Rural Digital Footprints
Information Gaps and Digital Representation

Dr. Jacob Thebault-Spieker
Assistant Professor, Information School
School of Computer, Data, and Information Science
University of Wisconsin – Madison
Rural Digital Footprints, Information Gaps, and Digital Representation

Three Main Goals

1. Introduce Digital Footprints and Information Gaps
2. One example of my work that helps us understand causes
3. Connect to conversations about AI and related technologies
Rural Digital Footprints, Information Gaps, and Digital Representation

Three Main Goals

1. Introduce Digital Footprints and Information Gaps
2. One example of my work that helps us understand causes
3. Connect to conversations about AI and related technologies
Popular Mechanics

Understanding the Information Superhighway

How You'll Shop, Bank, Learn, Be Entertained And More Via Interactive TV

January 1994 • $2.95

1994 Design & Engineering Awards
Best Products, Designs And Ideas Of The Year

The Greatest Social Revolution Since The Automobile

Printed in U.S.A.
Web 2.0 (also known as participative (or participatory)\(^1\) web and social web)\(^2\) refers to websites that emphasize user-generated content, ease of use, participatory culture and interoperability (i.e., compatibility with other products, systems, and devices) for end users.
<table>
<thead>
<tr>
<th>Website</th>
<th>Domain name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Search</td>
<td>google.com</td>
</tr>
<tr>
<td>YouTube</td>
<td>youtube.com</td>
</tr>
<tr>
<td>Facebook</td>
<td>facebook.com</td>
</tr>
<tr>
<td>Instagram</td>
<td>instagram.com</td>
</tr>
<tr>
<td>X</td>
<td>twitter.com</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>wikipedia.org</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>yahoo.com</td>
</tr>
<tr>
<td>WhatsApp</td>
<td>whatsapp.com</td>
</tr>
<tr>
<td>Amazon</td>
<td>amazon.com</td>
</tr>
<tr>
<td>Reddit</td>
<td>reddit.com</td>
</tr>
</tbody>
</table>
"The social web" is slightly euphemistic
Ok, but "footprints"? "Gaps"?
Black Earth, Wisconsin

From Wikipedia, the free encyclopedia

Black Earth is a village in Dane County, Wisconsin, United States. The population was 1,338 at the 2010 census. The village is located within the Town of Black Earth. It is part of the Madison Metropolitan Statistical Area.

History [ edit ]

In 1851, the town board renamed the town Farmersville, but the name Black Earth was readopted in 1857. [6]

Madison, Wisconsin

From Wikipedia, the free encyclopedia

This article is about the city. For the former town, see Madison (town), Wisconsin.

Madison is the capital city of the state of Wisconsin and the county seat of and largest city in Dane County. As of the 2020 census, the population was 269,840, making it the second most populous city in Wisconsin after Milwaukee, and the 80th most populous in the United States. Madison is named for American Founding Father and President James Madison.

Located on an isthmus and lands surrounding five lakes—Lake Mendota, Lake Monona, Lake Wingra, Lake Kegonsa and Lake Waubesa—the city is home to the University of Wisconsin–Madison, the Wisconsin State Capitol, the Overture Center for the Arts, and the Henry Vilas Zoo. Madison is home to an extensive network of parks and bike trails; it has the most parks and playgrounds per capita of any of the 100 largest U.S. cities and is one of five communities to have received a "Platinum Bicycle
Madison, Wisconsin

From Wikipedia, the free encyclopedia

This article is about the city. For the former town, see Madison (town), Wisconsin.

Madison is the capital city of the state of Wisconsin and the county seat of and largest city in Dane County. As of the 2020 census, the population was 269,840, making it the second most populous city in Wisconsin after Milwaukee, and the 80th most populous in the United States. Madison is named for American Founding Father and President James Madison.

Located on an isthmus and lands surrounding five lakes—Lake Mendota, Lake Monona, Lake Wingra, Lake Kegonsa and Lake Waubesa—the city is home to the University of Wisconsin–Madison, the Wisconsin State Capitol, the Overture Center for the Arts, and the Henry Vilas Zoo. Madison is home to an extensive network of parks and bike trails; it has the most parks and playgrounds per capita of any of the 100 largest U.S. cities and is one of five communities to have received a "Platinum Bicycle Friendly Community" rating from the League of American Bicyclists. Madison is also home to nine National Historic Landmarks, including several buildings designed by architect Frank Lloyd Wright, such as his 1937 Jacobs I House, which is a UNESCO World Heritage Site.

Residents of Madison are known as Madisonians. Madison has long been a center for progressive political activity, protests, and demonstrations, and contemporary Madison is considered the most politically liberal city in Wisconsin. The presence of the University of Wisconsin–Madison (the largest employer in the state) as well as other educational institutions has a significant impact on the economy, culture, and demographics of Madison.

