Using K-core to Find Needle In Haystack of Bot-net Data

Lieutenant Colonel David Beskow
dbeskow@andrew.cmu.edu

Motivation

**Problem:** Network viz is one of the most powerful means of exploratory network analysis, but is difficult if not impossible for large networks

**(One) solution:** K-Core analysis provides a way of reducing the network to a manageable size while retaining essential structure, facilitating network visualization
Data

- **Following** connections between Twitter bot accounts that use random alpha-numeric 15 digit screen names
- These accounts are spread out across the globe; deployed by numerous actors with widely varying purposes
- Node attribute includes account creator language setting
- We will use ORA and K-Core analysis to reduce this data to a manageable size while retaining essential structure in order to ‘triage’ the network

Import Table of Links
Large Meta-network

Import Node Attributes
Import Node Attributes

View Language Attribute
Generate K-Core Report

Evaluate K-Core Results

K-Cores

A K-core is a set of nodes in which each node is connected to at least K other nodes. There is at most one K-core for a network for each value of K. The K-cores are nested; for example, all nodes in the 3-core are in the 4-core. Isolates are only in the 0-core.

<table>
<thead>
<tr>
<th>K-Core</th>
<th>Number of nodes</th>
<th>Percent of nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>27913</td>
<td>43.00%</td>
</tr>
<tr>
<td>1</td>
<td>12918</td>
<td>19.55%</td>
</tr>
<tr>
<td>2</td>
<td>7114</td>
<td>10.87%</td>
</tr>
<tr>
<td>3</td>
<td>4236</td>
<td>6.48%</td>
</tr>
<tr>
<td>4</td>
<td>424</td>
<td>0.63%</td>
</tr>
<tr>
<td>5</td>
<td>151</td>
<td>0.23%</td>
</tr>
<tr>
<td>6</td>
<td>104</td>
<td>0.16%</td>
</tr>
<tr>
<td>7</td>
<td>140</td>
<td>0.21%</td>
</tr>
<tr>
<td>8</td>
<td>190</td>
<td>0.29%</td>
</tr>
</tbody>
</table>
View K-Core Attribute

Filter by 4-Core & Create New Meta-network
New Meta-network

Visualize Network
Reduce Network

The network you are displaying has 4470 nodes and 25033 links. For networks with more than 6000 entries and/or more than 10000 links, display might be very slow. Consider these options to pare down the meta-network before loading to improve performance:

- Remove links with weight less than or equal to 1
- Select networks to load
- Agent x Agent
- Remove components of size less than or equal to 1
- Hide labels
- Hide links
- Save as new meta-network in ORA

Continue  Cancel

Visualize bot-nets
Color by Language
Adjust Language Colors

Color by Louvain Grouping
Color by Louvain Grouping

Color by Dense Subgraphs
Building out the Dense Network

- Produce Tweets
  - in Russian
  - in Arabic

Initial Seed Nodes (Initial Random String U sernames)
Conclusion

• K-Core offers a one viable solution for network reduction and visualization

Questions