Macro-Level Information Transfer across Social Networks

Minkyoung Kim
Research School of Computer
Science, The Australian
National University
Canberra ACT 0200, Australia
minkyoung.kim@anu.edu.au

David Newth
Commonwealth Scientific and
Industrial Research
Organisation
Canberra ACT 2600, Australia
david.newth@csiro.au

Peter Christen
Research School of Computer
Science, The Australian
National University
Canberra ACT 0200, Australia
peter.christen@anu.edu.au

ABSTRACT

This study proposes a model-free approach to infer macro-level information flow across online social systems in terms of the strength and directionality of influence among systems.

Categories and Subject Descriptors

H.1.1 [**Systems and Information Theory**]: Information theory; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Information networks*; J.4 [**Computer Applications**]: Social and behavioral sciences

Keywords

Information transfer; social media; news diffusion

1. INTRODUCTION

Real diffusion networks are hard to obtain particularly in the case of interconnected heterogeneous social networks [4]. In this regard, our previous study [3] modeled macro-level diffusion across different populations, without requiring local network structures in detail. The proposed model was based on the assumption of a power-law degree distribution which real-world networks exhibit in general [5], and thus this model still needs to estimate its power-law exponent (α). Beyond such an assumption, this study aims to infer cross-population diffusion with a model-free approach by using information theoretic measures such as transfer entropy. This attempt can provide benefits when studying dynamics of large complex networks since it is independent of any specific assumptions on the interactions between individuals.

We investigate real-world news diffusion in social media as a working example of cross-population diffusion by using commonly interested topics across different types of social media, such as mainstream news, social networking sites, and blogs. To the best of our knowledge, this is the first attempt to study macro-level cross-population diffusion in social media using a model-free approach.

2. MACRO-LEVEL INFORMATION FLOW

Let us consider two discrete-time random processes, X and Y. Then, the *transfer entropy* (TE) from the process Y to X is defined as the conditional mutual information [7]:

$$TE_{Y \to X} = I(X_{t+1}; Y_t | X_t)$$

$$= \sum p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log \frac{p(x_{t+1}|x_t^{(k)}, y_t^{(l)})}{p(x_{t+1}|x_t^{(k)})}, \quad (1)$$

where x_t and y_t represent the states of X and Y at time t, respectively, and k and l denote the Markov orders for the previous states of X and Y. Thus, Equation (1) describes the predictability (reduction of uncertainty) of the state x_{t+1} of the process X, given the k previous states, by introducing the l previous states of the process Y, which provides the strength and directionality of information flow.

Our goal is now to estimate system-level information transfer using the transfer entropy. Let $a(t)$ be the number of new adopters at time t and $f(t) = a(t)/n$ be the proportion of new adopters, given the size of a population n. An online social system can be considered as a stochastic process which is a collection of new adoption rates at every time step, and the variation in the new adoption rates of one social system influences those of other social systems. Therefore, we consider the trend (-1: decrease, 0: transition, 1: increase) in new adoption rates, $f(t)$, as the discrete states of an online social system, X,

$$x_t = \begin{cases} -1, & \text{if } \Delta f(t) \in (-\infty, -\tau] \\ 0, & \text{if } \Delta f(t) \in (-\tau, \tau) \\ 1, & \text{if } \Delta f(t) \in [\tau, \infty) \end{cases} \quad (2)$$

where $\Delta f(t) = f(t) - f(t-1)$, and τ is introduced for the discretization of $f(t)$. The value of τ ($=0.006$) is determined so that the probabilities of three states are equally likely. For the calculation of the transfer entropy, it is necessary to estimate the probability distributions in Equation (1), but limited sample size brings about a biased estimation of the entropy. To resolve this issue, we subtract the Panzeri-Treves bias [6] from the estimated entropy with samples as [8].

3. EXPERIMENTS

We investigate real-world news diffusion in social media as a working example of cross-population diffusion, and we consider social media as stochastic communication channels across online social networks such as mainstream news (News), social networking sites (SNS), and blogs (Blog). We use the Spinn3r dataset from our previous study [3], which contains daily cumulative adopters from three different types of online social systems (News, SNS and Blog) during a month in early 2011. Initially, 284 news topics were identified by using the Wikipedia Current Events as a noteworthy
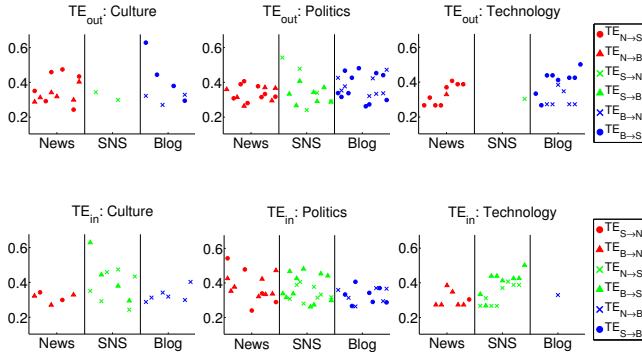
Figure 1: Transfer entropy for different news categories. TE_{out}/TE_{in} means outgoing/incoming transfer entropy. The x-axis indicates the news index for topics in a corresponding category, which is repeated for each media type (vertical lines). Media types are color-coded, and target (in TE_{out}) and source (in TE_{in}) media types are also distinguished by shapes. Note that only significant cases are shown for clarity.

news registry [4], and we chose 63 topics which have driven the largest diffusion across social media [3]. The selected topics cover representative categories of conventional news outlets such as politics (e.g., political protests in the Middle East), culture (e.g., multiculturalism failure), and technology (e.g., iPad2 release), and thus we can understand information flow with different context (see [2, 3, 4] for details).