As of 2021, Madison is the fastest-growing city in Wisconsin. Madison's economy features a large and growing technology sector, and the Madison area is home to the headquarters of Epic Systems, American Family Insurance, Exact Sciences, Promega, American Girl, Sub-Zero, End, Spectrum Brands, a regional office for Google, and the University Research Park, as well as many biotechnology and health systems startups. Madison is a popular visitor destination, with tourism generating over $1 billion for Dane County's economy in 2018.

History [edit]
Less Successful

More Successful


Popular Mechanics

UNDERSTANDING THE
INFORMATION SUPERHIGHWAY

THE GREATEST SOCIAL REVOLUTION SINCE THE AUTOMOBILE

How You'll Shop, Bank, Learn, Be Entertained And More Via Interactive TV
The social web is also "bypassing" rural communities.
On hand, this is somewhat expected, fewer people means fewer people with interest/able to contribute
Key aspect of these platforms: theoretically, anyone, anywhere, can contribute
Question becomes:
If anyone, anywhere can contribute, how and why are rural areas underrepresented in these important information resources?
Rural Digital Footprints, Information Gaps, and Digital Representation

Three Main Goals

1. Introduce Digital Footprints and Information Gaps
2. One example of my work that helps us understand causes
3. Connect to conversations about AI and related technologies
Three Main Goals

1. Introduce Digital Footprints and Information Gaps
2. One example of my work that helps us understand causes
3. Connect to conversations about AI and related technologies
Specifically: what kinds of places do current volunteer contributors focus their effort on?
Studied in OpenStreetMap
Continental US
Individual contributors
Two variables, over time:
(3mo intervals)

- "Focus" county: where contributors focus
- % Rural: what portion of contributors' effort happens in rural counties?
Categories of Volunteer Contributors

• Top 1% of contributors: produced 68% of the content
• Middle 9% of contributors: produced 27% of the content
• Bottom 90% of contributors: produced ~4% of the content
Focus County
Finding 1: People are consistent about the counties they contribute in
% Rural
Finding 2: People are consistent about the kinds of counties they contribute in as well (mostly not rural)
% Rural

- relative to:
  - county populations
  - # of rural/high-poverty counties
  - # of contributors focusing in rural/high-poverty areas
% Rural

- relative to:
  - county populations
  - # of rural/high-poverty counties
  - # of contributors focusing in rural/high-poverty areas

<table>
<thead>
<tr>
<th></th>
<th>population</th>
<th># counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>rural</td>
<td>15%</td>
<td>63%</td>
</tr>
</tbody>
</table>
Contextualizing Findings

- relative to:
  - county populations
  - # of rural/high-poverty counties
  - # of contributors focusing in rural/high-poverty areas

<table>
<thead>
<tr>
<th></th>
<th>avg. pct_rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>county populations</td>
<td>=</td>
</tr>
<tr>
<td># of counties</td>
<td>-</td>
</tr>
<tr>
<td># focused contributors</td>
<td>-</td>
</tr>
</tbody>
</table>

= proportional, - disproportionately low
Finding 3: In general, the rural contribution rates in this data are disproportionately low.
Finding 4: The most prolific contributors often focusing on "attractive" rural places, not average rural places
People who focus in **urban** areas remain in the system longer.

Consistency + longevity suggest that urban areas receive more content, longer.
To summarize, people are consistent in the places and kinds of places they contribute.
People who focus in urban places stick around longer
People who focus in rural places often focusing on "attractive" rural places
Take-away:
Default behavioral patterns of current contributors facilitate rural information gaps


In short: organic, volunteer-created information resources will have rural information gaps
Rural Digital Footprints, Information Gaps, and Digital Representation

Three Main Goals

1. Introduce Digital Footprints and Information Gaps
2. One example of my work that helps us understand causes
3. Connect to conversations about AI and related technologies
Rural Digital Footprints, Information Gaps, and Digital Representation

Three Main Goals

1. Introduce Digital Footprints and Information Gaps
2. One example of my work that helps us understand causes
3. Connect to conversations about AI and related technologies
Tesla Owners Can Edit Maps to Improve Summon Routes

It appears that Tesla is pulling data in realtime from OSM.
Apple Maps

Apple’s new maps for iOS6 make use of OpenStreetMap in some parts of the world. We’re not sure how extensive this use is, but it’s fair to say they are mostly using other sources. Apple have used TomTom as a key supplier of data for example. This means that inaccuracies in apple maps are probably not the fault of OpenStreetMap (contrary to some commentary!) However OpenStreetMap is mentioned in apple’s credits, and we have spotted some areas where we think we can see our data in use. For example here in Islamabad:
The Data Provenance Initiative:
A Large Scale Audit of Dataset Licensing & Attribution in AI

