The Internet as a networked system has been rendered more complex than ever before as human endpoints are grafted into the system via increasingly pervasive and personalized networked devices. According to the United Nations, the Internet is a transnational enabler of a number of human rights, and as such, access to the Internet has been proclaimed to be a basic right unto itself. Unfortunately, even as networked devices have become ubiquitous, access to the Internet has not.

In my thesis work, I have employed network analysis techniques to identify and understand problems surrounding Internet accessibility in a multitude of environments. In particular, I have focused on accessibility in communities where connectivity is not ubiquitous and network infrastructure fails to meet usage demands. By understanding network operation in these challenging contexts, we can design innovative network solutions that are centered on human objectives rather than infrastructural boundaries and constraints.

My research interests span the areas of network and data analytics and wireless network system design. Specifically, my thesis work can be categorized into the following topics:

1. Analysis of community network data to understand the interaction between network usage, performance, and underlying social structures [8, 6, 10, 9, 4, 5, 11, 12, 13, 14].
2. Data-driven design of community-centric systems operating in challenged environments [6, 9, 1, 7].

1 Network and Data Analytics

We are now living in an era where data is generated at unprecedented rates. Vast data sets detailing numerous interactions are readily available to those who seek deeper insights into the complex operations and behaviors of people and systems. Analysis of data enables us to predict trends over time and space, such as climate conditions, patterns of violence, and spread of disease. Moreover, synthesis of different data sets provides insight into hidden phenomena, such as the rise and development of social movements or information pathways through a community. Nuanced and meaningful mining of compound data sets requires research techniques that account for both the context of the data sets and the structures contained therein. Network analysis provides a set of techniques that meets both of these criteria. In my thesis work, I combined my background in wireless computer networks with data mining techniques to understand interactions between individuals, communities, and digital content. Pivotal to my research has been the Tribal Digital Village (TDV) network, a tribal-operated wireless Internet service provider that provides access to 13 different Native American reservations inhabited by 17 distinct tribes. By collecting traffic traces from this network and combining it with additional network data collected using Web scrapers and social media APIs, I have had the unique opportunity to study the Internet usage of communities connecting from a variety of cultural, geographical, and socioeconomic contexts. My findings have led to critical insights about the complex interaction among factors such as human usage, technical protocols, and wireless interfaces. Furthermore, I have leveraged these insights to design innovative human-centered information and communication systems.

Web usage: Web usage patterns illuminate challenges and opportunities that present themselves as connectivity becomes more pervasive. One of my studies represents the first trace-based characterization of Web usage in a Native American tribal context, where fewer than 10% of Native Americans had home Internet access (at the time of the study) []. Critically, this study provided initial insight into the strong digital-social connections that exist between members of tribal communities; these observations ultimately led to the design of several innovations in the design of community network systems. In this study, I collected and analyzed 5.5 TB of network traffic.
representing 52.8 billion packets. When examining the most-accessed Web domains in the TDV network, I found that in contrast to the general U.S. Web preferences, users in the TDV network engaged more frequently with social media (i.e., Instagram and Facebook) and online gaming (i.e., Xbox Live and PlayStation) services. A deeper investigation of the relationship between network performance and usage patterns of three of the most pervasive Web sites (i.e., Instagram, YouTube, and Netflix) revealed that smaller flows with shorter durations were more likely to complete successfully, likely because 98% of devices downloading this content were classified as mobile and were more prone to flow disruption due to mobility between spaces of high connection quality and low connection quality. Ultimately, this work pointed to the fact that platforms enabling social connectivity are important in tribal networks and there is a need for architectures that better support Web experience as users bridge spaces of high connectivity and low/no connectivity. This study was published in the proceedings of the world’s premier conference for research on the World Wide Web [8].

In a subsequent study of the TDV network, I collected and analyzed 37.8 TB of Web traffic with the goal of measuring the similarity in Web preferences among communities of various geographic scope. I examined preference similarity based on Web files downloaded and Web domains accessed by individual households, aggregate reservation communities, and the TDV network as a whole. Using a statistical rank comparison technique, I found that the Web domain preferences of households are not significantly different from the Web domain preferences of their corresponding reservation community when examining the top $k = 25$ Web domains (though similarity begins to decrease significantly larger values of $k$) and reservations as a whole accessed an average of 67% of any household’s preferred Web domains. Ultimately, this analysis demonstrated that by leveraging similar content interests, Web content accessed by online users in a community could be sufficiently representative of content desired by offline community members. This study is currently in submission to a top computer science conference [9].