News, SNS, and Blog are considered as three discrete-time random processes, N, S, and B, respectively. Accordingly, we estimate the transfer entropy of possible combinations, i.e., $TE_{N \to S}$, $TE_{N \to B}$, $TE_{S \to N}$, $TE_{S \to B}$, $TE_{B \to N}$, and $TE_{B \to S}$ for all 63 news topics. Figure 1 shows the estimated outgoing and incoming transfer entropy for each media type, separated by different categories, from time-series adopting trends (x_t in Equation (2)) of News, SNS, and Blog, when $k=l=1$ in Equation (1). We only considered significant transfer entropy whose value is greater than that of 1,000 shuffled data with 95% confidence interval. For the culture category, News and Blog have higher outgoing transfer entropy than that of SNS, while the incoming transfer entropy of SNS is generally higher than the others, i.e. N→S←B. Similarly, regarding the technology topics, we can find the directionality of information flow as N→S←B→N. On the other hand, political topics exhibits balanced interactions between different media types, compared with the other categories. These are consistent with outcomes from our previous study [1, 3].

Meanwhile, in order to examine the effect of the most recent k states of one social system on the future state of another social system, we experimented while varying $k(=l)$ from 1 to 5. Figure 2 shows the incoming and outgoing transfer entropy by media types with changing k. In this figure, the y-axis represents the number of significant cases whose transfer entropy is greater than that of 1,000 shuffled data. As the figure shows, in general the predictability of diffusion in one social system likely increases as the previous states of the other social system are known as much as possible. However, there are differences between media types. Regarding the outgoing transfer entropy, News and SNS likely influence other social systems for a longer period than Blog
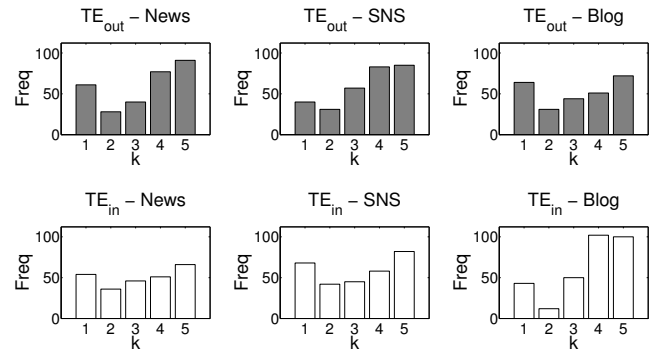

Figure 2: Transfer entropy for all news topics in our dataset by varying the length of k in Equation (1). Top and bottom rows are for outgoing and incoming transfer entropy, respectively. The y-axis indicates significant cases whose transfer entropy is greater than that of 1,000 shuffled data.

does. In terms of the incoming transfer entropy, Blog tends to be influenced by the longer trends of other social systems than the other media types, which corresponds to the fact that Blog users refer to relatively older documents compared to News and SNS. In particular, when it comes to $k = 1$ in Figure 2, the information flow of News is both ways, while that of SNS is incoming, and that of Blog is outgoing.

4. CONCLUSION

We proposed a way of understanding cross-population diffusion phenomena in a more principled way with a model-free approach using the transfer entropy to estimate the strength and directionality of information flow between online social systems at a macro level. This approach can be further used to give an insight into modeling diffusion across heterogeneous social networks.

5. REFERENCES

[1] M. Kim, D. Newth, and P. Christen. Modeling direct and indirect influence across heterogeneous social networks. In *ACM SIGKDD Workshop on Social Network Mining and Analysis*, Chicago, 2013. ACM.

[2] M. Kim, D. Newth, and P. Christen. Modeling dynamics of diffusion across heterogeneous social networks: News diffusion in social media. *Entropy*, 15(10):4215–4242, 2013.

[3] M. Kim, D. Newth, and P. Christen. Modeling dynamics of meta-populations with a probabilistic approach. In *CIKM*, San Francisco, 2013. ACM.

[4] M. Kim, L. Xie, and P. Christen. Event diffusion patterns in social media. In *ICWSM*, Dublin, 2012.

[5] M. Newman. *Networks: an introduction*. Oxford University Press, 2010.

[6] S. Panzeri, R. Senatore, M. A. Montemurro, and R. S. Petersen. Correcting for the sampling bias problem in spike train information measures. *Journal of Neurophysiology*, 98(3):1064–1072, 2007.

[7] T. Schreiber. Measuring information transfer. *Physical Review Letter*, 85(2):461–464, 2000.

[8] G. Ver Steeg and A. Galstyan. Information transfer in social media. In *WWW*, Lyon, 2012. ACM.