Repurposing FM: Radio Nowhere to OSNs Everywhere

Morgan Vigil
Dept. of Computer Science
UC Santa Barbara
Santa Barbara, CA
mvigil@cs.ucsb.edu

Elizabeth Belding
Dept. of Computer Science
UC Santa Barbara
Santa Barbara, CA
ebelding@cs.ucsb.edu

Matthew Rantanen
Southern California Tribal
Chairman's Association
Pala, CA
mrantanen@sctdv.net

ABSTRACT
While online social networks (OSNs) play a critical role in developing social capital [18], many communities are unable to utilize the benefits of OSNs due to lack of Internet accessibility. In this paper, we investigate the feasibility of the Radio Broadcast Data System (RBDS) associated with FM radio stations as a means to deliver social network content to OSN users who do not have access to Internet services. Using Instagram as a case study, we analyze data from 254 public Instagram users associated with the Tribal Digital Village (TDV) network in Southern California. Our analysis of over 1.2 million unique Instagram posts reveals that Instagram users in the TDV network interact with locally generated content $46.6\times$ more often than content generated by users from outside the network. We use our observations of OSN usage to compare five OSN content scheduling approaches. Our evaluation reveals that up to 81% of users received at least half of their content requests and 35.5% of the 1.1 million requested Instagram photos were transmitted to users.

ACM Classification Keywords
H.3 Information Storage and Retrieval: Information Search and Retrieval; H.4 Information Systems Applications: Miscellaneous

Author Keywords
Native American, RBDS, FM radio, OSN, social network analysis, locality

INTRODUCTION
By the end of 2014, only 58% of the U.S. American Indian/Alaska Native population [3] and only 10% of Native Americans living on reservations had Internet access [4]. This divide is not unique to North America–indigenous people around the globe are facing digital marginalization [8]. For many disconnected indigenous communities, lack of Internet access is largely determined by geographic obstacles, complex telecommunications policies, and socio-economic factors [16]. Several studies of networks that bring initial Internet services to communities have found that geographic remoteness and rugged geographic features comprise some of the major barriers to infrastructural deployments [8, 41, 16].

In order to bridge the information gap, many Native Americans living on reservations rely on FM radio to receive culturally relevant programming, emergency notifications, local news, and information on civic participation, health, and economics [8, 7, 38]. While FM radio operates at frequencies that have excellent penetration rates for distributing content over long distances and can be received by end users at little to no cost, it is not supportive of real-time, interactive, media-rich content that is pervasive on the web today.

Studies of indigenous broadband usage indicate that when Internet services are available, online social networks (OSNs) account for the most frequently accessed sites [10, 3, 34, 41]. While users cite reasons for using social media that are similar to the general population [39, 3], there is also a distinct emphasis on the use of OSNs for building cultural resilience through local social connections, staying connected to geographically distant friends and family [10, 34], and receiving news from alternative media sources [8, 23].

In this work, we propose and evaluate a content dissemination system that leverages the data subcarrier of FM radio transmissions (RBDS) in order to broadcast digital content to co-located communities of users who do not have access to point-to-point communications infrastructure due to issues of affordability or availability (or both). Based on the reported information needs of indigenous communities [8, 16, 10] and findings from previous studies of indigenous web usage [10, 34, 41], we find that content from online social networks (OSNs) is critical to indigenous efforts towards cultural revitalization and building community connections. Thus, our proposed system focuses on the efficient dissemination of OSN content in particular.

We describe the specific details of our contributions below.

Paper contributions
In order to assess the feasibility of OSN content delivery over RBDS, we use Instagram data generated by users of a Native American Tribal-operated WISP (wireless Internet service provider) in Southern California. This WISP provides Internet access to 17 different tribes located across 13 different reservations. In assessing the design and feasibility of our proposed system, we are guided by three research questions:

RQ1: How does a set of co-located users interact with OSN content and what implications does this activity have on bandwidth requirements?

CSCW '16, February 27-March 02, 2016, San Francisco, CA, USA 2016 ACM. ISBN 978-1-4503-3592-8/16/02...$15.00
DOI: http://dx.doi.org/10.1145/2818048.2820001

RQ2: How much overlap exists between the social networks of users who are from the same geographic community?

RQ3: Can we leverage social network relationships to reduce overall bandwidth usage while still providing adequate content coverage?

Previous studies of indigenous technology usage suggest that access to OSN content is helpful for building indigenous cultural identity, that there is limited empirical knowledge about North American indigenous OSN usage, and that there is a continuing digital divide affecting indigenous peoples living in remote and rural areas. In the interest of providing more people with access to these benefits, our work delivers three major contributions:

- We empirically characterize OSN usage and social interaction using Instagram data collected from a Tribal operated WISP. Our findings represent some of the first empirical evidence of strong locality of interest expressed through indigenous OSN usage.
- We propose a content distribution system that would provide access to OSN content without requiring access to pervasive broadband or cellular infrastructure.
- We evaluate the feasibility of such a system using Instagram data generated by a diverse set of users who access the Internet through a Tribal operated WISP. We demonstrate that it is possible to take advantage of social connections between co-located users in order to reduce high bandwidth content so that *users can receive the most relevant content via technology that is already ubiquitously available*.

The rest of this paper is organized as follows. First, we begin with a discussion of related work. Second, we provide an overview of a proposed OSN-to-RBDS system. Third, we describe the data sets used to assess the feasibility of the proposed system. Next, we analyze the data sets using research questions RQ1 and RQ2 as guides. We then introduce and evaluate four content scheduling approaches that could be used with the proposed system in order to address research question RQ3. Based on our evaluation of content scheduling approaches with respect to patterns identified in RQ1 and RQ2, we discuss the transferability of our study to other indigenous communities not represented by our data sets and conclude with a discussion of the system's impact on the CSCW research agenda.

RELATED WORK

Locality

A significant body of work has explored the relationship between social networks and the distance between the people that comprise a social network [33, 32, 30, 21]. There are two common approaches to understanding the correlation between social relationships among users and distance between users. Our work is most similar to work that explores themes of distance in terms of geography [32, 30, 21]. Given our specific interest in co-located communities, we focus more on work that explores the geographic relationship between users with respect to their social connections. While previous work has focused on characterizing the probability that two users are geographic friends based on online friendships [30], and the average geographic distance between friends in a social network based on the population density of communities where users live [21], we are interested in approaching the relationship between social closeness and geographical closeness from the reverse direction.

When looking specifically at locality as it relates to rural content interests, there have been numerous studies that reveal that locality is an important factor in determining the social relationships and content interests of rural users when it comes to their engagement with OSNs. Gilbert et al. examine the MySpace accounts of urban and rural users in the U.S. and determine that in general, rural users live 2.3× closer to their OSN friends than urban users [21]. Studies of specific rural [27, 28] and indigenous communities [34, 10, 41] have revealed that the strongest social ties of rural OSN users are with friends who live within their own geographic community. Moreover, when examining the content with which specific rural [22, 12, 28, 44] and indigenous [41] communities engaged, research found a high locality of content interest–with many users accessing locally generated content and locally viewed content.

While survey studies of indigenous content interests have suggested a strong locality of interest with respect to content, there have been very few empirical studies that corroborate this implication. Our work characterizes social network usage of Instagram users that originate from 17 different Native American tribes located across 13 reservations serviced by the TDV network in San Diego County. While our empirical analysis does not provide comprehensive coverage of all indigenous peoples, our findings reflect similarities between the OSN relationships found in the studied Native American community and those identified in previously studied rural communities. We discuss this in greater depth in the "Generalizability" section of this paper.

