Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap

Yair Grinberger, Moritz Schott, Martin Raifer, Rafael Troilo, Alexander Zipf
The Vision of Volunteered Geographical Information

- The vision of VGI – democratized and bottom-up geo-data production (Goodchild, 2007)

- The evolution of the vision:
  - Participation and data bias (Haklay, 2016)
  - Considering process with product (Sieber & Haklay, 2015)
  - Contextual effects on data (Fast & Rinner, 2014)

- OpenStreetMap is rich in contextual effects:
  - Mapping platforms
  - Interaction platforms (wiki, mailing lists, …)
  - Activity of organizations (Anderson et al., 2019; Palen et al., 2015; Poiani et al., 2016)
  - Data events

Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap
Grinberger, Schott, Raifer, Troilo, & Zipf
Defining events in OSM:
- The social perspective (Juhász & Hochmair, 2018; Mooney et al., 2015)
- The data perspective (Eckle & Albuquerque, 2015; Zielstra et al., 2013)

Large-scale data events:
- Can create lasting impacts on data and community
- **High** volume of contributions over a **short** period
- **Significantly** affect the data

The current study:
- Identifies events which show a **significant change**
- Analyzes spatio-temporal **patterns**
- Studies **impacts**
Identifying Large-Scale Events

- Assumption – a ‘normative’ model of data production (Gröching et al., 2014)
- Definition – events are sharp increases not predicted by the model
- Procedure:
  - Create cumulative series of contribution actions over time
  - Fit a logistic curve to the time series
Identifying Large-Scale Events

- Assumption – a ‘normative’ model of data production (Gröching et al., 2014)
- Definition – events are sharp increases not predicted by the model
- Procedure:
  - Create cumulative series of contribution actions over time
  - Fit a logistic curve to the time series
  - Compute lagged residuals
  - Find significant positive residuals

\[ C_t = \frac{\alpha}{1 + \beta e^{-\gamma t}} \]
Data Extraction

- Quad-tree spatial division by number of OSM entities
- Temporal resolution – one month
- Time period: 11-2007 to 03-2019
- Number of actions extracted using the OSHDB tool (Raifer et al., 2019)

Additional variables:
- Active users
- No. of contributions by type
- Maximal no. of actions by one user
- Actions per edited entity
Events

- Convergence errors for 700 cells (6.91%)
- Considered only events with no. of actions > 7,000
- 48,653 events identified
- Median of 5.00 events per cell (average: 5.16, std: 2.72)
Classifying Events

- K-means procedure used to differentiate between events (K=6)
- Variables used:
  - contributions by type (% of all contributions)
  - maximal volume of contribution by one user (% of all contributions)
## Weight of Events (% of All Actions/Contributions)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All</strong></td>
<td>48653.0</td>
<td>5468.89</td>
<td>40.42</td>
<td>38.94</td>
<td>48.62</td>
<td>45.52</td>
<td>30.65</td>
<td>35.98</td>
<td>21.07</td>
</tr>
<tr>
<td><strong>Local</strong></td>
<td>7394.0</td>
<td>459.26</td>
<td>3.39</td>
<td>4.03</td>
<td>2.03</td>
<td>2.54</td>
<td>5.79</td>
<td>2.68</td>
<td>5.84</td>
</tr>
<tr>
<td><strong>Early import</strong></td>
<td>14080.0</td>
<td>2301.58</td>
<td>17.01</td>
<td>15.94</td>
<td>21.85</td>
<td>25.48</td>
<td>5.65</td>
<td>0.81</td>
<td>1.69</td>
</tr>
<tr>
<td><strong>Tag import</strong></td>
<td>3216.0</td>
<td>570.95</td>
<td>4.22</td>
<td>0.64</td>
<td>13.89</td>
<td>0.36</td>
<td>1.34</td>
<td>26.42</td>
<td>0.71</td>
</tr>
<tr>
<td><strong>Remote</strong></td>
<td>6145.0</td>
<td>831.19</td>
<td>6.14</td>
<td>7.15</td>
<td>4.29</td>
<td>9.42</td>
<td>3.59</td>
<td>0.82</td>
<td>2.52</td>
</tr>
<tr>
<td><strong>Geom. import</strong></td>
<td>6008.0</td>
<td>374.57</td>
<td>2.77</td>
<td>3.43</td>
<td>1.4</td>
<td>1.09</td>
<td>2.0</td>
<td>2.2</td>
<td>6.22</td>
</tr>
<tr>
<td><strong>Late import</strong></td>
<td>11810.0</td>
<td>931.34</td>
<td>6.88</td>
<td>7.75</td>
<td>5.16</td>
<td>6.65</td>
<td>12.29</td>
<td>3.05</td>
<td>4.09</td>
</tr>
</tbody>
</table>
Temporal Patterns

Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap
Grinberger, Schott, Raifer, Troilo, & Zipf
Temporal Patterns

Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap
Grinberger, Schott, Raifer, Troilo, & Zipf
Spatial Patterns – Events’ Weights

Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap
Grinberger, Schott, Raifer, Troilo, & Zipf
Spatial Patterns – Events’ Weights

Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap
Grinberger, Schott, Raifer, Troilo, & Zipf
Spatial Patterns – Most Common Event Type

Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap
Grinberger, Schott, Raifer, Troilo, & Zipf
Spatial Patterns – Most Common Event
## Events’ Effects on Activity (6 months)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>897441.0</td>
<td>8.34</td>
<td>7.47</td>
<td>9.74</td>
<td>8.4</td>
<td>7.56</td>
<td>2.93</td>
<td>11.63</td>
<td>11.16</td>
</tr>
<tr>
<td>All</td>
<td>14623.0</td>
<td>12.96</td>
<td>7.69</td>
<td>10.86</td>
<td>12.31</td>
<td>4.74</td>
<td>20.94</td>
<td>16.84</td>
<td>15.12</td>
</tr>
<tr>
<td>Local</td>
<td>1840.0</td>
<td>9.56</td>
<td>6.69</td>
<td>9.33</td>
<td>11.8</td>
<td>5.8</td>
<td>8.46</td>
<td>14.27</td>
<td>10.34</td>
</tr>
<tr>
<td>Early import</td>
<td>4683.0</td>
<td>18.07</td>
<td>8.94</td>
<td>12.86</td>
<td>17.61</td>
<td>3.79</td>
<td>34.44</td>
<td>27.75</td>
<td>20.4</td>
</tr>
<tr>
<td>Tag import</td>
<td>888.0</td>
<td>-1.04</td>
<td>13.64</td>
<td>-5.51</td>
<td>-2.6</td>
<td>-15.2</td>
<td>25.49</td>
<td>4.08</td>
<td>7.12</td>
</tr>
<tr>
<td>Remote</td>
<td>1660.0</td>
<td>31.02</td>
<td>17.44</td>
<td>26.55</td>
<td>36.99</td>
<td>22.33</td>
<td>103.96</td>
<td>53.5</td>
<td>44.76</td>
</tr>
<tr>
<td>Geom. import</td>
<td>2041.0</td>
<td>12.2</td>
<td>4.95</td>
<td>11.84</td>
<td>7.14</td>
<td>6.85</td>
<td>1.27</td>
<td>6.06</td>
<td>9.54</td>
</tr>
<tr>
<td>Late import</td>
<td>3511.0</td>
<td>7.92</td>
<td>4.62</td>
<td>7.61</td>
<td>8.59</td>
<td>2.0</td>
<td>13.08</td>
<td>8.82</td>
<td>8.8</td>
</tr>
</tbody>
</table>
# Events’ Effects on Activity (12 months)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>619144.0</td>
<td>15.4</td>
<td>13.79</td>
<td>17.88</td>
<td>14.23</td>
<td>16.67</td>
<td>6.25</td>
<td>15.3</td>
<td>20.46</td>
</tr>
<tr>
<td>Local</td>
<td>919.0</td>
<td>15.18</td>
<td>11.68</td>
<td>15.29</td>
<td>14.94</td>
<td>13.33</td>
<td>6.85</td>
<td>15.72</td>
<td>13.72</td>
</tr>
<tr>
<td>Early import</td>
<td>2729.0</td>
<td>17.75</td>
<td>13.81</td>
<td>12.82</td>
<td>20.63</td>
<td>6.18</td>
<td>32.67</td>
<td>34.48</td>
<td>31.64</td>
</tr>
<tr>
<td>Tag import</td>
<td>606.0</td>
<td>9.75</td>
<td>18.78</td>
<td>9.52</td>
<td>3.29</td>
<td>-13.28</td>
<td>31.44</td>
<td>19.7</td>
<td>19.41</td>
</tr>
<tr>
<td>Remote</td>
<td>922.0</td>
<td>41.25</td>
<td>28.5</td>
<td>39.21</td>
<td>58.39</td>
<td>43.8</td>
<td>107.93</td>
<td>93.03</td>
<td>57.62</td>
</tr>
<tr>
<td>Geom. import</td>
<td>1135.0</td>
<td>12.54</td>
<td>7.49</td>
<td>13.58</td>
<td>5.89</td>
<td>10.97</td>
<td>-2.57</td>
<td>3.66</td>
<td>10.54</td>
</tr>
<tr>
<td>Late import</td>
<td>1912.0</td>
<td>13.44</td>
<td>11.02</td>
<td>14.03</td>
<td>10.35</td>
<td>5.55</td>
<td>16.73</td>
<td>12.38</td>
<td>13.12</td>
</tr>
</tbody>
</table>

Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap
Grinberger, Schott, Raifer, Troilo, & Zipf
## First Event and Probability for Future Events

<table>
<thead>
<tr>
<th></th>
<th># events</th>
<th>Local</th>
<th>Early import</th>
<th>Tag import</th>
<th>Remote</th>
<th>Geom. import</th>
<th>Late import</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>871.0</td>
<td>70.72</td>
<td>33.18</td>
<td>11.48</td>
<td>31.0</td>
<td>51.21</td>
<td>71.3</td>
</tr>
<tr>
<td>Early import</td>
<td>4047.0</td>
<td>19.08</td>
<td>66.17</td>
<td>26.19</td>
<td>25.95</td>
<td>24.64</td>
<td>42.43</td>
</tr>
<tr>
<td>Tag import</td>
<td>448.0</td>
<td>19.2</td>
<td>48.21</td>
<td>68.75</td>
<td>13.62</td>
<td>35.27</td>
<td>33.26</td>
</tr>
<tr>
<td>Remote</td>
<td>1662.0</td>
<td>36.4</td>
<td>29.18</td>
<td>6.26</td>
<td>54.15</td>
<td>16.43</td>
<td>24.79</td>
</tr>
<tr>
<td>Geom. import</td>
<td>636.0</td>
<td>60.22</td>
<td>42.92</td>
<td>15.25</td>
<td>28.93</td>
<td>62.11</td>
<td>75.63</td>
</tr>
<tr>
<td>Late import</td>
<td>1747.0</td>
<td>44.25</td>
<td>59.24</td>
<td>12.54</td>
<td>33.14</td>
<td>40.98</td>
<td>72.24</td>
</tr>
</tbody>
</table>
Conclusions

- Large-scale data events affect OSM in a meaningful way
- They are contextual **products** with contextual **impacts**:
  - Shifting trends related to the **maturity** of the data/community
  - ...but with **socio-geographical variability**
  - They may serve as a means for **exploration**
  - May **adversely** affect activity, but wrongs can make a right!
- Considering context as part of the production of events

**Further analysis:**
- Stability of event contributions
- Tagging schemes during and after events
- Changes in communities’ structures
Thank You!

yair.grinberger@mail.huji.ac.il
Analyzing the Spatio-Temporal Patterns and Impacts of Large-Scale Events in OpenStreetMap
Grinberger, Schott, Raifer, Troiilo, & Zipf

References