

IS620 – Assignment 4

Brian Chu | Sep 20, 2015

Dataset: Flickr user and image data via Flickr API -

<https://www.flickr.com/services/api/>

Reason: I am a fairly avid photographer and Flickr user

(<https://www.flickr.com/photos/bchu81/>) who is generally interested in how photographers connect and what photos they like or find interesting within the Flickr network.

Project: For the purposes of this assignment and my own learning curve, I will analyze the Flickr user network to identify 'central' users and their characteristics.

Nodes: Flickr users

Edges: Connections (either contacts, friends, or family)

Categorical variables: Sex, Country, Age (bin), Groups, # of favorited photos (bin)

Loading: I would load the data into Python via the Flickr API and parsing the retrieved JSON. Since the dataset will be very large, I would restrict the data to users with a certain minimum number of connections.

Analyzing: I would analyze the degreeness, closeness, and betweenness of users using the NetworkX package and functions. For identified central users, I would subset into the various categorical groups and re-measure centrality scores.

Potential outcomes:

- 1) Do male and female networks differ?
- 2) Are user's networks largely confined to their home country? Which countries are more likely to connect with each other?
- 3) Do users mostly interact with others in the same age range?
- 4) How important are Flickr groups in a user's overall network? Are there deep connections made in these groups or are they mostly weak connections?
- 5) Is user centrality just a byproduct of how many/often they favorite others' photos?