2023

MASTER’S THESIS

Analyzing Effects of Sociality of Social Media Users on Content Caching in ICN

ACADEMIC SUPERVISOR: KAMIYAMA Noriaki

Graduate School of Information Science and Engineering
Ritsumeikan University

MASTER’S PROGRAM
MAJOR in Advanced Information Science and Engineering

STUDENT ID: 6611220036-4
NAME: HONDA Kodai

Abstract

Nowadays, social media platforms such as X and Facebook have become essential infrastructure supporting our lives. As the network architecture to help reduce the large amount of content produced by social media users, Information-Centric Networking (ICN) has been drawing attention. Because ICN is one of the cache networks, it is important to carefully design a caching strategy to enhance the performance of ICN. If we assume that ICN is introduced as the network architecture for social media, the performance is improved by designing ICN considering the sociality of social media users: the existence of influential users, information propagation and community structure. Therefore, in this paper, I build a conceptual ICN-based platform for social media and model user's behavior. Moreover, I extensively analyze how the sociality of social media users affects content caching in ICN. My key findings are summarized as follows. First, according to the selection method and ratio of influential users, the cache hit ratio can significantly be improved. Second, while the propagation probability by users on social media rarely affects the cache hit ratio and the length of the delivery path, the community structure among users on social networks has positive effects on the content caching.

Contents

Chapter 1

Introduction

Nowadays, social media platforms such as X and Facebook have become essential infrastructure supporting our lives. Through these social media, we can communicate with our friends online and strengthen our social relationship. While social media has negative aspects, for instance, so-called ”flaming” and ”fake news,” it has essentially enriched our social and economic activities. However, with the growth of social media, the amount of content produced by users has dramatically increased, which accelerates the increase in traffic volume transferred across communication networks.

One of promising techniques to realize the efficient content distribution for social media is cache networks, two of which are Content Delivery Network (CDN) [1] and Information-Centric Networking (ICN) [2]. CDN today supports the Internet and enables users to retrieve content not only from a server storing the original content, i.e., origin server, but also from geographically-neighbor cache servers. Also, ICN has been drawing a large amount of attention as a next generation Internet architecture. In ICN, routers in the network can temporarily store the replica of content. Hence, content is distributed from intermediate routers as well as a server storing original content. As a consequence, users can retrieve content from nearby routers, thereby reducing the volume of traffic transferred across the network.

The key for cache networks to efficiently work is to appropriately design *caching strategy* and *cache replacement policy* that operate at caching nodes, e.g., cache servers in CDN and routers in ICN. The caching strategy is a method that when a caching node receives a content, it determines whether it inserts the content to its own cache or not. On the other hand, the cache replacement policy is to determine a content to be discarded from the caching node when the cache is fully occupied with other content. Because the performance of cache networks is dominated by whether I fully utilize the potential of caches within the network. For this reason, it is important to carefully design two mechanisms.

If we assume ICN is introduced as a communication network for social media, it is expected to improve the performance of the network by designing caching strategy and cache replacement policy taking into account sociality of social media users. Representative features of user’s sociality are summarized as follows [3, 4].

- Existence of influential users: Information sourced from an influential user, so-called *influencer*, is more likely to spread on social media compared to non-influential users.
- Information propagation behavior: Information sourced from a user spreads to other users via social relationship, i.e., social networks.
- Community structure: A community emerges, which is comprised of users with similar characteristics. Communication among users within a community is more active than that among users belonging to different communities.

One of the previous studies that have been tackling designing caching strategy focused on the existence of influential users among aforementioned sociality [5]. The authers proposed caching strategy considering influential users on social media, and revealed the caching strategy can contribute to the improvement of performance for ICN in the literature [5]. The core idea of the proposed caching strategy is that a router does not cache content uniformly, but caches content published by only influential users.

Even though several other studies have also been devoted to designing cache networks considering sociality of users, to the best of my knowledge, the interaction between specific features of social media and the content caching in ICN has not been fully understood. For instance, while the effectiveness of using caching strategy considering the existence of influential users was revealed, it is still unclear how we should identify influential users and how many users should be regarded as influential users. Moreover, the studies overlooked the other two aspects of sociality despite their importance. On the contrary, it remains uncertain whether I should take into account these two aspects of sociality when designing and controlling ICN.

Therefore, in this paper, I assume to introduce ICN particularly Content-Centric Networking (CCN) [6] and Named Data Networking (NDN) [7], as a content distribution platform for social media, and analyze the characteristics of socially-aware ICN. Specifically, I focus on the sociality of users on social media including the existence of influential users, information propagation and community structure. In order to analyze the sociality, I first tackle building a socially-aware ICN that is a content delivery model on ICN taking into account the sociality of users on social media. And then, I investigate the effects of the sociality of social media users on content caching in ICN.

The contributions of this paper are summarized as follows.

- I present a conceptual architecture to realize ICN-based content distribution for social media and its component, and build socially-aware ICN model that incorporates user's information propagation and community structure by extending an existing user's behavior model.
- And, I extensively analyze how the sociality affects content caching in ICN utilizing simulation and mathematical analysis. As a result, in terms of influential users I reveal that the selection method for influential users and the number of influential users affect cache hit ratio, which is defined as the probability that a user can retrieve content from any caching routers, and estimating influential users from partial structure of social network, which is obtained with sampling strategy, is effective to achieve as high cache hit ratio as the complete network topology can be available.
- Moreover, I conduct simulations to investigate the effects of the other two aspects of sociality: information propagation and community structure, on content caching in ICN. Consequently, I show that the information propagation behavior rarely affects the cache hit ratio and the length of delivery path, and the community structure among users on social networks has positive effects on the content caching.

The structure of this paper is organized as follows. First, Chapter 2 reviews previous research on cache networks incorporating specific features of social media. Chapter 3 presents a conceptual ICN-based platform for social media content distribution. Chapter 4 models socially-aware ICN that describes interaction among users on social media as the content delivery on ICN. Then, I extend an existing user's behavior model in order to incorporate user's information propagation behavior and community structure. Chapter 5, investigates how the three aspects of sociality affect the content caching in ICN through simulations. Specifically, I conduct simulations to reveal how selection of influential users affect the content caching in ICN and use mathematical analysis to examine the effect of the ratio of influential users on the content caching in ICN. Moreover, based on the socially-aware ICN modeled in Chapter 4, I investigate the effects of user's information propagation and community structure on content caching in ICN. Finally, Chapter 6 provides the summary of this paper.

Chapter 2

Related Work

Several types of approaches to reduce the large amount of network traffic on the Internet have been examined. One of the promising approaches is to use cache networks, two of which are CDN and ICN. In this chapter, I introduce some cache networks that take into account user's sociality on social media.

As a method for reducing the amount of network traffic caused by numerous users on social media, studies on CDN considering sociality of social media users have been made. In CDN, content is not centralized in one server, but the content that has a large popularity is cached in several servers that are determined according to some factors. By decentralizing the location of each content, CDN can help reduce content delivery delay and improve content availability. Zheng *et al.* [8] extracted geographical, social and time locality from content cascade pattern by measuring the actual large scale social network. As a result, they revealed the utilization of the framework based on these characteristics contributes to improving download ratio compared to conventional approach. Also, Hu *et al.* [9] tackled making it efficient to deliver content on CDN by classifying users into communities according to user's preference. To be precise, if a user is interested in a certain content, and another user is also interested in the same content, they are likely to be classified in the same community. These approaches contribute to reducing monetary cost compared to conventional methods.

In order to enhance the efficiency of content delivery for social media, the systems that use socially-aware ICN have been drawing attention. In ICN, each content is named uniquely so that routers can forward content based on the names. Each router has a cache capacity to temporarily store content, which enables users to retrieve content not only from the origin server but also from a caching router if the content is cached at the router. Therefore, a wide variety of methods considering sociality has been developed. For instance, by considering the geographical locality of users on social media, it is revealed that the amount of traffic and the burden of servers can be improved [10, 11]. Mathieu *et al.* [10] analyzed the behavior and locality of the network on Twitter, which is an actual online social network, and detected that the behavior of the current network does not match with the end users' behavior. Furthermore, they adjusted online social networks to ICN and then proposed ICN-based network architecture that reduces the burden for both network and servers. Truong*et al.* [11] proposed content distribution system based on Named Data Networking (NDN), which is one of the architectures to realize ICN, utilizing the fact that users on social media are likely to follow the users who are located in a near place geographically.

The most relevant study [5] to this paper proposed a caching strategy called Socially-Aware Caching Strategy (SACS), which incorporates the existence of influential users on social media. The concept of SACS is that a router does not cache content uniformly, but content only published by influential users. Simulation and experiment on a testbed revealed that the effectiveness of SACS, for instance, the cache hit ratio can be dramatically improved.

Chapter 3

ICN-Based Content Distribution Architecture for Social Media

In this chapter, I present a conceptual architecture realizing ICN-based content distribution for social media and its component. The key idea of my proposed architecture is summarized as follows.

- It realizes the content delivery among users who have social relationships on a physical network comprising ICN routers.
- It provides a way for the user to grasp a name of content that they seek by introducing two types of requests: pre-request and actual-request.
- It incorporates a mechanism for sharing a list of influential users among ICN routers. This enables ICN routers to operate the caching strategy considering influential users.

3.1 Preliminary

As a social media application, I assume a closed user-oriented application such as X and Facebook. In the social media application, a user can have social relationships with other users[1]; for instance, in X, this relationship corresponds to the follower and the followee. In addition, I assume that contents produced by a user can be shared among social neighbors of the producing user. The rationale behind this assumption is to focus on the effect of the social relationship among users on the content caching as an initial step to evaluate the socially-aware ICNs. Also, this assumption makes the implementation of our conceptual architecture easy; specifically, to achieve such a content retrieval, a user participating the social media application simply manages and updates his/her own friendship list. A substantial benefit of this mechanism is that a user does not need to grasp the entire structure of the social relationship, which is feasible for the social network evolving with time.

