Concept of Keystone Species in Web Systems: Identifying Small Yet Influential Online Bulletin Board Threads

Shota Ejima
Graduate School of Systems and Information Engineering, University of Tsukuba, Japan
eji@websci.cs.tsukuba.ac.jp

Mizuki Oka
Graduate School of Systems and Information Engineering, University of Tsukuba, Japan
ALTERNATIVE MACHINE Inc., Japan
mizuki@cs.tsukuba.ac.jp

Takashi Ikegami
Graduate School of Arts and Sciences, The University of Tokyo, Japan
ikeg@sacral.c.u-tokyo.ac.jp

ABSTRACT
Research is being conducted to understand social and innovative behavior in human interactions on the Web as a biological ecosystem. Keystone species in a biological ecosystem are defined as a set of species that significantly impacts the ecosystem if they are removed from the system, irrespective of its small biomass. Identifying keystone species is an important problem as they play an important role in maintaining diversity and stability in ecosystems. A human community in the web system also possesses keystone species. They can be influential users or contents on the web systems, even though their commitments to the web are relatively small. We use data from an online bulletin board, and identify keystone threads (= "species") that have a large impact if they are removed or become unpopular, despite their small population size. Here, the removal of threads can be regarded as a state in which there is no attention or actions by users on the thread. The multivariate Hawkes process is used to measure the degree of influence among all threads and calculate the overall activity level on the online bulletin board. Our analysis confirms that keystone threads do exist in the system. Apparently, the number of keystone species increases along with the service maturation. The keystone concept in online services proposed in this study gives a new viewpoint for their stable operation.

CCS CONCEPTS
• **Human-centered computing → Web-based interaction**; *Social network analysis*; • **Information systems →** *Web log analysis.*

KEYWORDS
online discussion threads, Hawkes process, keystone species

ACM Reference Format:
Shota Ejima, Mizuki Oka, and Takashi Ikegami. 2019. Concept of Keystone Species in Web Systems: Identifying Small Yet Influential Online Bulletin Board Threads. In *11th ACM Conference on Web Science (WebSci '19), June 30-July 3, 2019, Boston, MA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3292522.3326023

WebSci '19, June 30-July 3, 2019, Boston, MA, USA

ACM ISBN 978-1-4503-6202-3/19/06…$15.00
https://doi.org/10.1145/3292522.3326023

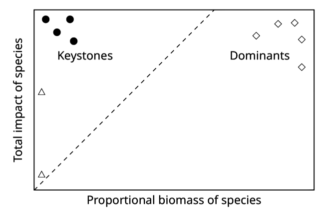

Figure 1: View of keystone species drawn with inference to [19]. Plots on the upper left are keystone species whose proportional biomass (population) is small but their impact is large, and plots in the upper right are dominant species whose proportional biomass (population) and impact are both large. The dashed line represents $y = x$, indicating a proportional relationship between population and impact.

1 INTRODUCTION
In recent years, various online communities, such as online bulletin boards, chat tools, and social network services, have been widely used. People's communication in an online community creates enormous log data on a daily basis and their interactions create new ideas that spread through social networks. Richard Dawkins called the spread of influence and information as memes analogous to genes [2]. This "meme evolution theory" has recently regained attention as large amounts of traces of individual actions and remarks are constantly recorded and accumulated on the Web, and they can be analyzed quantitatively [12, 20]. Furthermore, such an evolutionary biological perspective and its mathematical framework have been proposed for social media analysis [1, 13]. These studies suggest that an analogy can be borrowed from biological systems, and if correctly applied to social systems, it can open new perspectives on analyzing social systems.

