Scale Free: Twitter's Retweet Network

Ramine Tinati – rt506@ecs.soton.ac.uk – Web and Internet Science – University of Southampton – England

INTRODUCTION

A wide range of research have identified scale-free networks – a real-world and man-made phenomena – in networks such as protein networks (Jeong, Mason, Barabasi, & Oltvai, 2001), journal co-citation networks (A. Barabasi et al., 2002) and the World Wide Web (Albert, Jeong, & Barabasi, 1999). Common properties such as preferential attachment and growth enable these networks to be classified as scale-free (A.-L. Barabasi & Albert, 1999). These properties can clearly be identified in networks such as the World Wide Web; a complex man-man network of documents and links that grows in uncontrollable manner (Barabási, 2002), they produce the 'rich-get-richer' effect (A.-L. Barabasi & Albert, 1999), where nodes increase their connectivity at the expense of younger less well connected ones. By mapping the complex real-world and man-made networks, these studies are helping improve our knowledge on the "weblike" world we live in (A. Barabasi et al., 2002). However, as many of these scale-free networks still yet to be discovered, generalizing a scale-free model requires is still problematic (A.-L. Barabasi & Albert, 1999; Barabási, 2009).

Our research focuses on a network which is both a product of man-made networks, and real-life phenomena. Twitter, a micro-blogging social networking service provides a simple service to enable users to broadcast messages and form networks of 'friends' and 'followers'. Twitter has become a powerful tool for mass communication, with over 300 million tweets published each day. This study aims to examine the structure of the underlying network of users and tweet.

BACKGROUND

Studies have examine the structure of Twitter's static networks that form as a result of the friends and follower links between users (Choudhury, 2011; Teutle, 2010). There has also been a growing interest in exploring how it can be used to solve real-world problems (Paul & Dredze, 2011; Weng & Lee, 2011), and findings ways to classifying (Pennacchiotti & Popescu, 2011) and identifying influential users (Anger & Kittl, 2011; Weng, Lim, & Jiang, 2010).

As an alternative approach, we examined the dynamic network structures of Twitter conversations which form through the passing of messages between users, Our study builds upon other research examining the cascade of messages within a network (Bakshy, Hofman, Watts, & Mason, 2011) and places more emphasis towards examining the dynamic network structures that form within Twitter.

METHODOLOGY

In this study, a number of Twitter datasets were collected varying in size, region and topic. To capture a number of different conversation streams on Twitter, each dataset contained tweets related to a specific hashtag (#), shown in Table 1. Each of these datasets were then parsed, and the retweet edge – which represents a directed edge between one user to another – was extracted along with their timestamp. The process of a user retweeting acts as a mechanism to share a tweet of a specific user to more users, in effect amplifying the size of the audience. As Figure 1 shows, when a user U_1 retweets another user's tweet (U_2), U_1's followers (U_3 and U_4) will see the tweet as well.

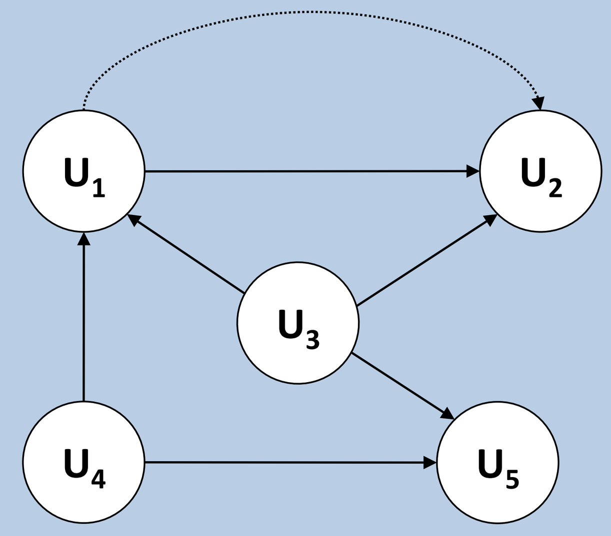

Figure 1. Twitter Retweet Functionality

FINDINGS

Constructing a directed graph of each of the datasets based on their respective extracted retweet edges, the initial findings of the analysis have shown that there exists a power law with similar exponents across all analyzed datasets in regards to the decay of 'retweeted' (or shared) messages between users, shown in Figure 2.

The exponents found – which ranged from 1.2 to 1.5 – are lower than similar scale-free networks such as the Web. Typically, such a low exponent would indicate a skewed and uncorrelated network as a result of the number of edges growing faster than the number of nodes (Seyed-allaei, Bianconi, & Marsili, 2006). However the Twitter networks examined exhibit scale-free properties such as preferential attachment and growth as networks of a higher exponent.

The findings of this study raises questions about the nature of communication in social networking sites, specifically the use of the retweet function and how it is changing not only the structure of the network, but the role that individuals have within a conversation.