The content in a social media application is delivered over an ICN network comprising multiple ICN routers with finite cache. Note that we refer an ICN router to a router used in CCN [6] and NDN [7]. In the following explanation, I assume that the forwarding table at each ICN router is appropriately configured according to the routing protocol. For this reason, the ICN router can forward a request packet and a response packet, i.e., content request and content, to a desired destination. It is worth of noting that this assumption is realized by using the routing protocol for ICNs, NLSR (Named-data Link State Routing) protocol [12]. More specifically, each router advertises its accommodating users as a prefix of contents in accordance with NLSR protocol; as a consequence of repeated advertisements among routers, the router can construct its own forwarding table.

[1] Hereinafter, I simply call these users as *social neighbors*.

3.2 Naming

Because a user is needed to uniquely identify its name when requesting a content, a unique content identifier must be assigned to each content in ICNs. In my conceptual architecture, I suppose that the content identifier includes the identifier of the user who published the content, e.g., the user's account name [11]. Specifically, I define the content identifier as the following notation: `/[SNS application]/[user account]/[content id]`.

3.3 Content Retrieval

A user participating in social media application, who is directly connected to a nearest ICN router, can perform two activities: *publish* and *retrieve* [5]. "Publish" indicates that a user newly produces a content, and "retrieve" indicates that a user obtains content from its social neighbors. For instance, in the context of X, "publish" and "retrieve" correspond to posting a new tweet and obtaining new tweets by followees, respectively.

In general ICN, the content retrieval is realized by a simple procedure — a user simply issues a request packet including the identifier of requesting content that is given as a priori knowledge; in contrast, in our network architecture, the content retrieval is realized by two types of requests, pre-request and actual-request. First, a user issues pre-requests to its social neighbors to retrieve lists of contents newly-published by each of social neighbors. Specifically, for every social neighbor, i.e., target user, the requesting user injects a single request packet that includes a timestamp when the requesting user lastly issued a pre-request for the target user. Then, the target user receiving the pre-request summarizes a list of contents that are published after the timestamp embedded in the pre-request, and returns the list describing identifiers of these contents as a response packet to the requesting user. After that, based on the returned list, the requesting user sequentially issues actual request packets to the target user. The request packet is relayed through ICN routers, and an intermediate ICN router immediately sends back the corresponding content if the ICN router caches it. Otherwise, the request packet will arrive at the target user, then it simply returns the requested content.

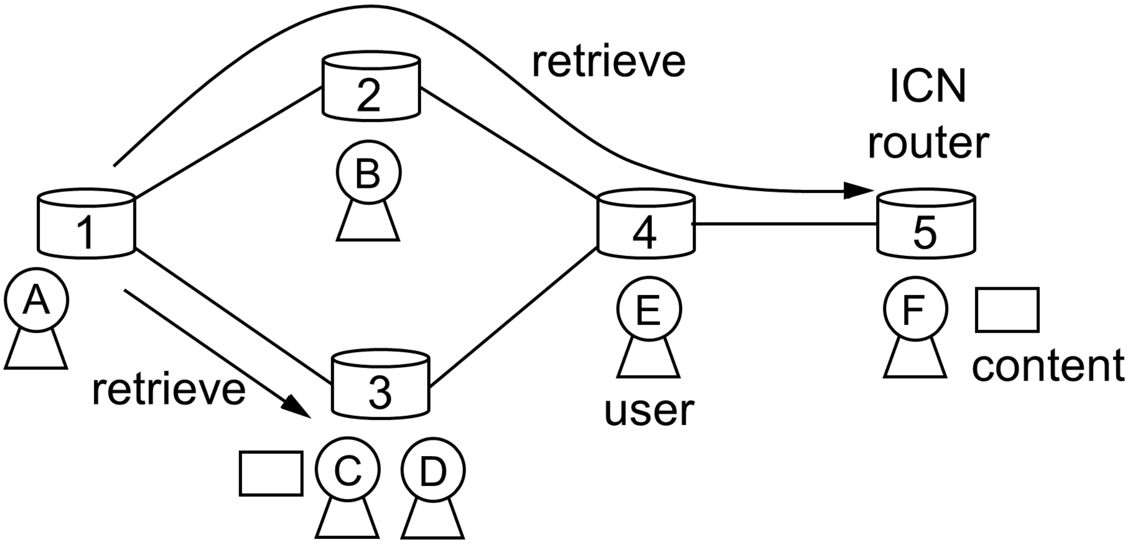
The advantage of using two types of requests is to enable users to retrieve content published by their social neighbors even though the users do not know their content's identifier. In contrast, its intuitive disadvantage is twofold: (i) increase in the time for a user to retrieve contents and (ii) increase in the traffic volume. In particular, regarding (i), using two types of requests causes that an additional delay due to a pre-request, which corresponds to one RTT (Round-Trip Time) between ICN routers accommodating users, is added to an actual content delivery delay.

An example of content retrieval on ICN-based social media application is illustrated in Fig. 3.1. Figure 3.1(a) depicts a physical network comprising of multiple ICN routers and users each of which is accommodated to an ICN router. Also, Fig. 3.1(b) depicts a social relationship among users; for instance, `user A` has relationship with `user C` and `user F`. As aforementioned, I assume that a request destined to a content published by a user is passed through ICN routers. In this example, `user A` retrieves contents published by `user C` and `user F`. In this case, a request from `user A` to the content published by `user C` is passed through `routers 1` and `3`; that to the content by `user F` is passed through `routers 1`, `2`, `4`, and `5`.

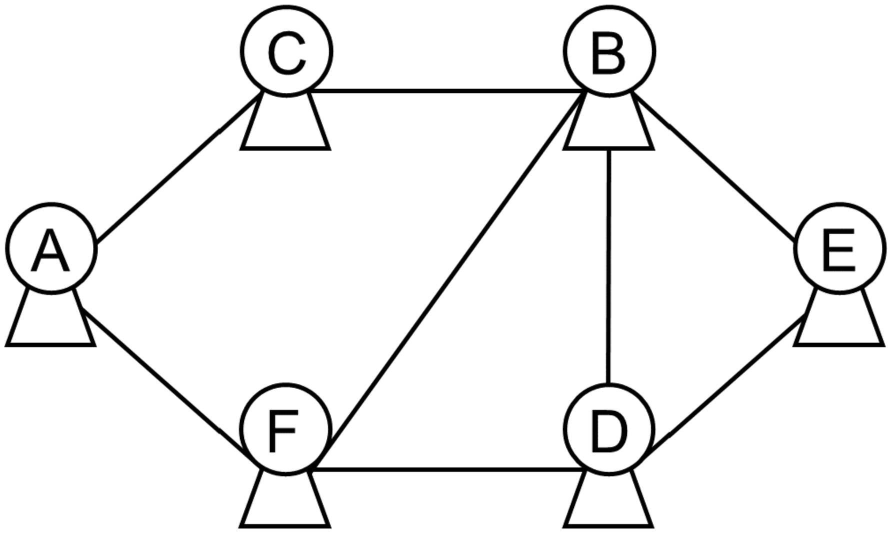
3.4 Content Caching

The ICN router controls its own cache according to the caching strategy and the cache-replacement policy when forwarding contents produced by SNS users. For instance, an ICN router following LCE (Leave Copy Everywhere) [6], which is a typical caching strategy, uniformly caches contents. Of course, other caching strategies, e.g., LCD (Leave Copy Down) [13], can operate at an ICN router.

Because I will use SACS [5] as a caching strategy in Section 5.1, we have to mention how SACS operates on our conceptual architecture. In the case of caching strategy SACS, an ICN router needs to judge whether the content publisher is an influential user or not, to determine whether to cache the content. To operate SACS, it is necessary for an ICN router to maintain a list of influential users in advance. This is realized that an administrator of the ICN network observes social relationship among users on SNS application, i.e., OSN, and then distributes a list of estimated influential users to each ICN router. Consequently, when an ICN router receives a response packet, it can determine whether it caches the packet by performing the following operations: (i) extracting a user account contained in the content identifier of the response packet, and (ii) collating the user account and a list of influential users. I should note that routers in ICN networks do not need to look up a content itself of response packet on determining their cache decision because I assume that the identifier of influential user is embedded in the content identifier of response packet as specified in [14].



(a) content retrieval between users on an ICN network



(b) social relationship among users

Figure 3.1: An example of content retrieval on ICN-based social media application

Chapter 4

Socially-Aware ICN Model

In this chapter, I model interaction among users on social media as the content delivery in ICN based on the conceptual architecture presented in Chapter 3. Then I extend an existing user's behavior model so as to incorporate the sociality of users on social media; information propagation and community structure.

4.1 Network Model

I model the content delivery among social media users accommodated in routers on an ICN network as a discrete-time system. I denote the time and the slot length in this system as t and Δ, respectively.

I express a network consisting of ICN routers with a finite cache as an undirected graph $G_I = (\mathcal{V}, \mathcal{E})$. The ICN router $r \in \mathcal{V}$ relays requests and contents. Also, it caches contents according to the caching strategy and the cache replacement policy.

Also, I express a network representing friendship among users on social media, which is so-called *social network*, as a directed graph $G_S = (\mathcal{U}, \mathcal{L})$. An edge $(u, v) \in \mathcal{L}$ in the directed graph G_S indicates that a user u follows a user v. Hence, a set $\mathcal{N}_u^+$ of users that a user u follows is $\{v \mid (u, v),\ v \in \mathcal{U}\}$. In contrast, a set $\mathcal{N}_u^-$ of users following user u, which is commonly known as *followers*, is $\{v \mid (v, u),\ v \in \mathcal{U}\}$.

