

Mining and Analyzing Social Media: Part 2

Dave King

HICSS 47 - January 2014



Agenda: Part 2



- Introduction to Social Network Analysis Metrics
 - Degrees of Separation
 - Hooray for Bollywood
- Standard Measures
 - Centrality
 - Cohesion
- Levels of Social Network Analysis
 - Facebook & Egocentric Analysis
 - Searching for Cohesive Subgroups

Social Network Analysis

Analyzing Social Networks – The World of Metrics



Social scientists, physicists, computer scientists, and mathematicians have collaborated to create theories and

algorithms for calculating novel measurements of social networks and the people and things that populate them.

These quantitative network metrics allow analysts to systematically dissect the social world, ***creating a basis on which to compare networks, track changes in a network over time, and determine the relative position of individuals and clusters within a network.***

Social Network Analysis

Wide Variety of Metrics Available



Measures Manager

Select Measures Set Measure Inputs

	Measure Title	Network Level	Node Level	Computation...	U...
<input type="checkbox"/>	Redundancy, Access	true	false	normal	fz
<input type="checkbox"/>	Actual Workload	false	true	normal	fz
<input type="checkbox"/>	Socio Economic Power, Agent	false	true	normal	fz
<input type="checkbox"/>	Redundancy, Assignment	true	false	normal	fz
<input type="checkbox"/>	Centrality, Authority	false	true	normal	tr
<input type="checkbox"/>	Characteristic Path Length	true	false	normal	tr
<input type="checkbox"/>	Speed, Average	true	false	normal	tr
<input checked="" type="checkbox"/>	Centrality, Betweenness	false	true	normal	tr
<input checked="" type="checkbox"/>	Network Centralization, Betweenness	true	false	normal	tr
<input type="checkbox"/>	Centrality, Bonacich Power	false	true	normal	tr
<input type="checkbox"/>	Capability	false	true	normal	tr
<input type="checkbox"/>	Clique Count	false	true	normal	fz
<input checked="" type="checkbox"/>	Centrality, Closeness	false	true	normal	tr
<input checked="" type="checkbox"/>	Network Centralization, Closeness	true	false	normal	tr
<input checked="" type="checkbox"/>	Density, Clustering Coefficient	true	true	normal	fz
<input type="checkbox"/>	Cognitive Demand	false	true	normal	tr
<input type="checkbox"/>	Cognitive Distinctiveness	false	true	normal	fz
<input type="checkbox"/>	Cognitive Expertise	false	true	normal	fz
<input type="checkbox"/>	Cognitive Resemblance	false	true	normal	fz
<input type="checkbox"/>	Cognitive Similarity	false	true	normal	fz
<input type="checkbox"/>	Breadth, Column	true	false	normal	fz
<input type="checkbox"/>	Count, Column	true	false	normal	fz
<input type="checkbox"/>	Centrality, Column Dearee	false	true	normal	tr

Select All Select Visible

13 / 160 Selected, 160 / 160 Visible

OK Close

Social Network Analysis

Best Known Social Metric

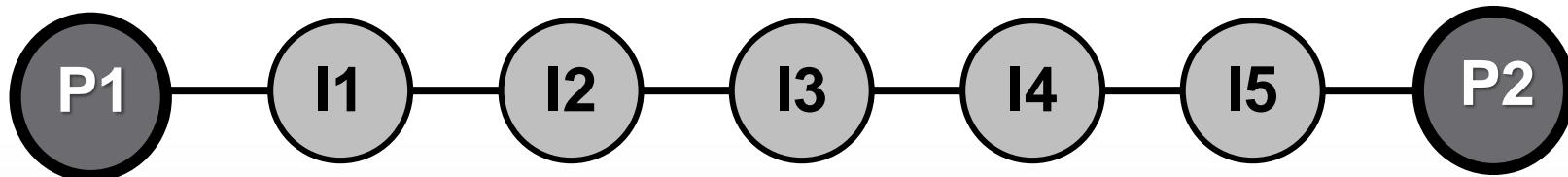
6

Social Network Analysis

6 Degrees of Separation

A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth—anyone, anywhere at all. He bet us that, using no more than *five* individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances.

Frigyes Karinthy , *Chains*, 1929



Degrees of separation ~ Average Path Length ~ Distance

Social Network Analysis

Why the notoriety?

Frigyes Karinthy
1929



John Guare
1990



Stanley Milgram
1967



Duncan Watts
1998



Six Degrees of Kevin Bacon (1994)

Social Networks

The Oracle of Bacon



THE ORACLE OF BACON

Robert Keeshan has a Bacon number of 2.

[Find a different link](#)



Kevin Bacon to [Robert Keeshan](#) Find link More options >>



Welcome
Credits
How it Works
Contact Us
Other stuff »



© 1999-2013 by Patrick
Reynolds. All rights reserved.