In biological ecosystems, the concept of *dominant* and *keystone* species exists to define influential species [19]. The dominant species are defined as species having both high biomass and large impact on the ecosystem. On the other hand, keystone species are the ones having large impact on the ecosystem despite their low biomass proportional to the ecosystem. Figure 1 depicts the relationship of dominant and keystone species with the abundance of species and their impacts on ecosystems. It has been reported that, if keystone species are removed from the ecosystem, the balance of the ecosystem collapses, causing extinction of many other species [17]. Many experiments on biological ecosystems have demonstrated that less abundant species also have strong effects on communities and ecosystems [3, 18], and finding and protecting those species is important from the perspective of the system's stability. In contrast to dominant species, keystone species are considered difficult to identify, as their population is small.

In online social-communication systems, identifying keystone species in ecosystems can be regarded as identifying influential users or contents, which have been extensively studies in the literature. These studies include identifying influential users based on follow-follower networks [10] or diffusion models of the spread of influence and information through social networks [9, 21]. The influential users or contents obtained from these studies can be regarded as identified *dominants* species in the biological ecosystem, and to the best of our knowledge, few studies have discussed keystone species in the context of online communities. Therefore, in this study, we introduce the concept of keystone species in an online communication system and propose a method to identify them. In contrast to the standard biological definition of keystone species, we extend it to more dynamic and network-based ones [8]. That is, we treat the keystone species as time-varying and context-dependent (i.e., the combinations of other species) processes.

Specifically, we apply the concept of keystone species to online bulletin boards by considering a thread as a species. On the online bulletin board, users create threads related to specific topics and themes to communicate. Threads can be said to influence each other because users move back and forth among threads. In other words, threads are in a competitive relationship to get a limited number of user resources. Such an environment can be considered similar to the competitive relationship between species in biological ecosystems. If keystone threads are identified, online bulletin board service providers can perform necessary interventions, such as protection of keystone threads, for stable operation.

We model each thread as an event sequence of all the actions performed by users (i.e., views, comments, replies, and claps) and estimate the influence among threads by using the multivariate Hawkes process [7], a self-exciting stochastic point process. The Hawkes process was originally used to analyze earthquakes [15] and applied to many other fields such as financial market modeling [5]. In recent years, it has been used for influential user identification [23] and recommendations [22], user interest modeling [11], and steering user activity by controlling dynamics [4] on social media.

For identifying keystone species, it is necessary to measure the degree of impact when the species is removed from the system. Gupta et al. proposed a method for characterizing the effect of removing selected events in the history of the Hawkes process [6]. We also employ a similar method in this study to simulate the degree of impact caused by thread deletion. Specifically, we use the fitted parameters of the multivariate Hawkes process and simulate the model to compute the degree of impact caused by deleting each thread from the system. This is done by computing the *activity level* of the entire bulletin board [16] and comparing the activity levels from before and after the removal and regard that as an *impact* of the removal of each thread. If a thread has a high degree of impact despite a small density (i.e., the number of users participating in the thread), then we identify the thread as a keystone thread.

The contributions of this study are summarized below.

- We introduce the concept of keystone species of the biological ecosystems in a web system to identify influential contents.
- We propose a method to identify keystone species in a web system via the multivariate Hawkes process and measure the difference of impact before and after the removal of a species (i.e., content).
- We apply our method to online bulletin board data and show the existence of keystone species on the Web, as well as the results of our analysis on the found keystone species.

2 DEFINITION OF KEYSTONE SPECIES AND HAWKES PROCESS

2.1 Keystone species in biological ecosystems

The keystone species has a large impact on the ecosystem despite its proportionally small biomass in the ecosystem. Starfish have been reported as a keystone species. When starfish, which is the apex of a food chain in a intertidal zone, was removed, the mussel population, on which starfish was feeding, increased abnormally and the balance of the ecosystem was destroyed [17]. Consequently, many other species living in the intertidal zone have become extinct. That is, keystone species play an important role in maintaining stability in ecosystems.