I denote a set of communities to which a user belongs as C, and assume that a user belongs to any one of communities $c \in C$. In this paper, since I focus on a simple community structure, a user does not belong to multiple communities at the same time. In other words, $\bigcup_c c = \mathcal{U}$ and $c \bigcap c' \neq \emptyset\, \forall c, c'(c \neq c') \in C$.

I map each user $u \in \mathcal{U}$ on the social network onto any one of ICN routers $r \in \mathcal{V}$ based on the user's community (see Fig. 4.1). Specifically speaking, this mapping procedure is composed of twofold: (i) for a given community $c \in C$, I determine an ICN router to be mapped with the community c and accommodate all users belonging to the community c into the router, and (ii) I transfer accommodated users to a randomly-chosen different ICN router with a given probability. The procedure (i), community-based accommodation of users, expresses a fact that users who are closely-related on the social network are expected to be closely-located on a physical network, i.e., an ICN network. The procedure (ii), probabilistic transfer of users, is intended to adjust *community strength*, i.e., the degree to which the community structure is preserved. Below, I denote an ICN router accommodating a user u as r_u and call the probability of transferring users from one router to another one *transfer probability*. The transfer probability with 0 means that the community structure is fully maintained; that with 1 means that the community structure is lost, which corresponds to random assignment of users to ICN routers.

A user on the social network publishes contents and then its friends, i.e., followers, request these contents. A request from a user u to a content published by a user v is supposed to be passing through ICN routers along a path $P_{u,v} = (r_u, \ldots, r_v)$ where r_u and r_v are routers accommodating users u and v, respectively. The content itself is supposed to be returned to a requesting user passing through a reverse path along which the request is forwarded.

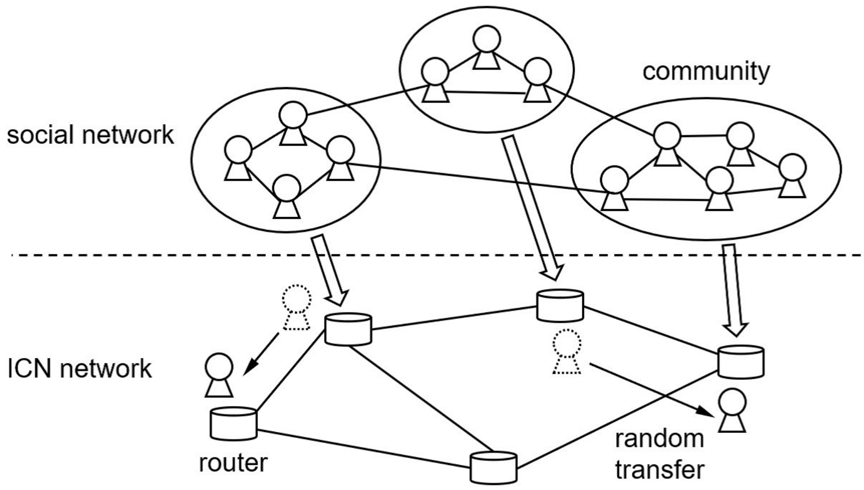


Figure 4.1: User accommodation to ICN router based on community structure.

4.2 User's Behavior Model

As a model of user behavior on social networks, I utilize the behavior model proposed in [15] and extend it so as to incorporate the information propagation behavior by users which is corresponding to, for instance, repost in X and retweet in Twitter. The user behavior model [15] specifies a user's behavior as follows; (i) the user initiates a session, (ii) within the session, the user takes either one of two types of actions: viewing contents published by following users or publishing a new content, (iii) after repeating the procedure (ii) with fixed intervals, the user terminates the session, (iv) the session, i.e., a series of procedures (i) to (iii), is repeated with fixed intervals. In this paper, I extend the procedure (ii) as follows. In the following, I describe the behavior of a user $u \in \mathcal{U}$ at time t.

- Request

  The user u retrieves contents that each of following users $v \in \mathcal{N}_u^+$ lastly published. At the same time, the user also retrieves contents advertised by following users. Namely, the user u issues requests to contents publisher to retrieve contents advertised during the interval $(t_{\mathrm{prev}} : t)$, where t_{prev} is time when the user u issued a last content request. I note that if the same content is advertised by different users, the content request is aggregated into a single one, i.e., the user never issues multiple requests to an identical content.

  After retrieving the content, the user u transitions to the *advertisement* behavior described below.

- Advertise

  The user u advertises retrieving contents to its followers $\mathcal{N}_u^-$ with a propagation probability p_u. This behavior is derived from IC (Independent Cascade) model [17].

- Publish

  The user u generates new content.

Chapter 5

Analysis of Effects of User's Sociality on Content Caching in ICN

In this chapter, I analyze how user's sociality affects content caching in ICN through simulation and numerical analysis. Specifically, as the sociality of social media users, I focus on three characteristics, which are the existence of influential users, information propagation and community structure.

5.1 Influential Users

In this section, I extensively investigate how the selection and the ratio of influential users affect the content caching in ICN using simulation and mathematical analysis. First, I conduct simulations to reveal the effect of the selection of influential users on social media. Next, I approximately derive the cache hit probability for a given degree distribution of a social network, and then analyze the relashonship between the ratio of influential users and the cache hit probability.

5.1.1 simulation

The experiments are composed of two parts. The first experiment aims to reveal what centrality measure should be used when selecting influential users. To accomplish this aim, I measure the cache hit ratio while changing various centrality measures. In the first experiment, I assume that the complete structure of a social network is given to calculate the centrality measure of nodes; however, it is generally known to be difficult to completely grasp the entire structure of a social network due to its dynamics, i.e., the structure of a social network dynamically changes. Therefore, the second experiment aims to verify that selecting influential users is still effective when the entire structure of a social network is unknown. I investigate the effectiveness of selecting influential users from a partial structure of the social network, which is obtained with a sampling strategy.

Method

I generated a random graph, which represents connection relationships among ICN routers, with ER (Erdõs-Rényi) model [26]. The number of nodes and links in the random graph wes set to 100 and 200, respectively. From two types of datasets: Last.fm[1] and Facebook[2], both of which are typical SNS, I obtained graphs which represent social relationship among OSN users. The number of nodes and links in Last.fm is 1,843 and 12,268. Also, the number of nodes and links in Facebook is 4,039 and 88,234. I should note that because a graph contained in Last.fm dataset is unconnected, i.e., a graph is composed of fragmented multiple graphs, I used the largest-connected component as a graph for my simulation. In My simulation, an OSN user is randomly accommodated to an ICN router.

[1]`https://grouplens.org/datasets/hetrec-2011/`
[2]`https://snap.stanford.edu/data/ego-Facebook.html`

According to a user's interaction model in OSN [5], a user repeatedly produces a content and issues a request for contents provided by users which have social relationship with the requesting user, i.e., adjacent nodes from the requesting node in OSN. Because I would like to focus on influential users, I use the existing user's interaction model, which doesn't consider information propagation and community structure. The parameter settings in the user's interaction model were the same with those in [5]. Please refer to [5, 15] for the details of the user's interaction model and its parameter setting.

As a caching strategy at a router, I used SACS proposed in [5]. Namely, a router following SACS only caches contents published by influential users. The cache size at an ICN router was set to 10 [contents], and the cache-replacement policy was LRU.

In this paper, I calculated the centrality of a user using a given centrality measure, and I regarded top $p\,(0 \leq p \leq 1)$ users with the highest centrality measure as influential users. I used eight types of centrality measures — degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, PageRank [18], k-core index [19], VoteRank [20], and CI (Collective Influence) [21]. Degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, PageRank, and k-core index are typical centrality measures used in the field of complex networks. In contrast, VoteRank and CI are recently proposed one, and it is known that these centrality measures are superior in terms of efficient information diffusion. In our simulation, I set the ratio of influential users p to 0.1. As described in Section 2, p is one of important parameters affecting our simulation results, but it is difficult for us to investigate the effect of p through the simulation due to its computational cost. For this reason, Section 5.1.2 will investigate the effect of the ratio of influential users p through mathematical analysis.

As mentioned above, I also conduct simulation assuming that the complete network structure of social media is not given. The overview of this experiment is as follows; (i) I obtain a partial structure of the social network, which is corresponding to a subnetwork comprised of the sampled nodes, with a sampling strategy; (ii) I select influential users from the subnetwork based on a given centrality measure; (iii) I measure the cache hit rate for a given influential users through a simulation. Except for the process of selecting influential users from a social network, the experiment methodology is same with that described above; hence, I shall explain only the selection of influential users from an unknown social network.

To obtain a partial structure of a social network, I used the following four types of sampling strategies [22, 23]:

- RandomWalk (RW) sampling

  This sampling strategy obtains a partial structure of a network topology with a randomwalk on a graph [24]. Specifically, a walker, called an agent, sequentially visits to an adjacent node, which is randomly chosen from neighbor nodes of the currently-visiting node with a uniform probability. This procedure is repeated until the number of nodes visited by the walker reaches a given number of sampled nodes.

- DFS (Depth-First Search)

  From a randomly-chosen starting node, DFS iteratively visits a one of visited nodes. DFS visits and explores unvisited neighbor nodes of the earliest visited node [22]. Similar to RW sampling, the exploration with DFS terminates when the number of explored nodes reaches a given number of sampled nodes.

- BFS (Breadth-First Search)

  Sampling with BFS is same with DFS except for the selection of node to be visited. BFS visits and explores unvisited neighbor nodes of the most-recently visited node [22].

