Trends of News Diffusion in Social Media based on Crowd Phenomena

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

Information spreads across social media, bringing heterogeneous social networks interconnected and diffusion patterns varied in different topics of information. Studying such cross-population diffusion in various context helps us understand trends of information diffusion in a more accurate and consistent way. In this study, we focus on real-world news diffusion across online social systems such as mainstream news (News), social networking sites (SNS), and blogs (Blog), and we analyze behavioral patterns of the systems in terms of activity, reactivity, and heterogeneity. We found that News is the most active, SNS is the most reactive, and Blog is the most persistent, which governs time-evolving heterogeneity of these systems. Finally, we interpret the discovered crowd phenomena from various angles using our previous model-free and model-driven approaches, showing that the strength and directionality of influence reflect the behavioral patterns of the systems in news diffusion.

Categories and Subject Descriptors

H.1.1 [**Systems and Information Theory**]: Information theory; H.3.4 [**Systems and Software**]: Information networks; J.4 [**Social and Behavioral Sciences**]: Sociology

Keywords

crowd phenomena; news diffusion; social media

1. INTRODUCTION

A wide range of information has become increasingly accessible and shared across heterogeneous online social networks, which enables these networks to be more interconnected than ever before [9, 10]. Such emergent phenomena have been investigated with a wide spectrum of diffusion spaces from a single social networking platform such as Twitter [13, 20, 25] to *n-Sphere*, consisting of two or more different kinds of social networks, such as blogosphere [16], blogosphere-to-news [5, 6, 15], blogosphere-to-YouTube [3], and blogosphere-SNS-news [8, 9, 10, 11, 12]. Accordingly, the underlying diffusion mechanisms have been studied by modeling diffusion processes at a micro (individual) or macro (population) level. Such models are all based on the assumptions of social structures from sampled snapshots of current social networks [3, 16], from site-specific actions (e.g., hashtag, mention, and retweet in Twitter) [20, 25], or from real-world network properties (e.g., a power-law degree distribution) [8, 9, 10, 17]. However, studies on diffusion have more focused on a single social platform alone than combined social media of different types.

To define real-world network structures is challenging in modeling diffusion, particularly for interconnected heterogeneous social networks. Regarding this issue, our previous studies [8, 10] modeled macro-level diffusion across different populations, requiring no detailed local structures. However, this model was based on the assumption of a power-law degree distribution and thus still needs to estimate the value of the scaling exponent for the power-law tail. Accordingly, our recent study [11] estimated macro-level information pathways with a model-free approach by using transfer entropy since it is independent of any specific assumptions on the interactions between individuals, providing benefits of studying dynamics of complex systems. This present study aims at analyzing crowd phenomena in news diffusion in terms of activity, reactivity, and heterogeneity, and then interpreting the discovered patterns with our model-free [11] and model-driven [8, 9, 10] approaches, which enables us to understand news diffusion trends in social media from diverse dimensions in a more consistent and systematic way.

We investigate the Spinn3r dataset from our previous study [10]. This dataset contains daily news adopters from different types of online social systems, such as mainstream news (News), social networking sites (SNS), and blogs (Blog), during a month in early 2011. Identified news topics are in common across the systems and cover eight representative categories of conventional news outlets (arts, culture, disasters, economy, politics, science, sports, and technology), which enables us to quantify and understand cross-population diffusion in a more principled way. With this dataset, we find that for real-world news topics, the most active 20% of SNS and Blog users generate 32% and 64% of documents, respectively, while the same percentage of News sites create over 90% of articles (the ratio of document percentage generated by each 20% is about 1 : 2 : 3 in that order). Also, the highly active top 1% of News sites are tightly connected to each other through hyperlinks in their main articles, which likely enhances the opportunity to be exposed to other social systems such as SNS and Blog. Regarding reactivity, SNS users are the most prompt to spread news even faster than News sites, while Blog users are the least reactive but the

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2579325.

most persistent in spreading old news. This result is consistent with the findings that there are more significant cases of incoming transfer entropy (TE) in SNS ($TE_{any \to SNS}$) and News ($TE_{any \to News}$) than Blog ($TE_{any \to Blog}$), and that the longer history of news adoption in News and SNS has more effects on diffusion in Blog. About heterogeneity, higher or lower heterogeneity of populations is likely attributed to balanced or unbalanced influence among social systems. Such macro-level cross-population diffusion from diverse angles, to the best of our knowledge, are seen for the first time.

