A First Look at Tribal Web Traffic

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

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

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

ABSTRACT

With broadband penetration rates of less than 10% per capita, Tribal areas in the U.S. represent some of the most underserved communities in terms of Internet access [4]. Although numerous sources have identified this digital divide, there have been no empirical measurements of the performance and usage of services that do exist in these areas. In this paper, we present the characterization of the Tribal Digital Village (TDV) network, a multi-hop wireless network currently connecting 13 reservations in San Diego county. This work represents the first traffic analysis of broadband usage in Tribal lands. After identifying some of the unique purposes of broadband connectivity in indigenous communities, such as language revitalization and cultural development, we focus on the performance of popular applications that enable such activities, including Youtube and Instagram. Though only a fraction of the bandwidth capacity is actually used, 30% of Youtube uploads and 24% of Instagram uploads fail due to packet loss on the relay and access links that connect the reservations to the TDV backbone. Although failure rates are prohibitive to the contribution of locally generated media (particularly videos), our analysis of Instagram media interactions and engagement in the TDV network reveals a high locality of interest. Residents engage with locally created media 8.2× more than media created by outside sources. Furthermore, locally created media circulates through the network two days longer than non-local media. The results of our analysis point to new directions for increasing content availability on reservations.

1. INTRODUCTION

Even as global broadband coverage continues to increase, Tribal communities remain largely unconnected to broadband services, with less than 10% of the population residing on Tribal lands having access to broadband–a stark contrast to the rest of the United States where the average broadband penetration rate is 70% [4]. The lack of broadband access for these communities is caused by a convergence of challenges, including geographic obstacles, funding, and complex network policy standards. Tribal communities represent some of the final frontiers of the digital divide in the U.S. and a disparity in the accessibility of critical opportunities, including education, health, and financial services. Even after issues of connectivity are resolved, content-related issues become apparent: web content does not reflect Tribal interests; indigenous languages are not represented in web media and are at risk of extinction; and indigenous people play a limited role in how they are represented in media in general. To this end, social media has been cited as a key factor in establishing digital sovereignty, identity, and cultural resilience in Tribal communities [25].

Despite the U.S. government's concern with making broadband more accessible to Tribal communities and anecdotal evidence supporting the importance of online services in preserving Tribal identity and culture, there have been no quantitative studies exploring how Tribal communities utilize existing broadband connectivity. Likewise, little has been done to understand Tribal use of social media from a networking perspective, even though this insight could motivate innovative technical solutions for bringing access to more people. This research represents some of the first work in characterizing web usage in Tribal communities.

Using network traffic collected from the Tribal Digital Village (TDV) network in rural San Diego County, we gain insight into web usage in Tribal America [10]. Our data represents Internet usage over a period of two months, from June 23 to August 20, 2014. The TDV network provides 13 previously unconnected Indian reservations[1] with affordable residential broadband services over a long-distance multi-hop wireless backbone that terminates in a high speed fiber connection to the Internet. Surprisingly, our work reveals that Instagram, not Facebook or Google, is the most requested application in the network, generating over 2 million requests in two months. Additionally, it is one of the top five most bandwidth-consuming applications in the network. Unlike other popular social networking platforms, such as Facebook or Twitter, Instagram focuses exclusively on the processes of creating, sharing, and interacting with media, which makes it especially relevant for understanding how social media functions within a Tribal context. Likewise, Instagram has been cited as one of the top social networking platforms of 2014 with one of the most diverse and engaged user bases [19]. While it has been studied in the context of data visualization [21] and participatory crowd-sensing [27],

[1] An area of land reserved for a Tribe or Tribes under treaty or other agreement with the United States for use as permanent tribal homelands. http://www.bia.gov/FAQs/

WWW 2015, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3469-3/15/05.
http://dx.doi.org/10.1145/2736277.2741645 .

to the best of our knowledge, there has been no study of its usage and performance within the context of a particular community. Thus, our work in Section 5 is the first to use Instagram as a lens through which to understand media prevalence and locality in a community context.

This paper leverages the Instagram social network along with traffic traces from a wireless network to better understand web usage and social media behavior in the unique context of Tribal America. We add the caveat that "Tribes" refers to a set of indigenous people who represent a wide variety of languages, cultures, and perspectives across the U.S.; our study provides measurements for only a small fraction of this population. However, with 13 individual Tribes represented in our study, our findings present a unique opportunity to get a diverse perspective on Tribal traffic. This work presents the following contributions:

- One of the first characterizations of web traffic in a Tribal community
- An evaluation of media performance in a network topology commonly used to bring initial access to developing regions
- An analysis of Instagram media popularity and locality within the context of the TDV network as a whole and within six individual reservations

Our analysis shows that while the usage patterns observed in the TDV network are consistent with reported usage patterns in other Tribal communities and content delivered over the TDV network is very socially and locally oriented, the technologies used to present, store, and share this content are not amenable for use in vastly disconnected environments.

We begin by reviewing related work in Section 2 and move to describing the TDV network and the uniqueness of the Tribal networking context in Section 3. We proceed to discuss general usage and media performance in Section 4. In Section 5, we explore media popularity and locality using Instagram network traces. We discuss our findings and future work in Section 6 and end with our concluding remarks.

2. RELATED WORK

There are several important measurement studies focused on initial analysis of usage and performance in developing contexts [22][14][18]. When it comes to measuring wireless networks in developing contexts, our work is most similar to [18] and [22]. While we look at the performance of media applications and approach network analysis through a cultural lens, our work is distinct in three major ways. First, the TDV network is unlike the networks studied in the aforementioned work in that it is not a highly latent or bandwidth-limited network. Rather, this work provides a first look at web traffic generated by a disenfranchised population that is unique in terms of political status, age demographics, and geographic context in a developed and well-connected nation. Second, our work focuses on the relationship between physical proximity and social closeness in an online social network, including the strength of social connections as well as the popularity of social media. Our focus on social media was informed by interview studies of First Nations (Tribal) broadband usage by Molyneaux et al.[25]. While Molyneaux et al. also focus on characterizing broadband use in Tribal communities, the study involves no analysis of measured network usage or performance and is based exclusively on survey responses. Our work is complementary to this approach and we find that our analysis supports many of the observations made by Molyneaux et al., including the significance of social media, the popularity of audio and video media as well as online gaming.