Content delivery systems

In essence, the system we propose to broadcast OSN content over RBDS is a content delivery system operating over constrained bandwidth. This is related to work exploring two other classes of content delivery systems–high traffic content delivery and content delivery over extremely low bandwidth. In both types of content delivery systems, a common bottleneck is located at the "last mile" network link that connects users to content [14, 19]. For both scenarios, numerous principles have been shown to improve the user experience, leading to faster delivery times and optimization of bandwidth resources.

In high traffic content delivery systems (including web content over HTTP and video over P2P networks), common solutions for avoiding bottlenecks include processes of monitoring network conditions to pinpoint congestion and failure points [14], identifying frequently requested content and replicating it to appropriate replication servers [14, 36], balancing content requests across a large number of servers containing replicated content [14], localized content caching [36], and splitting content [14, 11, 43, 29]. Our work is also similar to content delivery systems that allocate network resources to reflect the activities and preferences of system users [43, 29]. Principles for improving the user experience for content delivery systems designed for operation in low-bandwidth environments include content localization with distributed storage systems [28, 37] and web caching [36, 15, 12], transaction pre-scheduling [28, 13], and transmission scheduling based on content properties [9, 24].

Our work most closely resembles that of Ioannidis et al. and Zhao et al. Like Ioannidis et al., we focus on allocating bandwidth in order to maximize the fairness of content distribution across a community of users [24]. In contrast to this prior work, we allocate bandwidth specifically for OSN content. This allows us to take advantage of social relationships inherent to OSNs that can work as a heuristic for determining which content is most relevant to a co-located group of users. To the best of our knowledge, there has not been a content delivery system designed specifically for OSN data. Moreover, our work focuses on delivering OSN content via an extremely low-bandwidth broadcast technology. In order to maximize the ratio of relevant content to the bandwidth resources dedicated to an entire community, our work accounts for content relevance to a community of users when scheduling resources. In addition, our work is similar to that of Zhao et al. in that we determine a part of a larger collection of content that will be broadcast based on *a priori* information about users preferences [43]. While Zhao et al. focus on delivering video segments based on explicit user preferences, we rely on implied user preferences based on a record of their OSN interactions.

OSN CONTENT OVER RBDS

Leveraging the digital transfer capabilities of FM frequencies has not been explored in the context of bridging the digital divide. The FM radio broadcast data system (RBDS)[1] is a communications protocol that enables FM broadcast stations to embed small amounts of digital information into conventional FM radio broadcasts. RBDS functions as a subcarrier on the main FM radio transmission–its most familiar application is to carry digital information about current FM broadcasts (e.g. song title, artist name, station name) to display to listeners. While it operates over extremely low bandwidth (1.1875 kbps), it has many properties that are appealing for data transmission to areas lacking Internet infrastructure, including: large geographic coverage footprints [35], robustness to error [35], and operational infrastructure [6, 7]. While the bandwidth limitations of the protocol are prohibitive to certain types of data services (e.g. streaming media, interactive applications, real-time services), RBDS does have the potential to provide access to OSN content to areas where FM broadcast stations function as the sole arbiters of media content. This work is the first to characterize the feasibility of RBDS to provide data services to disconnected communities.

It is important to note that use of RBDS to deliver information from OSNs could be implemented in many ways, given the availability of FM receivers. In order to simplify our approach to understanding the feasibility of RBDS for OSN content delivery, we assume the architecture shown in Figure 1. Such an architecture would be able to actively download public OSN content and then use a protocol such as RDS-Link [35] to broadcast data to remote, offline repositories where users could access content over a local wireless network.

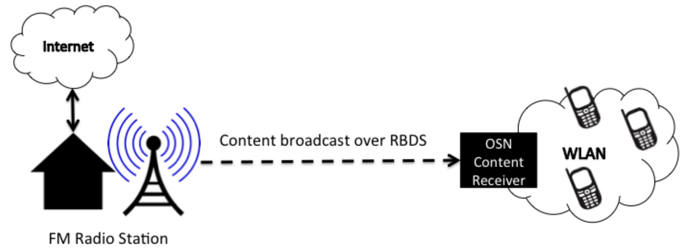


Figure 1: Proposed system for the propagation of OSN content over RBDS.

Our proposed system focuses on delivering OSN content to rural indigenous users over the FM radio broadcast data system. The focus on OSN content is motivated by studies of indigenous ICT usage which have shown OSNs to be major contributors to indigenous cultural revitalization and resilience. Prior surveys have revealed that social networking sites are the most popular means of everyday communication in the Sioux Lookout community [10, 34, 20]. Moreover, these studies have also found that OSNs play a potentially powerful role in indigenous cultural preservation and revitalization. Many surveyed users reported using OSNs to post information about upcoming cultural events [10, 34], to find information about cultural events and look at photos and videos of the events [10], to look at pictures of family and ancestral land [20], to provide and seek support for patients participating in indigenous telemedicine programs [10], to participate in political and social movements advocating for indigenous rights and justice [8, 31], and to communicate in their native language [10]. The most popular OSN may differ from community to community; in some of the studied indigenous communities Instagram is the most popular OSN, in others it is Facebook, and in still others, it is MyKnet[2]. We use Instagram to investigate the feasibility of delivering OSN content over RBDS and we believe that the general principles will hold for the delivery of content from any OSN. More important than the specific OSN is providing access to OSN content and the benefits it can provide to indigenous communities. Our OSN content delivery system over RBDS enables more people living in rural and remote areas without broadband coverage to gain access to this beneficial content *using a technology and infrastructure that is already ubiquitous: FM radio.*

We also note that our proposed system is not limited to delivering OSN content. In rural areas, it is still difficult to make reliable emergency announcements without ubiquitous cellular or telephone coverage. In cases of severe weather, natural disasters, and water contamination, our proposed system

[1] Radio Broadcast Data System in the U.S. and Radio Data System (RDS) internationally. `http://en.wikipedia.org/wiki/Radio_Data_System`

[2] A First Nations owned OSN based in northern Ontario.

could provide information about safety measures, evacuation plans, and access to alternative resources.

DESCRIPTION OF DATA SETS

The data used in this paper was collected as part of an ongoing study of the Tribal Digital Village (TDV) network that seeks to understand Tribal broadband use in order to create innovative technologies that bridge the digital divide on Tribal land. The TDV network is a wireless ISP operated by the Southern California Tribal Chairmen's Association in San Diego County. Established in 2001, the TDV network was created to bring Internet services to homes located on 17 of its member reservations. The TDV gateway is a 500 Mbps fiber link at the Pala reservation. This connectivity is extended over 11 and 18 GHz microwave backbone links and disseminated into residencies via 2.4 and 5 GHz WiFi. When it was established, there were no Internet providers in the area. Today, in spite of efforts towards mobile broadband deployment in the area, the TDV network continues to be the only provider available to the majority of residents in the area [1]. At the time of our study, 354 homes located across six different reservations utilized Internet services provided by the TDV network. For users who are unable to access the Internet using TDV's residential services[3], free Internet access is available at schools, health clinics, libraries, learning centers, and Tribal offices located on the reservations. Overall, our data set represents users who connect to the Internet from both private and public access points across several reservations in San Diego County in a variety of different contexts.