2. RELATED WORK

Identifying trending topics is important since diffusion patterns have significant variations by the context of information [3, 8, 9, 10, 11, 20, 25]. There have been attempts to identify specific events such as Twitter-only trends [2, 13] or approximated events by text clustering [7]. However, these recognized trending topics can be site-specific. For instance, the authors of [13] found that only 3.6% of Twitter's trending topics exist in hot search keywords from Google. This paper studies cross-population news diffusion, and thus commonly interested topics are important to truly understand diffusion trends across different social systems. In this context, we target news topics from the Wikipedia Current Events [1] as a noteworthy news registry.

Extensive research on information-theoretic measures has been conducted in diverse areas including computer science [4, 27, 28, 29], neuroscience [21, 22, 24], and economics [14, 18, 19]. When it comes to social media, the predictability of user behavioral patterns with a model-free approach has been common interests for predicting future interactions at individual and group levels [29], for estimating directed information pathways between Twitter users without follower-followee relationships [27, 28], and for classifying Twitter user behaviors [4]. These studies are all from the aspects of individuals or groups belonging to a single social platform alone. However, heterogeneous social networks are interconnected across a variety of social platforms (e.g., CNN, BBC, Facebook, Twitter, WordPress, and Tumblr) and they can be aggregated into different types of online social systems such as News, SNS, and Blog. In this context, our recent study [11] estimated macro-level information transfer using transfer entropy, and we found that the strength and directionality of information flow among these systems vary in different context of information. Based on outcomes mainly from this model-free approach [11] as well as from our previous model-driven approaches [8, 9, 10], this present study attempts to interpret crowd phenomena in real-world news diffusion from various angles.

3. CROWD PHENOMENA

In this section, we analyze behavioral patterns of different online social systems (News, SNS, and Blog) in terms of activity, reactivity, and time-evolving heterogeneity.

3.1 Data Description

We use the Spinn3r dataset from our previous study [10]. This dataset consists of daily cumulative adopters in each News, SNS and Blog systems for 63 real-world news topics, which have driven the largest diffusion (at least 50 users in each system for robust statistics), during a month in early 2011. We use document hyperlinks in the main text to generate accurate citation networks, and accordingly adopters are ones who explicitly cite to original documents through hyperlinks. We referred to the Wikipedia Current Events [1] to identify non site-specific but commonly interested news topics across these systems. Also, for identifying users who subscribe to the SNS or Blog system, we targeted 10 popular social platforms (5 for each) as they provided regular patterns of user identity in their web document addresses. Note that identifying borderline users, who have more than one account in different social platforms, is beyond the scope of this study. Regarding main stream news, we chose 9K News sites as they constitute the largest strongly connected news network through hyperlinks in their articles, and each News site is considered as one super user (see [10, 12] for details). For the rest of the paper, we use terms *user* or *individual* for indicating a principal agent in each system.

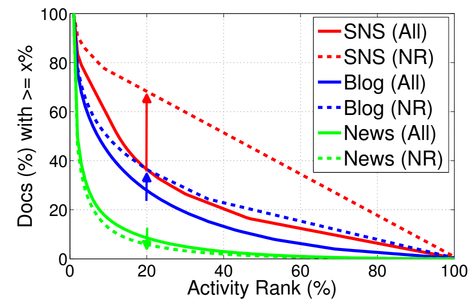


Figure 1: Activity levels of online social systems (News, SNS, and Blog) in news diffusion. Solid lines (*All*) indicate all documents regardless of topics, while dashed lines (*NR*: news-related) represent documents on identified news topics from the Wikipedia Current Events [1]. Arrows exhibit changing directions of activity levels from *All* to *NR* documents.