A significant body of work explores the diffusion of information in social networks, including characterization and modeling of diffusion patterns through various online social networks, including Twitter, Facebook, and Flickr [23] [12] [13]. Most relevant to our analysis is work focusing on information diffusion in social networks over time and space. Cha et al. study the role media popularity and the strength of social links play in the rate that media diffuses the Flickr social network [13]. In contrast, we study media diffusion and social bonds with respect to geographic locality. Also relevant to our work is research exploring the connection between location and social connections [16]. Cho et al. connect information from an underlying physical data network to those in an online social network. This work focuses on understanding the social motivation behind human mobility while our work seeks insight into the impact social bonds in geographic communities have on media in their online counterparts.

3. BACKGROUND

Most of the challenges for broadband accessibility on reservations stem from a centuries-long history of conflict between the U.S. government and indigenous Americans. This history includes removal of indigenous people from their ancestral homelands and forced assimilation and deculturization in Western boarding schools, resulting in a generation disenfranchised from their indigenous cultures and languages. However, Tribes have been resilient and view broadband technologies as a way to communicate, share, and develop indigenous identity as never before. Rejuvenation of disappearing languages and culture in addition to the growth and development of contemporary cultural identities are major motivators for broadband access in Tribal communities [25]. Demonstrating the breadth of the issue, Figure 1 highlights U.S. counties containing Tribal areas with absolutely no access to fixed or mobile broadband (i.e. 3G, 4G, LTE). As Tribes in the U.S. seek to cultivate broadband infrastructure within their communities, they must contend with national telecommunications policy, coordinate with neighboring deployment strategies and regulatory bodies, and find sustainable revenue sources. Issues of resource allocation and use are rendered even more complex by their status as sovereign nations. For instance, Tribes cannot use Tribal land as collateral on loans to subsidize the build-out of communications infrastructure as sovereign land cannot be repossessed. Additionally, attitudes regarding the ownership and control of digital content as well as the impact of Internet content on culture and community have caused some Tribal leaders to hesitate in the deployment of broadband infrastructure in their communities [17].

In spite of the challenges, several Tribes have taken action to increase broadband connectivity in their communities. Our research focuses on a high-speed wireless network established by the Southern California Tribal Chairmen's Association (SCTCA). As recently as 2001, none of the Tribal land represented by the SCTCA had access to broadband

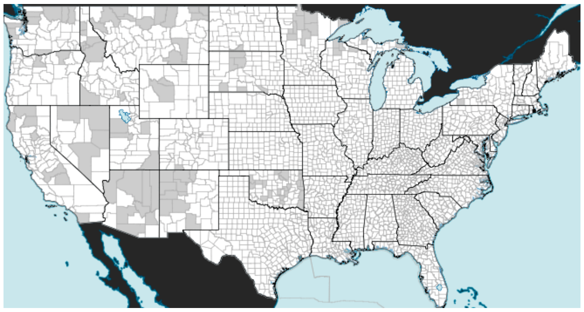
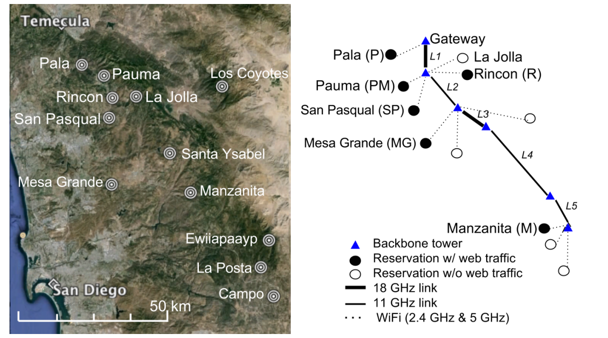


Figure 1: Counties shaded in gray contain one or more census blocks that do not have 3G or better mobile broadband coverage[2].



(a) (b)

Figure 2: Maps of (a) San Diego County with reservations covered by TDV network marked by circles and (b) TDV network topology.

services. As a solution, the SCTCA launched the Tribal Digital Village (TDV) network. The ultimate goal of the project is to provide affordable Internet access to the over 3,000 disconnected homes located on Tribal land. Our study of the TDV network provides insight into how initial connectivity is being used and how future technologies and information services might be designed to best serve these emerging networks.

3.1 Description of the network

Since its inception thirteen years ago, the TDV network has provided coverage to half of the 3,050 homes located on 13 of the 17 reservations that comprise the Southern California Tribal Chairmen's Association. While 576 homes have adopted Internet services at some point over the past thirteen years, only 354 homes currently utilize Internet provided by the TDV network. The lack of broadband adoption and attrition occurs for various reasons, including residents' inability to pay for services or lack of satisfaction with the quality of services. Currently, the TDV network offers two service packages: \$34.95 per month for 2 Mbps and \$64.95 per month for 3 Mbps. In addition to providing broadband access to homes, the TDV network provides access to Tribal municipalities, schools, libraries, and learning centers free of charge. Figure 2a gives an overview of the geographic location of the TDV network and the reservations it connects. Figure 2b shows the network architecture, where the TDV headquarters is denoted as 'Gateway' and tower masts associated with broadband access points are named after the reservations they serve. We note that the spatial positioning of relay tower masts in Figure 2b does not correspond with their actual geolocations, and instead serves to give the reader a sense of the network topology. The incoming access link provides 500 Mbps over fiber and terminates at the TDV headquarters, located on the Pala reservation. A wireless backbone comprised of five solar-powered, point-to-point microwave (11 GHz and 18 GHz) links connects 13 reservations to the gateway. Reservations connect to the backbone by way of point-to-point links over unlicensed 2.4 GHz and 5 GHz. Access links extending from relay towers into individual homes and municipality buildings are point-to-multipoint links operating over WiFi. Of the 500 Mbps of bandwidth at the gateway, 200 Mbps remain in Pala to support operations at headquarters and the Pala relay link while the remaining 300 Mbps are used to provide services to access points that connect to the Internet over the backbone. Overall, the TDV network consists of over 815 km of wireless links– the shortest link (L1) in the backbone spans 1.6 kilometers and the longest link (L4) spans 38.8 kilometers.

3.2 Data collection

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. We capture packet headers with tcpdump and use Bro to collect flow-level statistics for network applications [5]. Overall, we collected 5.5 TB of traffic representing 52.8 billion IP packet headers. Data collection methodology and subsequent analysis received IRB and SCTCA approval prior to collection. All MAC addresses and IP addresses are anonymized using TraceAnon [9].