Our initial study [41] of TDV web traffic and network performance between June and August 2014 revealed that Instagram traffic accounted for the largest number of web requests and that both Instagram and Facebook traffic were among the top ten applications with respect to traffic volume. Closer inspection of Instagram traffic in the TDV network revealed that the most engaging Instagram media (with respect to local engagement) circulating through the TDV network was generated by local users. Similarly, locally created Instagram media circulated for longer periods through the network and social connections between TDV Instagram users were stronger than social connections between TDV Instagram users and Instagram users from outside TDV.

By the end of 2014, Instagram was among the five most popular OSN platforms in the U.S., with 21% of the entire adult population owning an Instagram account [17]. After Facebook, Instagram has the highest number of users between ages 18-29 and has the highest level of daily engagement (49% of users report interacting with their Instagram account on a daily basis). Moreover, having an Instagram account made users more likely to have accounts on other OSN platform than any other type of account. Considering the general popularity of Instagram in the U.S., the frequency of engagement with the Instagram platform, and our observations of Instagram's popularity in the TDV network, we use Instagram as the lens through which to understand the feasibility of using co-located social structures to enable the transfer of OSN content via RBDS.

[3]Lack of residential access is typically due to the inability to pay monthly service bills or geographic remoteness from the main backbone and access towers of the TDV network infrastructure. These challenges are not unique to the TDV network. Tribal ISPs often face the challenge of connecting widely dispersed users living in remote regions with geographic features prohibitive to wireless signal penetration [16].

Data collection steps

In this work, we discuss observations made from data collected between June 23 and December 18, 2014. Over this period we collected 16.5 TB of data representing 137.2 billion packets. We use the Bro Network Security Monitor (Bro) to collect flow-level statistics for network applications [2]. For this work, we are specifically interested in HTTP Bro logs. These data logs include source and destination IP addresses, timestamp information, HTTP host, HTTP URL, flow size, and flow duration. Our point of collection is located at the Internet gateway of the TDV network. We collect data by attaching a traffic monitoring server to the switch that bridges the gateway and the TDV network. A mirror port is configured to capture all packets traversing the network. Data collection methodology and subsequent analysis received approval from UCSB IRB and the Southern California Tribal Chairmen's Association prior to collection. All IP addresses are anonymized using traceanon [5].

We use the Instagram API to identify which posts users can access during our collection period. We identified 254 public Instagram user identifiers (unique number corresponding to an Instagram account) via the URL field in the HTTP headers captured by the Bro web logs. For the remainder of this paper, we refer to these public users as TDV users. Because 99% of network traffic is produced at residential access points, we assume that all Instagram users we observe in the network are residents of the 13 reservations serviced by the TDV network. Using the Instagram API, we are able to use public Instagram user identifiers of users in the TDV network to identify the OSN objects they had subscribed to during our collection period. We collect this data in two steps. First, we identify the public user identifiers of public users followed by the TDV users. Users who are followed are called *content creators* for the remainder of this paper. Second, we identify all of the meta data associated with the public media posted by the content creators during our period of interest, including: the media identifier and the creation time of the OSN content. By combining the data collected in these two steps, we generate a final data set that includes: the user identifier of a TDV user, the user identifier of the content creator, the media identifier of the OSN content created by the content creator, and the time the OSN content was created.

In addition, we use "instagram.com/api/v1/media" tags in the HTTP URLs collected in the HTTP Bro logs to identify TDV Instagram users' interactions with Instagram content. These logs include: the timestamp associated with interaction, the media identifier of the OSN content that was interacted with, and the action ("like" or "comment") taken on the Instagram content. By coupling this activity data set with the data set representing the public Instagram content available to each user, we are able to create a combined data set that includes: the timestamp of when Instagram content was published, the

	Total	Local	% Local
Content objects	1,209,270	3,777	0.31
Content creators	47,645	184	0.39
Social interactions	12,615	1,607	12.7
Instagram users	NA	254	NA

Table 1: Overview of content interactions observed between June 23 and December 18, 2014.

Term	Definition
locality	Where in the network with respect to the Internet gateway a piece of content originates. Local content originates within the local subnetwork behind the Internet gateway and non-local content originates from beyond the Internet gateway.
publication time	Time an OSN content object was published to an OSN platform.
coverage	Portion of users who have access to a particular piece of content.
stimulated	OSN content object that has received likes or comments from TDV Instagram users.
dormant	OSN content object that has not received likes or comments from TDV Instagram users.
follow network	All the content creators that a particular user follows, meaning all the content creators from whom a particular user receives OSN content.

Table 2: Definition of terms used to describe properties of OSN content.

user identifier of the Instagram content creator, the identifier of the Instagram content object, the proportion of co-located users who have subscribed to the OSN content, and the number of times co-located users interact with the Instagram content.

ANALYSIS

An overview of the data set we analyze is presented in Table 1. We report on the total data set that includes both content generated within the TDV network ("local") and content generated outside the TDV network. Values that are labeled "Local" refer to content objects that were created by content creators from the TDV network. "Content objects" refer to the number of OSN content objects that were published by the content creators followed by public Instagram users in the TDV network. "Social interactions" refer to the number of TDV users' likes and comments on followed Instagram content. Both images and short videos can be published on the Instagram platform; overall, 93.6% of the followed content objects were images and 6.4% were videos.

Interaction patterns

To address our three research questions, we begin by examining interaction patterns with respect to the *locality* of OSN content (RQ1), *publication time* of OSN content (RQ1), the degree of *coverage* provided by OSN content (RQ2), and which OSN content is *stimulated* or *dormant* (RQ2). Definitions of the terms are provided in Table 2.

Reservation	Instagram users	Pop. size
Pala	71	1,573
Rincon	61	1,495
San Pasqual	54	752
Mesa Grande	48	75
Pauma	17	186
Manzanita	3	69

Table 3: Population[4] and Instagram statistics for six TDV reservations.

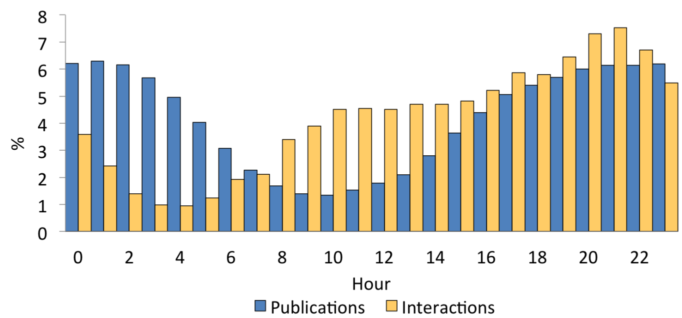


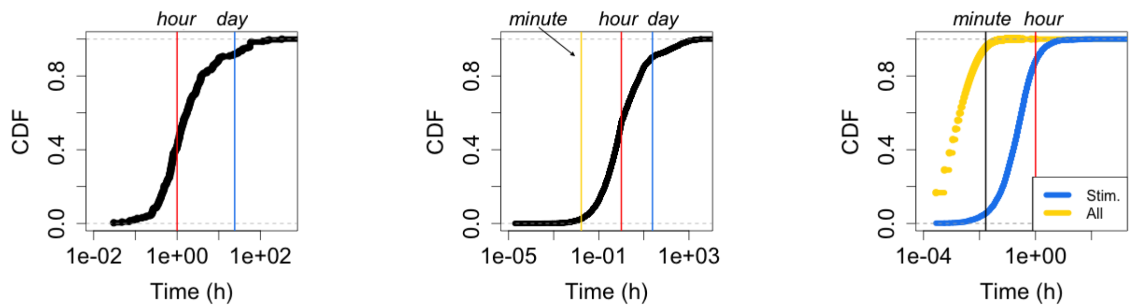
Figure 2: Percentage of overall activity that occurs per hour.