Social Network Analysis

Oracle of Bacon gone mainstream

The Oracle of Bacon

anna kendrick bacon number

THE ORACLE

Contact Us

Google News

anna kendrick bacon num

<https://www.google.com/search?q=anna+kendrick+bacon+number&oq=anna+ke>

Web Images Maps Shopping More Search tools

About 6,260,000 results (0.45 seconds)

Anna Kendrick's Bacon number is 2

Anna Kendrick and Ryan Reynolds appeared in *The Voices*.
Ryan Reynolds and Kevin Bacon appeared in *R.I.P.D.*.

Social Network Analysis

How the Oracle of Bacon Works

How the Oracle of Bacon Works

Every couple of weeks the Oracle downloads several database files from one of the [Internet Movie Database's FTP sites](#) containing around 1.5 million actors and actresses, around 1.2 million movies and TV shows, and around 200,000 nicknames. The Oracle builds a big map of actors and movies and stores it in a 252 MB database file.

There is a database server running at all times that stores the database file in memory. The server handles three different types of requests:

- Find the link from Actor A to Actor B.
- How good a "center" is a given actor?
- Who are all the people with an Actor A number of N?

There are several CGI programs -- one for each of the above types of queries -- that run on the Oracle of Bacon web server, which all connect to the database server using TCP.

The database server uses a breadth-first search (BFS) to find the shortest path between pairs of actors. If you want to dig further into how shortest-path algorithms work, I recommend the textbook [by Cormen, Leiserson, Rivest, and Stein](#) as an excellent place to start. Other algorithms textbooks are likely to cover the subject as well, if *Introduction to Algorithms* isn't available. You may also look at materials that I wrote to explain graph algorithms (including BFS) to Duke undergraduate CS students [here](#).

Whenever the Oracle answers a query, the results are cached so that future requests to link to the same actor will occur more quickly. About 80% of all queries can be served instantly from the result cache. The current contents of the cache (i.e., which actors can be linked quickly) can be found [here](#).

Social Network Analysis

Center of the Hollywood Universe

The Center of the Hollywood Universe

Click on a name to see that person's table.

1. [Harvey Keitel](#) (2.848635)
2. [Dennis Hopper](#) (2.849329)
3. [Robert De Niro](#) (2.855810)
4. [David Carradine](#) (2.857729)
5. [Martin Sheen](#) (2.858291)
6. [Udo Kier](#) (2.859489)
7. [Michael Madsen](#) (I) (2.860010)
8. [Donald Sutherland](#) (I) (2.860447)
9. [Michael Caine](#) (I) (2.862189)
10. [Eric Roberts](#) (I) (2.867675)
11. [Seymour Cassel](#) (2.869415)
12. [Malcolm McDowell](#) (2.870208)
13. [Max von Sydow](#) (I) (2.872338)
14. [Willem Dafoe](#) (2.873805)
15. [Samuel L. Jackson](#) (2.873819)
16. [Danny Trejo](#) (2.876002)
17. [John Hurt](#) (2.878378)
18. [Christopher Lee](#) (I) (2.879217)
19. [Harry Dean Stanton](#) (2.880725)
20. [Bruce Willis](#) (2.886364)
21. [Christopher Plummer](#) (I) (2.886928)
22. [John Malkovich](#) (2.888575)
23. [Morgan Freeman](#) (I) (2.891003)
24. [Christopher Walken](#) (2.894212)
25. [John Savage](#) (I) (2.894873)

Kevin Bacon Number	# of People
0	1
1	2799
2	313045
3	1078865
4	276680
5	22296
6	2361
7	251
8	24

Total number of linkable actors: 1696322
Weighted total of linkable actors: 5099799
Average Kevin Bacon number: 3.006

Kyra Sedgwick Number	# of People
0	1
1	1353
2	229226
3	1083255
4	350117
5	29167
6	2845
7	331
8	27

Total number of linkable actors: 1696322
Weighted total of linkable actors: 5275476
Average Kyra Sedgwick number: 3.110

Harvey Keitel (I) Number	# of People
0	1
1	4128
2	454260
3	1051685
4	169704
5	14709
6	1679
7	141
8	15

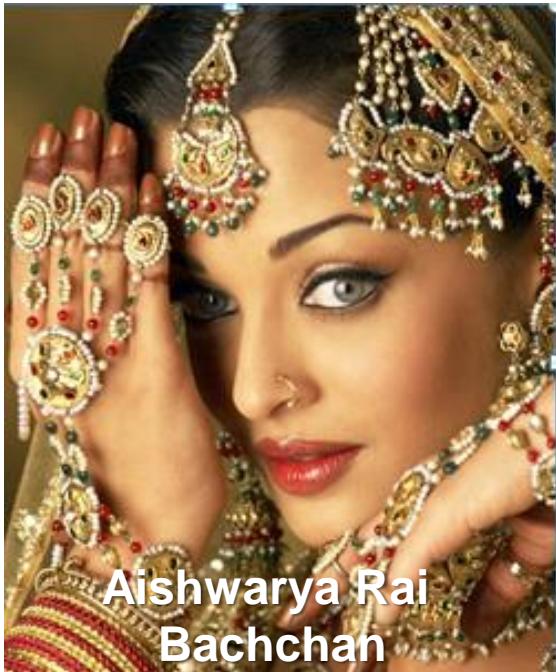
Total number of linkable actors: 1696322
Weighted total of linkable actors: 4831245
Average Harvey Keitel (I) number: 2.848