For the definition of keystone species, we use the population density p_i occupied by species i in the ecosystem and the *impact* I_i when i is removed from the ecosystem. The *impact* is defined as how a quantitative value representing the nature of the ecosystem, such as the population of all species in the ecosystem, will change before and after removing the species i. Let t_N be the quantitative value representing the nature of the ecosystem before removing species i, and t_D be the value after removal. The *impact* I_i of removal of the species i is shown below[19].

$$I_i = \frac{t_N - t_D}{t_N} \tag{1}$$

Intuitively, if you draw a scatter plot with population density p_i plotted on the horizontal axis and the impact I_i plotted on the vertical axis, the points on the upper left of the figure are keystone species (see Figure 1). Further, plots in the upper right in the graph, that is, the species whose impact and proportional population density are both large, are called dominant species.

In this study, let p_{max} and I_{max} be the maximum population density and maximum impact of all the target species, respectively. The species i that satisfies the following two conditions is defined as a keystone species [8]:

$$p_i \leq c_p \times p_{max} \tag{2}$$

$$c_I \times I_{max} \leq I_i \tag{3}$$

where c_p and c_I are the threshold coefficients, which are used to compute thresholds for population density and impact, respectively. Note that the threshold values $c_p \times p_{max}$ and $c_I \times I_{max}$ become different depending on the values of p_{max} and I_{max}, respectively. When the computed p_i is smaller than the threshold of population density and I_i is higher than the threshold of impact, the species i

is identified as a keystone species. We also define CI_i as the community importance of the species i, which indicates the keystonness of the species [14] . The higher this value, the higher is the keystonness of the species.

$$CI_i = \frac{I_i}{p_i} = \frac{t_N - t_D}{t_N}\frac{1}{p_i} \tag{4}$$

2.2 Multivariate Hawkes Process

Hawkes process[7] is a self-exciting stochastic point process, and is a probabilistic model that takes into account the influence of past event occurrences on future event occurrences. We consider *threads* in an online bulletin board as species and estimate the degree of influences between threads by modeling event sequences via the multivariate Hawkes process. Then, we calculate the *impact* on the online bulletin board when each thread is deleted by using the estimated degree of influence. By using this impact value and the density defined for each thread, a thread that has a high overall impact when deleted, despite its small density, is identified as a keystone. Event sequences of a thread are composed of user actions such as accesses, comments, and claps.

The intensity function $\lambda_i(t)$ of the event sequences at time t of thread i is represented as follows:

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{D} \sum_{t_k^j < t} \phi(t - t_k^j) \tag{5}$$

where D is the number of threads, μ_i is the base intensity of thread i, and t_k^j represents the k-th event sequence of thread j. ϕ is called the kernel function, and we use the exponential kernel function $\phi(t) = \alpha_{ij} e^{-\beta_{ij} t}$. α_{ij} represents the degree of the influence of thread j on thread i, and β_{ij} represents the decay rate of the intensity of thread i raised by the event sequences in thread j. These parameters can be obtained analytically by the maximum likelihood method, given an event sequence [15].

We define threads as nodes, set the estimated parameter α_{ij} as the edge representing the degree of influence from thread i to j, and construct a network. We consider this thread network as an ecosystem and identify the keystone thread that plays an important role in maintaining the stability of the network, that is, for stable communication within the online bulletin board.

2.3 Impact of removal of a thread

To calculate keystone threads, we need to calculate the impact I_i on the ecosystem, viz., the thread network, when thread i is removed. Here, we use cascading condition C, which represents the activity level of network [16]. Let $\boldsymbol{\alpha} = \{\alpha_{ij}\}_{i,j=1,\cdots,D}, \boldsymbol{\mu} = \{\mu_i\}_{i=1,\cdots,D}$ be the parameters estimated via the multivariate Hawkes process. Then C is defined as follows:

$$C \equiv \frac{\sum_{i,j}(L\Lambda L^T)_{ij}}{\sum_i \langle\lambda_i\rangle}, \tag{6}$$

where $L = \sum_{n=0}^{\infty} \boldsymbol{\alpha} = (I - \boldsymbol{\alpha})^{-1}, \Lambda = \mathrm{diag}(\langle\boldsymbol{\lambda}\rangle) = \mathrm{diag}(\{\langle\lambda_i\rangle\}_{i=1,\cdots,D})$ $= \mathrm{diag}(L\boldsymbol{\mu})$. The value of C becomes higher as the event sequence of each node (threads) of the network becomes more bursty. We define this burstiness to represent the activity level of the network.

