

Collecting data:

1. Select a target node. This should be a central node in the community you want to map.
2. Find the username of the root node. This is displayed on their page under their 'real' name and will be prefixed with an @ eg <https://twitter.com/#!/edlsupportgroup> -> @edlsupportgroup
3. Launch the harvesting script:
 - a. `python harvest.py edlsupportgroup`

Points of note:

- You may run this script on multiple different machines at the same time, provided that they have distinct external IP addresses.
- If it crashes, after any necessary debugging simply run the script again; it will resume from the last good state it was in before the crash.
- To start collecting data on a different group, remove the database:
 - `rm rettwit.db`
- To use a different account, follow the instructions at <https://dev.twitter.com/docs/auth/tokens-devtwittercom> to get the 4 tokens, then insert them into the consumer_key, consumer_secret, access_token and access_secret variables defined at the start of harvest.py. Nothing outside these 4 lines will need to be altered.

Transforming data

If the harvest script is still running, make a copy of the database. Otherwise, skip to step 2

1. Copy the database:

```
mkdir copy
cd copy
cp ../rettwit.db ./
```
2. Execute the script of your choice and pipe the output to a file

```
python generateGraphmlTweets.py > foo.graphml
```