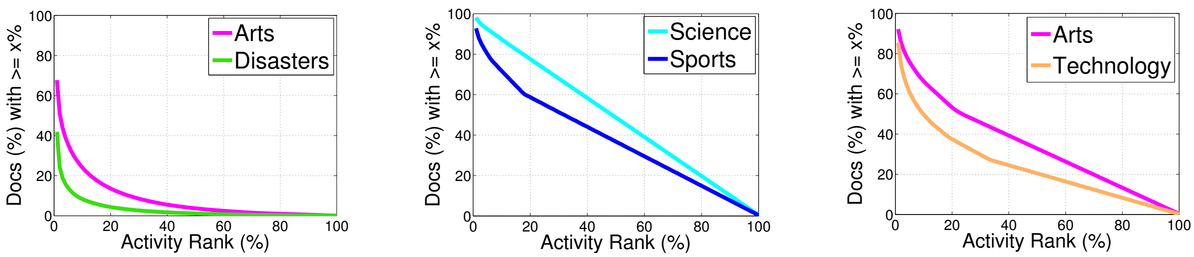
3.2 Activity

We define *activity* as the quantity of publication produced by an individual, i.e. frequency of an individual's web posting. We investigate the level of activity for each social system and analyze how it varies in different news categories.

We first examine the Pareto principle (the 80-20 rule) [23] in news diffusion, as shown in Figure 1. We rank users for each social system according to their activity levels and normalize the ranks to lie between 0 and 100. Thus, the x-axis indicates the percentage of users from the most active to the least active, while the y-axis represents the complementary cumulative percentage of documents created by all but x% of users (100-x%). For instance, the blue solid line shows that the most active 20% of Blog users published about 72% of all blog documents, which means that the Pareto principle in general applies to the activity of blogging arbitrary topics. This is consistent with the result from [3] in which the most active 20% of blogs account for 70% of all blog posts in 15 blog domains. However, when it comes to real-world news topics (blue dashed line), the same percentage of Blog users produced only 64% of news-related documents, decreased by 8% (from 72% to 64%). That is, the Pareto principle is not well applicable to the activity of blogging real-world news topics. Meanwhile, the top 20% of SNS users generate 64% of all micro-blogs (red solid line), while the most active

Table 1: The most cited top 5 news sites across online social systems (News, SNS, and Blog).

Rank	Cited by News Sites		Cited by SNS Users		Cited by Blog Users	
	News sites	#Sites	News sites	#Users	News sites	#Users
1	guardian.co.uk	1,288	huffingtonpost.com	3,719	guardian.co.uk	2,988
2	huffingtonpost.com	1,119	bbc.co.uk	3,404	bbc.co.uk	2,432
3	bbc.co.uk	987	yahoo.com	3,071	huffingtonpost.com	2,164
4	nytimes.com	978	msn.com	2,289	telegraph.co.uk	1,762
5	telegraph.co.uk	924	guardian.co.uk	2,254	nytimes.com	1,555



(a) News (b) SNS (c) Blog

Figure 2: Activity levels of online social systems (News, SNS, and Blog) for eight different news categories: Arts, Culture, Disasters, Economy, Politics, Science, Sports, and Technology. In each graph, only two distribution plots, positioning at the top and bottom, are shown for clarity.

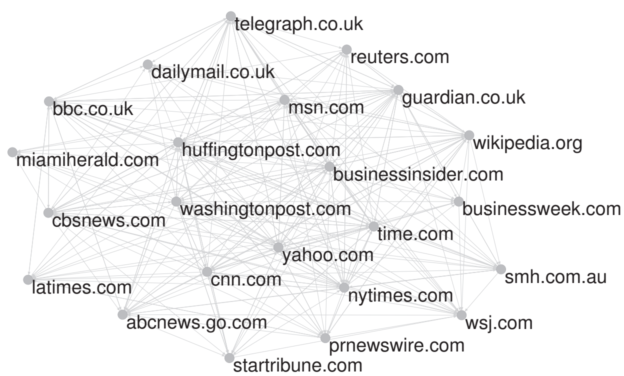


Figure 3: News media network. Every node in the network belongs to the most active top 1% of all news sites, and it is cited by at least nine other news sites in the network, i.e. 9-Core (in-degree).

20% of News sites create over 90% of all articles (green solid line). Regarding real-world news topics, the activity level decreases again by 32% for SNS users (red dashed line and upward arrow), but it increases by 4% for News sites (green dashed line and downward arrow). Consequently, the ratio of *NR* document percentages produced by the most active 20% of individuals from each SNS, Blog, and News systems is about 1 : 2 : 3 (32% : 64% : 90%). This means activity depends on online social systems, and thus the Pareto principle is not applied uniformly across all the systems.