4. GENERAL TRAFFIC ANALYSIS

We begin our analysis by looking at general usage and performance patterns. Relevant findings include:

- Performance bottlenecks are associated with relay and access links rather than the main backbone
- Packet loss leads to high rates of failure for media downloads and uploads

All tables and figures are derived from the entire two-month measurement period, with the exception of Figure 6.

As discussed in Section 3, major points of connectivity are located at municipality buildings including schools, libraries, and learning centers. The majority of these buildings close for the day between 4 pm and 6 pm, rendering their Internet connectivity inaccessible to the public. In order to gauge the impact of municipality connectivity on daily traffic load, we look at the average hourly traffic volume associated with web uploads and downloads across our entire sample (shown in Figure 3). We observe that even after the typical closing times (between hours 16 and 18), the hourly traffic volume maintains its level. We verify that the majority of traffic is attributable to residential access points by identifying traffic associated with municipality buildings. Overall, we find that 99% of the observed traffic volume is associated with residential access points.

[2]Tribal Mobility Fund Phase I Eligible Areas Map. Credit: FCC. http://www.fcc.gov/maps/tribal-mobility-fund-phase-1-eligible-areas

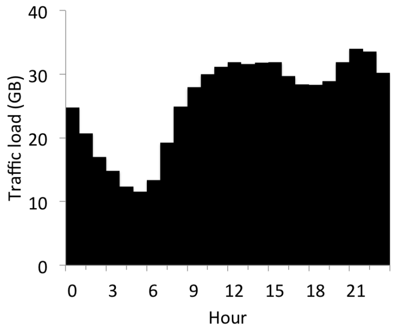


Figure 3: Average hourly traffic volume.

Table 1: Devices used in TDV network.

Mobile devices		**411**
	iOS	42.9%
	Android	35.2%
	Windows	11.5%
Desktop devices		**131**
	Mac OS X	23.1%
	Linux	11.9%
	Windows	45.0%

Table 1 provides an overview of unique devices used to access the Internet over the TDV network. These devices were identified using the user agent field of HTTP traffic headers. The majority of users access web content using mobile devices (smart phones, tablets, e-readers) as opposed to stationary devices (desktops, laptops, and gaming consoles). Additionally, gaming consoles account for 11% of desktop devices in the TDV network.

In order to characterize traffic, we examine network performance both as a whole and associated with each reservation. Figure 4 compares the traffic load averaged over the entire measurement period with the service level agreement (SLA). We see that the traffic load rarely exceeds the 500 Mbps SLA. This differentiates the TDV network from previously studied networks in developing contexts, as its bandwidth capacity is not a limiting factor on traffic usage and performance. In Table 2, we report performance statistics associated with each reservation relay link. The 'Link' column in the table is associated with the backbone link that connects to each relay tower and reservations are noted using their associated abbreviations from Figure 2b. Retransmission rate is calculated according to the number of retransmitted segments in a flow divided by the total number of segments transmitted. We observe high retransmission and failure rates at all relay links. Pala, which connects to the gateway via a single relay link, has the highest failure and retransmission rates. Exemplified by Pauma, Rincon, and San Pasqual, performance can vary in terms of packet loss and flow failure even for reservations connecting to the backbone via the same link. Likewise, reservations that are multiple hops away from the gateway, such as Mesa Grande and Manzanita, do not experience performance degradation in proportion to the number of backbone hops they must travel. Based on these observations, we conclude that performance degradation occurs either over the relay links between backbone towers and individual reservations or over the access links that extend connectivity from relay links into homes and municipality buildings. Our current measurement configuration does not allow us to pinpoint the exact location of packet loss, but based on our findings and our knowledge of the network topology, we suspect that packet loss is either due to low signal strength between the backbone towers and relay points or due to interference between access links connecting to the same relay point.

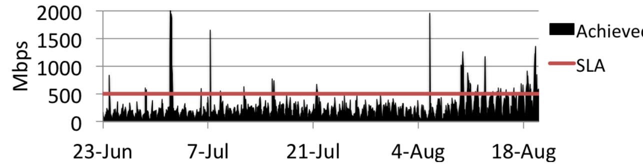


Figure 4: Achieved throughput compared to SLA.

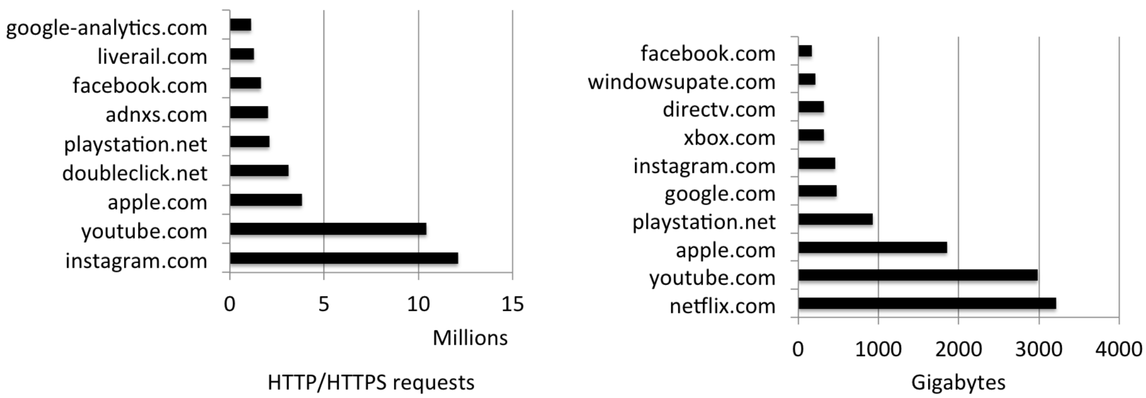
Table 2: TCP statistics for each reservation.

