(Hyper) local News Aggregation: Designing for Social Affordances

Andrea Kavanaugh[1,2]
Sloane Neidig[2]
Ankit Ahuja[1,2]
Samah Gad[2]
Manuel A. Pérez-Quiñones[1,2]
Naren Ramakrishnan[2]
John Tedesco[3]

[1] Center for Human-Computer Interaction
[2] Department of Computer Science
[3] Department of Communication Studies
Virginia Tech
Blacksburg, VA 24060

[kavan, sloane12, ankit88, samah, perez, naren, tedesco]@vt.edu

ABSTRACT
Being able to debate, reflect, form opinions, consider counter evidence and make informed decisions is a foundation of civic life in democratic societies. Government benefits from broad participation in collective decision making in terms of sustainable outcomes (e.g., greater consensus) and quality of life in our cities and communities. These collective decision making capabilities are undermined by a strange combination of diminishing (or extinct) local print media, especially local newspapers, and by its obverse, a plethora of information and communication opportunities that are scattered across numerous disparate and decentralized websites and resources (e.g., webpages, RSS feeds, social network software, email, listservs, podcasts, tweets). To address this problem, a number of news aggregators have emerged that capture local content from dispersed sites. To facilitate civic engagement, these sites also need to support social interaction and information exchange. In this paper we review the state of the art in local news aggregation in the US and their support for social affordances (social trust, networks and interaction) that are essential to civic participation. We present a prototype we have developed for local news aggregation that supports social affordances. We summarize briefly the design strategies and techniques (e.g., algorithms) we used to cluster topics and user generated content derived from existing local sources. The prototype should lead to a replicable model for other US communities.

Categories and Subject Descriptors
H5.m. [**Information interfaces and presentation (e.g., HCI)**]: Miscellaneous.

Digital Government Research Conference (dg.o'12), May 7–10, 2012, College Park, MD, USA.

General Terms
Algorithms, Design, Human Factors.

Keywords
Civic engagement, news aggregators, social affordances.

1. INTRODUCTION
Collective decision-making is central to the quality of life in US communities, towns, and city neighborhoods whether it is routine and long term planning or timely and critical follow up to crises. This painstaking and often divisive civic process is especially difficult for the many towns, counties and rural areas that have little or no coverage in print media, such as a local newspaper. For these small cities, city neighborhoods and towns, most relevant local information is only available by word of mouth or through electronic communication disseminated by information gatekeepers (i.e., various stakeholder and interest groups). How can local residents and organizations engage in effective collective problem solving when there is very limited information flow and limited opportunities for interaction with a broader spectrum of fellow citizens and organizations?

Increasingly, information technologies are playing a role in helping communities share information (and misinformation) and in fostering interaction and communication among local organizations and residents. The availability of web-based information and of opportunities for citizens to share ideas and opinions have generally led to increased civic awareness and for some citizens increased involvement [1-4].

Data from Pew's biennial news consumption survey indicated that Americans are spending more time following news – and online sources are increasing in prevalence as primary sources. The report indicates that slightly more than a third (34%) of the public went online for news "yesterday," putting online news ahead of the daily newspaper and at about the same level as radio. When the "online" definition is extended to include cell phones, email, and social networks, the percentage of Americans who indicate they received news yesterday from the Internet or a mobile source increased to 44% [5].

Furthermore, Pew's latest data regarding how people learn about their community demonstrates that the Internet plays an increasingly important role. According to Pew, for "79% of

Americans who are online, the internet is the first or second most relied-upon source for 15 of the 16 local topics examined. For adults under 40, the web is first for 11 of the top 16 topics—and a close second on four others" [6] (p. 2).

In fact, there has been such a plethora of online information sources, even at the local level, with the advent of user generated content and social software systems, that it has become increasingly difficult for interested citizens to identify and track all the appropriate information of interest and citizen comments or other responses to a given problem or set of issues. In a small town or large city neighborhood, for example, on the same topic, information has become widely scattered across multiple websites (e.g., government, schools, community groups, news organizations) as well as web-based user generated content (e.g., posts on public Facebook pages, individual blogs, Twitter messages or image collections).

To try to address this problem of the wide dispersion of sources for local information and citizen discussion, various computer-mediated systems and data mining techniques have begun to collect and update content automatically from diverse sources and to build in social affordances in 'news aggregation' websites.

Aggregators are playing an increasingly important role in news delivery for online audiences. Instead of going online to a single local news source, 86% of 18 to 29-year-olds "often" or "sometimes" rely on search engines and aggregators (e.g., Google News, Yahoo!) to search for news [7]. If this trend in news searching continues to characterize news consumption, communities have an opportunity to create a virtual space based on aggregator techniques where citizens not only gather for more comprehensive coverage of local news and information but also for social interaction (e.g., content sharing and discussion) about that news.

This paper presents an analysis of existing types of local and hyperlocal news aggregation, and examines the different social affordances that they support in order to help foster greater interaction and civic engagement.

Local news is the set of announcements, events and other information that pertains to one's city or town and environs; "hyperlocal news" relates to the immediate area around a resident, such as "my city neighborhood" or "these five square blocks" or "my area of town". Some websites have both local and hyperlocal information, especially when residents contribute content. For simplicity's sake, we use the term 'local' in the paper to mean both local and hyperlocal content. We also use parentheses -- (hyper) local -- for shorthand to designate local and/or hyperlocal.

