Social Networks 39 (2014) 1–11

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Efficiency of human activity on information spreading on Twitter

A.J. Morales, J. Borondo, J.C. Losada, R.M. Benito*

Grupo de Sistemas Complejos, Departamento de Física y Mecánica, Universidad Politécnica de Madrid, ETSI Agrónomos, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Complex networks
Social networks analysis
Information spreading
User behavior
Twitter
Influence

ABSTRACT

Understanding the collective reaction to individual actions is key to effectively spread information in social media. In this work we define efficiency on Twitter, as the ratio between the emergent spreading process and the activity employed by the user. We characterize this property by means of a quantitative analysis of the structural and dynamical patterns emergent from human interactions, and show it to be universal across several Twitter conversations. We found that some influential users efficiently cause remarkable collective reactions by each message sent, while the majority of users must employ extremely larger efforts to reach similar effects. Next we propose a model that reproduces the retweet cascades occurring on Twitter to explain the emergent distribution of the user efficiency. The model shows that the dynamical patterns of the conversations are strongly conditioned by the topology of the underlying network. We conclude that the appearance of a small fraction of extremely efficient users results from the heterogeneity of the followers network and independently of the individual user behavior.

1. Introduction

In the recent years, our society has experienced the rise of new ways to communicate and relate among each other through digital devices. The increasingly affordability of technology, together with the solutions brought, have turn mobile and Internet devices as one of the fastest growing markets worldwide (Infiniti, 2013). Specially in third world countries where the expanding projections of technological solutions double those found in the industrialized world (Cisco, 2013). Such technological revolution has given as a result, a massive amount of data provided by humans, as they interact with their digital devices on daily basis. The nowadays challenge is to turn these unstructured data into valuable information for policy makers to take better and more intelligent decisions (Lazer et al., 2009).

At the moment, traditional surveys have given important insights to our societal understanding. However, their cost in time and human efforts, makes it impossible for them to scale up and bring information of the structure of the social system behind their observation. Traditionally, the discovery of structural properties of social networks have been limited to the necessity of mapping a large amount of interactions between people. In this sense, online social networks, such as Twitter or Facebook, have become an ideal source of information to collect human-to-human interactions and unveil the social structures that people constitute, which opens an opportunity for researchers to characterize and model human behavior (Lewis and Christakis, 2008; Takhteyev et al., 2012). These web applications are used on daily basis by people to post opinions, propagate news and exchange information. As a result, several commercial, political and social organizations are increasingly exploiting these communication tools to advertise products, organize campaigns and disseminate updates on their respective fields.

Twitter, with over 200 million users, is the ideal tool to quickly propagate short text messages. It is an open debate that the data taken from Twitter are not necessarily representative samples of the outside world, as they are constrained to the population that participates in the online conversations (Mislove et al., 2011; Gayo-Avello, 2012). However, a social contextualization of the data, combined with a suitable computational and mathematical treatment, may provide important insights into how people behave. In fact, the activity performed by users on Twitter has brought information enough to understand a wide variety of phenomena, like the prediction of stock market variations (Bollen et al., 2011), the management of natural disasters (Sakaki et al., 2010), the understanding of epidemical diseases (Culotta, 2010) and the characterization of electoral processes (Borondo et al., 2012; Livne et al., 2011). The deeply understanding of these social processes is crucial to design better strategies and get optimal outcomes from the network potential.

Recent studies have revealed that most of the information posted on Twitter is hardly propagated through the network, as 71% of the messages do not travel any farther than the authors timeline (Cheng and Evans, 2009). Among other factors, this spreading

* Corresponding author. Tel.: +34 913365646; fax: +34 913365726.
E-mail address: rosamaria.benito@upm.es (R.M. Benito).

http://dx.doi.org/10.1016/j.socnet.2014.03.007
0378-8733/

inertia has been attributed to the fact that the novelty of the posted information decays quite rapidly, which stretches the effective time to attract the collective attention (Asur et al., 2011), in addition to the fact that most of the people on Twitter behave passively (Romero et al., 2011). However, in this context, there are people who do influence the rest of users and are able to get their messages spread through the network, in a wide variety of proportions.

The keys to success when propagating information on Twitter have been reported to be a combination of several factors, such as the popularity of the source, the posting frequency, as well as the novelty and resonance of the message content (Romero et al., 2011). In fact, the largest retweets cascades on Twitter, were found to be seeded by previously popular users, whose messages contained positive feelings (Bakshy et al., 2011). However, the efforts of each user to gain influence and get their information spread on the network is a subject that has not yet been explained. In the sense that although users may gain enough influence to transfer information on the network, this influence is not necessarily achieved with the same efficiency, in terms of the amount of efforts that had to be employed for this matter.

