Meeting Minutes

Term 1

8/11/11

PRESENT

- Matt, Isaac, James K, James M

MINUTES

- We discussed the areas we had chosen to research from the previous meeting
- Met with Phil Taylor to discuss his project last year, demonstration of it tomorrow (9/11/11) at 12pm
- Hopefully write a new project spec after meeting
- Phil mentioned research areas of influence, importance and ranking, and analysis of the text of messages

ACTIONS

- Read each others researched information
- Meet with Phil Taylor tomorrow

9/11/11

PRESENT

- Matt, Isaac, James K, James M

MINUTES

- Met with Phil for demonstration of the project he helped produce last year
- Code online at snat.googlecode.com

ACTIONS

- Organise meeting with Sarab

10/11/11

PRESENT

Matt, Isaac, James K, James M

MINUTES

- possible ideas
    - turn existing code into cluster (with hadoop?)
    - analytics as a web service
    - partitioning a social network into groups
    - converge multiple networks (creating a mapping of one set of nodes to another)
    - text mining messages
- could port the software to something else
- independent cascade model
    - complicated algorithm, takes a long time
    - currently uses heuristics to speed up
    - could implement a definitive cluster version, to compare results against

ACTIONS

- come up with project idea to present to Sarab in Tuesday's meeting 3:30

18/11/11

PRESENT
Matt, Isaac, James M

MINUTES
- Read document RE: police project
- Will divide the project into two parts:
    - Developing software to run on a cluster (lead by Matt and Isaac)
    - Implementing algorithms to run on the software (lead by James & James)

ACTIONS
- Read document and references
- Talk with Phil Taylor regarding integrating cluster/separating interface from snat
- Need to have design document down soon

22/11/11

PRESENT

Isaac, Matt, James M, James K, Simon, Sarab

MINUTES

- James & James will look into Modelling Emergent Behaviours
- Simon is not getting messages to the group email, need to fix that
- Sarab discussed the clustering side of the project with Isaac and Matt:
    - Sarab recommended looking up CouchDB, and other document-store databases
    - Adam Chester, a PhD student may become involved with the project
    - Discusses the issues with running graph problems on Hadoop. Partitioning the graph will be the main issue. Sarab recommended looking up the betweeness algorithm

- Sarab said the West Midlands Police could not give us data at short notice, but they recommended using Twitter data and looking at two groups, the EDL and Unite against Fascism
- Simon will look into SOAP interface
- Simon will look at getting twitter data
- The poster (due for week 9) should mainly cover a literature review of the project area

ACTIONS

- James & James Meeting with Nathan to discuss his papers
- Isaac will look up running the independent cascade model on a cluster
- Matt will look up running the graph partitioning algos on a cluster
- Isaac will fix Simon's group email
- Deadline for poster to be ready: Next Thursday

23/11/11

PRESENT
James M, James K, Nathan, Sarab

MINUTES
- Twitter data: how to define success
    - number of followers/following ratio centrality, RT (propagating data), tweeting at (looking at content?)
    - Use #ff follow fridays for additional info
    - failure: unfollows
- Learning
    - identify topics in tweets to determine what topics imply success
    - learn what profile to show (what topics to tweet about, who to follow...)
    - what you tweet can be different to what you read
        - e.g. security might follow edl, but not tweet edl content
        - following does not mean interested in a supportive way
    - Individual nodes would have their own learning rates
        - these are initialised randomly
        - there is a chance of mutation (with a given distribution)
        - could change in response to current success level
            - don't learn if the node is doing well
            - learn quickly if the node is doing worse
- Assume noisy observations
    - can be made worse by mutations in 'reproduction/learning'
- How to define a cheater on twitter
    - discreditors
- Reward/Punishment
    - retweeting = reward
    - unfollow = punishment
- Look into negative effects: break down the organisation, reduce trust
- Sometimes useful to see more than immediate neighbours
    - go out a number of levels, e.g. 1 or 2 additional levels
    - learn a lot from local neighbourhood, a bit from outer neighbourhood
- Gathering twitter data
    - get unfollow data

ACTIONS
- reimplement Nathan's work
    - allow learning rates to differ
    - allow non-synchronous generations (learning)
    - James K: Design the Tags and Image Scoring model
    - James M: Design the Changing Neighbours model
    - meet 11 on Monday to merge
    - have a design to present on Tuesday
- read paper on gossip matters
    - destablization of an organisation by injecting suspicion
    - http://www.bus.emory.edu/prietula/prietula-carley-chapter-rev1.pdf