Affordances are actionable properties between an agent (i.e., a person) and the world [8]. For example, a door handle affords the act of opening a door. In a computer system, various features, such as a keyboard, mouse, and screen display afford pointing, looking and selecting. It is not sufficient to build affordances into a design; it is necessary that a user actually *perceives* the affordances that are offered, e.g., that clicking with a mouse will result in a useful or meaningful outcome [9]. In human-computer interaction, affordances have been classified into subcategories including cognitive, physical, sensory and functional [10]. We seek to extend Hartson's classification into the social realm.

2. DESIGN for SOCIAL AFFORDANCES

Using a mouse for clicking on a computer screen is an individual affordance. *Social* affordance, by contrast, is the quality of artifacts in any physical or virtual space that invite or facilitate interaction among users [11]. Social affordances in a website could include tagging, ranking and sharing. Sometimes the user has specific individuals in mind (friends on Facebook) when they tag a photo, for example. Sometimes, the social interaction is with unknown people, such as correcting an entry in Wikipedia. Evaluations of social affordances in websites have focused on some of the following features [12-15]:

Tagging: web-based tagging systems let users annotate a particular resource, such as a web page, a blog post, an image, or any object with a freely chosen set of keywords. Tags usually facilitate finding and sharing content.

Activity Streams: flowing commentaries on users' actions on different sections of a site; this feature is helpful in discovering content.

User Profiles: self-disclosed information about the identity of individual users, which can aid in discovery of common interests and relationships.

Comments: the primary conversational medium on social network sites that often indicates social relationships; also present on some public sites, such as news media (e.g., New York Times) and Facebook pages of organizations (e.g., local government, community groups).

Ratings and Votes: reputation systems for users that can facilitate interaction and trust and limit aversive behavior.

These social affordances allow users to be aware of other users' opinions, thoughts and feedback and, in so doing, help to engage users and build social connections [12].

3. LOCAL NEWS AGGREGATORS

Aggregators compile information using Real Simple Syndication (RSS) feeds from diverse Internet sources, based on the purpose of the aggregator. For example, a blog aggregator site lists on one site all the blogs that bloggers have requested (and site owners have approved) to be listed there. News aggregators can collect news articles based on specified topics from various websites or Web applications into one site. Aggregators can also collect user-generated information, ranging from comments on news stories to user-written articles (e.g., blogs and other social media, such as messages, called 'tweets' from Twitter). Aggregators can pull in any content that is made available as an RSS feed. This compilation of sources gives users easy access to continuously updated local information from diverse sources.

Aggregation is an important feature of websites that seek to collect a lot of continuously updated content on an automatic basis. The automatic updating of content keeps the costs of maintaining and sustaining an aggregator site very low because the actual content is being generated elsewhere by others.

We reviewed and compared different types of self-proclaimed local (and hyperlocal) news aggregators in the US, with a view to determine the extent to which some of the most common and popular sites are in fact set up as aggregators, and are acquiring local content.

We found that for some of these websites to claim to be hyperlocal aggregators is, in fact, misleading. Some sites are not truly (hyper) local news aggregators because they are either just 'local or hyperlocal news' (and not 'aggregators') or just 'news

aggregators' and not local or hyperlocal. Examples of the misnomer types include Patch, EveryBlock, News360, Google News, iBrattleboro, and RVANews. We include sites that are limited to one location where the site itself (e.g., Blacksburg Electronic Village) is not aggregating the content it provides. In Section 4 we discuss true hyperlocal news aggregators and the extent to which social affordances are baked into their design.

3.1 (Hyper) local News but not Aggregator

A "(Hyper) local News" but not "Aggregator" site is a website that displays local news, but does not aggregate that content from existing local sources (Figure 1).

Figure 1. Misleading (Hyper)local News Aggregators

In 2010, Patch was delivering local news to about 800 towns in the US [16]. The company, based in New York, was founded in 2007 by Tim Armstrong, Chairman and CEO of AOL [17]. AOL purchased Patch for $7 million in 2009 [18] and has been paying local individuals to serve as editors to create and maintain most of each local site's content. Each of the 800 towns covered by Patch has its own editor. Users are allowed to write and submit their own articles; the local editor reviews and decides whether to approve these articles before they appear on Patch.

The AOL website "Patch" collects its content from editors and users, but not from official Internet sources (e.g., government, news organizations). It is a time commitment for the editors, who seem to write the majority of the news articles posted on the website every day. It is hard to imagine that this model is sustainable in the long run, since employing an editor for each town can become costly.

EveryBlock, another popular hyperlocal news aggregator, extends to 16 major cities in the US [19]. EveryBlock was founded in 2007 by Adrian Holovaty and, was acquired by MSNBC for an undisclosed price in 2009. EveryBlock heavily features user discussions regarding local topics. It allows users to sign up and post comments to their neighbors in the city. It obtains content by crawling websites of government departments, mentions of locations in the city in a subset of online media, local services including deals, meetups and real estate, and photos on third-party sites tagged with locations in the city.

Most discussions on EveryBlock revolve around local information and events posted by users. Besides commenting and posting articles and events, other forms of social affordances include subscribing to discussion threads and appreciating another user's contribution to a discussion by "thanking" the user.

3.2 Aggregator but not (Hyper)local News

The second type of misleading 'local aggregator' website is one that aggregates information that is not strictly local (*"News Aggregator", not "Hyperlocal"*). These websites include News360 [20] and Google News [21].

News360 is programmed to pull headlines from other news websites, such as CNN and The New York Times, and post a partial view of each article for the readers. If the readers find the snippet interesting, they can click a hyperlink that will take them to the original source of the article. News360 also has the ability to group articles by topic, from fashion to technology, and allow users to select topics that most interest them. Unfortunately, the focus on information is not local, so it cannot be used as an ideal hyperlocal news aggregator.