- Random sampling

  This sampling strategy simply obtains a set of nodes that is randomly-chosen from an entire network. I should note that this sampling strategy is not intended to obtain a partial structure of the network; I simply used this strategy to provide a baseline result.

In RW sampling, DFS, and BFS, I randomly selected a source node starting exhaustive exploration.

I identified the top pN nodes with highest degree from a sampled network as influential users, where p and N are a ratio of influential users and the number of nodes of a (complete) social network, respectively. Unlike the first experiment, I only used the degree centrality as a centrality measure. This is because, in this experiment, I focus on the effectiveness of the sampling technique to identify influential users rather than the effect of centrality measures.

I used our ICNSIM (ICN SIMulator) for the simulation, and measured the cache hit ratio, which is defined as the ratio of the number of corresponding contents returned from intermediate ICN routers along a path to the number of requests issued by users. I repeated a single simulation with 100,000 [slot] 10 times and computed the average cache hit ratio over an entire network.

Result

First, I discuss the effect of the centrality measure on the cache hit ratio. Figure 5.1 shows the average cache hit ratio with different centrality measures. Results of Last.fm and Facebook are shown in Figs. 5.1(a) and 5.1(b), respectively. For the sake of detail analysis, I additionally plotted results when randomly selecting influential users as "random".

From Fig. 5.1(a), I can find that selection of influential users based on centrality measures significantly improves the cache hit ratio. Specifically, among eight centrality measures, the highest cache hit ratio is achieved with betweenness centrality, PageRank, and VoteRank. In contrast, the lowest one is achieved with eigenvector centrality and k-core index. This tendency means the importance of appropriately selecting influential users, that is, the centrality measure is an important factor which affects the communication performance of socially-aware ICNs.

From Fig. 5.1(b), I confirm that the above observation is maintained in a different graph. Comparison between Figs. 5.1(a) and 5.1(b) implies that observed tendency is maintained in the case of Facebook. Namely, the cache hit ratio is dominated by the centrality measure regardless of the network topology of OSN. It is worth of mentioning that the cache hit ratio in the case of Facebook is overall smaller than Last.fm. This difference is caused by the fact that the number of nodes in Facebook is larger than Last.fm. Hence, there exist more contents generated by influential users, which become caching candidates at ICN routers, in a network. This results in degradation in the cache hit ratio.

I discuss which the centrality measure should be used while considering the computational cost of calculating centrality measures. Through our simulation, I confirmed that selection of influential users with betweenness centrality and PageRank significantly improves the average cache hit ratio compared to other centrality measures. However, I cannot ignore the computational cost required to calculate betweenness centrality and PageRank. Because the size of OSN is tremendously large in general, these two centrality measures whose computation depends on the network size might be unsuitable. From this viewpoint, degree centrality becomes an alternative approach because of the following reasons: (i) calculation of degree centrality does not need any computational overhead, (ii) the cache hit ratio of degree centrality is close to that of betweenness centrality and PageRank.

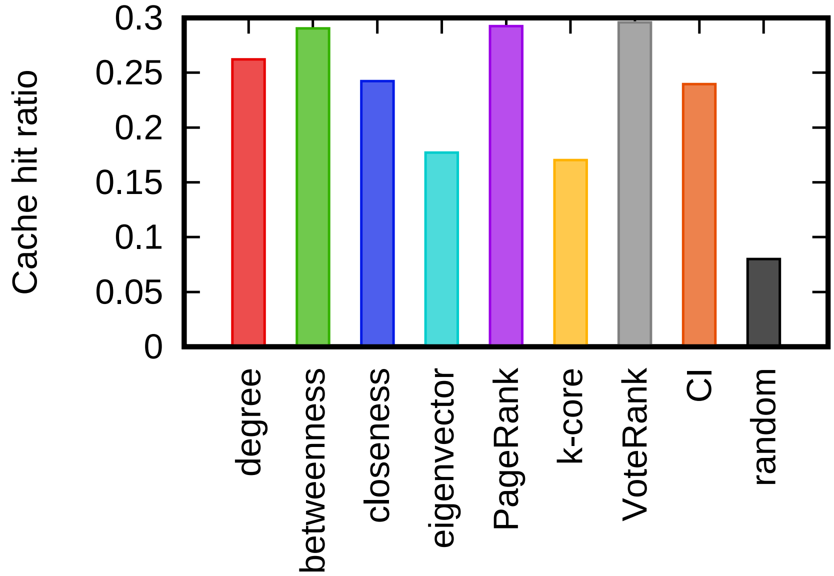
Next, I investigate whether estimating influential users from a partial structure of an OSN is effective. Especially, I focus on how much the number of sampled nodes is required to achieve the cache hit ratio when the OSN is completely given.

Figure 5.2 depicts the relationship between the sampling ratio, which is defined as the ratio of the number of sampled nodes to the number of nodes in a complete OSN, and the cache hit ratio. In these figures, results with different sampling strategies are plotted.

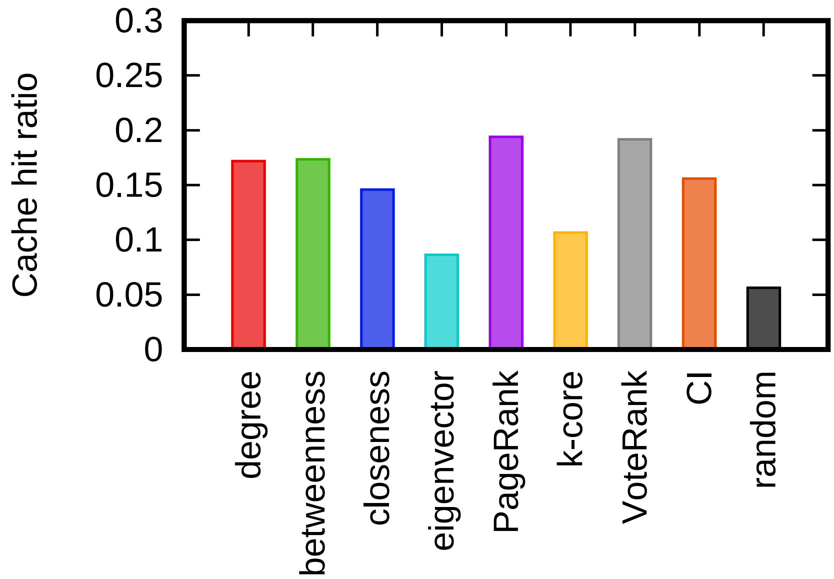
Figure 5.2 reveals that selecting influential users from a limited knowledge of an OSN is effective in socially-aware ICNs. In particular, the cache hit ratio with the sampling ratio of 0.3 almost reaches that of 1.0, i.e., the cache hit ratio when the complete structure of the OSN is known. This achievement stems from that the structural property of the OSN is (moderately) maintained in the sampled network, namely, the sampled network contains influential users corresponding to nodes with high degree. In particular, in the case of RW sampling, the degree distribution of the sampled network is biased towards high degree due to the characteristics of RW, i.e., the RW agent tends to visit nodes with high degree.

I shall refer to the effect of the sampling strategy, although our experiments focus on how much effectively the sampling strategy works for estimating influential users for socially-aware ICNs rather than which sampling strategies is the best. Results of Last.fm (Fig. 5.2(a)) indicate that DFS among four types of sampling strategies used in this experiment achieves the best performance; on the other hand, those of Facebook (Fig. 5.2(b)) indicate that the difference due to crawl-based sampling strategies, i.e., RW sampling, DFS, and BFS, is marginal. This means that I can freely choose a sampling strategy among these crawl-based sampling strategies, while considering the sampling cost[3].

[3]Please refer to, for instance, [25] for details of the sampling cost.

