Neurocomputing 172 (2016) 84–99

Contents lists available at ScienceDirect

Neurocomputing

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

Macro-level information transfer in social media: Reflections of crowd phenomena

Minkyoung Kim [a,*], David Newth [b], Peter Christen [a]

[a] Research School of Computer Science (RSCS), College of Engineering and Computer Science (CECS), The Australian National University (ANU), Canberra ACT 2601, Australia

[b] CSIRO Centre for Complex Systems Science, CSIRO Oceans and Atmosphere, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra ACT 2600, Australia

ARTICLE INFO

Article history:
Received 1 December 2013
Received in revised form
26 November 2014
Accepted 9 December 2014
Available online 11 May 2015

Keywords:
Macro-level information transfer
Transfer entropy
Crowd phenomena
Social media

ABSTRACT

Online social interactions have become more dynamic and far-reaching across multiple social media platforms than ever before. This is because of frequently changing online contact networks and increasing accessibility to diverse information sources outside of the networks. Accordingly, massive user-generated content spreads through heterogeneous social networks beyond a single social platform. The main goal of this paper is to propose a model-free approach for estimating macro-level information transfer across heterogeneous populations without any assumptions on such dynamic and complex social networks. With this approach, we estimate macro-level diffusion across mainstream news (News), social networking sites (SNS), and blogs (Blog), and the estimations are compared with outcomes from our previous model-driven approach. We also analyze crowd phenomena in diffusion for News, SNS, and Blog as online social systems in terms of activity, reactivity, and heterogeneity. We find that News is the most active, SNS is the most reactive, and Blog is the most persistent, which governs time-evolving heterogeneity. Discovered crowd phenomena are interpreted with respect to our proposed approaches. The strength and directionality of influence reflect reactivity, while topic-related diffusion patterns reflect heterogeneity. We expect that this study can provide a consistent way of understanding cross-population diffusion in diverse application domains.

1. Introduction

Interactions between real-world entities are heterogeneous and complex, since underlying network structures are not only far-reaching across different (social) systems or regions but also dynamically changing in accordance with the context of diffusion over the networks. For instance, from a neuroscience point of view, brain structures are not only functionally segregated but also interconnected across different cortical regions, and the underlying neuronal connectivity may vary by different input signals to the brain [1–5]. Thus, the cortical network structures are hard to collect and define, which makes it challenging to uncover diffusion mechanisms in the human brain.

When it comes to social media, diverse information sources have become increasingly accessible due in part to news organizations' active participations in social networking platforms and the help of web technologies such as feed readers and social media aggregators. That is, the underlying social interactions have become more reachable across multiple social media platforms, which leads heterogeneous social networks to be more interconnected than ever before [6]. Such emergent phenomena of complex online social networks, as shown in Fig. 1, have been investigated in a wide spectrum of diffusion spaces: from a single social networking platform such as Twitter [7–10] to *n-Sphere*, consisting of two or more different social platforms, such as blogosphere [11,12], blogosphere-to-YouTube [13], blogosphere-to-news [14,15], and blogosphere-SNS-news [6,16,17].

Accordingly, the underlying diffusion mechanisms have been studied by modeling processes of information diffusion over social networks. Most prior work has focused on micro-level diffusion within a single social platform alone, and their models are based on assumptions that social interactions are from sampled snapshots of current social networks [12,13], or from site-specific actions (e.g., metions, retweets, and hashtags in Twitter) [7,9,10]. In this regard, our previous study [6] modeled macro-level diffusion across heterogeneous social networks without the knowledge of local network structures in detail, but this model was based on the assumption of a power-law degree distribution. Thus, our

* Corresponding author.
E-mail addresses: minkyoung.kim@anu.edu.au (M. Kim), david.newth@csiro.au (D. Newth), peter.christen@anu.edu.au (P. Christen).

http://dx.doi.org/10.1016/j.neucom.2014.12.107
0925-2312/

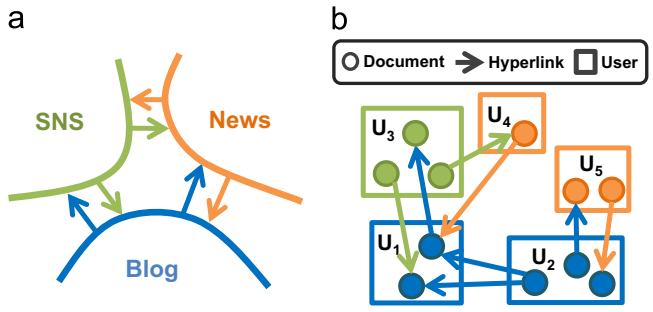

Fig. 1. Overview of cross-population diffusion on the Web. (a) Interactions across different types of social media. (b) Dual representations of networks: document and authorship networks.

previous model still needs to estimate the value of the scaling exponent for the power-law tail [18].