Google News is similar to News360 in that it is programmed to aggregate content from other major news websites. It can also group similar articles from various sources together, as a further step to collecting information for readers. Google News claims to favor local articles and searches [21] although non-local news is part of the mix.

Lastly, there are some sites that focus on local community news and information, but they do not automatically aggregate content that is, they are unique sites (*"Limited to One Location"*) such as iBrattleboro, the Blacksburg Electronic Village (BEV) and RVANews. The iBrattleboro website promotes local news "written by and for the people of Brattleboro" (Vermont) [22]. iBrattleboro relies only on users for information. This limitation may provide a good sense of community, but does not give users access to local information available through official sources (e.g., community organizations, town government, schools).

RVANews (serving the city of Richmond, Virginia) like iBrattleboro relies heavily on user-posted information [23]. However, RVANews also employs staff to post articles and information. The website tries to foster a sense of community by posting articles about local events such as the area's "Garden Wars" (a competition to see who has the greenest thumb). Nevertheless, users are still limited to a subset of information because the site does not aggregate additional articles from the many locally relevant Internet sources. In the case of the Blacksburg Electronic Village (BEV), focused on the town of Blacksburg, Virginia and environs, content is generated by a mix of BEV staff and users who maintain their own pages on the site (primarily local community groups) [24, 25]. The BEV site has links to many local official websites, but does not aggregate or display that content on its own site. Social interaction is limited to several local newsgroups (e.g., Education, For Sale,

Environment) most of which are fairly inactive, except for the 'For Sale' newsgroup[26].

4. LOCAL NEWS AGGREGATORS

Some sites are genuine hyperlocal news aggregators because they are both covering local news and do this through automatic aggregation. Examples include Topix, Fwix, and Outside.in (Figure 2). On the aggregator side, Topix collects information from over 50,000 sources and provides users (from over 360,000 locations) with local news [27]. Topix offers social interaction through discussions forums on the front page of each community website (See Figure 3 showing Topix Forum for Blacksburg).

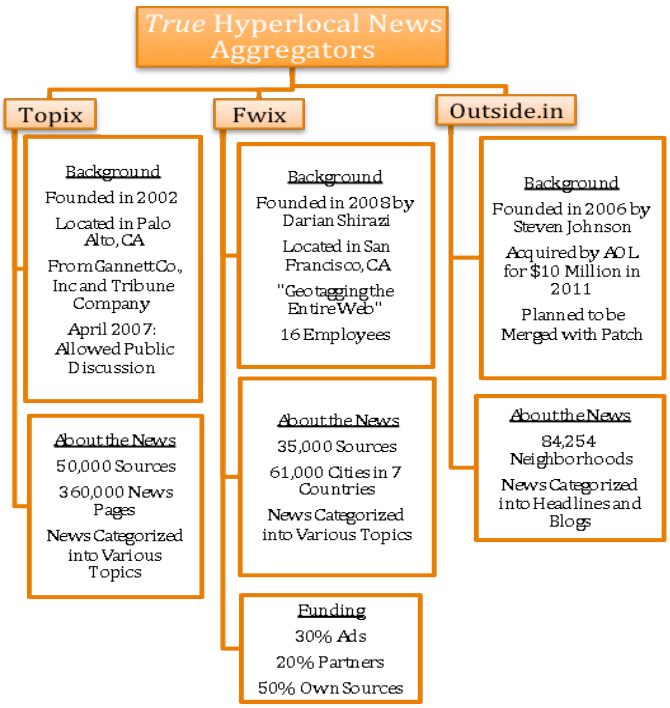

Figure 2. True Hyperlocal News Aggregators

The main drawback of the organization of the discussion forums on the site is that the posts are essentially a laundry list of the most recent topics or narrowly focused posts.

Figure 3. Discussion Forums on Topix.com for Blacksburg

Fwix is another example of a website that aggregates local content and encourages user participation. Fwix's slogan is "Geotagging the Entire Web" [28]. By geotagging available information from over 35,000 sources, Fwix is able to distribute local information to 61,000 cities in seven countries. Similar to News360, Fwix posts only part of an article and will direct users to the original source if they wish to continue reading. Besides news articles, Fwix also tries to capture other types of content from third-party services including photos, events, real estate and tweets geo-tagged with locations within the city. Fwix encourages user participation by posting events and allowing users to comment on articles.

Outside.in aggregates primarily content from local newspapers and blogs, but does not accept other types of user-content or social interaction (no comments, discussions, or events postings). The posted articles are divided into two categories: headlines and blog posts. This simple separation makes it harder for users to find information on specific topics because "headlines" and "blog posts" are not descriptive of each article's content. It also separates a news article from comments about that article by bloggers. Outside.in reached over 80,000 towns and neighborhoods in the US. Some of the stories that get caught up in the aggregation of RSS feeds (e.g., from the local television website) do not pertain to the local area itself.

In comparing these sites, we note that Topix sorts its news into very specific topics, from "Jobs" to "Politics". Fwix sorts news and information, but only into more general categories like "Photos" and "Events". Outside.in has essentially a laundry list of headlines pulled from individual news sources. That is, news stories are not grouped by topic or related blog posts. In the future, Outside.in may aggregate and present news stories differently if it merges with Patch. Both companies have been acquired by AOL and may be joining forces in order to control the hyperlocal media market.

On the social affordances side, each of these aggregators has some social affordances built in, primarily message posting. Yet, they do not offer a fuller range of features, such as tagging, ranking, activity streams and sharing that could help foster greater interaction and engagement over time.