	Link	% Failed requests	% Retransmitted	RTT (ms)	IAT (s)
P	–	25.6	5.9	140	0.37
PM	L1	24.0	5.7	89	0.26
R	L1	28.8	4.6	93	0.26
SP	L1	33.9	4.9	94	0.30
MG	L2	25.7	4.8	71	0.15
M	L5	23.1	5.1	79	0.27

4.1 Overview of web applications

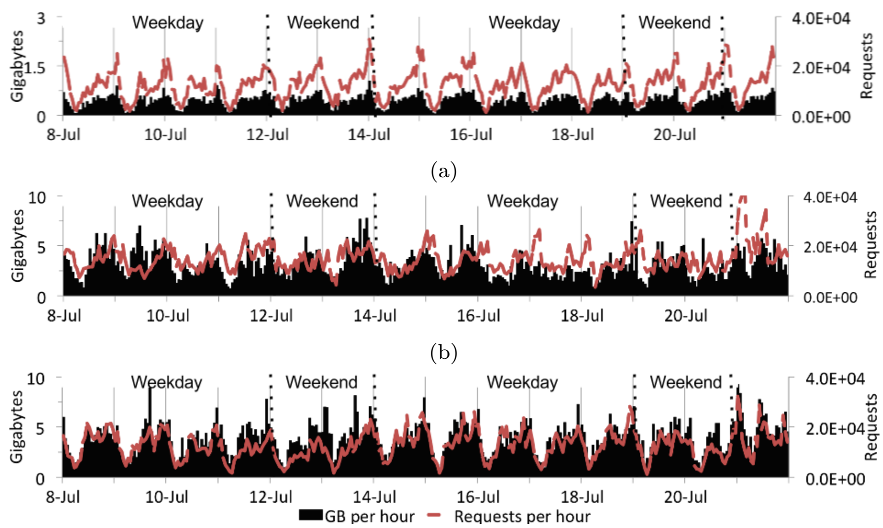
Figure 5 reveals that Instagram is the most requested web application, with over 12.1 million requests over the measurement period. This is surprising given current measurements of web usage in the U.S., which have revealed Facebook as the most dominant social media presence, followed by Twitter, LinkedIn, and Pinterest [8]. We are also surprised by the high levels of gaming traffic represented by PlayStation and Xbox. While Xbox is not ranked as one of the top 10 most requested applications in the network, it is ranked in the top 15. Similar to what we observed with the rank of Instagram traffic, the popularity of PlayStation is unexpected; it ranks as the sixth most popular web site in the TDV network compared to 1,089th in the U.S. [8]. In terms of popularity, Youtube is second only to Instagram with 1.8 million requests during the measurement period. This is similar to the rest of the U.S. where Youtube is ranked as the third most popular web site [8]. Surprisingly, e-commerce sites like Amazon and Ebay are not even ranked in the top 30 most accessed websites in the TDV network, despite ranking as the 4th and 8th most accessed websites in the U.S. When searching for an explanation for this dearth, we find that sovereignty plays a role. Many e-commerce sites rely on U.S. Postal Services for shipping; however, as reservation roads are not maintained by the county, U.S. Postal Services will not deliver to homes that must be accessed through these roads. Similarly, other major shipping companies (i.e. UPS and FedEx) reserve the right not to deliver to all areas and do not guarantee delivery to all addresses. In short, online ordering and shipping can be a significant challenge on reservations and the lack of e-commerce traffic in the TDV network reflects this.

Figure 5b shows the top 10 most bandwidth-consuming web sites observed in the TDV network. We find that streaming media sites represent the greatest bandwidth consumers in the network, accounting for 44% of the overall web traffic volume. This is consistent with streaming media usage in the rest of the U.S., where streaming media accounts for 34-50% of peak traffic bandwidth [6]. Likewise, the composition of streaming media mirrors that of the U.S., where Netflix accounts for 60% of streaming media. We also notice that video gaming sites such as playstation.com and xbox.com



(a) (b)

Figure 5: Overview of the most popular URLs by (a) number of HTTP requests and (b) traffic volume.



(c)

Figure 6: Hourly traffic demand for (a) Instagram, (b) Youtube, and (c) Netflix from July 8 to July 21.

rank among the top 10 web sites in terms of bandwidth and overall, gaming traffic accounts for 8.9% of web traffic volume. Social media comprises a much smaller portion of traffic volume, counting for only 4.5% of the total bandwidth consumed by web traffic. Finally, media and software downloads from online stores such as the Google PlayStore and the iTunes Store account for 23% of web traffic volume. This is in contrast to global mobile traffic patterns, which measure app store downloads as accounting for less than 18% of mobile traffic [6]. In total, media represents over 70% of observed web traffic by volume. Given the popularity of media-oriented applications in terms of HTTP requests and proportion of traffic volume, as well as Tribal interests in leveraging broadband for the development of cultural media, we focus the remainder of this paper on understanding media performance and usage in the TDV network.

4.2 Media performance

Previous studies of emergent wireless networks have observed connections between usage patterns and network performance [22][28][15]. Based on the high levels of media traffic in the network in addition to the popularity of media-based sites, we look specifically at the performance of media applications in the TDV network to identify the presence of performance bottlenecks that might impact user behavior. In order to do this, we look to the performance of three applications: Youtube, Netflix, and Instagram. These three applications are representative of the predominant media transaction types (streaming and bulk transfer). We select each application based on its high traffic volume in the network and the different ways that it allow users to interact with media. We begin by studying the daily usage of each application. In Figure 6 we show the daily traffic volume and web requests generated by Instagram, Netflix, and Youtube over a two-week period which we verify as representative of the entire sampling period (note the different y-axis scales). Traffic volume per hour was calculated by summing the total number of bytes per hour; the number of HTTP requests per hour was calculated by summing the total number of HTTP requests per hour. While weekend and weekday traffic are not significantly different for any of the applications, all three applications exhibit anthropocentric patterns in their usage over time.

We now look at media performance for Instagram, Youtube, and Netflix. Table 3 reports performance statistics for each application including the retransmission rate, download fail-

Table 3: TCP statistics for Instagram photos (I-P), Instagram videos (I-V), Youtube (Y), and Netflix (N).

	% Retransmitted	% Failed downloads	% Failed uploads	RTT (ms)	IAT (s)
I-P	3.2	11.6	24.9	91	0.22
I-V	3.8	31.0	25.0	91	0.29
Y	2.9	32.3	30.1	73	0.18
N	3.4	75.0	NA	94	0.34

ure rate, upload failure rate (when applicable), round trip time (RTT), and packet inter-arrival time (IAT). Round trip times and packet inter-arrival times associated with each application were calculated by taking the average round trip time and packet inter-arrival time for each TCP flow and averaging them over all TCP flows. The retransmission rate for each application represents the percentage of segments retransmitted per TCP flow averaged across all TCP flows.