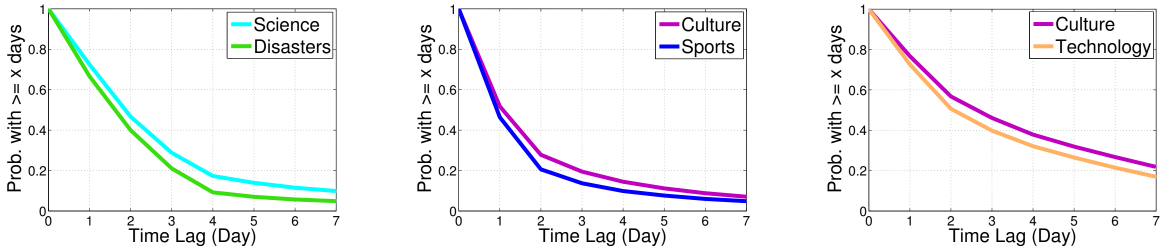
In addition, different news categories lead to different levels of activity across these systems, as shown in Figure 2. The activity level of active News sites is the highest for the disasters category and the lowest for the arts category. In SNS, the activity level increases the most when posting sports topics and the least for science topics whose linear plot exhibits that active and non-active users produce almost the same number of posts on science news. On the other hand, Blog users' activity level grows the most when blogging technology topics and the least for arts topics.

The News system shows the most distinct activity level among these systems. In more detail, the most active 5% of news sites generate over 80% of all documents during a month. Among the top 1%, we found a very tightly connected news media network where every news site receives citations from at least nine other news sites in the network, as shown in Figure 3. These closely connected news sites also include the most cited top 5 news sites across all online social systems in Table 1. As the table shows, popular news sites are very similar across the systems. The most cited top 10 news sites belong to 25-Core (in-degree) and 74-Core (degree), which means that every news site in a connected network receives citations from at least 25 other news sites, and also interacts with 74 other news sites. It can be interpreted that such strong connectivity of news media enables them to be more frequently exposed and connected to other online social systems compared to isolated news media.

3.3 Reactivity

We define *reactivity* as an individual's response time to new information since it was created. Accordingly, we quantify the degree of reactivity with time difference of posting between a reference and a citing document through creating an explicit hyperlink in the main text. We also examine the level of reactivity for each social system and analyze how it varies in different news categories, as in the case of activity.

Figure 4 shows the cumulative distribution function (CDF) of time lags during 30 days corresponding to our dataset period. In order to reduce the time zone effects, we use one day resolution. Zero on the x-axis indicates that a reference document is cited on the day of its publication. As the figure shows, in general, SNS users are the most responsive to exposed information even faster than News sites, while Blog users are the least reactive. It is often reported that some



(a) News (b) SNS (c) Blog

Figure 5: Reactivity levels of online social systems (News, SNS, and Blog) for eight different news categories: Arts, Culture, Disasters, Economy, Politics, Science, Sports, and Technology. In each graph, only two distribution plots, positioning at the top and bottom, are shown for clarity.

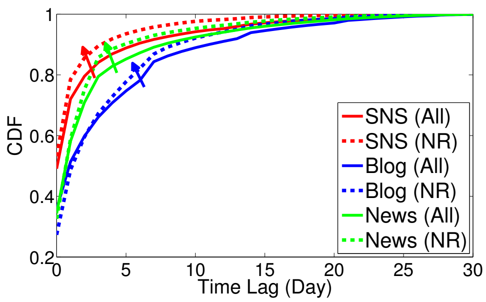


Figure 4: Reactivity levels of online social systems (News, SNS, and Blog). ***All*** **and** ***NR*** **denote the same as in Figure 1. The same notation is used for** ***All*** **and** ***NR*** **as in Figure 1.**