We have examined news aggregators with a view to building on or modifying the designs and strategies that most effectively seem to address our initial concern: the plethora of local content online that is scattered across many local websites and user

generated feeds (e.g., tweets, Facebook posts, photos). In contrast to the examples above, we seek to aggregate content from existing individual local websites, blogs, social network systems or image collections and automatically cluster that content based on a common topic classification.

5. DATA MINING RESEARCH

Two broad classes of data mining research are applicable to social computing contexts.

5.1.1 Network Analysis

With the rapid penetration of social networking sites, many analytical techniques have been used to study both explicit networks (e.g., a friendship network on Facebook.com) and implicit networks as captured through indirect information such as ratings, feedback, and views (e.g., recommender systems such as Amazon.com [29]. Algorithms for mining community structures in networks [30, 31] are now considered mainstream and have been generalized significantly, e.g., to find "signed" networks featuring both positive and negative interactions [32]. The dynamics of interactions, network effects, and cascades of information propagation [33] have been studied from both a sociological standpoint and through an algorithmic lens, as a means to characterize large dynamic networks. A key insight from accompanying empirical studies is the densification of the networks over time and the simultaneous shrinking of the average diameters of the graphs [34]. Diffusion processes have been used to characterize the spread of rumors and postings over online social media and contrasted with diffusion processes in physical organizations and other traditional media [35]. It is hence fair to say that data mining techniques have significantly extended the tools available for studying and characterizing networks.

5.1.2 Topic Modeling

A second class of data mining techniques addresses the large availability of text and aims to characterize it to understand trends and events in social contexts. A fundamental building block to such research is the use of topic modeling on harvested news sources (including newspaper entries, blogs, Twitter feeds, and public Facebook pages [8, 9, 12, 13, 36] obtained via RSS feeds and HTML scraping.

One of the basic algorithms available for topic modeling is Latent Dirichlet Allocation (LDA) [37] which prefers a statistical model for discovering the abstract topics in a collection of documents. In LDA, a document is considered as a distribution over topics and a topic is in turn a distribution over words.

LDA posits that documents are generated in two stages. In stage one, a distribution over topics is chosen. In stage two, a document is generated by first choosing a topic, and then selecting a word from the chosen topic's distribution over words. This can be expressed mathematically as a joint probability distribution over the observed documents, topic structure, per-topic-document topic distribution, and the per-document per-word topic assignment. The following equation is a formulation of the generative process of LDA.

$$p(\beta_{1:k},\theta_{1:D},z_{1:D},w_{1:D}) = \prod_{i=1}^{k} p(\beta_i)\prod_{d=1}^{D} p(\theta_d)\left(\prod_{n=1}^{N} p(z_{d,n}|\theta_d)\,p(w_{d,n}|\beta_{1:K},Z_{d,n})\right) \quad (1)$$

Here, ß denotes the topics, K represents the number of topics, ɵ is topic proportions, D represents the number of documents, z is topic assignments, w is the words, and N represents the number of words.

Inference in LDA involves estimating the posterior probability or conditional probability for a given set of documents, using Equation 2. Here, we divide the joint probability from equation 1 by the probability of seeing the observed corpus under any topic model.

$$p(\beta_{1:K},\theta_{1:D},z_{1:D}|w_{1:D}) = \frac{p(\beta_{1:K},\theta_{1:D},z_{1:D},w_{1:D})}{p(w_{1:D})} \quad (2)$$

In our research, topics are discovered from data collected daily. Two key parameters of the topic modeling approach are the number of topics (5) in the LDA formulation and the number of terms (20 per topic) used for representing documents. The latter was selected by applying traditional IR cutoff thresholds for significance and the former was selected by a manual inspection of the LDA output by domain experts. More systematic model selection procedures are underway. A sample output in the form of a tag cloud is shown in Figure 5. A tag cloud represents a topic and the size of a word in the cloud is proportional to its relative frequency within the topic.

Figure 5. A sample tag cloud denoting a topic about news articles pertaining to the 2011 Virginia Tech football season.

5.1.3 Ongoing work

We are currently addressing four problems toward greater integration with the objectives of VTS. The first is the dynamic tracking of topics over time, in particular the capture of fleeting and transient topics as public sentiment changes rapidly. The second is an automatic labeling algorithm to be able to comprehend discovered topics for the average user.

A third idea we are pursuing is the understanding of how information flows through a small community such as ours and what factors contribute to the spread (or decline) of an idea. Such a modeling of information flow can be addressed through information genealogy methods [38] that is, the problem of automatically uncovering the origination of ideas. Solving this problem will help us answer questions like: which documents (e.g. article or blog) influenced each other, how did ideas spread over time, and which documents (or authors) were most influential? It will also help in identifying opinion leaders in our community. There is also the possibility that people who are not

opinion leaders may gain influence nonetheless, due to domain expertise.

The final problem is to characterize public perception and sentiment about an issue in the community. Current sentiment analysis techniques [39] are very narrow in how they define sentiments (e.g., positive or negative moods). We propose to model *affec*t, by which we mean not an emotional state of mind, but a broader vocabulary than sentiments. For instance, affect modeling can capture a broader range of dichotomies e.g. liberal vs. conservative, small government vs. big government, and public school vs. private school. The main goal behind detecting affect is identifying multiple and possibly contradicting aspects associated with an idea. Using affect modeling, we can obtain a richer understanding of public sentiments about an idea.