Shayne Longpre\textsuperscript{1,\dagger} Robert Mahari\textsuperscript{1,2} Anthony Chen\textsuperscript{3} Naana Obeng-Marnu\textsuperscript{1,4}
Damien Sileo\textsuperscript{5} William Brannon\textsuperscript{1,4} Niklas Muennighoff\textsuperscript{6} Nathan Khazam\textsuperscript{7}
Jad Kabbara\textsuperscript{1,4} Kartik Perisetla\textsuperscript{5} Xinyi (Alexis) Wu\textsuperscript{8} Enrico Shippole\textsuperscript{5} Kurt Bollacker\textsuperscript{7}
Tongshuang Wu\textsuperscript{9} Luis Villa\textsuperscript{10} Sandy Pentland\textsuperscript{1} Sara Hooker\textsuperscript{11}

\textsuperscript{1}MIT \hspace{0.5cm} \textsuperscript{2}Harvard Law School \hspace{0.5cm} \textsuperscript{3}UC Irvine \hspace{0.5cm} \textsuperscript{4}MIT Center for Constructive Communication
\textsuperscript{5}Inria, Univ. Lille Center \hspace{0.5cm} \textsuperscript{6}Contextual AI \hspace{0.5cm} \textsuperscript{7}ML Commons \hspace{0.5cm} \textsuperscript{8}Olin College
\textsuperscript{9}Carnegie Mellon University \hspace{0.5cm} \textsuperscript{10}Tidelift \hspace{0.5cm} \textsuperscript{11}Cohere For AI
<table>
<thead>
<tr>
<th>Name</th>
<th>Pct</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Encyclopedias</strong></td>
<td>21.5%</td>
</tr>
<tr>
<td>wikipedia.org</td>
<td>14.6%</td>
</tr>
<tr>
<td>wikihow.com</td>
<td>2.7%</td>
</tr>
<tr>
<td>dbpedia</td>
<td>1.4%</td>
</tr>
<tr>
<td><strong>Social Media</strong></td>
<td>15.9%</td>
</tr>
<tr>
<td>reddit</td>
<td>6.2%</td>
</tr>
<tr>
<td>twitter</td>
<td>4.0%</td>
</tr>
<tr>
<td>quora</td>
<td>1.6%</td>
</tr>
<tr>
<td><strong>Entertainment</strong></td>
<td>8.5%</td>
</tr>
<tr>
<td>opensubtitles.org</td>
<td>2.5%</td>
</tr>
<tr>
<td>imdb.com</td>
<td>1.6%</td>
</tr>
<tr>
<td>travel guides</td>
<td>1.3%</td>
</tr>
<tr>
<td><strong>Code</strong></td>
<td>5.7%</td>
</tr>
<tr>
<td>stackexchange.com</td>
<td>2.0%</td>
</tr>
<tr>
<td>github</td>
<td>1.2%</td>
</tr>
<tr>
<td>opus software projects</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Website</th>
<th>Domain name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Search</td>
<td>google.com</td>
</tr>
<tr>
<td>YouTube</td>
<td>youtube.com</td>
</tr>
<tr>
<td>Facebook</td>
<td>facebook.com</td>
</tr>
<tr>
<td>Instagram</td>
<td>instagram.com</td>
</tr>
<tr>
<td>X</td>
<td>twitter.com</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>wikipedia.org</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>yahoo.com</td>
</tr>
<tr>
<td>WhatsApp</td>
<td>whatsapp.com</td>
</tr>
<tr>
<td>Amazon</td>
<td>amazon.com</td>
</tr>
<tr>
<td>Reddit</td>
<td>reddit.com</td>
</tr>
</tbody>
</table>
Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research

Artificial-intelligence tools are transforming data-driven science – better ethical standards and more robust data curation are needed to fuel the boom and prevent a bust.

By Brooks Hanson, Shelley Stall, Joel Cutcher-Gershenfeld, Kristina Vrouwenvelder, Christopher Wirz, Yuhan (Douglas) Rao & Ge Peng
When it comes to AI,
information gaps in,
information gaps out
AI tools are very likely to be ineffective and risky for rural communities, because AI tools do not have rural information
Airline held liable for its chatbot giving passenger bad advice - what this means for travellers

23 February 2024

By Maria Yagoda, Features correspondent
So to wrap up
The social web is also "bypassing" rural communities
information gaps in, information gaps out

digital footprints in, digital footprints out

"a futuristic, AI-looking highway going around a city with farmland on one side", created with Dall-E 3
Thank you

Questions?

Dr. Jacob Thebault-Spieker, Information School, UW-Madison
e: jacob.thebaultspieker@wisc.edu
w: http://jacob.thebault-spieker.com
mastodon: @jts@hci.social / Twitter(X): @jaketangosierra