactions within the Web graph and in complex social networks in general.

1 Why Network Content Matters

Research on the Web graph has been very influential in social science research regarding the structure and function of complex social networks. While the structure and emergence of networks has been a long-standing theme in disciplines such as sociology [1][2], political science [3][4], and economics [5], coupling theoretical models with rigorous models of network self-organization (e.g., [6] and [7]) is still an emerging area of research. Indeed, research on complex networks, especially in the context of the Web graph, has broad applicability in the social sciences and can help to inform methods to unpack the complexity that characterizes many social systems.

At the same time, the social sciences can also contribute to modeling work in mathematics and computer science, since it offers concrete theories about the factors that drive network relationships. Thus, social science theory can help to discipline researchers’ focus on particular models that are likely to be more realistic in particular contexts. In research on the Web graph it is important to focus on social drivers of network structure since the Web is, after all, a self-organizing network created and manipulated by human beings. Interactions within the Web...
graph are both a direct reflection of human behavior (e.g., when organizations decide to reference one another due to shared interests or resources), but also hold promise as indicators for latent forms of socially-relevant relations such as trust or agreement [8].

In this paper we argue that modeling work on social networks should take seriously the role of network content—meaning the inherent attributes of network actors [9]—in driving network self-organization. Some research has begun to do this by asking, for example, how network structures are influenced by the fitness of actors [10], strategies [11], or the spatial positioning of agents [12]. These types of “content” models of network structure are important supplements to classical modeling approaches that emphasize the importance of structural drivers such as node degree or other measures of centrality. This is because many social science theories are ultimately concerned with attributes of individual actors—why are some powerful and others marginalized, why do political organizations behave the way they do, and how are behaviors, norms, or beliefs learned from others within a network. Thus, to understand complex social networks one must consider structure, but also how structure is dependent upon, and co-evolves with, network content. This will allow researchers to move towards coherent theories of emergent behaviors within social networks.

We contribute to this endeavor in two ways. First, this paper posits a simple, mathematically tractable, yet reasonable model of network self-organization that accounts for the ways in which network content drives network structure. This is a contribution in of itself, and builds heavily upon earlier modeling work in this area by Luczak, Prałat, Wormald (e.g., [13], [14], [15], [16]) and especially by Prałat and Janssen (e.g., [17] [18]). The model outlined here is a “rank” model where link formation probabilities are based on externally-determined prestige labels relative to other agents in the system; this general approach was first proposed by Fortunato, Flammini and Menczer in [19]. Thus, this paper is concerned with at least preliminary models of network self-organization.

Second, and more importantly, we investigate how this network model may be used to estimate network content—that is, the rank of nodes—based on observed structure alone. This is an important area for research, since in many applications of social network analysis we may know the structure of the network (for example, if networks amongst organizations are measured using hyperlink data), but attributes of actors remain a latent, unobserved variable. Our research builds on prior work to estimate node attributes from observed structure [12], although this research involved a different model and was focused on predicting distances between nodes rather than the attributes of the nodes themselves. We find that making inferences about node ranks is eminently doable, and this research establishes a baseline for methods of statistically inferring node attributes from network structure only. We illustrate the use of this approach through computational simulation, which provides a starting point for future work emphasizing mathematical proof.

The progression of this paper is as follows. We first discuss how the content of a network may be thought of in terms of the rank of vertices—while this is