6. DESIGN of VIRTUAL TOWN SQUARE

In prior work leading up to the design of the Virtual Town Square (VTS), we modified software available through social network (SN) systems [40, 41] to create a local conversation system we called Colloki.[i] VTS will use a similar set of social and organizational features available in SN sites to support discussion [42-44] derived from dispersed blogs, tweets, and SN sites or posts within VTS. VTS also provides RSS feeds by sections, recent stories, particular users, tags, and many of the other organizational schemes in the system. This allows users to be alerted when new information is available in whichever way they find most useful. We will also provide options to receive updates via emails, SMS, Twitter, and even systems like '(Blacksburg) First Alert' that use emails and tweets to disseminate public announcements. A mock up of a VTS page appears in Figure 4.

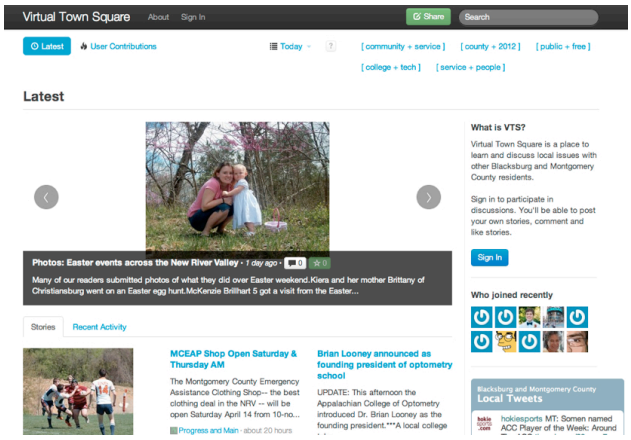

Figure 4. Mock up of VTS

VTS uses a framework based on social affordances and social actions, key aspects of social networks. As noted above, we accept the general definition of social affordance as the quality of an artifact in any space which invites and facilitates social actions by the participants in that space [14, 42]. Social action is closely tied to the concept of social capital [45]; that is, essentially any behavior that facilitates individual or collective action, generated by networks of relationships, reciprocity, trust, and social norms (e.g., information sharing, commenting, and discussion). VTS extends the early capabilities of Colloki by including content aggregation functions, network analysis, data mining and personalization features. The social affordances in VTS have the potential to increase heterogeneity in discussion networks, allow for greater flow and wider distribution of information, and to incorporate a broader, more diverse user population.

Our emphasis is on local issues where finding news sources and online discussion is often hard for several reasons. First, noted above, there are fewer resources devoted by news agencies to local issues. Second, online discussion at the local level often occurs in closed community groups and it is difficult for other citizens to join and become active participants. Third, social software systems that allow users to gather online and discuss their interests (e.g., Digg, Slashdot) work in part due to the large number of people participating. The effectiveness of these sites is that they attract a lot of people, thus increasing their popularity. But for local participation, the number of participants will always be lower, as only people with local concerns are participating. Automated solutions and aggregators are not sensitive enough to pick up material that is truly relevant. Often automated services are too simplistic, doing mostly "surface" checks, for example matching "Blacksburg" to identify local news. Simple search aggregators often return stories where "Blacksburg" is mentioned by coincidence, but the story is about another topic (e.g., "Joe worked as manager of a restaurant in Blacksburg").

A solution is needed that: a) does not depend on thousands of users participating in the sites, b) does not depend on superficial ways of identifying relevant information, c) provides support not only for politically active citizens, but for the less politically active and for lurkers, and d) makes use of Web 2.0 concepts (i.e., content syndication, tagging, user-provided content and organization). VTS seeks to aggregate news and local information in such a way that it becomes the "hub" of local interaction and discussion. The goal of aggregating information is to have a self-sustaining, self-organizing system with minimal oversight and maintenance.

7. PROTOTYPE OF VTS

The VTS prototype captures news articles daily from a finite set of RSS feeds that include local news websites, government sites, local groups' and organizations' blogs and individual blogs. This list is incrementally growing as we interact with local groups and influential individuals to discover more locally relevant content. After capturing these news articles, we run the topic modeling algorithm on them to get a set of topics for the day. These topics and their articles are then posted on the VTS website.

7.1 VTS Website

The VTS website posts the news articles collected each day and displays the topics discovered using topic modeling for the day. Each topic is labeled using the top two words belonging to that topic. This way of labeling topics is a placeholder, until we come up with a better, user-friendly way to label topics. The landing page of the site shows all the articles displayed in the order of publishing or user participation on the site. Another tab shows the recent activity of users on the site (Figure 6).

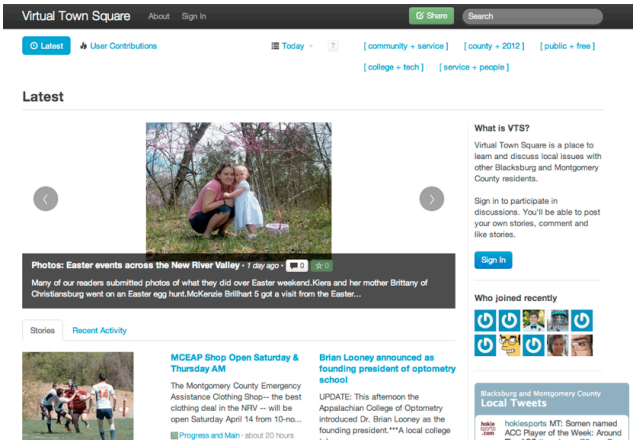

Figure 6. Main landing page of VTS website

Individual topic pages show all the articles belonging to the topic, along with topic keywords as a word cloud in the sidebar (Figure 7).

