Social Media network analysis: What can SNA offer?

Malcolm Alexander

Griffith University

What are the questions?

○ Rob's question: I have very large databases recording social media users' communications. What does Social Network Analysis (SNA) tell me about analysing this data?

○ Immediate responses:

- I am a sociologist, I want to think about the qualitative dimensions of this data
- ('Whole network') WN-SNA (where most people start) is good for small (up to 100) datasets but not for large datasets
- I'll get back to you!

Overview

- Common problems
  - Analyse the network data; graph-theoretic diagrams are a tool not the goal
  - Remember where the data comes from – the people behind it
  - WN-SNA measures are wrong starting point, look at egonet SNA
- Positive messages
  - Databases capture online activity, not people's 'virtual' social worlds
  - Egonet data analysis: treat relations as data points
  - WN-SNA modelling (PNet) builds and assess complex network models

Data mining for network analysis

- Examples of 'network data'
  - The internet: Packages between routers
  - The WWW (Rob and VOSON)
    - Hyperlinks between pages
    - Blogs and blog references
  - Biological databases: Protein gene interactions
 - SNA datasets

What is network data?

- What constitutes network data
  - FROM 'A' TO 'B' where A and B are labels (IDs) of nodes (points)
  - Tie data lists, (or 'edge lists') with information about the designated tie
  - Descriptive information about nodes (not essential) can be stored in a node data list.
- Network and non-network data
  - My credit card spending profile
    - Items classified by type
    - Relational/ network format is possible: Amazon.. also purchased

Basic NodeXL file and diagram
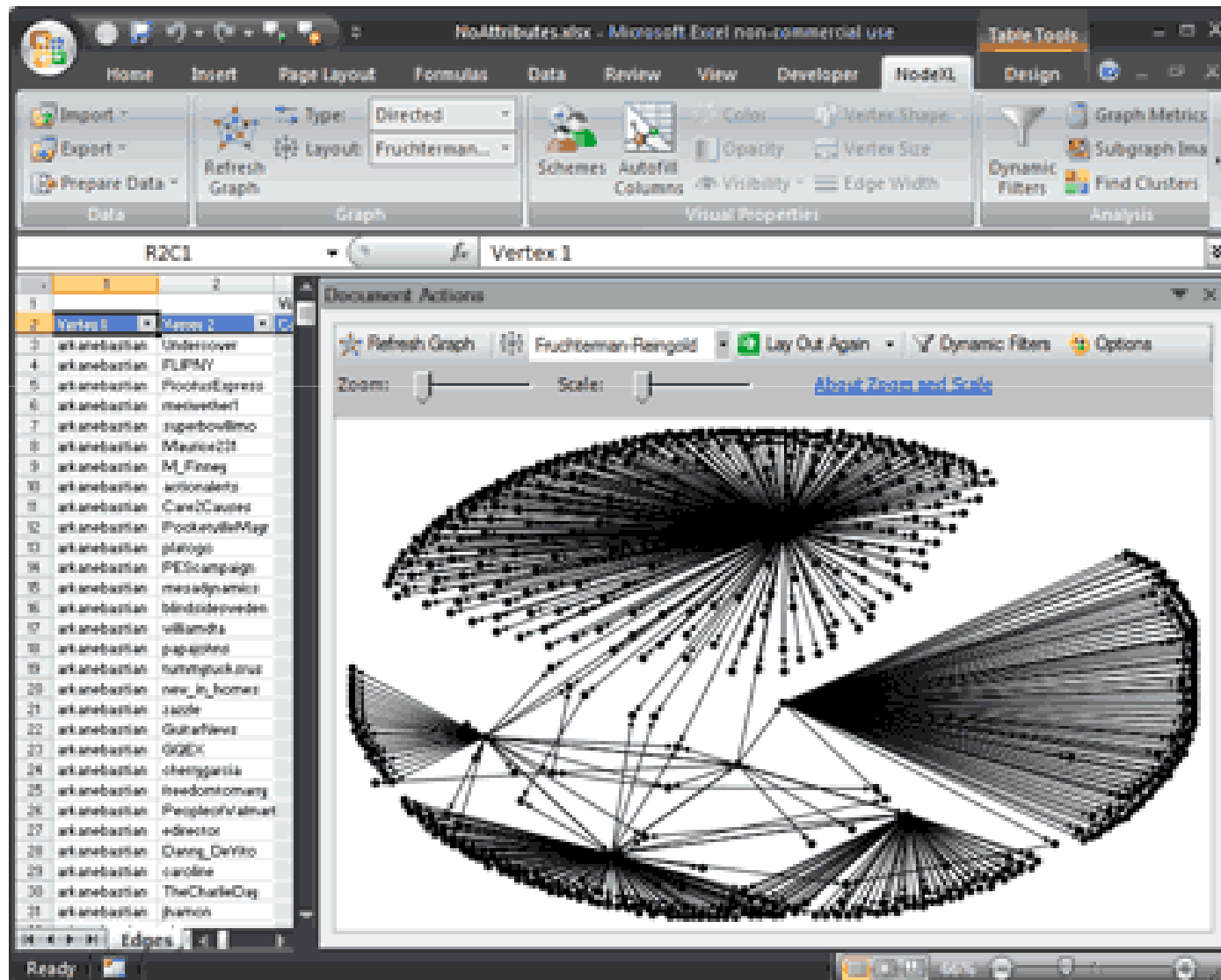
A sociometric questionnaire

Sociometric survey of XXXX office

Your Information:
Name: …………………………………… Age: ……… Gender: ……….
Level: ……………………………………Department: …………………………..
Please indicate the people on this list whom you would consider to be a friend. (Leave your own row blank. You may add comments if you wish.)

