**Task:** Analyze an authentication (authN) dataset with 708M authN events over a nine month period. An event is anonymized and represented as: 
\(<\text{time}>, <\text{U#}>, <\text{C#}>\)
where \(<\text{time}>=\text{secs}\), offset from the start, 
\(<\text{U#}>\) is a User ID, and \(<\text{C#}>\) is a Computer ID.

**Approach:**
Install Anaconda Python (3.x) which includes: pandas, NetworkX, matplotlib, IPython modules.

Download original data, graph, scripts:
http://csr.lanl.gov/data/auth/
http://trustedci.org/data
https://github.com/rheiland/authpy

**Step 1:** Use the pandas module to read the entire dataset, in chunks, and generate two files with:
1. only time values
2. the bipartite graph (as an adjacency list)

**Step 2:** Sanity check: read time values; plot a histogram, looking for a temporal pattern.

**Step 3:** Sanity check: read the graph, verify it’s bipartite, and check # of Users and Computers.

**Step 4:** Perform additional analysis/visualization of the graph, e.g. looking for hubs, connected components, etc.

**Acknowledgements:** LANL for providing the authN dataset. The Python community. The NSF (OCI-1234408) for supporting this work.