4.2.1 Video downloads

Understanding the user experience with regards to the network performance of these platforms can highlight areas of improvement in terms of network infrastructure and application design. Overall, we find relatively high rates of download failure for Youtube, Netflix, and Instagram videos. We find that Netflix has the highest failure rate of all, with 75.0% of video downloads ending in a failure. In comparison, 32.3% of Youtube downloads fail, and Instagram video downloads experience a failure rate of 31%. In exploring failures associated with video downloads, we find that the predominant cause of failure for all applications is a TCP RST sent by the client. This type of failure causes 63% of download failures for Youtube, 60% of download failures for Netflix, and 58% of download failures for Instagram videos. This type of failure is indicative of a poor user experience, often due to packet loss [22].

To assess the impact of retransmission rate on download performance, we compare the distributions of successful video downloads to failed video downloads for Netflix (N), Youtube (Y), and Instagram (I) in Figure 7. As an example of our notation in Figures 7 and 8, we use "N-S" to signify successful Netflix flows and "N-F" for failed Netflix flows. We find that for all three applications, retransmission rates are higher for failed downloads than for successful downloads, and on average, failed flows experience 8.2-9.7% loss. For streaming video applications, such as Youtube and Netflix, this type of loss rate would negatively impact the user quality of experience and termination of download mid-stream (triggering a client-sent RST) is consistent with our findings of cause of failure. To assess how failure impacts user interaction with streaming video applications, we study the distributions of flow size and flow duration for successful and unsuccessful Netflix and Youtube downloads using Tstat [1]. Figure 8a graphs the distribution of flow sizes for failed and successful flows for both Youtube and Netflix and Figure 8b shows the distribution of flow duration for failed and successful flows for the two applications. For the size of a flow, we report the goodput rather than the total size including retransmitted data. We find that for both applications, failed flows are on average 83% larger and last 28% longer than their successful counterparts. While we have established that failed downloads are associated with higher retransmission rates, we note that these failures correspond to longer, lengthier downloads, which are more likely to experience losses resulting in

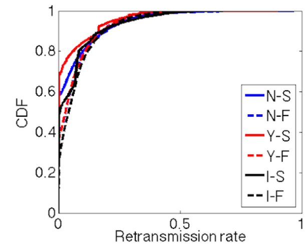
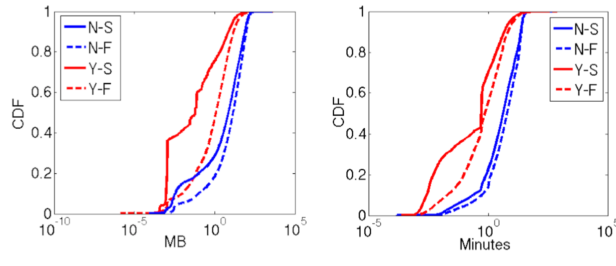


Figure 7: Distribution of retransmission rates for downloads.



(a) (b)

Figure 8: Distributions of (a) flow size and (b) duration for Netflix and Youtube downloads.

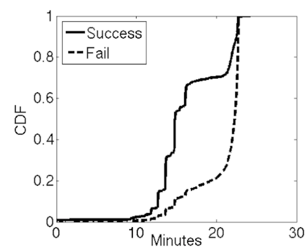


Figure 9: Distribution of flow duration for Instagram video downloads.

a poor user experience. However, we are surprised to find that although Instagram videos are smaller than Youtube and Netfix downloads and experience shorter flow duration, this application experiences the highest retransmission rate for both failed and successful downloads. One reason for the high retransmission rate for successful Instagram video downloads is that it downloads media in bulk, rather than as a stream–so quality of experience is not impacted by the number of packet losses. However, packet loss impacts the length of time it takes to download a video before a user can watch it. In Figure 9, we compare the distributions of flow duration for successful and failed Instagram video downloads. We find that download times are 62% longer for failed downloads than for successful downloads. This is expected as packet loss leads to more retransmissions that result in longer download times. Overall, 40% of downloads require over 15 minutes.

4.2.2 Video uploads

We now investigate upload performance for Youtube and Instagram videos. Upload performance in these applications is particularly important in the context of the TDV network given many of the goals Tribal communities have for broadband connectivity, including cultural content creation, dissemination, and engagement. Overall, 504 video files were uploaded to Instagram compared to 444 uploaded to Youtube. 25% of Instagram video uploads failed and 30% of Youtube uploads failed. When examining the predominant cause of failure for uploads, we find that for Instagram, 85% of failures were due to an unresponsive client (no data packets or control packets were observed coming from the client) and 11% of failures were caused by an RST sent by the server. For Youtube, 55% of upload failures are caused by a timeout-triggered RST sent by the server and 36% of failures are caused by an unresponsive client.

While packet loss over relay and access links can be a contributing factor to failure, we also consider that all Instagram uploads and 98% of Youtube uploads are initiated from mobile devices. Uploading from a mobile device increases the likelihood of a user inadvertently moving from a space of high connection quality to low connection quality, particularly if upload times are extensive. On average, successful Instagram video uploads take 1.4 minutes and successful Youtube video uploads take 4.9 minutes. Failed Instagram uploads take 2.1 minutes before termination and failed Youtube uploads take 5.7 minutes before termination. We believe that this short upload duration, facilitated by lower retransmission rates for upstream traffic, is what allows for Instagram video uploads to be more successful than video downloads. The average uploaded Instagram video is only 0.24 MB compared to the average uploaded Youtube video, which is 11.2 MB. This difference is unsurprising given the restrictions Instagram places on video length (15 seconds), dimension (640×640 pixels), and resolution. This is in contrast to Youtube, which limits video uploads at 11 hours or 128 GB. Increased video sizes have negative consequences for upload failure, as they typically take longer to upload and are more likely to experience packet loss. We also find that on average, failed Instagram uploads are retried 1.3 times and failed Youtube uploads are retried 1.7 times. 3% of retried Instagram video uploads and 7% of retried Youtube uploads are never successfully completed. For each of these failed retries, we find that each final attempted retry takes 10 minutes on average before the attempt is terminated. We also find that repeatedly failed Instagram video uploads require 3 more minutes than video uploads that eventually succeed; repeatedly failed Youtube uploads require 3.6 more minutes than video uploads that eventually succeed.