29/11/11

PRESENT
Isaac, Matt, James M, James K, Simon, Sarab
MINUTES

- clustering
- modelling emergent behaviours
- twitter data acquisition
- web service
- poster
    - will be ready for printing by Thursday, week 9
    - presentation on Thursday, week 10
- meeting with police, probably Friday week 10, maybe Thursday week 10

ACTIONS

- finish implementation of Nathan's algorithm
- do poster
- research more on Giraph
- have a finalised design for the SOAP message interface before the end of term

5/12/11

PRESENT
Isaac, Matt, James M, James K
MINUTES

- Poster is being printed tomorrow by Warwick Print and delivered back to dcs
- Decided which section each person will talk about
- We should be able to talk/present the poster for at least 30 minutes without any questions being asked by other people

ACTIONS

- practice section each person will talk about in the presentation

6/12/11

PRESENT
Isaac, Matt, James M, Simon
MINUTES

- Simon has got in contact with Twitter, they may be able to help with Data Acquisition
- Isaac has read some papers on Influence propagation, he will look into how to distribute it efficently and starting implementation
- Matt has looked at Giraph, will need to check if the graph is partitioned in a decent manner
- James showed the Java implementation, will look at import/exporting graphs, different network topologies, directed vs. undirected graphs, real world restrictions on rewiring and how it affects topologies, look at getting results, mapping twitter to a graph
    - relevant
        - JUNG - Java Universal Network/Graph framework
        - networkx
        - GraphML
        - SNAP - Stanford Network Analysis Project
- Discussed poster presentation, format will likely be 10 minutes of us presenting the poster, any additional time can be used for questions

ACTIONS

-

Term 2

13/1/12

PRESENT

- Matt, Isaac, James K, James M, Simon, Sarab

MINUTES

- Progress made over Christmas break
    - Matt got Giraph to work with PageRank algorithm, and now understands how Giraph jobs work
    - Hadoop is a bit inefficient when it comes to graph algorithms as it needs to model iteration in an algorithm using multiple map/reduce jobs, which write to file system after each job
- Post-Christmas Progress and fault tolerances of what needs to be done
- James and James to focus more on project once other coursework is done
- Simon looked at last year's interface

ACTIONS

- Research into how to adapt algorithms for Giraph paradigm
- James and James to start working from Tuesday
- Simon to start getting Twitter data

3/2/12

PRESENT

- Matt, Isaac, James K, James M, Simon, Nathan, Sarab

MINUTES

- Simon has made progress with previous work on GUI
- James and James presented results from implementing one of Nathan's papers
- Isaac been looking at implementing influence propagation on Hadoop
- Matt has been struggling to use Giraph
    - Errors when running example code
    - No obvious solution

ACTIONS

- Simon to work on Twitter data
- James and James to start adapting previous GUI for use
- Isaac to implement an influence propagation algorithm
- Matt to try and get a community detection algorithm working in either Giraph or Hadoop

10/2/12

PRESENT

- Matt, Isaac, James K, James M, Simon, Nathan, Sarab

MINUTES

- James and James have converted the graphs they've made using Jung and Peersim into SQL and inserted into the existing databases
    - Visualiser doesn't seem to display these properly
    - Talk to Phil about the database
- Simon worked on the visualiser
- Isaac is implementing an Influence propagation algorithm for hadoop
    - Not sure how to parse the graph data into Hadoop
- Matt has implemented a community detection algorithm for identify triangles within a graph
    - Not outputting correctly, but should be simple to fix
    - Needs to run on some other data sets
    - Analysis of algorithms on large dataset would be good with varying amounts of nodes to establish if distributing computation actually improves runtime of algorithms

ACTIONS

- James M to talk to Phil about the database
- James K to start gathering twitter data
- Matt to finish the community detection algorithm
- Matt to parse existing graphs into format suitable for Hadoop/Giraph
- Matt to run community detection algorithm on more data sets
- Matt to talk to Olly about Hadoop cluster in dcs
- Isaac to look at Stanford Network Analysis Project data
- Isaac to finish implementing influence propagation algorithm
- Simon to finish work on the visualiser

17/2/12

PRESENT

- Matt, James K, James M, Nathan, Sarab

MINUTES

- James M presented results of rewiring on JUNG generated graphs
- Matt has got Hadoop working on a set of livejournal data
    - but it runs out of memory
- Matt produced a parser for the SNAP twitter and livejournal dataset for use on Giraph
- James K presented twitter data

ACTIONS

- James M to email results to Nathan and Sarab
- James M to repeat experiment with different numbers of iterations
- James M to gather metrics about how the graph changes over time
- Matt to talk to Oliver about getting a Hadoop cluster set up
- James K to work on getting more twitter data, have a look at visualising the data