Table 1: Descriptive statistics of online bulletin board

	pregnant women	mothers
periods	03/2017 ∼ 09/2018	
# of unique users	10,687	10,945
# of unique threads	425	399
# of accesses	112,868	105,821
# of comments	22,777	23,978
# claps	136,211	97,820
average # of actions per thread	640	570

The change in activity level of the network ΔC_i when deleting a certain thread i in the network is defined as follows:

$$\Delta C_i = \sum_{j \neq i}(C - C'_j), \tag{7}$$

where C'_j is the cascading condition calculated by setting the influence between threads i and $j(\neq i)$ (i.e., α_{ij}, α_{ji}) to 0. Performing the same calculation for all $j \neq i$, and calculating the differences with the original C and all the C'_j, we obtain ΔC_i by adding the differences. This ΔC_i is equivalent to $t_N - t_D$ in the equation (1). We can thus define the impact I_i by removing the thread i from the network as follows:

$$I_i = \frac{t_N - t_D}{t_N} = \frac{\Delta C_i}{C} \tag{8}$$

3 EXPERIMENTS

3.1 Data

The data used by us was provided by QON Inc., which is an online community service in Japan [1]. More than 100 companies use this service as an online bulletin board for communication between a client company and consumers, and the total number of users exceeds 1 million. Each company is provided a community in which the company can run its own bulletin board. We analyzed two communities among them[2]. One targeted women before or during pregnancy, and the other targeted mothers who were raising children. Table 1 shows the statistics of the data of the two communities of the online bulletin board. On the online bulletin board, users can communicate with other users by posting threads and replying to the threads by commenting and clicking claps (corresponds to likes) for threads and/or comments.

We analyzed the event sequences composed of user actions, such as views, comments, and claps, on threads. The event sequences are divided into two-week windows (30 windows were obtained for the period between March 2017 and September 2018), and the parameters are fitted using the multivariate Hawkes process for each window. We disregard threads having less than 50 actions in the event sequences, as fitting cannot be performed if the number of actions is too small. Accordingly, 218 unique threads out of 435 threads are targeted for our analysis (51.3 % of the total number of unique threads). The number of overlapping target threads for analysis in all windows is 449. For the other community, the number of unique target threads is 213 out of 399 threads (53.3%). The number of overlapping target threads in all windows is 427.

[1] https://www.q-o-n.com

[2] The data we used in this study is publicly available at https://github.com/AlternativeMachine/keystone

There are two types of threads, viz., threads created by community service administrators (called thread by administrators) and by general users (called thread by users). In the thread by administrators, because questionnaires and gift campaigns for users are often conducted, many users gather, and therefore the number of actions is larger compared to thread by users. On the other hand, thread by users is mainly used for chatting and consultation about daily life. The number of users who participated per thread as well as the number of actions are smaller compared to thread by administrators.

For the population density, which we need to identify for calculating keystone threads, we use the number of users participating in the thread. Let N be the number of target threads and u_i be the number of unique users who have taken action at least once in thread i in a window. Then the population density p_i of thread i can be calculated as follows:

$$p_i = \frac{u_i}{\sum_{j=1}^{N} u_j}. \tag{9}$$

The threshold coefficients c_p and c_I are set to 0.25 and 0.5, respectively, with reference to [8]. That is, the thread i whose p_i is less than one-fourth of the maximum population thread c_{max} and I_i is more than half of the maximum impact thread I_{max} in a window, is identified as a keystone thread. It is noteworthy that thresholds are computed in relation to the maximum values for population and impacts, respectively; the threshold values become different for each window. Figure 2-left shows the temporal development of keystone thresholds for p_i and I_i computed for each window and the number of threads for each window in the bar graph. We can see that the number of threads maximizes at window ID = 15 and the threshold values correlate roughly with the number of threads for each window. Figure 2-right shows the number of windows selected as keystones for each thread sorted in descending order. One of the threads is selected the maximum number of times, viz., six times, over the period.