4.2.3 Images

With the largest number of uploaded files and a platform that lends itself to media sharing and collaboration, Instagram embodies much of the community enrichment potential of broadband connectivity. Unlike Youtube, it enables users to upload images as well as videos. Because the Instagram app was designed exclusively for mobile devices, it imposes limitations on its media formats and as a result provides a higher likelihood of successful media transfers. This primarily manifests as limited upload file dimensions (640×640 pixels). Once an image has been uploaded to Instagram, three standardized versions are created: small image ($\geq$10 KB), medium image ($\geq$20 KB), and large image ($\geq$100 KB). For uploaded videos, small, medium, and large images are generated from the first frame of the video. When the Instagram app is active, images are downloaded from content servers where large versions of the image correspond with images that are more relevant to users, medium images correspond with images that are less relevant, and small images are used for metadata reports. In Table 4, we show the failure rates associated with small, medium, and large downloaded, as well as uploaded Instagram images. 76% of download failures and 86% of upload failures are caused by unresponsive clients after a TCP connection has been established.

Table 4: Failure rates for Instagram images.

	Total #	Failure rate (%)
Small images	5,450,807	13.9
Medium images	2,880,253	15.0
Large images	2,345,020	17.6
Uploaded images	10,677	23.6

Table 5: Overview of TDV Instagram data.

	Total	Local
Media objects	150,368	4,807
Content creators	1,180	164
Media interactions	277,309	19,099
Social interactions	144,721	11,159
Instagram users	NA	238

5. SOCIAL MEDIA AND LOCALITY

Discussions from Section 3 highlight the importance of digital media to Tribal interests in preserving and sharing cultural knowledge. Moreover, studies of emergent networks in under-served communities note the importance of local content sharing for establishing a digital culture that emphasizes locally relevant information. In the previous section, we evaluated the performance of media transactions in three of the most popular web applications in the network. As we would expect, packet loss has a negative impact on both streaming and bulk media downloads alike. Media uploads fail due to timeouts and unresponsive clients. We are surprised that Instagram is so vulnerable to loss, given its design for mobile environments. In this section, we analyze media practices in the TDV network to determine whether we can distinguish localized usage patterns. Due to its surprising popularity in the TDV network and its publicly accessible API, we use Instagram network traces to answer questions about media locality in a Tribal context. Our analysis reveals that the locality of interest in the TDV network is very high compared to non-indigenous communities, with the top 10% of users' strongest social connections spanning less than 10 km [20]. This is expected given the social values observed in indigenous communities that focus heavily on developing intra-community relationships [26]. In Section 6, we discuss the implications of these findings for improving connectivity.

Table 5 provides an overview of the Instagram data we observe, including information about media interactions and users involved in the TDV Instagram network. In Table 5, "content creators" refer to Instagram users who have uploaded media to Instagram. We note that media interactions differ from social interactions: a media interaction includes user-explicit actions such as liking, commenting on, or viewing a media object, while a social interaction only includes user-explicit actions that are announced on the acting user's social feed, such as liking and commenting on a media object.

We find that media views comprise nearly half of all media interactions. This means that only about half of the media interactions that occur in the network are publicly broadcast through the Instagram social network. We also find that only 7% of media interactions occur between a user from the TDV network and media created by a user from the TDV network; only 8% of social interactions are between a TDV user and media created by a TDV user. This is consistent with the fact that only 3% of the media objects we observed were created by local users.

Table 6: Definitions of Instagram interaction types.

Media interactions	User-explicit actions (i.e. liking, commenting on, and viewing a media object).
Social interactions	User-explicit actions that are broadcast to a user's followers (i.e. liking and commenting on a media object).

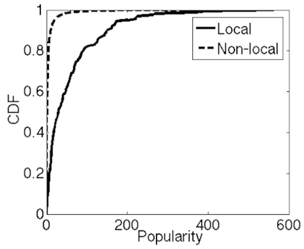


Figure 10: Distributions of media interactions associated with local and non-local content creators.

At first glance, this low proportion of local interaction seems to indicate a low level of local interest. However, when we look to interactivity based on content creator locale, we find a much higher proportion of social interactions associated with content creators from within the TDV network. We are able to identify the media associated with a content creator based on the media's identifier, which contains a unique ten digit identifier concatenated to the user identifier of its creator. In Figure 10, we show the "popularity" of each content creator we observe in the TDV network. Here, we define "popularity" as the total number of media interactions (see Table 6) associated with media created by each observed content creator. On average, the TDV community interacts with locally created media 57 times while it interacts with non-locally created media only five times. So while there are far more non-local media objects to interact with, TDV users are much more engaged with content created by local content creators than with content created by non-local content creators.

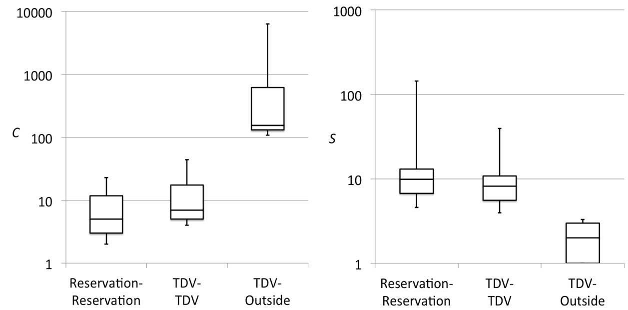
Based on previous work exploring the connection between social media and cultural resilience in indigenous communities, we study the underlying social structure of the TDV Instagram network [25].We begin by defining a social connection between a given pair of Instagram users u_i and u_j as:

$$C_{u_i} = \sum_{j=0}^{n} P_{u_i}(u_j) \quad (1)$$

where $P_{u_i}(u_j) = 1$ if there exists any social interaction between user u_i and media created by user u_j and $P_{u_i}(u_j) = 0$ if no such social interaction exists between user u_i and media created by user u_j. In Figure 11a, we show the number of social connections (C) between each user and all other users in the same reservation ($n = \{11, 17, 25, 69, 114\}$[3] [4]), in the TDV network ($n = 238$), and outside the TDV network ($n = 33, 183$). As users interact with other users from broader circles of the Instagram network, the number of social connections associated with individual users increases. In addition to the number of social connections, we measure the strength (S) associated with a connection, or the

[3]The value of n is dependent on which reservation u_i is associated.