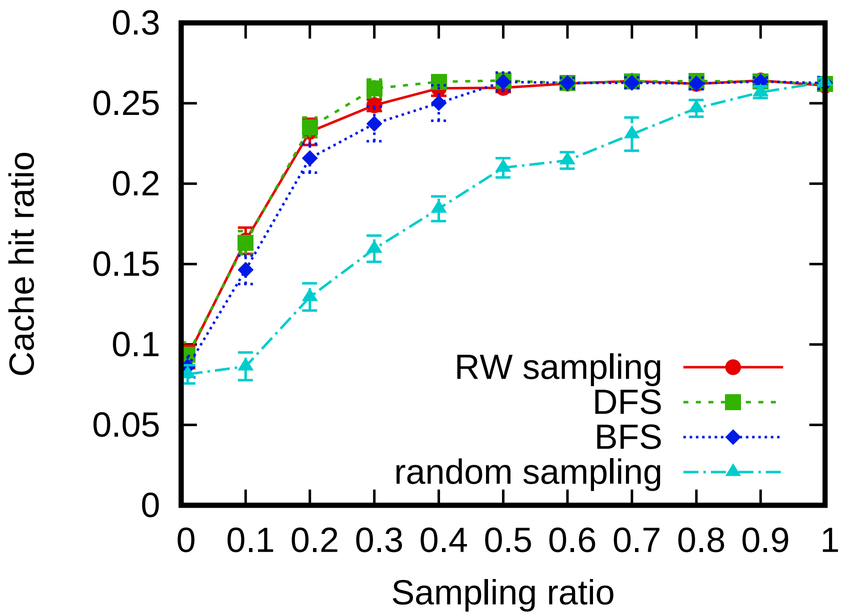


(a) Last.fm

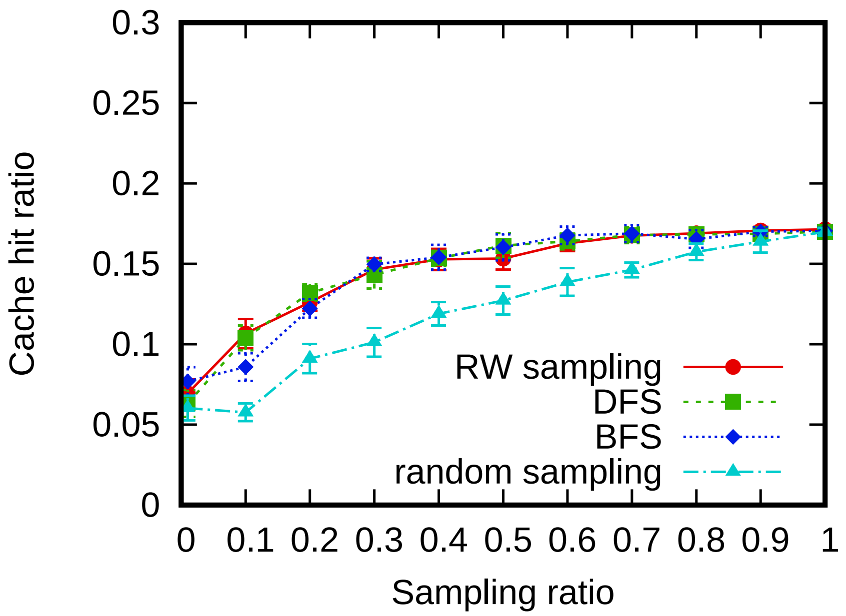


(b) Facebook

Figure 5.1: Average cache hit ratio with eight types of centrality measures



(a) Last.fm



(b) Facebook

Figure 5.2: Relationship between sampling ratio and cache hit ratio

5.1.2 Analysis

I approximately derive the cache hit probability for a given degree distribution of social network, and then analyze the relationship between the ratio of influential users and the cache hit probability. Section 5.1.1 focuses on the selection of influential users using centrality measures; the mathematical analysis focuses on the ratio of influential users on a simplified analytic model.

Analytic Model

In this paper, I consider content delivery between OSN users with social relationship on a cache network comsisting of multiple ICN routers.

I denote the number of users and the degree distribution of OSN representing social relationship among users as N and $P(k)$, respectively.

Each user is accommodated into a randomly-chosen ICN router, and it repeatedly issues a request to retrieve contents produced by its adjacent users in OSN. The content request issued by a user is forwarded through ICN routers, and the discovered content in the network is returned to the requesting user.

I assume that the caching strategy at an ICN router is SACS [5], so the ICN router caches contents only produced by influential users. In this paper, I select influential users based on a centrality of a user and on the ratio of influential users in OSN. Unlike Section 5.1.1, we focus on *degree centrality* as a centrality measure. Hence, we regard top $p\,N$ users with the highest degree in OSN as influential users, where $p\,(0 \leq p \leq 1)$ is the ratio of influential users. It is worth of noting that SACS involves the conventional caching strategy LCE (Leave Copy Everywhere) [6]. This is because SACS with the ratio of influential users p of 1 is equivalent to LCE, i.e., ICN routers uniformly cache contents without considering influential users.

The cache size at an ICN router is denoted as B [content]. Also, the cache-replacement policy is supposed to be LRU.

Derivation of Cache Hit Probability

In this paper, I derive the expected value of the cache hit probability, which is the probability that when a user requests content, the corresponding content is returned from any ICN routers along a path. I denote the expected value of the cache hit probability as H.

Because a user issues multiple requests to all adjacent users including influential users and non-influential users, the cache hit probability H is composed of twofold: the cache hit probability H^+ for contents produced by influential users and the cache hit probability H^- for contents produced by non-influential users.By letting q be the fraction of *influential neighbor* that is a set of influential users which is adjacent to a user, the cache hit probability H is given by

$$H = q\,H^+ + (1 - q)\,H^-. \tag{5.1}$$

Recall that I assume the caching strategy at an ICN router as SACS. Hence, ICN routers cache contents only produced by influential users and it never caches contents produced by non-influential users. Therefore, I obviously have $H^- = 0$ and I can rewrite Eq. (5.1) as

$$H = q\,H^+. \tag{5.2}$$

In the following, I first derive the fraction of influential neighbors q. The fraction of influential neighbors q is corresponding to the probability whether an adjacent user is influential user. To determine whether an adjacent user is belong to influential users, it must be satisfied that the degree of the adjacent user is larger than or equal to the minimum degree of influential users. Hence, by denoting the degree distribution of an adjacent user and the the minimum degree of influential users as $Q(k)$ and $k^+_{\min}$, respectively, the fraction of influential neighbors q is given as follows.

$$q = \sum_{k=k^+_{\min}}^{\infty} Q(k) \tag{5.3}$$

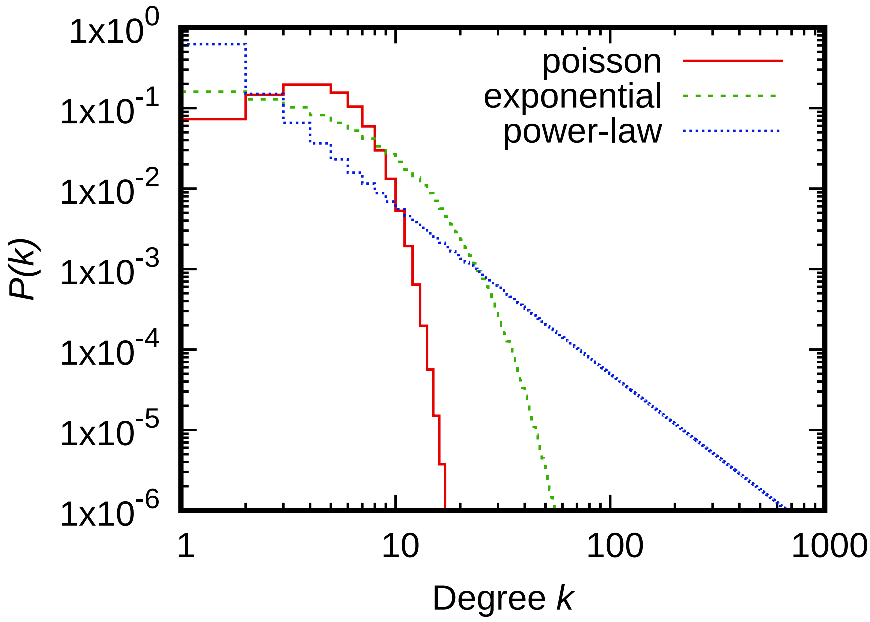


Figure 5.3: Degree distributions used in our numerical examples

In the above equation, degree distribution $Q(k)$ is known as an excess degree distribution, and Newman [18] formulates $Q(k)$ as follows.

$$Q(k) = \frac{(k+1)P(k+1)}{\langle k \rangle} \tag{5.4}$$

Here, $\langle k \rangle$ is the average degree of OSN and $\langle k \rangle = \sum_{k=1}^{\infty} k\,P(k)$. To derive the minimum degree of influential users, I use a cumulative distribution function $F(k)$ of degree distribution $P(k)$. From an inverse function of cumulative distribution function $F(k)$, I have

$$k^{+}_{\mathrm{min}} = F^{-1}(1-p). \tag{5.5}$$

Next, I derive cache hit probability H^{+} for contents produced by influential users. The cache hit probability H^{+} is the probability that when a user issues a content request, a cache hit occurs at any ICN routers along a path. By denoting the cache hit ratio at a single ICN router as h, H^{+} is given by

$$H^{+} = 1 - (1-h)^{\langle \ell \rangle + 1}, \tag{5.6}$$

where $\langle \ell \rangle$ is the average path length of a cache network comprising of ICN routers.

In the case of SACS, only $p\,N$ types of contents are candidate for cached contents within the cache network.

Hence, using cache size B at an ICN router, the cache hit probability h is approximately given by

$$h \simeq \min(\frac{B}{pN}, 1). \tag{5.7}$$

In the above equation, I implicitly assume that the cache size is sufficiently small compared to the number of users in OSN, i.e., $B \ll N$.

Numerical Example

Through several numerical examples, I investigate the relationship between the ratio of influential users and the cache hit probability.

The degree distribution of OSN is given by three types of probability mass function — Poisson, exponential, and power-law distributions [27].

- Poisson distribution

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{5.8}$$

- Exponential distribution

$$P(k) = (1 - e^{-\mu}) e^{-\mu k} \tag{5.9}$$

- Power-law distribution

$$P(k) = \frac{k^{-\alpha}}{\zeta(\alpha)} \tag{5.10}$$

In Eqs. (5.9) and (5.10), μ and α are parameters, and $\zeta(\alpha)$ is also Riemann's zeta function. We determined parameters μ and α such that the mean of the probability mass function is equal to average degree $\langle k \rangle$. In our numerical example, the number of users N in OSN was 10,000 and the average degree $\langle k \rangle$ was 4. I show degree distributions with three types of probability mass functions as Fig. 5.3. Note that the x-axis and the y-axis of this figure are logarithmic. This figure clarifies the difference in the three types of mass probability function. In particular, "power-law" exhibits difference tendency from "poisson" and "exponential"; namely, the probability that nodes with degree k exist $P(k)$ linearly decreases with the increase in the degree k. This is because, as Eq. (5.10) implies, the degree distribution is proportional to a power of the degree.

Unless explicitly stated, I used the following parameter settings: cache size at an ICN router $B = 100$ [content] and average path length of the cache network $\langle \ell \rangle = 3$.