Locality

Table 1 provides an overview of the OSN content objects that were visible to TDV Instagram users during our observation period. We find that less than 0.31% of content objects were created by TDV Instagram (local) users. However, on average, TDV Instagram users interact with locally generated OSN content objects 46.6× more often than non-locally created OSN objects, demonstrating a strong preference for locally created content.

Next, we examine locality with respect to users' *follow networks*, or the content creators each user follows. To analyze the impact of locality at finer granularity, we examine interactions between Instagram users connecting from the same reservation. Table 3 provides an overview of the number of public Instagram users we observe at each reservation and the corresponding census population. By mapping Instagram users to their corresponding reservation, we are able to calculate the portion of each user's follow network that is comprised of content creators from the same reservation, a different reservation in the TDV network, or outside the TDV network. Given that the majority of observed content creators are from outside the TDV network, it is unsurprising that 97% of users have follow networks that contain more content creators from outside the TDV network than from inside the TDV network. When considering only TDV users that comprise a given user's follow network, the median percentage of content creators from the same reservation is 2.9× the median percentage of content creators from another reservation in the TDV network. This indicates the potential impact that community boundaries have on the success of sharing limited bandwidth. The broader the boundaries that define a "co-located" group of users, the less likely they are able to capture meaningful connections between that set of users.

[4]Population sizes from the 2010 U.S. Census. **http://www.sandiego.edu/nativeamerican/reservations.php**



(a) (b) (c)

Figure 3: Distribution of the time interval between (a) received OSN content objects, (b) publication time and time of first interaction, and (c) publication times between incoming OSN content objects.

Publication time

By understanding how frequently OSN content is published, received, and interacted with, we develop a temporal sense for how frequently OSN content broadcasts would need to occur in order to sustain a relevant information flow across a shared low-bandwidth link. While we observe significantly more publications than interactions (see Table 1), we normalize the frequency of publication and interaction activities by reporting the percentage of activity that occurs per hour in Figure 2. We note that the percentage of activity is based on all activities observed during our six month period of interest. As shown in Figure 2, the majority of publications occur during the evening and early morning, while the frequency of interactions is consistent with daily web traffic patterns observed in [41].

To understand post frequency from the perspective of the user, we examine the distribution of the average number of hours between each user's reception of a new OSN post in Figure 3a. This relates to how frequently a user's newsfeed is updated with new content. The median frequency of reception experienced by users is one received post per 1.23 hours; 40% of users receive a new post more frequently than once per hour. We observe the number of hours between the initial publication of a post and its first TDV-originated interaction (like or comment made on the post by a TDV Instagram user) in Figure 3b. The mean number of hours between a post's publication and the initial interaction on the TDV network is 1.04 hours; 90.2% of posts that are interacted with receive their initial interaction within 24 hours of publication. We take these values into consideration when evaluating potential approaches for scheduling OSN content broadcasts over RBDS.

In order to conceptualize how the rate of OSN content publication would impact the bandwidth required to update TDV OSN users in real-time, we examine the distribution of time intervals between all viewable posts assuming a naive First Come, First Serve (FCFS) approach. In Figure 3c, we plot the time interval between all unique viewable content in blue (note the logarithmic x-axis). We find that the rate of content publication translates into a median interval of 2 seconds between OSN content posts that would traverse a shared link. Depending on the size of the post, the observed data rate could require bandwidth capacities of 2.24 kbps for the most bandwidth-light OSN platforms (Twitter) or beyond 80 kbps for media-based OSN platforms (Instagram), assuming an average data size of 560 bytes and 20 KB, respectively. Since the RBDS protocol operates at only 1.1875 kbps, transferring all OSN content on the current FCFS basis is clearly infeasible. However, there are reasons to remain optimistic about transferring OSN content to disconnected communities via RBDS. Only 0.55% of all viewable published content receives interactions from TDV users. We plot the distribution of time intervals between the publication times of stimulated OSN content posts using FCFS scheduling in Figure 3c as a yellow line. For the stimulated OSN content objects, the median interval between publication times of OSN content is 13.66 minutes. This rate of publication translates into bandwidth requirements of 0.00546 kbps for the most bandwidth-light platforms (Twitter) and approximately 0.195 kbps for media-based OSN platforms (Instagram). The lower data rates suggest the feasibility of transmitting the most relevant media over RBDS. The challenge lies in the identification of relevant media, which we discuss in the following section.

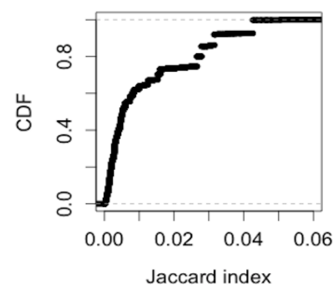


Figure 4: Distribution of the Jaccard similarity indices associated with pairwise comparisons of each user's follow network.

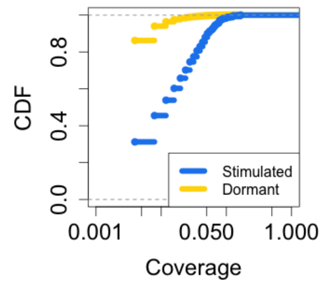


Figure 5: Distribution of coverage provided by stimulated and dormant content.

Coverage

We next investigate whether the high locality of interest identified in our previous study of the TDV network will be reflected in the amount of overlap observed between TDV Instagram users' follow networks. Using the Jaccard index [25],

$$jaccard(A, B) = \frac{|A \bigcup B|}{|A \bigcap B|} \quad (1)$$

where A is the set of users that comprise one user's follow network and B is the set of users that comprise another user's follow network, we compare users' follow networks and report the distribution of similarity indices in Figure 4. Contrary to our expectations, the Jaccard similarity between users' follow networks is quite low, with similarity indices ranging from 0 to 0.051 with a median similarity index of 0.0052.

Patterns of interaction with OSN content provide nuance to our understanding of coverage and relevance. As illustrated by Figure 5, we observe a significant difference ($p < 2.2 \times 10^{-16}$) between the coverage associated with stimulated content and dormant content. While stimulated content does not represent the entire set of content that would be considered relevant to the TDV users, the higher coverage values associated with stimulated content signifies a potential heuristic for determining content relevance to the group of TDV users and can be used as a prioritization value for scheduling content broadcasts over RBDS. We explore this in more detail in the following section.

CONTENT-ORIENTED BROADCASTING

Now that the analysis has shown that social connections between more geographically proximate users are stronger and that stimulated content tends to be viewed by a higher number of co-located users, we address RQ3 by evaluating different approaches to scheduling OSN content over RBDS. In this section, we compare and evaluate five different scheduling approaches: random scheduling, first come first serve (FCFS), coverage-based priority scheduling (P-Coverage), cluster-based priority scheduling (P-CP), and round robin scheduling. The software used to generate the resulting scheduled content is available at `http://github.com/mvigi190/RBDS_Scheduling`. While all of the approaches investigated in this paper have been evaluated in the context of networks, these prior works assume a one-to-one connection between one content provider and one receiver. Prior to our evaluation context, there has not been a data-based evaluation of how these approaches perform in a broadcast scenario where broadcast content can be relevant to multiple receivers.