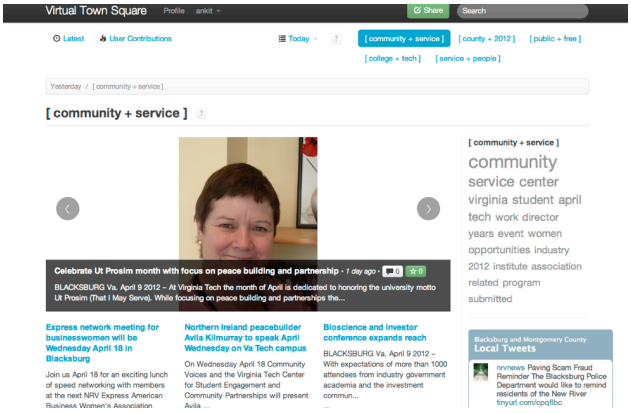

Figure 7. Individual topic page for "community + service" topic

Individual article pages show the article's content and a link back to the original article. The sidebar displays other articles belonging to the same topic and articles belonging to the same source of the original article. Users can comment on the article, like it, and share it on other services including Facebook, Twitter, Google+ and email. (Figure 8).

Figure 8. Individual article page

Users can post their own articles by clicking the site-wide "Share" button in the header of the site and entering the title, text and optionally uploading a photo (Figure 9).

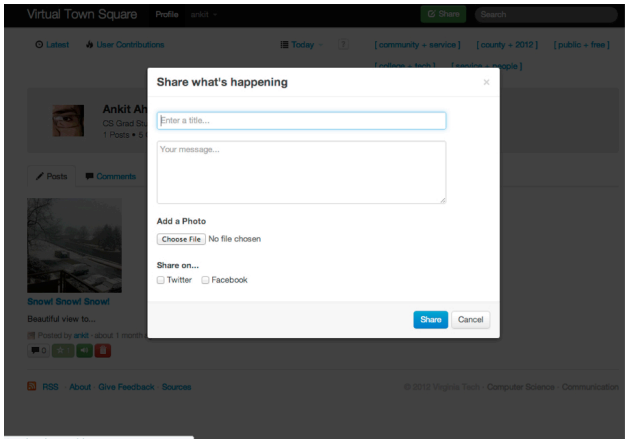

Figure 9. Posting an article on VTS

Each user has a profile page that lists the user's contribution on the site including their posts, comments, and likes (Figure 10).

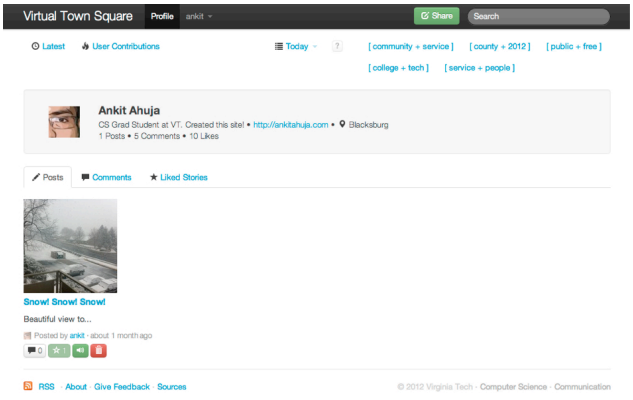

Figure 10. A user's profile page on VTS

We performed a basic evaluation of our system using the Cognitive Walkthrough [46, 47] approach to identify usability problems. We identified the key tasks for our system and answered the following four questions for each task:

- Will the user know how to perform the task?
- Will the user notice that the correct action is available?
- Will the user match the action to the task?
- Will (did) the user get the correct feedback?

Several of the identified design issues were fixed in the next version of the prototype. A couple of examples of identified issues include:

- Lack of location breadcrumbs which will aid the user to keep track of her location while browsing the site
- Ability to create posts was not obvious to the user

We plan to perform future evaluation of our system by inviting a focus group to use it as well as deploying it to a wider audience in the local community of Blacksburg.

We are also experimenting with different ways to present the topics to users including visualizing them (Figure 11). In the figure, each bubble is a topic and each node in a bubble indicates a keyword that belongs to the topic. The size of the central bubble indicates the number of news items belonging to the topic.

Figure 11. Force directed graph used to visualize top topics

A close up of one of the topic bubbles (Figure 12) shows the nodes in the bubble indicating keywords that belongs to the topic. The size of the keyword nodes indicates its frequency of occurrence in the topic.

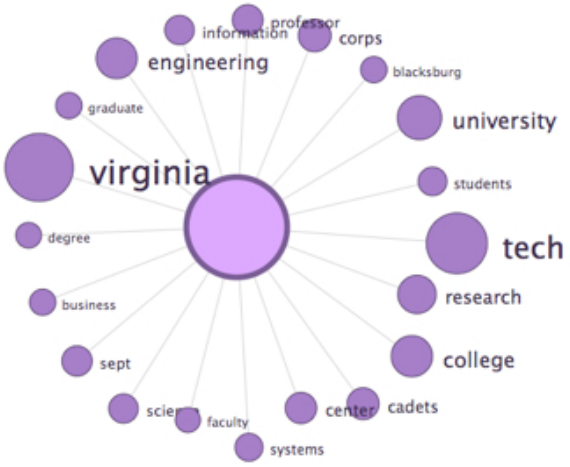

Figure 12. Force directed graph with nodes in single topic

In addition to the auto-generation of topics using data mining, we are exploring alternate ways to navigate the site, including news and discussions overlaid on a map, so that users are able to quickly navigate to articles related to their immediate neighborhood. Another possibility we are exploring is to use stream graphs to visualize the evolution of topics over a period of time. We also plan to capture and integrate relevant discussions from local Facebook groups, Twitter and other relevant pieces of content from third-party services into the VTS website. Through these and similar socially oriented features, VTS seeks to leverage a wide range of social affordances that would help build trust, interaction and engagement.

