1 Analyzing the Wikipedia voters network [27 points]

Download the Wikipedia voting network `wiki-Vote.txt.gz`: Dataset1 in bundle file.

Using one of the network analysis tools above, load the Wikipedia voting network. Note that Wikipedia is a directed network. Formally, we consider the Wikipedia network as a directed graph $G = (V, E)$, with node set V and edge set $E \subset V \times V$ where (edges are ordered pairs of nodes). An edge $(a, b) \in E$ means that user a voted on user b.

To make our questions clearer, we will use the following small graph as a running example: $G_{\text{small}} = (V_{\text{small}}, E_{\text{small}})$, where $V_{\text{small}} = \{1, 2, 3\}$ and $E_{\text{small}} = \{(1, 2), (2, 1), (1, 3), (1, 1)\}$.

Compute and print out the following statistics for the `wiki-Vote` network:

1. *The number of nodes in the network.* (G_{small} has 3 nodes.)
2. *The number of nodes with a self-edge (self-loop), i.e., the number of nodes $a \in V$ where $(a, a) \in E$.* (G_{small} has 1 self-edge.)
3. *The number of directed edges in the network, i.e., the number of ordered pairs $(a, b) \in E$ for which $a \neq b$.* (G_{small} has 3 directed edges.)
4. *The number of undirected edges in the network, i.e., the number of unique* unordered *pairs (a, b), $a \neq b$, for which $(a, b) \in E$ or $(b, a) \in E$ (or both).* If both (a, b) and (b, a) are edges, this counts a single undirected edge. (G_{small} has 2 undirected edges.)
5. *The number of reciprocated edges in the network, i.e., the number of unique unordered pairs of nodes (a, b), $a \neq b$, for which $(a, b) \in E$ and $(b, a) \in E$.* (G_{small} has 1 reciprocated edge.)
6. *The number of nodes of zero out-degree.* (G_{small} has 1 node with zero out-degree.)
7. *The number of nodes of zero in-degree.* (G_{small} has 0 nodes with zero in-degree.)
8. *The number of nodes with more than 10 outgoing edges (out-degree > 10).*
9. *The number of nodes with fewer than 10 incoming edges (in-degree < 10).*

Each sub-question is worth 3 points.

Finding Experts on the Java Programming Language on StackOverflow [40 points]

Download the StackOverflow network `stackoverflow-Java.txt.gz`: Dataset2 in bundle file. An edge (a, b) in the network means that person a endorsed an answer from person b on a Java-related question.

Using one of the network analysis tools above, load the StackOverflow network. Note that StackOverflow is a directed network.

Compute and print out the following statistics for the `stackoverflow-Java` network:

1. *The number of weakly connected components in the network.* This value can be calculated in Snap.py via function `GetWccs`.

2. *The number of edges and the number of nodes in the largest weakly connected component.* The largest weakly connected component is calculated in Snap.py with function `GetMxWcc`.
3. *IDs of the top 3 most central nodes in the network by PagePank scores.* PageRank scores are calculated in Snap.py with function `GetPageRank`.
4. *IDs of the top 3 hubs and top 3 authorities in the network by HITS scores.* HITS scores are calculated in Snap.py with function `GetHits`.