



Large-scale evaluation framework for local influence theories in Twitter



Magdalini Kardara^{a,*}, George Papadakis^b, Athanasios Papaoikonomou^a, Konstantinos Tserpes^a, Theodora Varvarigou^a

^a National Technical University of Athens, Greece

^b IMIS, Athena Research Center, Greece

ARTICLE INFO

Article history:

Received 4 September 2013

Received in revised form 18 June 2014

Accepted 19 June 2014

Available online 22 July 2014

Keywords:

Evaluation framework

Social influence

Topic communities

ABSTRACT

Influence theories constitute formal models that identify those individuals that are able to affect and guide their peers through their activity. There is a large body of work on developing such theories, as they have important applications in viral marketing, recommendations, as well as information retrieval. Influence theories are typically evaluated through a manual process that cannot scale to data voluminous enough to draw safe, representative conclusions. To overcome this issue, we introduce in this paper a formalized framework for large-scale, automatic evaluation of topic-specific influence theories that are specialized in Twitter. Basically, it consists of five conjunctive conditions that are indicative of real influence exertion: the first three determine which influence theories are compatible with our framework, while the other two estimate their relative effectiveness. At the core of these two conditions lies a novel metric that assesses the aggregate sentiment of a group of users and allows for estimating how close the behavior of influencers is to that of the entire community. We put our framework into practice using a large-scale test-bed with real data from 75 Twitter communities. In order to select the theories that can be employed in our analysis, we introduce a generic, two-dimensional taxonomy that elucidates their functionality. With its help, we ended up with five established topic-specific theories that are applicable to our settings. The outcomes of our analysis reveal significant differences in their performance. To explain them, we introduce a novel methodology for delving into the internal dynamics of the groups of influencers they define. We use it to analyze the implications of the selected theories and, based on the resulting evidence, we propose a novel partition of influence theories in three major categories with divergent performance.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In the context of a social network, *influencers* are prominent individuals with special characteristics that enable them to affect a disproportionately large number of their peers with their actions. Their special characteristics are related to their individual activity and social background as well as to their position in the network (i.e., their connections with the other members). These influencers typically play a crucial role in a variety of scientific and business domains (Bakshy, Hofman,

* Corresponding author. Address: National Technical University of Athens, 9 Iroon Polytechniou Str., 15773 Zografou Campus, Athens, Greece. Tel.: +30 2107772568.

E-mail address: nkardara@mail.ntua.gr (M. Kardara).

Mason, & Watts, 2011). For example, marketing campaigns could gain value from this process, since individual customers are prone to imitate their highly influential peers with respect to product adoption (Keller, Fay, & Berry, 2007). Instead of overwhelming the entire customer base with massive, but blind advertisements, marketing campaigns could target a small number of influential people. This cost-effective alternative is called *viral marketing* and is capable of achieving similar levels of product diffusion with traditional approaches (Brown & Hayes, 2008).

To facilitate such applications, a lot of research has focused on the identification of influencers. This effort led to the development of *influence theories*,¹ which constitute formal models that estimate for every member of a social network the influence she exerts on her peers according to one or more criteria. Some theories study influence on a *global* scale, considering the activity and the user base of the entire social network. Most commonly, though, an individual's influence is *local*: she may be considered expert in a specific domain, but her opinion usually holds little weight outside this particular area. Based on this principle, *local influence theories* aim at identifying influencers among the members of individual communities, which are usually formed around a particular topic. Typically, the local theories are more accurate and efficient than the global ones, as they exclusively consider the activity and the dynamics inside the individual communities.

A major issue in the study of local influence theories is their evaluation. Over the recent years, *On-line Social Networks* (OSNs) have provided researchers with powerful tools for studying the dynamics of influence diffusion. They contain a vast amount of user-generated content as well as explicit connections among their members, thus allowing for the analysis of social influence on an unprecedented scale. Still, influence constitutes a subjective concept and, as such, it is very hard to measure and track. Most works actually lack a formal methodology for evaluating the results produced by their theories. Instead, they typically resort to selecting a small sample of the top ranked users in order to assess their authority in the real world (e.g., their fame or the quality of their content) (Cha, Haddadi, Benevenuto, & Gummadi, 2010; Weng, Lim, Jiang, & He, 2010). This manual procedure, however, cannot scale to large volumes of data and, thus, is incapable of yielding representative, reproducible and generalizable results.

In this work, we aim to overcome this shortcoming, by establishing a principled framework that is capable of evaluating local influence theories for OSNs on a large-scale. It receives as input the groups of influencers they define – called *prominent groups* in the following – along with the rest of the community and the corresponding activity. The goal of our framework is to estimate the relative accuracy of influence theories in predicting activity patterns that denote an imitation by the rest of the community. Internally, our framework encompasses five conditions that should be satisfied by a prominent group with real influence over the other members of the community. These conditions can be summarized as follows:

1. real influencers comprise a small subset of the community,
2. they are able to affect their fellow members with limited cost, i.e., by accounting for a limited portion of the community's overall activity,
3. their activity is highly correlated with that of the remaining community with respect to an objectively measured metric,
4. their activity that is relevant to this metric chronologically precedes that of their peer community members, and
5. the volume of their activity that is relevant to this metric corresponds to a mere fraction of the overall activity this metric takes into account.

Conditions 1, 2 and 5 actually correspond to the pre-processing requirements of our framework. Their goal is to ensure that a prominent group is compatible with it, accounting for a limited portion of the activity and the user base of the underlying community. These are fundamental prerequisites for drawing safe conclusions from the analysis performed by our framework. The remaining two conditions encapsulate the real functionality of our framework. At their core lies an objectively measurable metric that correlates the activity of a prominent group with the rest of the community. To elucidate its functionality, consider a metric that assesses the aggregate sentiment of a group of users; a high correlation between the prominent group and the rest of the community members indicates that the stance of the former coincides with the overall “mood” of the latter. Failure with respect to either of these conditions indicates a theory that is inadequate in identifying real influencers. In contrast, an influence theory is effective if the individuals it marks as influencers satisfy both conditions. The stronger these conditions hold for them, the more effective the theory is. For instance, among two theories with similar performance, the one that exhibits higher correlation with the rest of the community is preferred.

Given that all five conditions rely on objectively measurable metrics, our framework allows for comparing the performance of local influence theories on a large-scale, without manual intervention. To put it into practice, we form a large-scale benchmark dataset that consists of real-world data. We actually draw our data from Twitter,² which was selected for several reasons (Bakshy et al., 2011; Cha et al., 2010; Weng et al., 2010): it is one of the most popular OSNs in the field, it conveys ad hoc, yet clear and manageable rules for social interaction among its members, it abounds in dynamic topic communities and finally, it provides easy access to large volumes of user-generated content. In total, our test-bed comprises 75 topic communities

¹ It should be stressed at this point that the term “influence theory” is used abusively in this work, since it does not refer to actual theories that provide insights into the behavior of social network users, explaining why some of them imitate the behavior of others. Instead, the term refers to *influence ranking methods*, which associate every user with a score that is proportional to an estimation of the influence she exerts on her peer members. By convention, though, these methods are termed influence theories in the literature. This convention is followed in this work, as well.

² <https://twitter.com/>.

Download English Version:

<https://daneshyari.com/en/article/515387>

Download Persian Version:

<https://daneshyari.com/article/515387>

[Daneshyari.com](https://daneshyari.com)