breaking news first spreads through SNS, and thus this distribution plot can be one of supportive clues showing prompt behavior of SNS users. These behavioral patterns can be also found in other studies in a separated way: (1) News vs Blog and (2) News vs SNS. For instance, news sites have a tendency to spread news much faster than blogs [5], and some news spreads on Twitter before the announcement of main stream news agencies such as CNN [13]. In Figure 4, the cumulative diffusion rates of *NR* documents (dashed lines) increase across all media types (upward arrows), compared with *All* documents (solid lines). That is, users are more reactive to real-world news than other arbitrary topics. On the day of publication of *NR* documents, over 50% of citations to the documents are found in SNS, while 30% in both Blog and News. Also, within five days at least 70% of all citations are generated across the systems for all different news categories. That is, hyperlinks are mostly created within a week since an original document is published. Thus, one week is a meaningful period for tracking news cascades regardless of types of online social systems and topics of news. Note that such reaction is one of ways to spread information in social media, not diffusion itself. In terms of persistence of adoption or citations to old documents, Blog users show the most lasting spreading behavior in contrast to SNS users.

Reactivity also exhibits variations in different news categories. To be more specific, Figure 5 shows the upper and lower bound time lag distributions of different news categories for each system. As the figure shows, News sites are the most and the least reactive to the disasters and science categories, respectively. In common, SNS and Blog users are the least responsive to the culture category, but they are most prompt to the sports and technology categories, respectively. The science and culture categories are not time-sensitive topics compared with other topics such as natural disasters, sports games, and new product releases. Thus, they tend to drive less responsive but more persistent behavior than the other topics. Regarding persistence, SNS users exhibit the least persistent behavior of referring to old news since it is likely forgotten only after two days. On the other hand, Blog users still have a more than even chance to spread news after two days, while the reactivity level of News sites lies between SNS and Blog.

In terms of activity and reactivity, SNS and Blog users tend to be less active but more reactive in posting news content. Also, the most active news category in each system corresponds to the most reactive news category.

3.4 Heterogeneity

We examined that noteworthy real-world news spreads across online social systems, not limited to a specific media type, leading to common and distinctive behavioral patterns among the systems. Then, how heterogeneous are the participants in news diffusion? Does the degree of heterogeneity change as time evolves? If it changes, how different is it for various news categories? In order to answer these questions, we use *relative entropy* for measuring the dispersion of heterogeneous populations compared with their uniform distribution as conducted in our previous study with a different dataset [8]. We quantify the time-evolving heterogeneity of populations participating in information diffusion as:

$$J(t) = \frac{H(t)}{H_{max}} = \frac{1}{H_{max}} \left(-\sum_{i=1}^{m} p_i(t) \log_2(p_i(t)) \right), \qquad (1)$$

where the entropy $H(t)$ is a measure for quantifying the dispersion of different types of populations at time t, and $p_i(t)$ is the proportion of cumulative population of type i among m types at time t. When the proportions are uniformly distributed, the heterogeneity (relative entropy) is the highest (H_{max}=$log_2 m$), while it is the lowest (H_{min}=0) when only one population exists.

In our case, we consider SNS and Blog users as heterogeneous populations participating in news diffusion. As Figure 6 shows, the heterogeneity (J) generally increases as time evolves, showing that populations tend to be more uniformly distributed, and thus it is more uncertain which population will be dominant in the diffusion. From the beginning, the heterogeneity rapidly increases due to small number of users

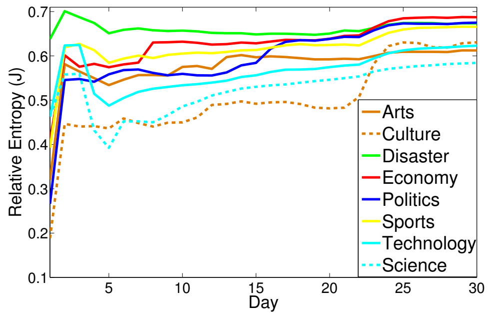


Figure 6: Time-evolving heterogeneity (relative entropy) of populations from the SNS and Blog systems participating in news diffusion during a one-month period. News categories are color-coded.

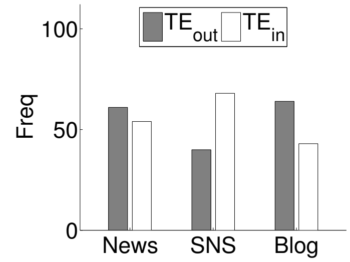


Figure 7: Number of significant cases of transfer entropy of online social systems (News, SNS, and Blog) when k=l=1 in Equation (2). TE_{out}/TE_{in} denotes outgoing/incoming transfer entropy.

and fast responses of early adopters in each system. However, the heterogeneity decreases soon due to different levels of reactivity between SNS and Blog users. That is, SNS users tend to response to new information much earlier than Blog users, which may bring in larger number of SNS users than Blog and consequently lead to the disproportionate distribution of the populations for one to three days.