[4]We exclude users from Manzanita in these calculations as $n = 2$ for this reservation and skews the distribution.



(a) (b)

Figure 11: (a) Number of social connections and (b) strength of social connections per user.

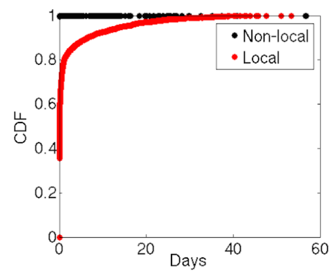


Figure 12: Circulation times of local and non-local media.

number of social interactions that exist between user u_i and media created by user u_j:

$$S_{u_i,u_j} = I_{u_i,u_j} + I_{u_j,u_i} \quad (2)$$

where I_{u_i,u_j} is the number of social interactions between u_i and media created by u_j. Figure 11b illustrates the distribution of S for each user in association with other users from the same reservation, from the TDV network, and from outside the TDV network. While we observe a greater number of social connections between local and non-local users in Figure 11a, Figure 11b reveals that the strength of these connections is weak. In contrast, Figure 11b shows that the more proximate users are in the network, the stronger the social connections are between them. Thus, we find that the TDV Instagram network is composed of a dense core of a few strong local connections and expands out via numerous weak connections. Based on this finding, we expect to see a high locality of interest with respect to media.

5.1 Media prevalence

In addition to the popularity of media, we examine interactions over time to identify how long media circulates in the TDV Instagram network. We measure this by identifying the time delta between a media object's initial appearance in the TDV network and its final appearance in the network across all media interactions (defined at the beginning of Section 5). We have already established that the number of non-local media far exceeds the number of local media present in the network (see Table 5); however in Figure 12, we see that locally created content circulates for much longer within the TDV network than non-local media. Looking more closely, we find that 99.6% of the 145,561 non-local media occur only once in the TDV network. Half of the 4,807 local media objects circulate for over 4.2 hours and are liked or commented on an average of 5 times, while 382 local media objects circulate for over a week and are liked or commented on an average of 7 times. Intuitively, longer circulation times should correspond to an increased level of social engagement; however, circulation time is based on media interactions, not just social interactions. This means that media views, which account for 50% of all media interactions, contribute significantly to a media object's circulation time. Therefore, not only is local media more socially popular than non-local media (see Figures 10 and 11), but it is relevant over a longer period of time.

6. DISCUSSION

6.1 Bonds and bridges

Work in the social sciences has identified the importance of online social networks (OSNs) to Tribal identity and cultural preservation [25]. In particular the formation of "bonds" and "bridges" within Tribal OSNs has significant impact on the strength and resilience of indigenous culture [26]. Bonding relations within a community help members make sense of negative experiences and look forward to a more positive future, creating cultural continuity that connects the traditional culture of the past to the development of a new culture of the present. Bridging connections between different communities allows for communal empowerment and influence on the wider society. In Section 5, we investigated various dimensions of local versus non-local media prevalence and propagation in the TDV Instagram network. With respect to media interactions on Instagram, our findings show that there are many weak bridging connections and a dense core of strong bonding connections. This leads us to believe that social media in Tribal functions to strengthen and maintain bonding connections; however, it is unclear whether bridging connections exert any quantifiable influence over outside communities.

The high locality of interest with respect to locally created media presents opportunities to improve bonding connections in several ways. Current connectivity can be improved by storing locally created content locally, which would prevent users from uploading media over lossy relay and access links that extend upload completion times, improving upload success rates. This strategy would also improve download performance (particularly for videos), as user quality of experience could be improved by avoiding lossy links. Moreover, if locality of media interest corresponds to spatial closeness, social bonds could be used as a means of extending connectivity via opportunistic encounters with peers. This way, users who have access to spaces of ubiquitous connectivity might serve as media vectors, transferring relevant media over Bluetooth, WiFi Direct, or NFC to other users as they come into contact geographically.

6.2 Nations within a nation

While previously studied networks in developing contexts typically reside within greater regions of development, Tribal communities in the U.S. are positioned within a developed nation. In many ways, the effects of the digital divide are amplified in this context and citizens without broadband access are marginalized at an accelerated rate. With high broadband penetration rates in the non-Tribal U.S., many services now assume ubiquitous broadband accessibility, and with this assumption come expectations that cannot be met on Tribal lands with the current state of broadband accessibility. As funding and infrastructure licensing processes move online, Tribes are increasingly alienated from the means by which necessary infrastructures are established. In Section 4, we show that even with sufficient bandwidth capacity, the wireless links used to extend connectivity over long distances are prone to average packet loss rates of 5%, which degrade media performance in both the up-link and down-link directions. Packet loss is caused by limitations in 802.11, which make it sub-optimal for transmitting over long-distance wireless links. Even though the WiFi spectrum is inappropriate for the distance requirements of the TDV network, the spectrum is unlicensed and helps reduce the cost of using the network. If Tribes were able to project their sovereignty over radio frequency in Tribal land, software defined networks leveraging unlicensed or opportunistically available spectrum could significantly increase the penetration and affordability of broadband services. These spectrum rights have been recommended by the FCC and are currently under discussion as becoming part of spectrum access policy in the U.S. [2].

6.3 Attitudes towards technology

Broadband is generally recognized as a means to increase the number of available communication pathways. However, there are various concerns surrounding the Internet and Internet technologies. Some of these concerns exist in both Tribal and non-Tribal contexts, including online engagement detracting from face-to-face interactions, the abundance of "negative" and "pornographic" content online, and the dissonance between offline and online identities [25]. Other concerns are more specific to the Tribal context; these include issues of acculturation and protecting American Indian intellectual property on the Internet. Age demographics also play an important role in Tribal broadband usage. Reservations have younger populations than the surrounding U.S.; the median age on reservations is 26 years, whereas the median age in the U.S. is 37.2 years [7]. Reasons for the skew in ages on reservations include: higher mortality rates due to health conditions, accidents, and suicide; and lack of employment and educational services on reservations which cause working-age populations to seek school and job opportunities outside reservations [3] [11]. In Figure 5, we observe the impact of the age skew in the types of applications that are popular, such as the prevalence of online gaming sites (Playstation and Xbox) and niche social media platforms (Instagram). Dissolving the digital divide in a way that benefits Tribal communities involves understanding inter-generational Tribal attitudes and concerns regarding broadband technologies and allowing Tribes to be actively involved in navigating how those concerns should impact network deployment and technological design in their communities. In this context, we note that localized content storage not only presents a solution that reduces the impact of packet loss, but also presents an opportunity for Tribes to impose tighter control on their intellectual content.