A Cross-Comparison of Twitter Hashtag Retweet Networks

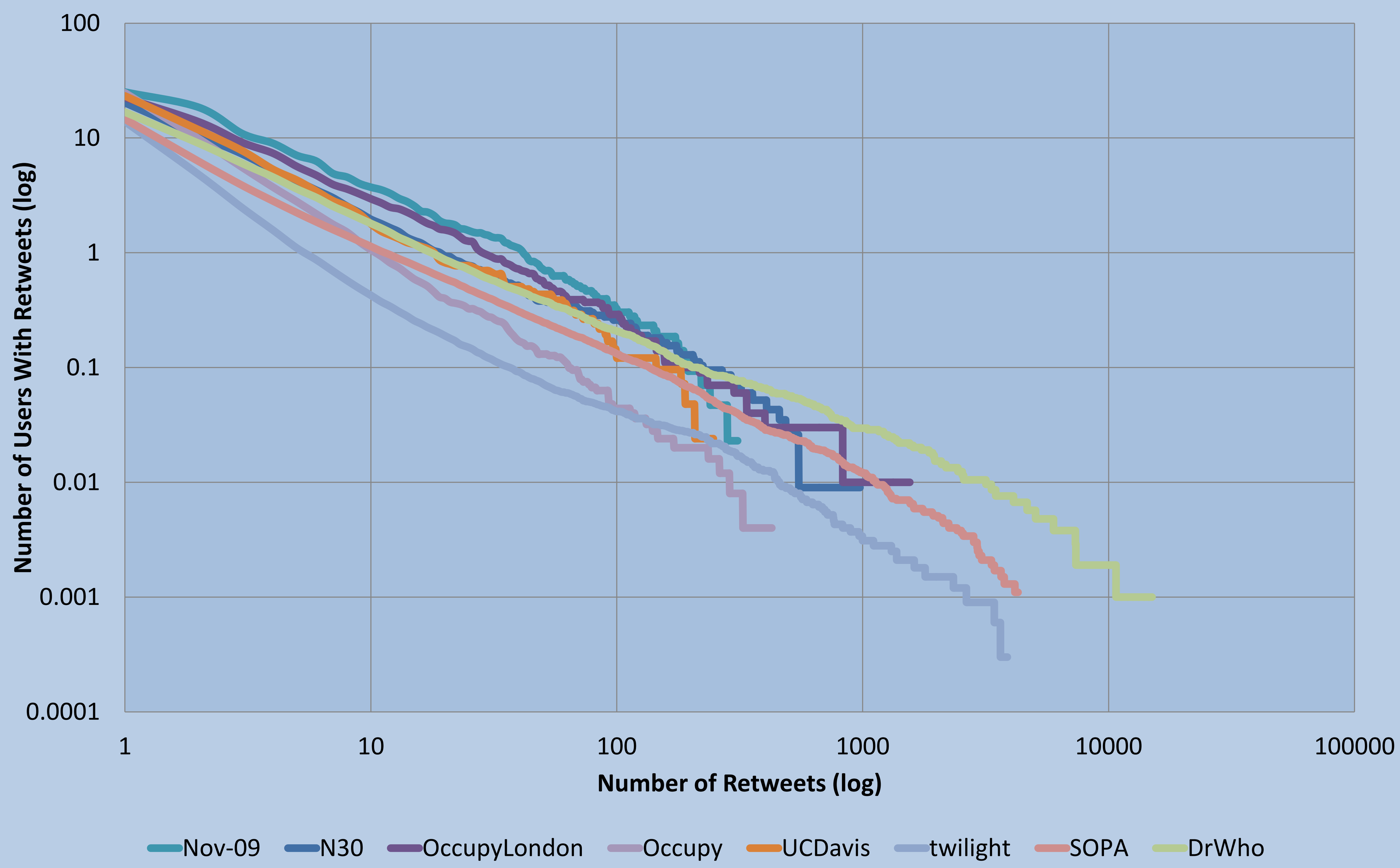

Figure 2. Log/Log Distribution of retweets amongst users within a Twitter conversation

Dataset	Tweets	Retweets	Users	k
#twilight	529530	139441	336446	1.233
#DrWho	709093	204301	104688	1.233
#SOPA	1004482	438894	485692	1.236
#Occupy	41568	16673	29025	1.527
#OccupyLondon	19128	9834	7548	1.218
#Nov9	12831	7188	4737	1.389
#Nov30	22054	14243	12330	1.398
#UCDavis	7950	3895	4523	1.298

Table 1. Harvested Twitter Datasets and their power law exponent (k)

Although the datasets varied in topic, size, and geographic relevance, the distribution of retweets amongst users were similar, with a mean exponent(k) of 1.310