The level of heterogeneity varies in different news categories. The disasters, economy, and politics categories lead to higher heterogeneity (more even distribution of populations), while the arts, culture, science, and technology categories lead to lower heterogeneity (relatively disproportionate distribution). Also, higher heterogeneity comes up with larger number of adopters in diffusion, which reconfirms that large diffusion is attributed to collective behavior of heterogeneous populations, not limited to a single social platform alone [8, 9, 10]. Thus, analyzing behavioral patterns in cross-population diffusion would be a step towards a more accurate and consistent understanding of real-world diffusion.

4. INFORMATION PATHWAYS BETWEEN ONLINE SOCIAL SYSTEMS

In this section, we analyze the discovered crowd phenomena in news diffusion from various angles using our previous model-free [11] and model-driven [8, 9, 10] approaches so that we can understand cross-population diffusion in a more consistent and principled way.

4.1 Macro-Level Information Transfer

From our recent study [11], we estimated macro-level transfer entropy by defining an online social system as a stochastic process, which is a collection of time-series adopting trends (e.g., a chronological sequence of states such as *decrease*, *transition*, and *increase*), influencing other social systems.

When considering two discrete-time random processes, X and Y, the *transfer entropy* (TE) from the process Y to X is defined as the conditional mutual information [26]:

$$\begin{aligned} TE_{Y\to X} &= I(X_{t+1};Y_t|X_t) \\ &= \sum p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log \frac{p(x_{t+1}|x_t^{(k)}, y_t^{(l)})}{p(x_{t+1}|x_t^{(k)})}, \quad (2) \end{aligned}$$

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 (2) 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.

4.2 Reflections of Crowd Phenomena

Figure 7 shows significant cases of transfer entropy with the Markov order set to be one day (k=l=1). The distribution of significant cases of incoming transfer entropy (TE_{in}) reflects the reactivity levels of online social systems. For instance, SNS and News have more significant cases of incoming transfer entropy than Blog in that order, which means that they are more influenced by the one-day recent past of the other social systems than Blog. This result is consistent with the findings in the previous section that SNS and News are more reactive than Blog in that order, and over 60% of news citations in SNS and News are generated only after one day during a one-month period of observation.

In terms of the directionality of information flow, SNS is more incoming, but Blog is more outgoing, while News exhibits both ways in Figure 7. This can be interpreted that News introduces new information as well as spreads exposed information in a balanced way (diligent creator and diligent adopter), while SNS rather propagates than introduces information (lazy creator but diligent adopter), and Blog has a tendency the other way around (diligent creator but lazy adopter). Thus, even the same level of activity can have very different effects on diffusion, depending on a document's propensity to introduce or spread information.

Regarding the Markov order (the length of k and l), longer activity sequences of online social systems tend to influence other social systems more, from the fact that the number of significant cases are largest at k=l=5 when varying k (=l) from 1 to 5 [11]. In particular, when comparing two cases between k=l=1 and k=l=5, Blog shows the biggest change in the ratio of significant cases, unlike the ignorable changes for News and SNS. That is, longer histories of adoption trends in News and SNS have a stronger effect on diffusion in Blog than the others. Concerning diffusion rate, we estimated cumulative adopter proportions for each system from our previous model-driven approach [9, 10], showing more rapid spreading trends in SNS and News than Blog during one-month diffusion. These are consistent with the prompt behavior of SNS and News and the persistent behavior of Blog.

From our previous model-free [11] and model-driven approaches [8], the strength and directionality of influence between systems vary in different news categories. In [11], the culture and technology categories show an unbalanced distribution of incoming and outgoing transfer entropy, while the politics category exhibits relatively balanced distribution. In [8], the distribution of estimated influence between systems is more balanced for the politics and disasters categories and less balanced for the art and sports categories. These are consistent with the heterogeneity in Figure 6, showing higher relative entropy for the economy, disasters, and politics categories and lower relative entropy for the arts, culture, science, sports, and technology categories after a month. These can all be interpreted that balanced interactions between populations more likely lead to higher heterogeneity (more evenly distributed populations).