The main purpose of this paper is to propose a model-free approach to estimate macro-level information transfer across heterogeneous populations without any assumptions on underlying social structures by using information-theoretic measures. This is achieved by considering a social system as a stochastic process and by defining the system's signals at a population level. Here, we follow the definition of a *social system* by Talcott Parsons [19], which consists of individuals who interact and influence each other's behavior. Identifying a system's signals at its population level enables to catch invisible information flow between social systems at a macro level. For this identification, we use the transfer entropy which provides the strength and directionality of influence between two random processes. For a more accurate estimation of information transfer, we do not ignore time-delay and memory effects on diffusion.

As one of the real-world applications, we focus on news diffusion in social media and apply our approach for estimating global diffusion across different types of social media such as mainstream news (News), social networking sites (SNS), and blogs (Blog) as representative online social systems. The estimated diffusion patterns with this model-free approach are compared with outcomes from our previous model-driven approach [6] in terms of the strength and directionality of influence. This comparison helps to obtain a more consistent understanding of cross-population diffusion in social media. Finally, we analyze common and distinct collective diffusion behavior of the News, SNS, and Blog systems in terms of activity, reactivity, and heterogeneity, which is interpreted with respect to our current and previous proposed approaches.

For this study, we use the Spinn3r dataset [20] which was preprocessed in our previous work [6] for studying diffusion mechanisms. This dataset contains daily news adopters in the News, SNS, and Blog systems for each during a one month period in early 2011. Identified news topics are in common across the systems and cover eight different information categories of conventional news outlets (i.e. Arts, Culture, Disasters, Economy, Politics, Science, Sports, and Technology), which enables us to quantify and understand topic-related diffusion patterns in a more principled way. As a result, our current model-free and previous model-driven approaches exhibit similar topic-related diffusion patterns in terms of the strength and directionality of influence, but with different perspectives on diffusion. As discovered crowd phenomena in diffusion, News is the most active, SNS is the most reactive, and Blog is the most persistent, which governs time-evolving heterogeneity of diffusion. When we interpret the crowd phenomena with respect to our approaches, systems' reactivity levels are reflected in estimated memory effects and diffusion rates from our model-free and model-driven approaches, respectively. Heterogeneity is reflected in topic-related diffusion patterns from our both approaches. That is, the Politics and Disasters categories tend to drive more balanced influence than the Culture and Technology categories, which is consistent with higher heterogeneity of populations (more evenly distributed populations) for the Politics and Disasters categories than the others. Such macro-level cross-population diffusion from various angles, to the best of our knowledge, is studied for the first time.

The main contributions of this study are as follows. First, we propose a model-free approach to estimate macro-level diffusion across heterogeneous (social) systems without any assumptions on complex and dynamic interactions between entities, which not only provides benefits of studying dynamics of complex systems but also presents a macroscopic view of global diffusion. Second, this study provides a consistent understanding of diffusion by comparing outcomes from our current model-free and previous model-driven approaches, which suggests alternative or complementary options to apply to different research fields. Finally, the way of interpreting crowd phenomena suggests diverse aspects of diffusion analytics. We expect that this study can help to quantify and understand cross-population diffusion in a wide range of research areas such as neurocomputing and social science for estimating brain connectivity patterns across functional cortical regions and dynamics of influence among social organizations, respectively.

The rest of this paper is structured as follows. We begin with related work in Section 2 and then describe our dataset and prepocessing steps in Section 3. We propose a model-free approach to estimate macro-level information transfer across different social systems and apply this approach to inferring news diffusion in social media in Section 4. In Section 5, we analyze crowd phenomena in news diffusion in terms of activity, reactivity, and heterogeneity. In Section 6, we thoroughly discuss findings by comparing our current and previous studies and by interpreting the discovered crowd phenomena with respect to our proposed approaches. Finally, Section 7 concludes this study with a discussion of future work.

2. Related work

Identifying trending topics is important since diffusion patterns have significant variations in accordance with the context of information [6,9,10,13,16,22]. There have been attempts to identify specific instant events such as Twitter-only trends [8,23] or approximated events using text clustering [24]. However, these recognized trending topics can be site-specific. For instance, the authors of [8] found that only 3.6% of Twitter's trending topics exist in the hot search keywords from Google. This paper, on the other hand, studies cross-population diffusion, and thus topics of common interest are important to investigate the trends of information diffusion across different online social systems. In this context, we use the Wikipedia Current Events [21] as a pertinent and noteworthy news registry.

Extensive research on information-theoretic measures has been conducted in diverse areas including computer science [25–28], neuroscience [29–33], and economics [34–36]. When it comes to social media, the predictability of user behavioral patterns has drawn attentions such as individual or group level future interactions [28], information pathways between Twitter users without explicit follower–followee relationships [26,27], and classifications of Twitter user behaviors [25]. These studies are all from the aspects of intra-relationships within a single social platform alone.

However, as discussed in the previous section, heterogeneous social networks are increasingly interconnected across multiple social media platforms (e.g., CNN, BBC, Facebook, Twitter, WordPress, and Tumblr). Such emergent complex networks demand