In this work we address the question of which factors, like the individual behavior or the underlying substratum, determine the users efficiency to have their messages spread through the network. More specifically, we propose a measure to characterize the user efficiency to influence the emergence and growth of retweets cascades, by means of the relationship between the activity employed by the users and the emergent collective response to such activity, measured in terms of the number of retransmissions gained. On this basis, we propose a model to understand the emergence of the user efficiency distribution, based on independent cascades taking place on networks (Goldenberg et al., 2001), biasing the probability of retransmission among nodes, in order to decay as we move farther from the message source, as we see in the empirical data.

The results indicate that some regular users may gain a similar amount of retransmissions as the popular ones, but far less efficiently, as they must employ a much larger amount of activity. Furthermore, we have seen that the emergent distribution of users according to their efficiency, is strongly conditioned to the underlying network where information is being propagated. As a matter of fact, it actually represents a reflection of the dynamical rules behind the spreading process.

The paper is organized as follows. First, we introduce the system of our study in Section 2, as well as the datasets that we have built and analyzed. Then in Sections 3–5 we focus on the empirical measurements that lead us to state the dynamical rules of the propagation process. After this, in Section 6 we propose a simple model to verify the dynamical processes reported. Finally, we discuss the effects of the underlying topology and initial user activity behavior in the emergent dynamical patterns, which we found to be universal on Twitter conversations.

2. System

The system under study is based on human activity taking place around specific topics of conversation on Twitter. In this section we give some background on the user interaction mechanisms provided by Twitter, as well as describe the datasets that we have built and analyzed.

2.1. Twitter background

Twitter is a microblogging service where people are able to post and exchange text messages limited by 140 characters either from personal computers or mobile devices. There are several mechanisms for users to interact on Twitter. The first of these is the ability to follow and be followed by other persons. This is a passive mechanism that allows users to receive all the messages posted by those who follow, as well as to deliver their own messages to their own followers. In this sense, it establishes the Twitter followers network, where the users are connected among each other, through links that determine the explicit ways where messages are delivered. Previous studies have reported complex properties in this network (Kwak et al., 2010), like degree distribution with power law behavior, small mean distance between nodes and modular structure. However, it has been observed that individuals do not actively interact with all of the declared contacts, but only with a small fraction of them (Huberman et al., 2009). Among these active mechanisms to interact, the *retweet* (or retransmission) is the most popular one to propagate the received messages throughout the network. By retweeting a message, users deliver specific information to their own followers, at the same time that endorse ideas and gain visibility in the network (Boyd et al., 2010). The study of the retweets cascades has served to characterize user profiles (Galuba et al., 2010), measure influence (Cha et al., 2010) and propose spreading models (Xiong et al., 2012). At last, all messages on Twitter, may be identified using keywords called *hashtag*. This mechanism organize conversations and individuals use it to exchange ideas on specific subjects. Recently, the statistical analysis of the hashtags usage has let prediction on social relations (Romero et al., 2011) and collective attention (Lehmann et al., 2012).

Table 1
Properties of the studied datasets and their resulting user efficiency distribution properties.

Keyword	Messages	Users	μ_η	σ_η
Andreafabra	35, 835	23, 498	0.15	1.05
Gingrich	93, 063	43, 061	−0.08	1.13
Leones	142, 808	46, 608	−0.08	1.09
20N	389, 988	123, 710	−0.49	1.08
SOSInternetVE	421, 602	77, 706	−0.79	1.21
Obama	6, 818, 782	2, 265, 799	0.14	1.15
Egypt	7, 433, 542	1, 180, 715	−0.80	1.33

2.2. Datasets

Using the Twitter Search API version 1.0,[1] we have built several datasets from public access messages. This API provides data from a temporal index of recent tweets, posted within a lapse of a week from the time the query is made. The limitations of this API are not specified as a relative volume of messages, nor a fixed number of queries, but instead a combination of the queries' complexity and frequency. The datasets were built querying for messages with specific keywords related to topics of conversation that captured a significant part of the collective attention. Their sizes vary from 10^4 to more than 10^6 messages or participants, as may be seen in Table 1.

First, we considered an online Venezuelan political protest as a case study. This event took place exclusively on Twitter on December 16th, 2010. Two days before the protest, the convoker asked his followers to post messages identified with the hashtag #SOSInternetVE, who responded massively and the conversation propagated becoming trending topic. We collected up to 421,602 messages, identified with the protest hashtag, which were posted by 77,706 users, between December 14–19, 2010 (two days before and after the protest). In our previous work (Morales et al., 2012), we found that some influential users acted as information producers, providing messages that are received by the passive large majority of information consumers. Besides, we found that users

[1] https://dev.twitter.com/docs/using-search.