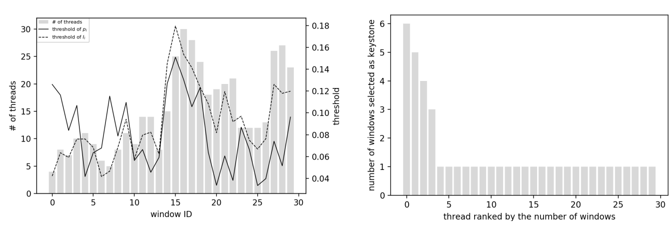

Figure 2: (left) Temporal development of keystone thresholds for p_i and I_i computed for each window and the total number of threads in each window denoted by the bar graph. (right) The number of windows selected as keystones for identified keystone threads.

4 RESULTS

Because of the limitation of space in the paper, we mainly show the results of one community (pregnant women). First we show the time development of population density p_i of all the threads in all the windows during the whole period in Figure 3. If a thread in a window is selected as a keystone thread, it is colored red. We can see that keystone threads tend to appear more as the time elapses or the community matures.

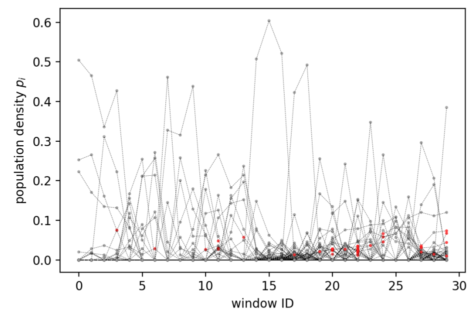

Figure 3: Temporal development of population density p_i over windows for all the threads. Each line indicates the development of each thread and it is colored red when selected as a keystone thread.

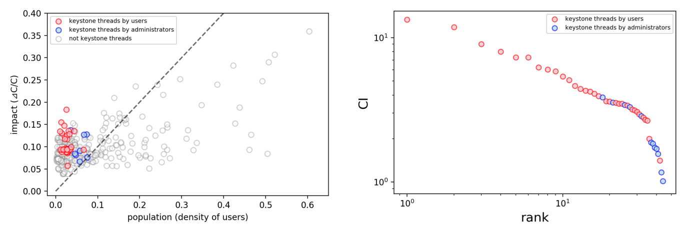

Figure 4: (left) Scatter plot with p_i on the horizontal axis and I_i on the vertical axis of all windows. The dashed line represents $y = x$ indicating a proportional relationship between population and impact. Keystone threads by users (red=32), administrators (blue=12), and other non-keystone threads (white=405). (right) CI values of keystone threads are arranged in rank order. Red: keystone threads by users, blue: keystone threads by administrators.

Figure 4-left shows a scatter plot with population density p_i and impact I_i on the horizontal and vertical axes, respectively. The results of all threads are plotted here. This figure shows the results of all windows. Because the thresholds for keystones are computed considering the maximum values for population and impacts in each window, the threshold values become different for each window. The dashed line in the figure indicates $y = x$, and threads being on this line indicates that their impacts on removal are proportional to the population density. The colored (red and blue) circles are threads identified as keystones. If a thread is by users, it is colored red and is blue if it is by administrators. Other non-keystone threads are in white. The number of threads identified as keystones is 44 among 449 threads. Of the 44 threads identified as keystone threads, 32 (72.7%) are threads by users. This indicates a possibility that the thread by users is more likely to be selected as a keystone thread.

Figure 4-right shows the keystone threads sorted by community importance CI_i. Reds are keystone threads by users and blues are keystone threads by administrators. The top 16 ranked CI_i are of threads by users. Among keystone threads, the average CI_i of keystone threads by users is 4.88, and the average CI_i of keystone threads by administrators is 2.33. The analysis using CI_i also confirms that keystoneness tends to be higher for threads created by users than by administrators.