5. CONCLUSION

We analyzed crowd phenomena across online social systems (News, SNS, and Blog) in terms of activity, reactivity, and heterogeneity, and interpreted the discovered behavioral patterns of the systems from various angles based on our previous model-free and model-driven approaches. Main trends of news diffusion in social media are summarized as below.

- News is the most active, SNS is the most reactive, and Blog is the most persistent.
- SNS and Blog users are less active but more reactive for real-world news than other arbitrary topics.
- Regarding activity, the Pareto principle is not applied uniformly across different online social systems.
- Active news media are tightly connected, enhancing the opportunity to be exposed to other social systems.
- One week is a meaningful period for tracking news cascades regardless of system types and news topics.
- The most active news category in each system corresponds to the most reactive news category.
- Larger diffusion exhibits higher heterogeneity.
- News is a diligent creator and diligent adopter, SNS is a lazy creator and diligent adopter, and Blog is a diligent creator and lazy adopter.

6. REFERENCES

[1] Wikipedia Current Events in January, 2011. `http://en.wikipedia.org/wiki/January_2011`.

[2] H. Becker, M. Naaman, and L. Gravano. Beyond trending topics: Real-world event identification on twitter. In *ICWSM*, Barcelona, Spain, 2011.

[3] M. Cha, J. Pérez, and H. Haddadi. Flash floods and ripples: The spread of media content through the blogosphere. In *ICWSM*, San Jose, USA, 2009.

[4] R. Ghosh, T. Surachawala, and K. Lerman. Entropy-based classification of'retweeting'activity on twitter. *arXiv preprint arXiv:1106.0346*, 2011.

[5] M. Gomez Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *KDD*, pages 1019–1028, Washington, D.C., 2010. ACM.

[6] M. Gomez Rodriguez, J. Leskovec, and B. Schölkopf. Structure and dynamics of information pathways in online media. In *WSDM*, Rome, Italy, 2013. ACM.

[7] V. Ha-Thuc, Y. Mejova, C. Harris, and P. Srinivasan. Event intensity tracking in weblog collections. In *ICWSM*, San Jose, USA, 2009.

[8] M. Kim, D. Newth, and P. Christen. Modeling direct and indirect influence across heterogeneous social networks. In *SIGKDD SNA Workshop*. ACM, 2013.

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

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

[11] M. Kim, D. Newth, and P. Christen. Macro-level information transfer across social networks. In *WWW Companion*, Seoul, Korea, 2014.

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

[13] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW*, Raleigh, USA, 2010.

[14] O. Kwon and J.-S. Yang. Information flow between composite stock index and individual stocks. *Physica A: Stat. Mech. Appl.*, 387(12):2851–2856, 2008.

[15] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *KDD*, Paris, France, 2009. ACM.

[16] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst. Cascading behavior in large blog graphs. In *SDM*, Minneapolis, USA, 2007.

[17] M. Luu, E. Lim, T. Hoang, and F. Chua. Modeling diffusion in social networks using network properties. In *ICWSM*, Dublin, Ireland, 2012.

[18] E. Maasoumi and J. Racine. Entropy and predictability of stock market returns. *Journal of Econometrics*, 107(1):291–312, 2002.

[19] R. Marschinski and H. Kantz. Analysing the information flow between financial time series. *EPJ B-Cond. Matter & Complex Sys.*, 30(2):275–281, 2002.

[20] S. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *KDD*, Beijing, China, 2012. ACM.

[21] L. Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6), 2003.

[22] 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.

[23] V. Pareto. *Cours d'Economie Politique*. F. Rouge, Droz, Geneva, 1896.

[24] R. Q. Quiroga and S. Panzeri. Extracting information from neuronal populations. *Nature Reviews Neuroscience*, 10(3):173–185, 2009.

[25] D. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics. In *WWW*, Hyderabad, India, 2011. ACM.

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

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

[28] G. Ver Steeg and A. Galstyan. Information-theoretic measures of influence based on content dynamics. In *WSDM*, Rome, Italy, 2013. ACM.

[29] C. Wang and B. A. Huberman. How random are online social interactions? *Scientific Reports*, 2, 2012.