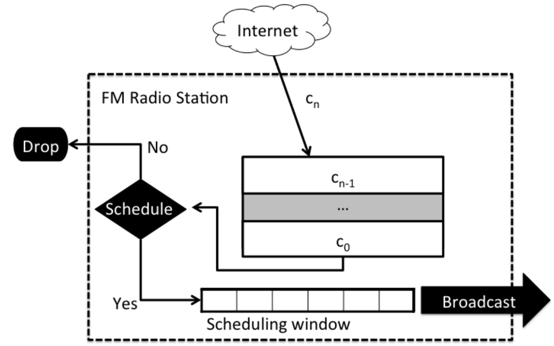


Figure 6: General scheduling approach, where published content objects, c, are identified by a unique identifier and placed in an array until the end of the scheduling period when content is scheduled for broadcast.

Description of scheduling approaches

We provide an illustration of the general scheduling approach in Figure 6. For all scheduling approaches, we assume that there is sufficient storage at the FM radio station to store all incoming content prior to scheduling it for broadcast. This collection of content includes all content that was published during the current scheduling period. The amount of content that is broadcast per scheduling period is determined by the length of the scheduling period using

$$W = T \times B \quad (2)$$

where W is the size of the broadcast window, T is the length of the scheduling period, and B is the bandwidth capacity. For our evaluations, we use a scheduling period of 3600 seconds (one hour). This means that we schedule enough content to fill one hour's worth of broadcasts and drop the remaining unscheduled content in order to provide timely delivery of transmitted content. We assume the RBDS transmission operates without error, yielding a net bandwidth capacity of 1.1875 kbps. This leads to a broadcast window of 534 KB. Because video content objects are over $10\times$ larger than image content and only account for 6.4% of the overall content objects that occur in our evaluation set, we only evaluate scheduling approaches based on how they schedule image content. Since we assume that no transmission errors occur, all content that is scheduled into the broadcast window is actually broadcast.

Random

To provide a standard against which to evaluate the other approaches, we schedule content using a random approach. For each scheduling period, the random scheduling approach selects a random content object from the queue of published content and then either schedules it or drops it with equal probability. Once the scheduling window is full for the broadcast period, any remaining content is dropped.

Approach	Median coverage per user (%)	Mean coverage per user (%)	Standard deviation coverage per user (%)	Jain fairness index
Random	33.1	33.8	9.5	0.88
FCFS	32.3	33.3	9.2	0.88
P-Coverage	52.2	53.6	18.5	0.86
P-CP	77.8	65.5	30.1	0.77
Round robin	54.3	55.0	14.5	0.89

Table 4: Overview of the performance of scheduling approaches.

FCFS

We evaluate a naive first come, first serve approach to schedule data for broadcast in each scheduling window. Our implementation of the FCFS approach fills the scheduling window with the most recent content every scheduling period, then drops the remaining unscheduled content. In the context of broadcast over RBDS, large content flows would correspond to users following prolific content creators, as well as users who follow many content creators.

Coverage-based Priority

Priority scheduling is similar to the FCFS approach, but instead of scheduling content based on the time it was published, we prioritize content based on the coverage provided by the creator of the content. The more followers within a set of co-located users that a content creator has, the higher its content is prioritized. We call this priority-coverage (P-Coverage). We expect this approach to bias in favor of users who follow content creators with high coverage values. Conversely, we expect priority-coverage to bias against users who follow more obscure content creators.

Cluster-based Priority

The cluster-based priority scheduling approach uses the clustering potential (CP) associated with content creators in order to prioritize content. We find the clustering potential associated with a content creator, c, by calculating the summated clustering coefficient [42] of each user who would receive a given content object using the following equation:

$$CP_c(x_0, x_1, ..., x_n) = \sum_{i=0}^{n} \frac{T_{x_i}}{A_{x_i}} \tag{3}$$

where x_i represents a user that follows content creator c , n is the total number of users that follow content creator c, T_{x_i} is the number of creators that form a triadic closure with x_i, and A_{x_i} is the number of creators in x_i's follow network. In the context of the TDV Instagram social network, a user's clustering coefficient determines how embedded a user is within the TDV network. While triadic closures, or three nodes connecting to each other, provide some clue as to how embedded a user is within their local network, the clustering coefficient normalizes this value so that relative embeddedness can be compared across different subgraphs of the network. Given the publish-subscribe nature of the Instagram platform, we consider single-directional links to count towards a triadic closure. The clustering coefficient in the context of co-located users provides an indication of how many users might find information destined for a particular user to be relevant. The approach that prioritizes based on content creators' CP values is called priority-CP (P-CP). Priority-CP is expected to bias in favor of highly embedded users and users with smaller follow networks as it is easier to have a high clustering coefficient with a smaller overall follow network. We expect both priority approaches to result in starvation of users who are dissimilar from the majority of co-located users with respect to content interests and users who are not deeply embedded in their follow network.

Round robin

In order to ensure that each of the co-located users receives some portion of their followed content, we evaluate a round robin approach to content scheduling. For the round robin approach, we create a queue for each user. As content is published, if a user follows that content creator, the content is added to the user's queue. Content in user queues is ordered based on the community coverage provided by the content creator. The round robin approach cycles through the user queues at the end of each scheduling period. While there is room available in the scheduling window, it adds the content at the front of each user queue to the scheduling window. When a content object is added to the scheduling window, it is removed from the queue of any other user that has the object in their queue. This would occur when multiple users are following the same content creator. Round robin provides an equal portion of bandwidth to each user. Some users opportunistically benefit from other users' access to bandwidth because they follow the same content. Since some users follow significantly more content than others, round robin is expected to bias against users that require an overall larger portion of the bandwidth resources.

Performance evaluation

We evaluate each of the scheduling approaches by scheduling one representative[5] ($p < 2.2 \times 10^{-16}$) month of the data set we collected, with respect to the number of content objects posted by creators and the distribution of coverage values associated with posted content objects. We use a one hour scheduling window for each approach. Based on the specifications of RBDS, this window size enables a maximum of 534 KB to be broadcast per hour. Our prior study of the TDV network revealed that the median size of downloaded Instagram image content was 20 KB [41]. As no file size is provided as part of the content meta data we collected, we make the simplifying assumption that all image content is 20 KB. An overview of each approach's performance is provided in Table 4, which reports on the distribution of the percentage of followed content that was received per user and the fairness of the approach with respect to the coverage provided to each user.

[5] Based on the Two-sample Kolmogorov-Smirnov test comparing the distribution of coverage values and the number of content objects posted per month.

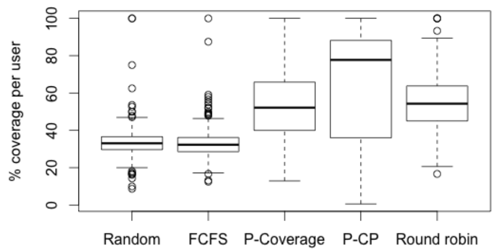


Figure 7: The proportion of content coverage provided to each user.

Completeness

We examine completeness from the perspective of individual users and the group of users. We measure the total coverage provided by a particular scheduling approach by identifying the number of OSN content objects that the approach scheduled over our evaluation month relative to the total number of OSN content objects that were published during that month. Each scheduling approach was able to provide a total coverage of 35.5%. We measure the coverage per user by identifying the total number of content objects that would be received by each user during the evaluation month, then identifying what percentage of that content was actually broadcast using each scheduling approach. The distributions of the coverage per user provided by each approach are plotted in Figure 7. On average, individual users received the highest percentage of their followed content when using the round robin scheduling approach. These results are unsurprising for several reasons. By scheduling content based on user demand in a fair way, the round robin scheduling approach is able to address the content needs of individual users. Because several content creators are followed by more than one user in the TDV network, the round robin scheduling approach was able to opportunistically satisfy the needs of several users when scheduling content based on the needs of a single user.