Before discussing the relationship between the ratio of influential users and the cache hit probability, I confirm how the fraction of influential neighbors (Eq. (5.3)) is varied according to the degree distribution. Figure 5.4 shows the relationship between the ratio of influential users p and the fraction of influential neighbors q. In this figure, results with three degree distributions are plotted as "poisson", "exponential", and "power-law". From this figure, I can find that when the degree distribution follows power-law, the ratio of influential neighbors rapidly increases as the ratio of influential users increases because of the existence of a few users with extremely-high degree. Let us provide the rationale behind this tendency more specifically while focusing on the region where the ratio of influential users p is small. I recall that the ratio of influential neighbors q is dominated by two factors: $k^+_{\rm min}$ and the partial sum of $Q(k)$, i.e., $\sum_{k=k^+_{\rm min}}^{\infty} Q(k)$, as shown in Eq. (5.3). In the region of small p, the minimum degree of influential users $k^+_{\rm min}$ becomes relatively-large; this indicates that the ratio of influential neighbors q is determined by the partial sum of $Q(k)$ in the region of high degree. Based on the fact that $Q(k)$ is derived from the degree distribution $P(k)$ and that $P(k)$ of "power-law" is biased towards high degree, the partial sum of $Q(k)$ of "power-law" becomes large compared to "poisson" and "exponential". As a consequence, the ratio of influential neighbors q of only "power-law" exhibits rapid increase even though the ratio of influential users p is small.

Finally, as shown in Fig. 5.5, I present the relationship between the ratio of influential users and the cache hit probability. Similar to Fig. 5.4, results with three types of degree distributions are plotted in Fig. 5.5. From this figure, we can easily observe that the cache hit probability varies significantly according to the degree distribution of OSN. In the case of "poisson" and "exponential", there is almost no difference between the peak cache hit probability and the cache hit probability when $p = 1$. This is mainly due to the fact that, in the small regime of the ratio of influential users p, the ratio of influential neighbors q is not sufficiently increased; consequently, as Eq. (5.2) implies, the cache hit probability H remains low with respect to p. These results imply that the benefit to the content caching by selecting influential users is limited. In contrast, in the case of "power-law", the cache hit probability exhibits convex upward around $p = 0.02$ with respect to the ratio of influential users p. This is mainly caused by the following two reasons: (i) even if the ratio of influential users p is small, the fraction of influential neighbors becomes large as shown in Fig. 5.4, and (ii) restricting the number of influential users contributes to improve the cache hit probability (Eq. (5.7)) at an ICN router. From these observations, I conclude that by appropriately selecting the ratio of influential users in OSN, I can expect a significant improvement in the cache hit probability compared to the conventional caching strategy.

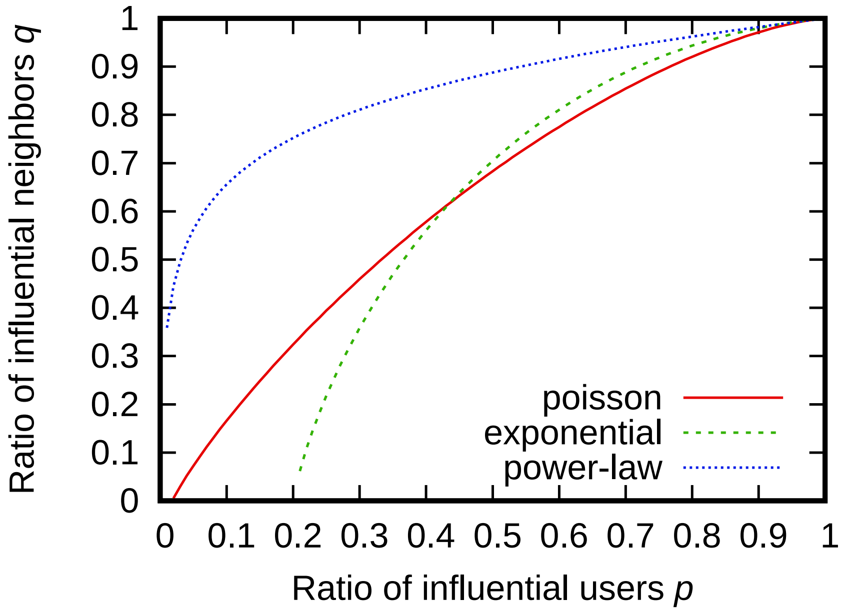


Figure 5.4: Relationship between ratio of influential users and ratio of influential neighbors

5.2 Information Propagation

In this section, I describe the methodology to investigate the effects of information propagation by users on the content caching of ICN.

5.2.1 Method

I generated a random graph with ER (Erdös-Rényi) model [26] and used it as the ICN network G_I. The number of nodes, i.e., routers, and the average degree of the ICN network are 25 and 3, respectively. A path between routers through which requests and contents are passing was set to a shortest path.

Also, I generated a scale-free network with community structure using Li-Maini model [28] and used it as the social network G_S. Recall that the social network represents friendship among users on social media. The number of nodes, i.e., users, and the average degree of the social network are 1,000 and about 12, respectively. I set the number of communities M, a parameter for the Li-Maini model, to 25. This is because I equalize the number of communities in the social network with the number of routers in the ICN network, The other parameters are as follows: $(t, m_0, m, \alpha, n) = (850, 6, 6, 0.05, 1)$. For the definitions of these mathematical symbols, please refer to [28].

As described in Section 4, I accommodated users of the social network to ICN routers based on their communities, and then transferred them to other routers according to the transfer probability. In our simulation, because the number of communities and that of ICN routers were equal, I mapped communities to ICN routers in a one-to-one manner.

Users on the social network perform the following behaviors: request, advertisement, and publication of contents, according to the user's behavior model described in Section 4 In this paper, I varied the propagation probability p_u from 0 to 0.1 regardless of the user. The propagation probability is a probability whether a user u advertises a content or not.

To focus on information propagation, I set the transfer probability to 1, which means all users are accommodated to a randomly-chosen router.

The setting of the cache of ICN routers is as follows. The cache size of a router was set to either 5 [content] or 10 [content]. As the caching strategy, I used LCE (Leave Copy Everywhere). LCE is a typical caching strategy; a router following LCE caches contents uniformly. Also, as the cache replacement policy, I used LRU (Least-Recently Used). LRU is a discipline that evicts the oldest-referred content from the cache.

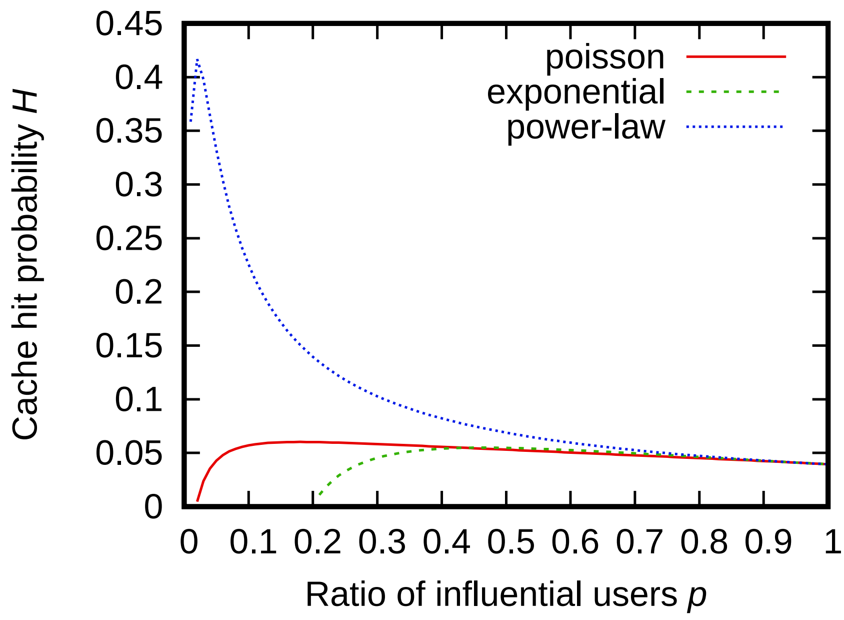


Figure 5.5: Relationship between ratio of influential users and cache hit probability

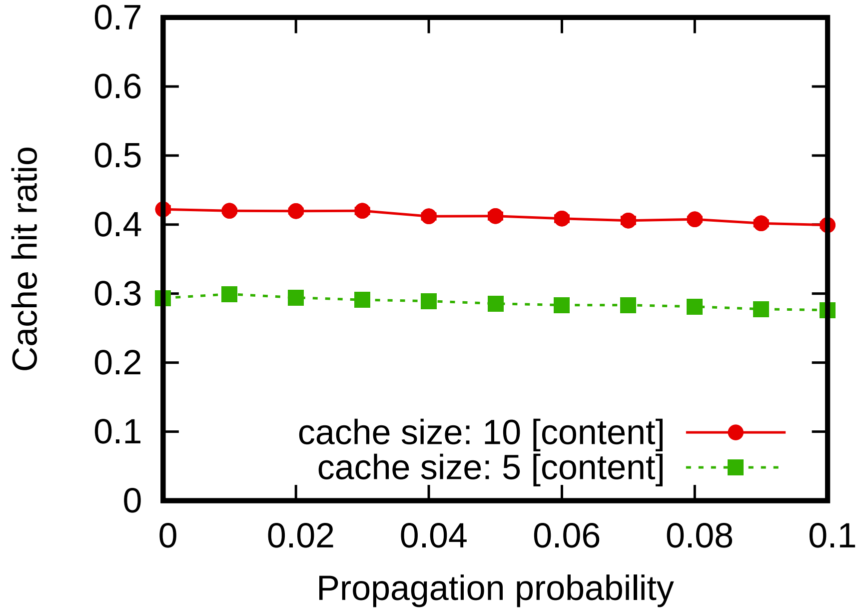
Performance metrics is twofold: the cache hit ratio and the path length for content delivery. The cache hit ratio is defined as ratio of the number of contents returned from caches on a path to the number of user-issued requests. In contrast, the path length for content delivery is the length of a path that a content traverses as a response to a user's request. Formally, it is defined as the length of a path from an ICN router accommodating a requesting user to that returning a corresponding content. The entity that returns contents is either the cache of an intermediate ICN router or an ICN router accommodating a content publisher.