7. CONCLUSION

Internet usage in the TDV network is distinct from the U.S. context in which it resides, especially with respect to the prevalence of niche social media traffic, as well as the influence utility infrastructures have on traffic patterns, exemplified in the lack of online shopping traffic due to unreliable delivery infrastructure. Our work empirically supports trends identified in survey studies regarding Tribal broadband usage: we show that social media is the most popular application in the network in addition to other media-oriented applications like Netflix, Youtube, and iTunes. While packet loss negatively impacts media performance, we use Instagram traffic to identify strong social connections between users within the same reservations and we find that locally created media receives significantly more interactions than non-local media. By exploiting these patterns, architectures that combine local storage, user mobility, and offline social encounters can improve and extend current connectivity.

Tribes in the U.S. are not the only indigenous groups facing issues of disenfranchisement. While issues of sovereignty, funding, lack of political power, and geographic isolation exist for indigenous people around the world, where it is available, broadband connectivity has been shown to strengthen indigenous identity and further collective goals as cultural practices and struggles are communicated between indigenous groups at a transnational scale [24]. By understanding indigenous usage of broadband as well as the technical barriers to broadband deployment, indigenous communities can focus on developing and extending connectivity in ways that make sense for their unique political, geographical, and community contexts.

8. ACKNOWLEDGMENTS

This work was funded through NSF Network Science and Engineering (NetSE) Award CNS-1064821. We would like to thank Joseph Peralta and Geoff Herrin of the Southern California Tribal Digital Network for assisting us with information collection and analysis.

9. REFERENCES

[1] TCP Statistics and Analysis Tool. `http://tstat.tlc.polito.it/index.shtml`, 2008.

[2] Taking Additional Steps to Make U.S. Spectrum Policy More Comprehensive. `http://www.broadband.gov/plan/5-spectrum/#s5-3`, March 2010.

[3] American Indian and Alaska Native Populations. `http://www.cdc.gov/minorityhealth/populations/REMP/aian.html`, July 2013.

[4] Tribal Initiatives. `http://transition.fcc.gov/indians/`, March 2013.

[5] The Bro Network Security Monitor. `https://www.bro.org/`, October 2014.

[6] Mobile broadband, smartphones, apps, fixed networks. `https://itunews.itu.int/En/4958-Mobile-broadband-smartphones-apps-fixed-networks.note.aspx`, February 2014.

[7] National Congress of American Indians, Demographics. `http://www.ncai.org/about-tribes/demographics`, July 2014.

[8] Top Sites in United States. `http://www.alexa.com/topsites/countries/US`, October 2014.

[9] Traceanon. `http://www.wand.net.nz/trac/libtrace/wiki/TraceAnon`, May 2014.

[10] Tribal Digital Village. `http://sctdv.net/`, July 2014.

[11] A. Austin. Native Americans Are Still Waiting for an Economic Recovery. `http://www.epi.org/publication/native-americans-are-still-waiting-for-an-economic-recovery/`, October 2013.

[12] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The Role of Social Networks in Information Diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 519–528, Lyon, France, May 2012.

[13] M. Cha, A. Mislove, and K. P. Gummadi. A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 721–730, Madrid, Spain, May 2009.

[14] J. Chen, D. Hutchful, W. Thies, and L. Subramanian. 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, pages 443–452, Hyderabad, India, May 2011.

[15] J. Chen, L. Subramanian, and J. Li. RuralCafe: Web Search in the Rural Developing World. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 411–420, Madrid, Spain, May 2009.

[16] E. Cho, S. A. Myers, and J. Leskovec. Friendship and Mobility: User Movement in Location-based Social Networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 1082–1090, San Diego, California, USA, 2011.

[17] T. Davis and M. Trebain. Shaping the Destiny of Native American People by Ending the Digital Divide. In *EduCause Review*, pages 38–46, January 2001.

[18] B. Du, M. Demmer, and E. Brewer. Analysis of WWW Traffic in Cambodia and Ghana. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, pages 771–780, Edinburgh, Scotland, May 2006.

[19] M. Duggan and A. Smith. Social Media Update 2014. `http://www.pewinternet.org/2015/01/09/social-media-update-2014/`, January 2015.

[20] E. Gilbert, K. Karahalios, and C. Sandvig. 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, pages 1603–1612, Florence, Italy, April 2008.

[21] N. Hochman and R. Schwartz. Visualizing Instagram: Tracing Cultural Visual Rhythms. In *Proceedings of the Workshop on Social Media Visualization (SocMedVis) in conjunction with the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM '12)*, pages 6–9, Dublin, Ireland, June 2012.

[22] D. L. Johnson, V. Pejovic, E. M. Belding, and G. van Stam. Traffic Characterization and Internet Usage in Rural Africa. In *Proceedings of the 20th International Conference Companion on World Wide Web*, WWW '11, pages 493–502, Hyderabad, India, May 2011.

[23] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a Social Network or a News Media? In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 591–600, May 2010.

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

[25] H. Molyneaux, S. O'Donnell, C. Kakekaspan, B. Walmark, P. Budka, and K. Gibson. Social Media in Remote First Nation Communities. *Canadian Journal of Communication*, 39(2), 2014.

[26] D. Narayan-Parker. *Bonds and Bridges: Social Capital and Poverty*, volume 2167. World Bank Publications, 1999.

[27] T. Silva, P. de Melo, J. Almeida, J. Salles, and A. Loureiro. A Picture of Instagram is Worth More Than a Thousand Words: Workload Characterization and Application. In *2013 IEEE International Conference onDistributed Computing in Sensor Systems (DCOSS)*, pages 123–132, Cambridge, MA, USA, May 2013.

[28] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. The Increased Bandwidth Fallacy: Performance and Usage in Rural Zambia. In *Proceedings of the 4th Annual Symposium on Computing for Development*, ACM DEV-4 '13, pages 11–21, Cape Town, South Africa, December 2013.