The FCFS scheduling approach performed the worst with respect to coverage per user. Given its prioritization based on publication time, it is unsurprising that FCFS would not provide the optimal approach for scheduling bandwidth in an overloaded environment. FCFS is unable to discern community value associated with content because it operates without any concept of relative content demand. This causes users who follow highly productive content creators, users who follow many content creators, and users who follow more video content to receive a much larger portion of the available bandwidth.

We also note that priority-CP had a high coverage rate with respect to the individual, but varied significantly between users and ultimately provided the most unfair division of resources (we discuss this further in the section "Fairness"). Priority-CP tends to favor content arriving for small, tightly knit sub-groups of users and leads to starvation of less embedded users. While it was not the best scheduling approach for the Instagram context, it could be useful in other scenarios. For instance, if broadcast were the first part of an information relay, targeting content towards a clustered group of individuals may lead to more effective information dissemination throughout the community.

Fairness

In order to evaluate the fairness of each scheduling approach, we calculate the Jain fairness index [26] over the number of content objects each user would receive. The Jain fairness index is calculated as follows:

$$jain(x_0, x_1, ..., x_n) = \frac{(\sum_{i=0}^{n} x_i)^2}{n \cdot \sum_{i=0}^{n} x_i^2} \quad (4)$$

where x_i represents the percentage of user i's content requests that were broadcast (coverage per user), given n co-located users to receive scheduled broadcasts.

The Jain fairness index for each scheduling approach is reported in Table 4. It is expected that the round robin approach would be the fairest of those evaluated on the data, as it is an inherently fair approach because it provides equal opportunities for all users to receive content that is relevant to them. Because of the broadcast nature of RBDS and the fact that multiple co-located users can follow content published by a single content creator, scheduling content from a single user's queue could lead to multiple users receiving content outside of what would be scheduled from their own queues. This causes co-located users with overlapping follow networks to have a slight advantage over users who do not follow networks that are similar to other co-located users. The least fair approach is priority-CP. This is expected since prioritizing content destined for a highly clustered set of users would starve out a significant portion of content given the lack of observed overlap and clustering between TDV users (the mean clustering coefficient is 0.039). We also find that FCFS is less fair than the random scheduling algorithm, which is unsurprising given the inherent unfairness of FCFS, which favors users who follow highly productive creators and tends to starve out smaller flows. Even though the random approach and FCFS provide significantly lower coverage per user, they still have high fairness values. These approaches are fair despite the lower coverage values because the coverage they provide has very low variance. The variance in coverage per user is 0.7% for the random approach and 0.8% for FCFS. Round robin provides a higher level of fairness than both priority-coverage and priority-CP because round robin provides a much lower variance in coverage per user. The variance for the round robin approach is 2.1%, the variance for priority-coverage is 3.4%, and the variance for priority-CP is 9.1%.

Timeliness

By only scheduling content that arrives during the current scheduling period, the delay between the time content was published and the time it is broadcast is always under one hour[6]. While this delay is reasonable given our observations of users' responsiveness to content in the section "Publication

[6]Only scheduled content has associated delay times; unscheduled content is not considered when calculating delay.

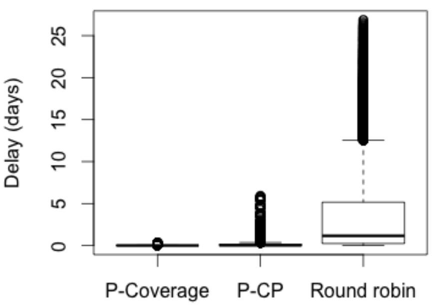


Figure 8: The distribution of delays incurred by scheduling stored content.

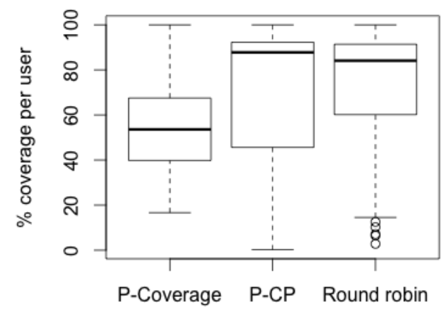


Figure 9: The distribution of content coverage provided by approaches using stored content.

time," we note that if we were to relax this delay requirement, certain approaches would be able to provide a higher level of coverage to individual users. We test the impact of relaxing delay constraints by storing unscheduled data for future scheduling periods (instead of dropping unscheduled content as described in the beginning of the section "Description of approaches"). This means that scheduling approaches would be able to fill the scheduling window with content published during the current scheduling period and unscheduled content from previous a scheduling period. We then evaluate performance over the same representative month of content for approaches that could increase coverage by scheduling stored data: priority-coverage, priority-CP, and round robin. We plot the distributions of delay incurred by scheduling stored content in Figure 8. Although we expected the delay to increase for all approaches that schedule stored content, priority-CP was able to maintain an average delay time below an hour, with a median delay of 46 minutes and a standard deviation of 11.5 hours. Since the scheduling window is limited to 534 KB per hour, only 26 content objects with the highest CP values are broadcast. When we examine the distribution of the CP values associated with incoming content across time, we find that on average, 36 of the content objects published per hour have a CP value greater than the median CP ($CP = 0.01558$) value and 18 of the content objects published per hour have a CP value greater than the third quartile CP value ($CP = 0.05556$). This means that each hour, content that is broadcast is likely to have been published during the most recent hour, leading to generally low delays. The median delay for priority-coverage is 4.9 days (standard deviation = 6.1 days) and the median delay for round robin is 1.2 days (standard deviation = 5.0 days).

Despite the longer delay times associated with broadcast content, we do find that the coverage provided by the approaches increases. Figure 9 plots the distribution of coverage per user provided by each approach scheduling stored content. Overall, each approach is able to provide a total coverage of 36.4%. This is an increase compared to the approaches that schedule only content published in the current scheduling period because some hours of the month do not have enough content publications to fill an entire scheduling window. With the ability to schedule content stored from previous scheduling periods, the approaches are able to fill the entire scheduling window during every scheduling period because there is always a backlog of unscheduled content to fill any gaps. In addition, the round robin approach and priority-CP approach are able to increase the coverage per user (the coverage per user provided by priority-coverage remains within 1% of what was provided for scheduling without stored content). The median coverage per user for priority-CP is 87.8% (standard deviation=32.1%) and the median coverage per user for round robin is 84.1% (standard deviation=23.8%). Consequently, the Jain fairness index changes for these two approaches, and when operating over stored content, the Jain fairness index for the priority-CP approach decreases to 0.75 and the Jain fairness index for the round robin approach increases to 0.91.

There is a clear trade-off between timeliness and coverage. Storing unscheduled data for future scheduling periods ensures that content has more opportunities to be broadcast. However, each content object has to compete against a growing reservoir of content in order to be scheduled. This can lead to longer delays between publication time and broadcast time. One way this trade-off can be negotiated is by regularly flushing unscheduled content after it has been stored for longer than a certain period of time. The flush period would be based on the delay requirements and the average volume of data published per hour. Another way to balance delay and coverage would be to schedule only content published during the scheduling period and at certain times, consider stored content for broadcast. Our analysis in the section "Publication time" reveals that Instagram users have an eight hour trough in activity, meaning they are not likely to be checking their phone for incoming posts during this period. During the hours of the day that users are checking content more frequently, it would make sense to only broadcast the most current published content. Conversely, the period of inactivity might be an opportunity to broadcast stored content.

