Online Urbanism: Interest-based Subcultures as Drivers of Informal Learning in an Online Community

Ben U Gelman, Chris Beckley, Aditya Johri, Carlotta Domeniconi, and Seungwon Yang*

George Mason University, *Louisiana State University

Acknowledgements:
We appreciate the Lifelong Kindergarten group at MIT for publicly sharing the Scratch datasets. This work is partly based upon research supported by U.S. National Science Foundation (NSF) Award#DUE-1444277 & EEC-1408674
Purpose

• Understand the growth of online communities of creators over time

• Observe the effects of population growth on user interests and network structure

• Place our research in the context of real life urbanism as well as online community design that supports an evolving population
Overview

• Review terms – informal learning and online communities of creators

• What is Scratch? Data and community structure

• Clustering method and validation

• Temporal analysis of clusters
Informal Learning

• Contrasts a classroom setting

• “Predominately unstructured, experiential, and non-institutional”
  • Victoria Marsick*

• Lack of standardized examination or measurement

How I Learned to Type—Runescape 2004-2006
Online Communities of Creators

• A specific type of informal learning community

• Social network sites where the core activity of members is sharing personal and original creations
Scratch

• An online community of creators – our focus in this study

• Visual programming interface

• Users are able to:
  • Post projects
  • View and edit source code of other projects – “remixing”
  • Follow other users
  • Chat, comment, and post on the forums
Scratch Project Interface

- (a) – Project window
- (b) – Block selection
- (c) - Script
Scratch Data

• March 2007 – April 2012

• 1,056,950 registered users

• 1,928,699 projects

• 170 Programming blocks
Scratch Data: Long-tail Phenomenon
Community Clustering

• Users with 25 followers

• Gephi
  • Graphing and visualization tool

• OpenOrd
  • Multi-level
  • Average-link clustering
Clustering Results
Validating Cluster Themes

• Text mining approach

• TFIDF
  • Term Frequency Inverse Document Frequency

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Top Words (TFIDF Values as percentage of Total Terms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>comput (5.3); time (4.7); make (4.4); space (4.4); random (4.0); new (3.8); stupid (3.6); first (3.2)</td>
</tr>
<tr>
<td>B</td>
<td>waffle (19.3); mario (17.6); game (12.3); super (9.5); tag (9.3); fun (8.5); pokemon (8.0)</td>
</tr>
<tr>
<td>C</td>
<td>funni (16.2); waz (15.0); wuz (15.0); tag (11.8); pokemon (10.4); good (8.6); add (8.6); sonic (8.5)</td>
</tr>
<tr>
<td>D</td>
<td>mario (22); epic (20.7); tag (16.3); press (12.2); sonic (9.8); super (8.4); add (6.2); scratch (4.2)</td>
</tr>
<tr>
<td>E</td>
<td>warrior (11.5); wolf (9.6); yay (8.8); cat (7.9); contest (7.7); pleas (7.0); art (5.7); pie (5.7)</td>
</tr>
</tbody>
</table>
Temporal Emergence of Subgroups

- Sliding window approach
- Each window shows users from the start of the data set up to some point in time after the start
- Each window clustered by OpenOrd
Scratchage and Scratchend

• **Scratchage** – Age of a user in years starting from March 5, 2007
  • A user who created their account on March 5, 2007 will have a Scratchage of 1 on March 5, 2008.

• **Scratchend** – Time in years between a user’s final activity and the end of the data set on April 1, 2012
  • A “final activity” is either following another user or posting a project.
  • A user whose final activity was on April 1, 2011 will have a Scratchend of 1.
Cluster Scratchages and Scratchends

- Time was not a factor in the clustering
- Temporal trends emerged anyways

<table>
<thead>
<tr>
<th>Cluster</th>
<th>SA: Avg.</th>
<th>SA: SD</th>
<th>SE: Avg.</th>
<th>SE: SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4.85</td>
<td>0.25</td>
<td>3.80</td>
<td>1.28</td>
</tr>
<tr>
<td>B</td>
<td>4.11</td>
<td>0.46</td>
<td>2.43</td>
<td>1.25</td>
</tr>
<tr>
<td>C</td>
<td>2.79</td>
<td>0.68</td>
<td>1.19</td>
<td>0.86</td>
</tr>
<tr>
<td>D</td>
<td>1.69</td>
<td>0.89</td>
<td>0.37</td>
<td>0.46</td>
</tr>
<tr>
<td>E</td>
<td>1.25</td>
<td>0.72</td>
<td>0.23</td>
<td>0.30</td>
</tr>
</tbody>
</table>
Migration Patterns

- Size of circle is proportional to size of cluster
- Grey sections of each circle are new users
Limitations and Future Work

• Lack of demographic data

• Cumulative Analysis

• Future work: individual user development
Discussion

• Fischer’s* Subcultural Theory of Urbanism

• Views subcultures as a “set of interconnected social networks”

• Argues that population growth provides opportunities for similar people to group

Discussion

• Support newcomers by targeting their interests, help them find place in the community

• Normally done by curating interesting projects, targeting user groups or experience levels

• Interest-based clusters allow for strategic project curation
Online Urbanism: Interest-based Subcultures as Drivers of Informal Learning in an Online Community

Ben U Gelman, Chris Beckley, Aditya Johri, Carlotta Domeniconi, and Seungwon Yang*

Georgetown University, *Louisiana State University

Acknowledgements:
We appreciate the Lifelong Kindergarten group at MIT for publicly sharing the Scratch datasets. This work is partly based upon research supported by U.S. National Science Foundation (NSF) Award#DUE-1444277 & EEC-1408674