8. CONCLUSIONS and FUTURE WORK

This paper has presented and compared different local news aggregators, with a view to determining how these are designed and implemented. We have examined which sites and designs are integrating different social affordances that are important for supporting and fostering citizen interaction and engagement.

Several popular websites were true hyperlocal news aggregators (e.g., Topix, Fwix, and Outside.in). In comparing these sites, we note that they impose a topic category on news articles, such as Jobs, Politics, Photos and Events. Each news article was a separate headline that linked to the original news source (e.g., WSLS-TV). These news aggregators provide a single source for each headline and do not cluster together news articles from different sources reporting on the same story. This results in a simple laundry list of different top stories, each with a unique single source for content. Some of the sources for articles are other news websites; others are editors working specifically for a given aggregator to write articles.

Each of these true aggregators has some social affordances built in, primarily message posting. However, user messages or blogs that are included are often not displayed in relation to the news stories they are commenting on. Rather, they are listed as a separate set of content, thereby disconnecting them from the news or issue they may relate to. Thus, the online sources for local news and information and user discussion about goings on about town have proliferated independently of each other, thereby splintering the coverage, as well as citizen discussion, of the same topics.

The social affordances on existing local news aggregators are limited. The sites do not offer many features beyond blog posts or comments, such as tagging, ranking, activity streams and sharing, that could help focus interaction among local residents, organizations and groups around specific topics and thereby help to increase engagement over time.

The VTS website by contrast compiles local news articles from both user and official Internet sources as RSS feeds and automatically determines topics provided by Internet sources or discussed by the public through clusters. That is, the clustered content is displayed in VTS as a topic, covered by multiple sources and including user comments, tags, or photos as part of the topic-clustered content. This design is intended to optimize social affordances in order to encourage both online and offline user participation.

Our approach for the Virtual Town Square (VTS) is to use state-of-the-art data mining techniques to alleviate the need for a "local editor" and to obtain the benefits of a network effect in a small (local) population (or data sample). The result is that the content is continuously and automatically updated from existing sources and is based on collections of articles from diverse sources on the same topic. The automatic updating from existing sources helps to reduce the ongoing costs of providing content. By aggregating articles on the same topic VTS offers more in-depth coverage on a topic from diverse perspectives and emphases.

VTS is designed to be self-organizing and self-sustaining, since content is generated elsewhere by citizens and local organizations. This design should allow users to find content of interest at a single point of entry (VTS), as well as to discuss issues by sharing, commenting, tagging and rating content. Users can also link back to content generated from the original sources to continue discussions that are displayed on VTS eventually as part of the aggregated content.

We seek to strike a balance of a centralized (one-stop source) location for local news while at the same time supporting (and possibly enhancing) the social affordances that lead to online participation. As VTS becomes more stable, following further development and usability testing, it should be replicable in other cities, towns and communities across the US.

9. ACKNOWLEDGMENTS

We are grateful for support from the National Science Foundation for part of this research [SES-1111239]. We would also like to thank our research assistant Kumbie Madondo.

10. REFERENCES

1. Shane, P.M., *Democracy online: the prospects for political renewal through the Internet* 2004, New York, NY: Routledge.
2. Chadwick, A., *Internet politics : states, citizens, and new communication technologies* 2006, New York, NY: Oxford University Press.
3. Coleman, S. and J.G. Blumler, *The Internet and Democratic Citizenship: Theory, Practice and Policy*2009, New York: Cambridge University Press.
4. Kavanaugh, A., et al., *Net gains in political participation: Secondary effects of Internet on community.* Information, Communication, and Society, 2008. **11**(7): p. 933-963.
5. Kohut, A., et al., *Americans Spending More Time Following the News*, 2010, The Pew Research Center for the People & the Press: Washington, DC.
6. Rosensteil, T., et al., *How People Learn about their Local Community*, in *Pew Research Center, Project for Ecellence in Journalism* 2011, Pew Internet & American Life Project and Knight Foundation: Washington, DC.
7. Peer, L., et al., *The Local TV News Experience: How to win viewers by focusing on engagement*, 2007, Northwestern University Media Management Center: Evanston, IL.
8. Gibson, J.J., *The Theory of Affordances.*, in *Perceiving, Acting and Knowing*, R.E. Shaw, Editor 1977, Lawrence Erlbaum Associates: Hillsdale, N.J.
9. Norman, D., *The Design of Everyday Things*1990, New York: Doubleday.
10. Hartson, R., *Cognitive, physical, sensory and functional affordances in interaction design.* Behaviour & Information Technology, 2003. **22**(5): p. 315-338.
11. Howarth, J., et al., *Designing a conference for women entering academe in the sciences and engineering.* Advancing Women in Leadership Online Journal, 2007. **23**.
12. Ahuja, S., *A Tale of Two Sites: An Explorative Study of the Design and Evaluation of Social Network Sites*, in *Computer Science*2009, Virginia Tech: Blacksburg, Virginia
13. Hanrahan, B., et al. *Evaluating Software for Communities Using Social Affordances*. in *International Conference on Human Factors in Computing Systems (CHI '11)*. 2011. Vancouver, BC, Canada: ACM.
14. Wellman, B., et al., *The Social Affordances of the Internet for Networked Individualism.* Journal of Computer-Mediated Communication, 2003. **8**(3).
15. Ali-Hasan, N. and L. Adamic. *Expressing social relationships on the blog through links and comments*. in *Proceedings of the ICWSM*. 2007.
16. Patch. *Patch Crunchbase Profile.* [cited 2011 July]; Crunchbase, The Free Tech Company Database]. Available from: http://www.patch.com.
17. Patch. [cited 2011 August]; Available from: http://www.patch.com.
18. Kopytoff, V.G. *AOL Bets on Hyperlocal News, Finding Progress Where Many have Failed*. New York Times; Available from: http://www.nytimes.com/2011/01/17/business/media/17local.html?pagewanted=all.
19. Everyblock. *EveryBlock CrunchBase Profile*. [cited 2011 July]; CrunchBase, The Free Tech Company Database]. Available from: http://www.crunchbase.com/company/everyblock.
20. *News360*. Available from: http://www.news360.com.
21. Google. *About Google News*. [cited 2011 July]; Available from: http://www.google.com.
22. iBrattleboro. *Brattleboro, Vermont Citizen News.* [cited 2011 July]; Available from: http://www.ibrattleboro.com.
23. RVANews. *About RVANews - Richmond News, Local Sports, Entertainment & Richmond Virginia Events.* [cited 2011 July]; Available from: http://rvanews.com/about.