Social networks: Previous survey-based studies of Internet usage in tribal communities revealed the critical importance of online social media platforms for preserving and revitalizing Indigenous culture. My own studies of Web usage in the TDV network highlight the prevalence of social media in tribal networks. By mining 6 months of Instagram and Web traces generated by 254 Instagram users connecting from the TDV network, I was able to collect data about 1.2 million Instagram content objects, 47,645 content creators, and 12,615 social interactions. I found that TDV users interacted with content generated by other users in the TDV network 46.6× more frequently than content generated outside the TDV network. An interesting finding of the study was that TDV Instagram users only interacted with 0.55% of the Instagram content available to them. Moreover, most users typically engaged with content 1–24 hours after its publication. In light of this, I found that the bandwidth required to support only content engaged by users was 0.195 kbps, a 99.7% reduction over the bandwidth required to support all accessible content. These findings demonstrated that only a small portion of accessible content is relevant to users and normal usage involves some delay in the receipt of relevant content by interested users. In addition to examining content interactions, I also investigated the overlap between TDV Instagram users’ social networks. I discovered that while the similarity between the users’ social networks was very low, content that received interactions from TDV users was accessible to a significantly larger portion of the community than content that received no interactions. This indicates that even when social networks of community members are dissimilar as a whole, users are generally more interested in the social content that is interesting to their local community of users. Ultimately, this work was published and presented at a top computer science conference for understanding community collaboration and social media analysis [6].
In addition to studying social network usage in a specific tribal community, I led the first network scientific study of Native American engagement with political content on a social media platform. Motivated by the lack of Native American participation in civic elections, this collaboration crossed multiple disciplines and provided a unique mixed methods approach to understanding political content diffusion across community and topical networks on Twitter through the lenses of connective action and media richness theory. One of the key findings from this study was that certain social communities comprised of Native American advocates and their followers remained relatively stable over time despite the ephemerality of specific causes, issues, and initiatives they gathered around. The connective tissue that enabled the adhesion of these communities was an identity-based hashtag, #indigenous, which I discussed as the mechanism for instigating a larger online social movement. Thus this work points to the idea that content annotations support community-building over time. This work will appear in the proceedings of a top computer science conference [10].

2 Data-driven System Design

By the beginning of 2016, approximately 51.3% of the global population lacked access to the Internet [3]. Even in a developed country like the U.S., only 15% of Native Americans living on tribal lands have Internet access [2]. Geographic obstacles, low-population density, socioeconomic factors, and historic oppression are all contributors to this digital divide. Motivated by these challenges and the findings of my work with network characterization, I design innovative network systems that are centered on human usage and content needs. Critically, these systems have been evaluated using traces of actual network usage. While I focus specifically on designing communication systems for resource-poor communities, my research contributes to best practices for general contexts where systems seek to become more community- and user-centric.

**Repurposing FM radio:** Inspired by my findings of the importance of social media sharing, overlapping social media content interests, and the lack of ubiquitous Internet connectivity on tribal lands, I proposed a social media content delivery system that broadcasts the most relevant content via the radio broadcast data system (RBDS), which is the data subcarrier of FM radio. The proposed system leverages the advantages of RBDS technology: it propagates data over long distances, it is robust to transmission errors, and FM infrastructure is already ubiquitous on tribal lands. Since RBDS is an extremely low-bandwidth broadcast technology, I proposed a content scheduling algorithm which leveraged users’ social connectivity and round robin scheduling in order to allow for fair sharing of the broadcast medium. Specifically, I introduced an algorithm that prioritizes content based on the cumulative clustering coefficient associated with the creators of content. Using six months of Instagram traces generated by TDV users, I evaluated the fairness and coverage provided by the system. My scheduling algorithm was able to provide half of the users in the community with 81% of their Instagram content requests and 35.5% of the 1.1 million requested Instagram photos were delivered to users over the six month simulation period. The key finding of this work is that community social networks can be leveraged to select relevant content for distribution over extremely low bandwidth. This work was published and presented at a top-tier computer science venue [6].