DISCUSSION

Here we discuss some of the issues of practical information dissemination via social RBDS, including user interaction, moving beyond social content, and bootstrapping. Although this paper has focused on a Native American community, it is important to note that the proposed social broadcast over RBDS could be used beyond the reservation context in developing areas where there is limited broadband coverage or where communication services are prohibitively expensive.

Generalizability

The data used to answer our research questions (RQ1, RQ2, and RQ3) was derived from Instgaram users from six different Native American reservations in San Diego County. While the OSN usage data that we characterize in our analysis are specific to TDV Instagram users, research exploring the social media practices of both indigenous and rural communities has revealed similar characteristics of high locality of interest with respect to content and social connections [34, 27, 28, 44, 21]. Here we discuss how the observed locality of interest in other communities might be leveraged by our proposed OSN content delivery system over RBDS. We provide two examples of how our findings would transfer to other communities based on their OSN usage patterns. The first example describes OSN usage in the Sioux Lookout region, which is home to indigenous Ojibway, Oji-Cree, and Cree peoples. The second example describes OSN usage in a rural village in Africa.

Survey-based studies performed by Molyneaux et al. on the social networking practices of indigenous peoples in the Sioux Lookout region of Northwestern Ontario reveal that OSNs are the most frequently used web applications and that members of the community use technology to help preserve their indigenous culture by posting cultural content to OSNs, posting announcements about upcoming cultural events, and reading and listening to other OSN users' culturally relevant posts [34]. In this study, 60.2% of the 588 surveyed users accessed OSNs to daily interact with co-located OSN users. This is similar to our finding that TDV Instagram users interacted with content from co-located users $46.6\times$ more often than content generated by users outside the TDV network. The locality of interest suggests that OSN content transmitted over RBDS can provide high coverage to this community of users, particularly if content pertains to local cultural topics.

Looking to increasingly global populations, locality of interest (particularly in small, rural communities) is strong. Previous studies of network traffic locality of a rural village in Macha, Zambia reveal that users are highly involved with online social networks and that the majority of Facebook instant message exchanges are between co-located users [28]. Further study of OSN usage in the Macha community revealed that 46% of downloaded Facebook images were not unique images [28] and the top 10 monthly downloaded Youtube videos received an average of 541 views per month with a standard deviation of 2,580 views per month [44]. This implies that the Macha community could be well-served by our proposed OSN-to-RBDS system given the high rate of overlapping content interests exhibited by co-located users.

Interaction and feedback

The scheduling approaches we evaluated had access to users' follow networks prior to scheduling content. Since follow networks are dynamic, there would need to be a mechanism that allows users to update their social network information in order to receive content from all of the content creators they follow. Additionally, the data sets used to evaluate the feasibility of broadcasting OSN content via RBDS did not incorporate information on relevance based on ratings or preferences–from the TDV Instagram users as a group or as individuals. If users had access to a text-message based interface, they would be able to assign priority preferences to different types of content or to provide feedback on the relevance of content that was broadcast. With this integrated information, it would be possible to develop models that predict relevance of incoming content to a specific user and to a group of users.

Access

In order to deploy a system like the one described in Figure 1, there are a few bootstrapping concerns that must be addressed. If users wish to receive OSN content over RBDS, there needs to be an interface for them to grant access to their social network information or subscribe to certain content. While this is possible with the current APIs for most OSN platforms (e.g. Instagram, Facebook, Twitter) and sites that publish RSS feeds, it still requires users to have some form of Internet access over which to grant these permissions or requests. With federal grant programs focusing on providing broadband access to schools and libraries on reservations, there may be public Internet access available in the community that would allow users to provide initial information to an RBDS content broadcast system. However, on reservations that span a significant geographic area, the nearest Internet connection may be tens of miles away, and access may require significant time and fuel resources even if the Internet service itself is free [40]. One way to address this challenge involves a text-message based application that could interface with a service that grants access to a users' social network information on their behalf. Although high speed mobile broadband services such as 3G, 4G, and LTE are not pervasive in Tribal lands, 2G services (which work predominantly for voice and text messaging) have significant penetration and most service plans include unlimited text-messaging or inexpensive "voice and text-only" plans.

CONCLUSION

While it is easy to imagine that issues of the digital divide are endemic to developing nations, there are still communities in developed nations that do not have access to the Internet due to lack of availability. In this study, we focus predominantly on the digital divide present on Native American reservations in the U.S. [4]. Prior studies investigating the informational needs of reservation communities have emphasized the importance of OSN content for connecting users together through indigenous language [34, 20], cultural practice [10], experience sharing [10], participation in political and social movements [8, 31], and daily communication [34]. Understanding the information habits of indigenous communities enables researchers to develop new solutions that help users that live beyond the borders of Internet service connect to each other through existing platforms for computer-supported collaboration (e.g. OSNs). Our characterization of OSN usage by TDV Instagram users reveals a high locality of interest that is consistent with reports of OSN usage in both indigenous [34, 10, 20] and rural [28, 21] communities around the world. Based on these observations, we propose a system that disseminates OSN content via RBDS. We show that RBDS holds unlocked potential for content dissemination in communities that do not have Internet access, and our evaluation of various content scheduling approaches suggests that RBDS bandwidth can be shared fairly amongst co-located users even when they have relatively dissimilar content interests. Through the implementation of our system, users who have been historically excluded from the unique collaboration opportunities afforded by OSNs could participate using infrastructure and technologies that are already ubiquitous and affordable.

ACKNOWLEDGMENTS

We would like to thank Joseph Perralta and Geoff Herrin of the Southern California Tribal Digital Village Network for their help with the collection of data used in this analysis. This work was funded by NSF Graduate Research Fellowship Program under Grant No. DGE-1144085 and NSF Network Science and Engineering (NetSE) Award CNS-1064821.