24. Cohill, A. and A. Kavanaugh, eds. *Community Networks: Lessons from Blacksburg, Virginia* 1997, Artech House: Norwood, MA.
25. *Blacksburg Electronic Village (BEV)*. Available from: http://www.bev.net.
26. Kavanaugh, A., *Community in the Information Age: The Blacksburg Electronic Village in Perspective (1993-2010)*, in *Leadership in Science and Technology: A Reference Handbook*, W. Bainbridge, Editor 2011, Sage: Thousand Oaks, CA. p. 593-601.
27. Topix. *About Topix: Your Town. Your News. Your Take.*; Available from: http://www.topix.com/topix/about.
28. Fwix. *Fwix-Geotagging the Entire Web.* [cited 2011 July]; Available from: http://fwix.com/about.
29. Linden, G., B. Smith, and J. York, *Amazon.com Recommendations: Item-to-Item Collaborative Filtering.* IEEE Internet Computing, 2003. **7**(1): p. 5.
30. *Virginia Tech Students daily newspaper*. Available from: http://www.collegiatetimes.com.
31. Leskovec, J., et al., *Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters.* Internet Mathematics, 2009. **6**(1): p. 95.
32. Leskovec, J., D.P. Huttenlocher, and J.M. Kleinberg, *Signed networks in social media*, in *Proceedings of the 28th International Conference on Human Factors in Computing Systems*2010. p. 1361-1370.
33. Papagelis, M., N. Bansal, and N. Koudas. *Information Cascades in the Blogosphere: A Look Behind the Curtain*. in *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009*. 2009. San Jose, CA: AAAI Press.
34. Leskovec, J., J.M. Kleinberg, and C. Faloutsos, *Graph evolution: Densification and shrinking diameters.* ACM Transactions on Knowledge Discovery from Data, 2007. **1**(1).
35. Gruhl, D., et al. *Information diffusion through blogspace*. in *Proceedings of the 13th international conference on World Wide Web, WWW 2004*. 2004. New York, NY.
36. Gibson, J.J., *The Ecological Approach to Visual Perception*1979, Boston: Houghton Mifflin.
37. LaRose, M. and R. Eastin, *A social cognitive theory of internet uses and gratifications: Toward a new model of media attendance.* Journal of Broadcasting and Electronic Media, 2004. **48**(3): p. 358-377.
38. Shao, G., *Understanding the appeal of user-generated media: A uses and gratification perspective.* Internet Research, 2009. **19**(1): p. 7-25.
39. Pang, B. and L. Lee, *Opinion Mining and Sentiment Analysis.* Foundations and Trends in Information Retrieval, 2007. **2**(1--2): p. 135.
40. Godara, J., A. Kavanaugh, and P. Isenhour. *The Efficacy of Knowledge Sharing in Centralized and Self-Organizing Online Communities: Weblog Networks vs. Discussion Forums* in *Hawaii International Conference on System Sciences (HICSS)*. 2009. Kona, Hawaii: IEEE Computer Society Press.
41. Tauro, C., et al. *Deliberation in the Wild: A Visualization Tool for Blog Discovery and Citizen-to-Citizen Participation*. in *9th Annual International Conference on Digital Government Research (dg.o 2008)* 2008. Montreal, Canada: ACM.
42. Ahuja, S., M. Pérez-Quiñones, and A. Kavanaugh, *Rethinking Local Conversations on the Web*, in *Online Deliberation: Design, Research and Practice*, T. Davies and S.P. Gangadharan, Editors. 2009, CSLI Publications: Palo Alto. p. 123-129.
43. Java, A., et al., *Why We Twitter: Understanding microblogging usage and communities*, in *Joint 9th WEBKDD and 1st SNA-KDD Workshop '07*2007, ACM: San Jose, California.
44. Ahuja, S., *A Tale of Two Sites: An explorative study of the design and evaluation of social network sites*, in *Computer Science*2009, Virginia Tech: Blacksburg.
45. Putnam, R.D., *Bowling Alone: The collapse and revival of American community*2000, New York, NY: Simon & Schuster.
46. Lewis, C., et al., *Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces*, in *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*1990, ACM: Seattle, Washington, United States. p. 235-242.
47. Blackmon, M.H., et al., *Cognitive walkthrough for the web*, in *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*2002, ACM: Minneapolis, Minnesota, USA. p. 463-470.

[i] We deployed Colloki for a small community when we created *Slurp!* in support of graduate students in the Computer Science Department at Virginia Tech. Colloki was implemented with Ruby on Rails; its source code is available as an open source project (http://github.com/colloki/colloki).