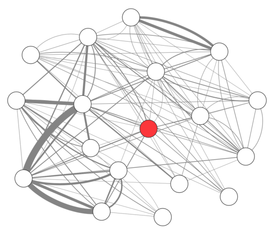

Figure 5: Example of the thread network (window ID = 19) in which a keystone thread is colored red.

To investigate what types of threads by users are identified as keystones, we also categorized each thread according to its content and found five major categories such as chatting (12 threads), word-chain-game (9), consultation (7), picture sharing (2), and reports on winning the lottery (1). One interesting category is keystone threads with word-chain-games. The most keystone-selected thread (six times) as well as the top CI_i ranked thread is also a word-chain-game thread. Although the number of participating users is small, a thread with active interaction between users is more likely to be identified as a keystone thread.

Figure 5 shows a thread- network (α_{ij}) of in a window (window ID = 19) in which the word-chain-game thread exists[3]. Each node corresponds to a thread and the edge width corresponds to the influential strength among nodes. The red-colored node is the keystone thread that is the only one selected as the keystone thread. Moreover, the nodes connected with strong influences are those of dominant threads. Network analysis on centralities (degree, betweenness, and pagerank) show that the discovered keystone node has neither high nor low values, but rather, median values. Nevertheless, their impacts are large enough to collapse the entire system.

From this result, it became clear that thread by users is important for the stable operation and autonomy of an online bulletin board. The results on the other community (mother community) showed very similar results, that is, the keystoneness is higher in threads by users than threads by administrators. Among 427 threads, 64 threads are identified as keystone threads, and 38 threads (59.4%) are threads by users, more than the majority. The average values of CI_i of keystone threads are 10.4 and 4.82 for threads by users and administrators, respectively.

5 CONCLUSION

In this study, we introduced the concept of keystone species, which is important for keeping the biological ecosystem stable, on the online bulletin board, and proposed a new influential-content identification method in the online community. We considered threads in online bulletin boards as species in a biological ecosystem and modeled influences among threads by using the multivariate Hawkes process. We constructed a thread network with threads as nodes, estimated the influence as edges, and calculated the degree of impact on the activity level of the entire network when a thread is removed. Accordingly, we showed that there are threads that play keystone roles. An analysis of threads from the viewpoint of thread by users and thread by administrators shows that all dominant threads are occupied by thread by administrators, and in contrast, thread by users is more likely to be identified as a keystone thread.

Keystone species plays an important role in preserving the stability of biological ecosystems. When considering an online community as an ecosystem, its stability is realized by its high level of activity and engagement by the user. Identifying content that may work to increase the activity of the community despite its small population or event occurrence is difficult in the existing method, so the proposed method gives a new perspective for stable operation of the service.

In this study, we only analyzed two communities, both of which showed similar tendencies, but in the future, we would like to analyze all communities (more than 100) in the bulletin board service and verify the overall tendency. Moreover, we would like to analyze how the types of keystone threads or their numbers differ when grouping threads with semantics. Finally, our method is general and can be applied for identifying, for example, keystone users in other online communities such as social network services.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP17H01821 and JP19H04214.

REFERENCES

[1] Lada A Adamic, Thomas M Lento, Eytan Adar, and Pauline C Ng. 2016. Information evolution in social networks. In *Proc. of the 9th ACM International Conference on Web Search and Data Mining*. 473–482.

[2] Richard Dawkins. 1976. *The selfish gene.* Oxford University Press New York. 224 pages.

[3] James A. Estes, Alexander Burdin, and Daniel F. Doak. 2016. Sea otters, kelp forest, and the extinction of Steller's sea cow. *Proc. of the National Academy of Sciences of the United States of America* 113, 4 (2016), 880–885.