**Community content delivery:** Based on my findings of similarity in the Web browsing preferences within reservation communities, I proposed a community-based CDN that delivers content opportunistically [9]. This system was designed specifically to serve the households lacking access to the Internet by delivering content to mobile users on behalf of their entire household. Using statistical models to simulate opportunistic data rates and a state machine to model rural user mobility, I was able to play a week’s worth of Web traces to evaluate 1,680 unique configurations of the proposed system. In particular, I sought to evaluate the performance of various content
filtering mechanisms as means of prioritizing content delivery over opportunistic connections. My study revealed that even when the data rate available to a household via opportunistic connectivity averages at 0.2 kbps (2.2 MB per day), 37.1%–69.8% of a household’s daily expected content could be delivered using the proposed system configured to schedule content based on a combination of collaborative filtering and user preferences.

3 Future Work

Envisioning my future research, I am interested in pursuing the design of networks and information propagation mechanisms that support community building and pervasive content engagement. My near-term research agenda includes the following:

**People-centric networks:** Computer networks have long been device-centric in their design. Web-based services and applications assume personal devices operating in spaces with ubiquitous, high-speed connectivity. However, these assumptions are far from universally applicable: over half the global population lacks access to Internet connectivity, bandwidth capacity is unevenly distributed over heterogeneous networks, many people have multiple personal networked devices, and in some contexts personal devices are used communally. To better facilitate human-centric content transactions in this context, I am interested in synchronizing heterogeneous community networks to form people-centric networks. While the system architectures I propose in my thesis have focused on coordinating infrastructure to provide access to disconnected users, they have emphasized interaction with personal devices rather than people and have focused on content download and delivery rather than upload and annotation. I intend to explore the nature of an integrated personal identity that interacts with a heterogeneous community network consisting of Wi-Fi and cellular endpoints, microwave and fiber backhauls, and community FM broadcasts. By emphasizing identification at the granularity of “person” rather than “device,” it becomes possible to tailor network interactions according to explicit and implicit personal preferences as well the context of network connectivity. Consequently, this requires collection, mining, and analysis of individual usage data across numerous media platforms, devices, and networks in a variety of contexts. Given the personalized nature of people-centric networks, I also plan to investigate the unique privacy measures required to protect personal identity and information in such systems in a way that is sensitive to potential network limitations such as high latency, low bandwidth capacity, and partitions. Moreover, I am interested in investigating the role of using low-bandwidth, FM radio as a way to coordinate heterogeneous networks over a wide area of partially connected community infrastructure.

**Community media platform:** In the mass media platforms of the past century (e.g., broadcast radio, newspapers, magazines), content was generated for a community audience by community members, celebrities, experts, and organizations. As an increasing volume of our content needs are met online, content is increasingly curated by proprietary algorithms and curation is tailored towards individuals. I am interested in combining the efficiency of algorithmic content selection with the wisdom of location-based communities through the design of a community media platform, which mediates the content interests of individual community members with the content agendas asserted by organizations representative of the community. Specifically, I plan to mine individual and organizational media usage in order to coordinate to community content goals (rather than isolated individual/organizational content interests). Additionally, I am interested in developing methods by which geospatial annotations can be integrated into community media platforms as another source of information and how appropriate information generated on a community platform can be integrated into physical space by creating such annotations where they might be helpful to community building and community development.
Social learning networks: In many poor and rural communities, access to the Internet corresponds to access to educational opportunities. While MOOCs and online curricula provide access to high quality educational content, classically under-served demographics (e.g., rural, poor, lack of education in family) are failing to experience high rates of success with these programs. I am interested in investigating how social networks can assist individuals in their achievement of learning goals. Using network analytics to analyze participation in educational forums and social groups, it is possible to characterize the formation of sub-communities around educational content, the propagation of information within and between those sub-communities, and the influence that operant conditioning mechanisms have on information diffusion and success propagation. Moreover, I am interested in understanding how the roles of various individuals in a social network encourage or discourage an individual’s progress towards their goal. Furthermore, I am interested in designing an adaptive feedback algorithm for individuals that maximizes their progress towards their goal based on historic responses to feedback and explicit user preferences.

References