Using ICNSIM (ICN SIMulator) developed by our research group, we repeated a single simulation with 100,000 [slot] 10 times for a given condition and obtained mean and 95% confidence interval of each performance metrics.

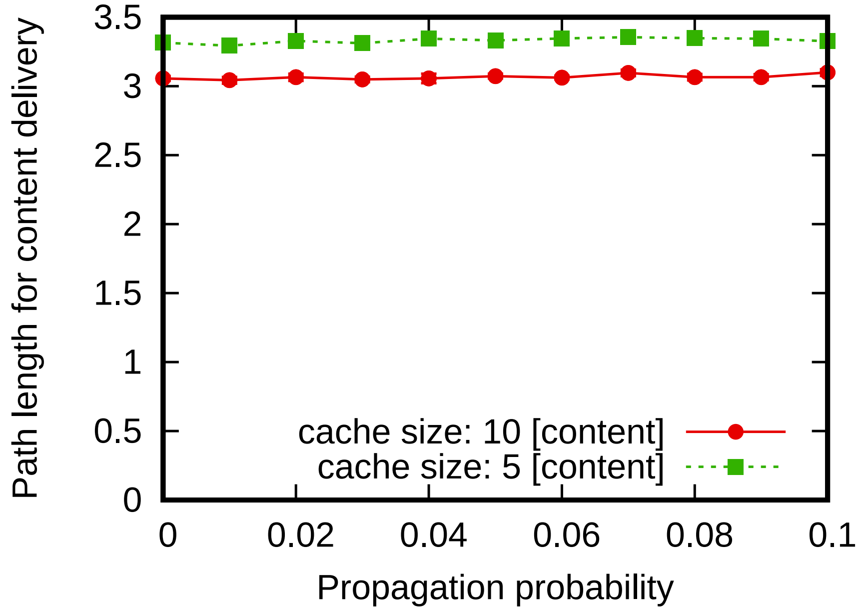
5.2.2 Result

I focus on the user's information propagation behavior and verify its effect on the cache hit ratio and the length of delivery path. Figure 5.6 shows the cache hit ratio and the length of delivery path when the propagation probability is varied from 0 to 0.1. In this figure, I plot results for the cache size at ICN routers of 5 [content] and 10 [content]. To make results interpretable, I intentionally accommodate users in ICN routers randomly, i.e., set the transfer probability to 1, so as to exclude the effect of community structure.

I can find from these results that even though contents are advertised by users, this effect on the content caching is marginal, i.e., the cache hit ratio and the length of delivery path are slightly degraded and also rarely affected. Intuitively, the information propagation behavior might degrade the effectiveness of caching in the ICN router. This is because a user requests contents not only published by its neighbors in the social network but also advertised by them; this makes types of contents retained within the network more diverse. In fact, however, the reason why results contradict our intention is explained with a phenomenon called *temporal locality* of content requests. Specifically speaking, I first recall the user behavior model used in this letter; on retrieving a content, a user consecutively transitions to advertise the content to its neighbors and then neighbors request it. Here, if the interval between the time a user advertises a content and the time its neighbors actually request it is short, the advertised content is expected to be still cached at any one of ICN routers in the network. As a consequence, the cache performance is not degraded.



(a) cache hit ratio



(b) path length for content delivery

Figure 5.6: Effect of information propagation by users

5.3 Community

In this section, I focus on community structure, which is one of the important features of social media, and investigate how community structure affects the cache hit ratio and the length of delivery path while the transfer probability, which adjusts the strength of community structure, is varied.

5.3.1 Method

Because the user's behavior model described in section 4 represents community structure, I use the same model to investigate the effects of community structure on the content caching in ICN. In the model, users are accommodated into a router based on the community of the social network, and according to transfer probability, users are transferred to randomly-chosen router. By varing the transfer probability, I change the strength of community structure, which means how much community structure of social network reflects to users' accommodation.

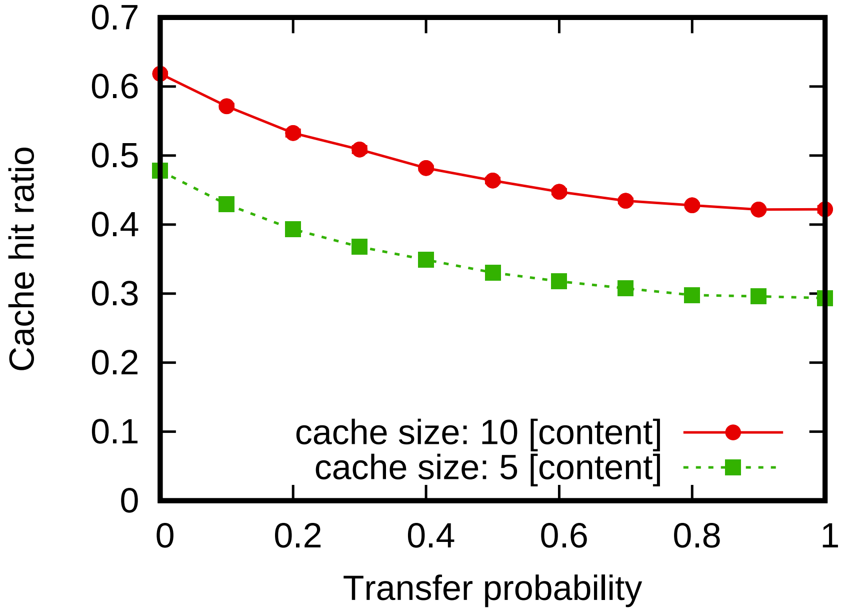
Other settings are the same as the simulation conducted in section 5.2. I use the same network as that in section 5.2. Also, in terms of caching method, caching strategy, cache replacement policy and cache size are set to the same parameter.

Using ICNSIM, I repeated a single simulation with 100,000[slot] 10 times for a given condition and obtained the mean and 95% confidence interval of each performance metric: cache hit ratio and the length of a path.

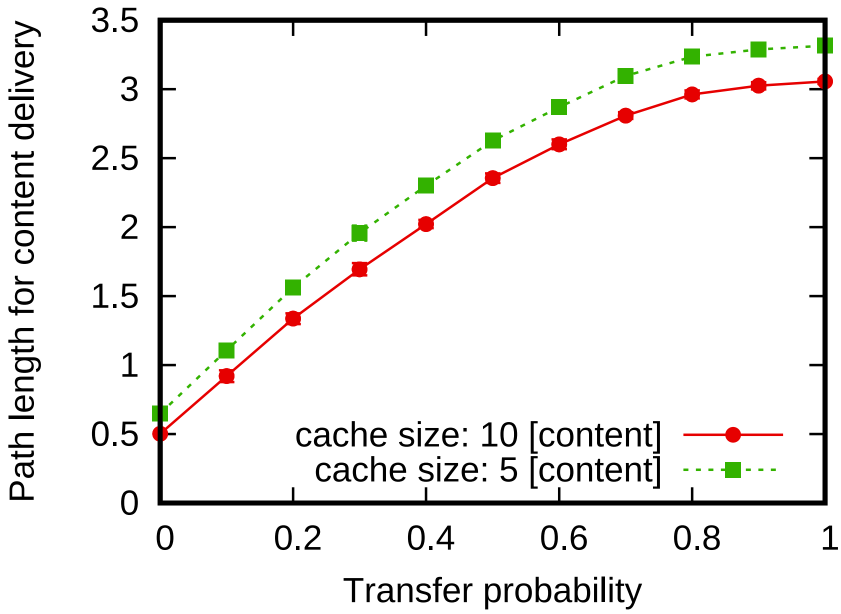
5.3.2 Result

I examine the effect of the user's community structure. Figure 5.7 shows the cache hit ratio and the length of delivery path while the transfer probability, which adjusts the strength of community structure, is varied. Similar to Fig. 5.6, results with different cache sizes are shown in this figure. I should note that I disable the user's information propagation behavior by setting the propagation probability to 0.

These results show that the cache hit ratio and the length of delivery path significantly differs in accordance with keeping the community structure. Let us look at results when users are accommodated while keeping the community structure, i.e., when the transfer probability is set to 0. In this case, the content delivery among users within a same router becomes more frequent than that among users accommodated in different routers, which improves the efficiency of the cache utilization and reduces the length of delivery path. But this tendency is weakened as the transfer probability increases, i.e., as the community structure is lost. This observation that the cache hit ratio and the length of delivery path is strongly affected by the transfer probability suggests the importance of considering the community structure as user's sociality.



(a) cache hit ratio



(b) path length for content delivery

Figure 5.7: Effect of community structure

Chapter 6

Conclusion

In this paper, I assumed that ICN is introduced as a network architecture for social media, and presented a conceptual architecture to realize ICN-based content distribution for social media. Furthermore, I modeled the ICN that takes into account the sociality of social media users, which is the existence of influential users, the information propagation behavior and community structure, by extending the existing user's behavior model. I conducted several simulation and mathematical analysis to reveal the effects of the sociality of social media users on the content caching of ICN. Specifically, I first investigated when I use SACS as a caching strategy for ICN, how the influencer should be decided and how many users should be regarded as influential users through simulation and mathematical analysis. As a result, I revealed the cache performance differs according to centrality measure. Furthermore, I investigate the effects of the other sociality of social network on content caching in ICN. Even though propagation probability by users on social media rarely affects the cache hit ratio and the length of delivery path, the community structure among users on social networks has positive effects on the content caching.