[4] Mehrdad Farajtabar, Xiaojing Ye, Sahar Harati, Le Song, and Hongyuan Zha. 2016. Multistage Campaigning in Social Networks. In *Proc. of the 29th International Conference on Neural Information Processing Systems (NIPS 2016)*. 4725–4733.

[5] Vladimir Filimonov and Didier Sornette. 2012. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E* 85, 5 (2012), 056108.

[6] Amrita Gupta, Mehrdad Farajtabar, Bistra Dilkina, and Hongyuan Zha. 2018. Discrete Interventions in Hawkes Processes with Applications in Invasive Species Management. In *Proc. of the 27th International Joint Conference on Artificial Intelligence*. 3385–3392.

[7] Alan G Hawkes. 1971. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)* (1971), 438–443.

[8] Takashi Ikegami. 2005. Neutral phenotypes as network keystone species. *Population Ecology* 47, 1 (2005), 21–29.

[9] Tomoharu Iwata, Amar Shah, and Zoubin Ghahramani. 2013. Discovering latent influence in online social activities via shared cascade poisson processes. In *Proc. of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 266–274.

[10] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media?. In *Proc. of the 19th International Conference on World Wide Web*. 591–600.

[11] Andrew S Lan, Jonathan C Spencer, Ziqi Chen, Christopher G Brinton, and Mung Chiang. 2018. Personalized Thread Recommendation for MOOC Discussion Forums. In *Proc. of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 725–740.

[12] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. 2009. Meme-tracking and the Dynamics of the News Cycle. In *Proc. of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 497–506.

[13] Yasuko Matsubara, Yasushi Sakurai, and Christos Faloutsos. 2015. The web as a jungle: Non-linear dynamical systems for co-evolving online activities. In *Proc. of the 24th International Conference on World Wide Web*. 721–731.

[14] L Scott Mills, Michael E Soulé, and Daniel F Doak. 1993. The keystone-species concept in ecology and conservation. *BioScience* 43, 4 (1993), 219–224.

[15] Yosihiko Ogata. 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association* 83, 401 (1988), 9–27.

[16] Tomokatsu Onaga and Shigeru Shinomoto. 2016. Emergence of event cascades in inhomogeneous networks. *Scientific reports* 6 (2016), 33321.

[17] Robert T Paine. 1966. Food web complexity and species diversity. *The American Naturalist* 100, 910 (1966), 65–75.

[18] Robert T. Paine. 1969. A note on trophic complexity and community stability. *American Naturalist* 103 (1969), 91–93.

[19] Mary E Power, David Tilman, James A Estes, Bruce A Menge, William J Bond, L Scott Mills, Gretchen Daily, Juan Carlos Castilla, Jane Lubchenco, and Robert T Paine. 1996. Challenges in the Quest for Keystones: Identifying keystone species is difficult—but essential to understanding how loss of species will affect ecosystems. *BioScience* 46, 8 (1996), 609–620.

[20] Elad Segev, Asaf Nissenbaum, Nathan Stolero, and Limor Shifman. 2015. Families and networks of internet memes: the relationship between cohesiveness, uniqueness, and quiddity concreteness. *Journal of Computer-Mediated Communication* 20, 4 (2015), 417–433.

[21] Greg Ver Steeg and Aram Galstyan. 2012. Information transfer in social media. In *Proc. of the 21st International Conference on World Wide Web*. 509–518.

[22] Ali Zarezade, Utkarsh Upadhyay, Hamid R Rabiee, and Manuel Gomez-Rodriguez. 2017. Redqueen: An online algorithm for smart broadcasting in social networks. In *Proc. of the 10th ACM International Conference on Web Search and Data Mining*. 51–60.

[23] Ke Zhou, Hongyuan Zha, and Le Song. 2013. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Proc. of the 16th International Conference on Artificial Intelligence and Statistics*. 641–649.

[3] A threshold is set to draw the edges of the network. The threshold value is set using an average value of α_{ij} excluding the diagonal components, and the edges are drawn if the values are larger greater than the this threshold.