REFERENCES

1. Tribal Fund Phase I Potentially Eligible Areas. `http://www.fcc.gov/maps/tribal-mobility-fund-phase-i-potentially-eligible-areas`, March 2013.
2. The Bro Network Security Monitor. `https://www.bro.org/`, October 2014.
3. Exploring the Digital Nation: Embracing the Mobile Movement. `http://www.ntia.doc.gov/files/ntia/publications/exploring_the_digital_nation_embracing_the_mobile_internet_10162014.pdf`, October 2014.
4. Native Nations Consultation and Policy. `http://www.fcc.gov/encyclopedia/native-nations-consultation-and-policy`, December 2014.
5. traceanon. `http://www.wand.net.nz/trac/libtrace/wiki/TraceAnon`, May 2014.
6. World Fact Book. `https://www.cia.gov/Library/publications/the-world-factbook/fields/2213.html`, June 2014.
7. Ahtone, T. Radio on the Reservation. `http://projects.aljazeera.com/2014/reservation-radio/index.html`, March 2014.
8. Alia, V. *The New Media Nation: Indigenous Peoples and Global Communication*, vol. 2. Berghan Books, 2010.
9. Ben-David, Y., Vallentin, M., Fowler, S., and Brewer, E. JaldiMAC: Taking the Distance Further. In *Proceedings of the 4th ACM Workshop on Networked Systems for Developing Regions*, NSDR '10 (San Francisco, California, June 2010), 7–13.
10. Carpenter, P., Gibson, K., Kakekaspan, C., and O'Donnell, S. Social Media in Remote First Nation Communities. *Journal of Rural and Community Development 39*, 2 (2014), 275–278.
11. Castro, M., Druschel, P., Kermarrec, A., Nandi, A., Rowstron, A., and Singh, A. SplitStream: High-Bandwidth Content Distribution in Cooperative Environments. In *Peer-to-Peer Systems II*, M. Kaashoek and I. Stoica, Eds., vol. 2735 of *Lecture Notes in Computer Science*. 2003, 292–303.
12. Chen, J., Hutchful, D., Thies, W., and Subramanian, L. Analyzing and Accelerating Web Access in a School in Peri-urban India. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11 (Hyderabad, India, March 2011), 443–452.
13. Chen, X., and Zhang, X. Coordinated Data Prefetching by Utilizing Reference Information at Both Proxy and Web Servers. vol. 29, ACM (Cambridge, MA, USA, September 2001), 32–38.
14. Dilley, J., Maggs, B., Parikh, J., Prokop, H., Sitaraman, R., and Weihl, B. Globally Distributed Content Delivery. *IEEE Internet Computing 6*, 5 (September 2002), 50–58.
15. Du, B., Demmer, M., and Brewer, E. Analysis of WWW Traffic in Cambodia and Ghana. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06 (Edinburgh, Scotland, May 2006), 771–780.
16. Duarte, M. *Network Sovereignty: Understanding the Implications of Tribal Broadband Networks*. PhD thesis, University of Washington, November 2013.
17. Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., and Madden, M. Social Media Update 2014. `http://www.pewinternet.org/2015/01/09/social-media-update-2014/`, January 2015.
18. Ellison, N., Steinfield, C., and Lampe, C. The benefits of Facebook friends: Social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication 12*, 4 (2007), 1143–1168.
19. Fan, L., Cao, P., Lin, W., and Jacobson, Q. Web Prefetching Between Low-bandwidth Clients and Proxies: Potential and Performance. In *Proceedings of the 1999 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems* (Atlanta, GA, USA, May 1999), 178–187.
20. Gibson, K., Kakekaspan, M., Kakekaspan, G., O'Donnell, S., Walmark, B., and Beaton, B. A History of Everyday Communication by Community Members of Fort Severn First Nation: From Hand Deliveries to Virtual Pokes. In *Proceedings of the 2012 iConference* (Toronto, ON, Canada, February 2012).
21. Gilbert, E., Karahalios, K., and Sandvig, C. The Network in the Garden: An Empirical Analysis of Social Media in Rural Life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08 (Florence, Italy, 2008), 1603–1612.

22. Guo, S., Derakhshani, M., Falaki, M., Ismail, U., Luk, R., Oliver, E., Ur Rahman, S., Seth, A., Zaharia, M., and Keshav, S. Design and Implementation of the KioskNet System. *Computer Networks 55*, 1 (2011), 264 –281.

23. Guskin, E., and Mitchell, A. Innovating News in Native Communities. `http://www.stateofthemedia.org/2012/native-american-news-media/`, January 2012.

24. Ioannidis, S., Chaintreau, A., and Massoulie, L. Optimal and Scalable Distribution of Content Updates Over a Mobile Social Network. In *IEEE INFOCOM 2009* (April 2009), 1422–1430.

25. Jaccard, P. The Distribution of the Flora in the Alpine Zone. *New Phytologist 11*, 2 (1912), 37–50.

26. Jain, R., Chiu, D., and Hawe, W. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. *Technical Report, Digital Equipment Corporation, DEC-TR-301* (1984).

27. Johnson, D., Pejovic, V., Belding, E., and van Stam, G. Traffic Characterization and Internet Usage in Rural Africa. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11 (Hyderabad, India, March 2011), 493–502.

28. Johnson, D., Pejovic, V., Belding, E., and van Stam, G. VillageShare: Facilitating Content Generation and Sharing in Rural Networks. In *Proceedings of the 2nd ACM Symposium on Computing for Development*, ACM DEV '12 (Atlanta, GA, USA, March 2012), 61–70.

29. Krishnamoorthi, V., Carlsson, N., Eager, D., Mahanti, A., and Shahmehri, N. Quality-adaptive Prefetching for Interactive Branched Video Using HTTP-based Adaptive Streaming. In *Proceedings of the ACM International Conference on Multimedia, MM '14* (Orlando, Florida, USA, November 2014), 317–326.

30. Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., and Tomkins, A. Geographic Routing in Social Networks. *Proceedings of the National Academy of Sciences of the United States of America 102*, 33 (2005), 11623–11628.

31. McCaskill, D., and Rutherford, J. Indigenous peoples of South-East Asia: poverty, identity and resistance. In *Indigenous Peoples and Poverty: An International Perspective*, R. Eversole, J.-A. McNeish, and A. D. Cimadamore, Eds. Zed Books, London, UK, December 2005, 126–157.

32. McPherson, M., Smith-Lovin, L., and Cook, J. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* (2001), 415–444.

33. Milgram, S. The Small World Problem. *Psychology Today 2*, 1 (1967), 60–67.

34. Molyneaux, H., O'Donnell, S., Kakekaspan, C., Walmark, B., Budka, P., and Gibson, K. How Women in Remote and Rural First Nation Communities are Using Information and Communication Technologies (ICT). *Canadian Journal of Communication 8*, 2 (2013), 79–97.

35. Rahmati, A., Zhong, L., Vasudevan, V., Wickramasuriya, J., and Stewart, D. Enabling Pervasive Mobile Applications with the FM Radio Broadcast Data System. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems and Applications*, HotMobile '10 (Annapolis, MD, USA, February 2010), 78–83.

36. Saroiu, S., Gummadi, K. P., Dunn, R. J., Gribble, S. D., and Levy, H. M. An Analysis of Internet Content Delivery Systems. vol. 36 (December 2002), 315–327.

37. Shah, S., and Joshi, A. COCO: A Web-based Data Tracking Architecture for Challenged Network Environments. In *Proceedings of the First ACM Symposium on Computing for Development*, ACM DEV '10 (London, United Kingdom, December 2010), 48–55.

38. Siegler, K. Radio Station KYAY is Lifeline For Apache Tribe. `http://www.npr.org/2013/09/03/218455207/radio-station-kyay-is-lifeline-for-apache-tribe`, September 2013.

39. Smith, A. Why Americans Use Social Media. `http://www.pewinternet.org/2015/01/09/social-media-update-2014/`, November 2011.

40. Smith, G. On Tribal Lands, Digital Divide Brings New Form of Isolation. `http://www.huffingtonpost.com/2012/04/20/digital-divide-tribal-lands_n_1403046.html`, April 2012.

41. Vigil, M., Rantanen, M., and Belding, E. A First Look at Tribal Web Traffic. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 (Florence, Italy, May 2015), 1155–1165.

42. Watts, D., and Strogatz, S. Collective dynamics of small-world networks. *Nature 393*, 6684 (1998), 440–442.

43. Zhao, Y., Eager, D. L., and Vernon, M. K. Scalable On-demand Streaming of Nonlinear Media. *IEEE/ACM Transactions on Networking 15*, 5 (October 2007), 1149–1162.

44. Zheleva, M., Schmitt, P., Vigil, M., and Belding, E. Internet Bandwidth Upgrade: Implications on Performance and Usage in Rural Zambia. *Information Technologies & International Development 11*, 2 (2015), 1–17.