Name of person	Friend (Tick)	Comment
Al-1		
Bru-2		
Chas-3		
Dave-4		
Ed-5		
Frank-6		

Questionnaire to network data

Sociometric survey of XXXX office

Your Information:

Name: ……………………………………   *Age: ……… Gender: ……….*

Level: ……………………………………   *Department: ………………………..*

Please indicate the people on this list whom you would consider to be a friend. (Leave your own row blank. You may add comments if you wish.)

Name of person	Friend (Tick)	Comment
Al-1		
Bru-2		
Chas-3		
Dave-4		
Ed-5		
Frank-6		
Gra-7		
Hal-8		
Ian-9		

	A	B	C	D	E
1	FROM	TO	Friendship		
2	Al-1	Al-1	0		
3	Al-1	Bru-2	1		
4	Al-1	Chas-3	0		
5	Al-1	Dave-4	1		
6	Al-1	Ed-5	0		
7	Al-1	Frank-6	0		
8	Al-1	Gra-7	0		
9	Al-1	Hal-8	1		
10	Al-1	Ian-9	0		
11	Al-1	Jo-10	0		
12	Al-1	Ken-11	0		
13	Al-1	Leo-12	1		
14	Al-1	Mal-13	0		
15	Al-1	Ned-14	0		
16	Al-1	Ollie-15	0		
17	Al-1	Pat-16	1		
18	Al-1	Quinn-17	0		
19	Al-1	Ron-18	0		
20	Al-1	Sam-19	0		
21	Al-1	Tom-20	0		
22	Al-1	Unwin-21	0		
23	Bru-2	Al-1	1		
24	Bru-2	Bru-2	0		
25	Bru-2	Chas-3	0		
26	Bru-2	Dave-4	0		
27	Bru-2	Ed-5	0		
28	Bru-2	Frank-6	0		
29	Bru-2	Gra-7	0		
30	Bru-2	Hal-8	0		
31	Bru-2	Ian-9	0		
32	Bru-2	Jo-10	0		
33	Bru-2	Ken-11	0		
34	Bru-2	Leo-12	0		
35	Bru-2	Mal-13	0		

NodeData / OrigTieData / FullTieDataList /

Ready

Start

Cognitive dimensions of network data

- Recorded data of activities has no cognitive dimension
  - Email from A to B
  - Packages between routers
- Questions/ responses involve cognitive dimension
  - Who do you consider a friend? has a deliberate, subjective, cognitive dimension
  - Who do you have lunch with? tries to eliminate the subjective; Cognition still impacts through recall

Qualitative dimensions of network data

- Qualitative researchers explore the content of relations
  - Friendship:? What do you do together? Do you trust them? Etc.
  - Other relations:
    - Family: how often seen? etc. Friends ('keep in touch with')
    - Interaction contacts: help out with jobs, have to dinner, socialise with etc.

Positives: 1. Remember the humans behind the activity

- Lessons from the qualitative and cognitive aspects of SNA practice
  - Recorded activity is surface phenomena – 'traffic'
  - Underneath is (network) infrastructure
  - Surrounding each relation are cognitive perceptions and qualitative ambiguity
  - Sustained patterns of traffic suggest inter-subjective agreement
  - We have maps of activity patterns (like tracking the weather)
- Are we getting inside people's heads
  - Early years of the web seemed so; the online world was the 'virtual' world. Not so now.
  - Underlying process is people building and maintaining social relations and relationships

WN-SNA measures

- WN-SNA has an arsenal of measures
  - Centrality, density, closeness, betweenness
  - Cliques, clans, K-cores, N-cores…
  - Path length (degrees of separation)
- Origins of these measures
  - Designed to normalise features of small observed networks to allow comparison
  - In fact, the separate elements of (small) network size (number of nodes AND number of ties) make this difficult
- Network science alternatives make better sense
  - Global (sampled) measures; average path lengths, clustering coefficients

Positive 2. Egonet data analysis

- Ego-centric network analysis (egonet SNA)
  - Starts with open name generator questions:
    - Who are the people you …
    - Content of the relation is defined by question
  - Once a name is given, the data point is the relation or relationship (Respondent TO 'A'). A single interview/questionnaire generates information about many relations.
  - A data bank of relations can be analysed with normal statistical methods (if categorisations are valid)
- Social media application
  - Download information about sampled relationships
  - Analyse for quantitative (frequency etc) or qualitative (content of messages) properties

Positive 3: SNA network modelling

- Modelling ('Exploratory data analysis')
  - It is systematic, rigorous and 'scientific'
  - Analogy: It is like an aircraft (model) in a wind tunnel; adjustments are made and assessed
  - Goodness of fit ('tests against randomness') assess the model, not specific causes
- PNet (UniMelb)
  - World best network modelling practice
  - How do we link to this? How do conceptualise online databases as these complex (network) systems?