Acknowledgement

I am deeply grateful to my supervisors, Professor Noriaki Kamiyama and Instructor Ryo Nakamura, Fukuoka University. I would like to express my appreciation for their helpful, continuing and considerable support, which enabled me to write and finish my master's thesis.

I would also like to thank all lab members, with whom I spent almost all my lab life sharing our thoughts and ideas, discussing research, and helping each other.

References

[1] A. Vakali and G. Pallis, ”Content delivery networks: status and trends,” in IEEE Internet Computing, vol. 7, no. 6, pp. 68-74, Nov.-Dec. 2003, doi: 10.1109/MIC.2003.1250586. keywords: Network servers;Delay;File servers;Collaboration;Availability;Bandwidth;Telecommunication traffic;Terrorism;Throughput;Scalability,

[2] G. Xylomenos *et al.*, “A survey of information-centric networking research,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 1024–1049, 2014.

[3] S. Tsugawa, “A survey of social network analysis techniques and their applications to socially aware networking,” *IEICE Transactions on Communications*, vol. E102–B, pp. 17–39, Jan. 2019.

[4] F. Xia, L. Liu, J. Li, J. Ma, and A. V. Vasilakos, “Socially aware networking: A survey,” *IEEE Systems Journal*, vol. 9, pp. 904–921, Sept. 2015.

[5] C. Bernardini, T. Silverston, and O. Festor, “Socially-aware caching strategy for content centric networking,” in *Proceedings of the 2014 IFIP Networking Conference*, pp. 1–6, June 2014.

[6] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, “Networking named content,” in *Proceedings of the 5th International Conference on emerging Networking EXperiments and Technologies (CoNEXT 2009)*, pp. 1–12, Dec. 2009.

[7] Lixia Zhang, Alexander Afanasyev, Jeffrey Burke, Van Jacobson, kc claffy, Patrick Crowley, Christos Papadopoulos, Lan Wang, and Beichuan Zhang. 2014. Named data networking. SIGCOMM Comput. Commun. Rev. 44, 3 (July 2014), 66–73. https://doi.org/10.1145/2656877.2656887

[8] K. Zheng, Y. Cui, X. Liu, X. Wang, X. Jiang and J. Tian, ”Asymptotic Analysis of Inhomogeneous Information-Centric Wireless Networks With Infrastructure Support,” in IEEE Transactions on Vehicular Technology, vol. 67, no. 6, pp. 5245-5259, June 2018, doi: 10.1109/TVT.2018.2810312.

[9] H. Hu, Y. Wen, T. -S. Chua, J. Huang, W. Zhu and X. Li, ”Joint Content Replication and Request Routing for Social Video Distribution Over Cloud CDN: A Community Clustering Method,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 7, pp. 1320-1333, July 2016, doi: 10.1109/TCSVT.2015.2455712.

[10] Bertrand Mathieu and Patrick Truong: A CCN-based Social Network application optimising netowrk proximity, IEEE Conference on Computer Communications Workshops (INFOCOM Wksps) 2014

[11] Patrick Truong and Bertrand Mathieu and Jean-Francois Peltier: Networking Impact of a local-aware content-based delivery for twitter-like applications, 2015 18th International Conference on Intelligence in Next Generation Networks 2015

[12] A K M Mahmudul Hoque, Syed Obaid Amin, Adam Alyyan, Beichuan Zhang, Lixia Zhang, and Lan Wang. 2013. NLSR: named-data link state routing protocol. In Proceedings of the 3rd ACM SIGCOMM workshop on Information-centric networking (ICN ’13). Association for Computing Machinery, New York, NY, USA, 15–20. https://doi.org/10.1145/2491224.2491231

[13] Nikolaos Laoutaris, Hao Che, and Ioannis Stavrakakis. 2006. The LCD interconnection of LRU caches and its analysis. Perform. Eval. 63, 7 (July 2006), 609–634. https://doi.org/10.1016/j.peva.2005.05.003

[14] M. Mosko, I. Solis, and C. A. Wood, ”Content-Centric Networking(CCNx) Messages in TLV Format,” in Request for Comments (RFC) 8609, July 2019.

[15] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, “Characterizing user behavior in online social networks,” in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (IMC 2009)*, pp. 49–62, Nov. 2009.

[16] M. Zhang, H. Luo, and H. Zhang, “A survey of caching mechanisms in information-centric networking,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1473–1499, 2015.

[17] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD 2003)*, pp. 137–146, Aug. 2003.

[18] Newman, Mark, Networks: An Introduction, 1st edn (Oxford, 2010; online edn, Oxford Academic, 1 Sept. 2010), https://doi.org/10.1093/acprof:oso/9780199206650.001.0001, accessed 22 Jan. 2024.

[19] Seidman, Stephen. (1983). Network Structure And Minimum Degree.Soc Netw 5:269-287. Social Networks. 5. 269-287. 10.1016/0378-8733(83)90028-X.

[20] Zhang, JX., Chen, DB., Dong, Q. et al. Identifying a set of influential spreaders in complex networks. Sci Rep 6, 27823 (2016). https://doi.org/10.1038/srep27823

[21] Morone F, Makse HA. Influence maximization in complex networks through optimal percolation. Nature. 2015 Aug;524(7563):65-68. DOI: 10.1038/nature14604. PMID: 26131931.

[22] Arun S. Maiya and Tanya Y. Berger-Wolf. 2011. Benefits of bias: towards better characterization of network sampling. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD ’11). Association for Computing Machinery, New York, NY, USA, 105–113. https://doi.org/10.1145/2020408.2020431

[23] S. Tsugawa and H. Ohsaki, ”Benefits of Bias in Crawl-Based Network Sampling for Identifying Key Node Set,” in IEEE Access, vol. 8, pp. 75370-75380, 2020, doi: 10.1109/ACCESS.2020.2988910. k

[24] László, Lovász. (1993). Random Walks on Graphs: A Survey, Combinatorics, Paul Erdos is Eighty. Bolyai Soc. Math. Stud.. 2.

[25] M. Coscia and L. Rossi, ”Benchmarking API Costs of Network Sampling Strategies,” 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 663-672, doi: 10.1109/BigData.2018.8622486.

[26] P. Erdös and A. Rényi, “On random graphs I.,” *Mathematicae*, vol. 6, pp. 290–297, Nov. 1959.

[27] N. Kawabata, Y. Yamasaki and H. Ohsaki, ”Modeling Restrained Epidemic Routing on Complex Networks,” 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), Milwaukee, WI, USA, 2019, pp. 285-290, doi: 10.1109/COMPSAC.2019.00049.

[28] C. Li and P. K. Maini, “An evolving network model with community structure,” *Journal of Physics A: Mathematical and General*, vol. 38, pp. 9741–9749, Oct. 2005.

Published Works

[1] 本田紘大，中村遼，上山憲昭，ソーシャル性を考慮した ICN のためのキャッシュ制御方式に関する一検討，電子情報通信学会ソサイエティ大会，2021 年 9 月

[2] 本田紘大，中村遼，上山憲昭，ソーシャル性を考慮した情報指向ネットワークの特性分析，電子情報通信学会インターネットアーキテクチャ(IA) 研究会，2022 年 3 月

[3] 本田紘大，中村遼，上山憲昭，ソーシャルネットワークのトポロジ構造が ICN におけるコンテンツキャッシングに与える影響に関する一検討，電子情報通信学会総合大会，2022 年 3 月

[4] 本田紘大，中村遼，上山憲昭，ソーシャル性を考慮した情報指向ネットワークにおける影響力の強いユーザの推定に関する一検討，電子情報通信学会ソサイエティ大会，2022 年 9 月

[5] 本田紘大，中村遼，上山憲昭，ソーシャルメディア上のユーザによる情報拡散行動を考慮した ICN におけるコンテンツ配送モデルに関する一検討，電子情報通信学会総合大会，2023 年 3 月

[6] K. Honda, R. Nakamura and N. Kamiyama, "Characteristic Analysis of Socially-Aware Information-Centric Networking," NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium, Poster Session, Miami, FL, USA, 2023, pp. 1-5, doi: 10.1109/NOMS56928.2023.10154426.

[7] 本田紘大，中村遼，上山憲昭，ソーシャルメディア上のユーザによる情報拡散行動が ICN のコンテンツキャッシングに与える影響に関する一検討，電子情報通信学会ソサイエティ大会，2023 年 9 月

[8] 本田紘大，中村遼，上山憲昭，ユーザのソーシャル性が ICN のコンテンツキャッシングに及ぼす影響–情報拡散行動とコミュニティに着目して–，電子情報通信学会コミュニケーションクオリティ(CQ) 研究会，2023 年 11 月

[9] 本田紘大，中村遼，上山憲昭，ユーザのソーシャル性が ICN のコンテンツキャッシングに及ぼす影響に関する一検討–情報拡散行動とコミュニティに着目して–，電子情報通信学会情報指向ネットワーク (ICN) 研究会, 2023 年 12 月

[10] K. Honda, R. Nakamura and N. Kamiyama, "Analyzing Effects of Social Media User's Influence on Contents Caching in ICN," in IEEE Access, vol. 11, pp. 127679-127688, 2023, doi: 10.1109/ACCESS.2023.3330850.

[11] K. Honda, R. Nakamura and N. Kamiyama, "A Preliminary Study on the Impact of User's Sociality on Content Caching of ICN," accepted in IEICE Communications Express, Jan. 2024