John Savage (I) Number	# of People
0	1
1	3766
2	408573
3	1073294
4	192718
5	16000
6	1764
7	195
8	11

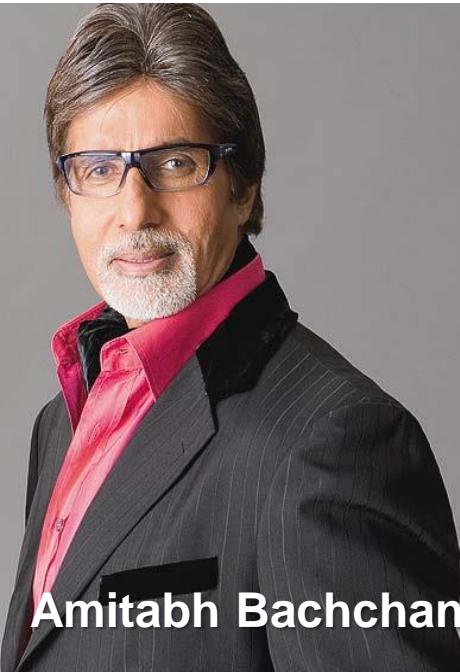
Total number of linkable actors: 1696322
Weighted total of linkable actors: 4903703
Average John Savage (I) number: 2.891

Social Network Analysis

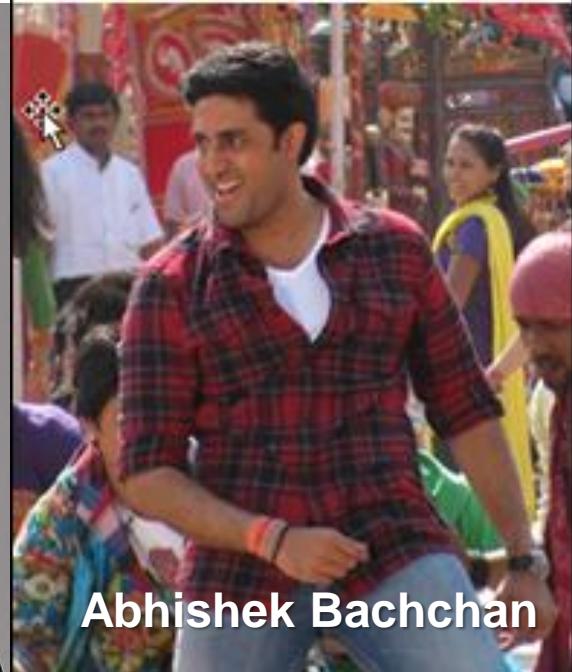
ओरेकल *of Bachchan*



Aishwarya Rai
Bachchan



Amitabh Bachchan



Abhishek Bachchan

जुदाई की छह डिग्री

judā'ī kī chaha digrī

Social Network Analysis

There's a new game in town

Amitabh Bachchan - Google

List of Bollywood films - Wikipedia

en.wikipedia.org/wiki/List_of_Bollywood_films#2010s

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
Print/export

Languages
हिन्दी
Português
ଓଡ଼ିଆ
Edit links

List of Bollywood films

From Wikipedia, the free encyclopedia

This article does not cite any references or sources. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (September 2012)

This is a list of films produced by the Bollywood film industry of Mumbai ordered by year and decade of release and also contains the top ten or forty superhit films of respective years as the case may be. Although "Bollywood" films are generally listed under the Hindi language, most are in mixed Hindi, Urdu and Punjabi and occasionally other languages. There is a range of mixtures from mostly Urdu to mostly Hindi to mostly Punjabi. Speakers of Hindi, Urdu, and Punjabi understand the mixed language usage of Bollywood thus extending the viewership to people all over the Indian subcontinent (throughout India and its neighboring countries). Here are some examples - Partly Hindi: *Om Shanti Om*, *Dhoom 2*, *No Entry* and *Kabhi Alvida Naa Kehna*, Partly Urdu: *Jodhaa Akbar*, *Fanaa*, *Saawariya* and *Kurbaan*, Partly Punjabi: *Singh Is Kinng*, *Jab We Met*, *Patiala House* and *Rab Ne Bana Di Jodi*. The film *Veer Zaara* is an equal mix of Hindi, Punjabi and Urdu.

Contents [hide]

- 1 2010s
- 2 2000s
- 3 1990s
- 4 1980s
- 5 1970s
- 6 1960s
- 7 1950s
- 8 1940s
- 9 1930s

2010s [edit]

- List of Bollywood films of 2010
- List of Bollywood films of 2011
- List of Bollywood films of 2012
- List of Bollywood films of 2013
- List of Bollywood films of 2014

2000s [edit]

- List of Bollywood films of 2000


Alam Ara (1931), the first Indian sound film

Social Network Analysis

The Center of Bollywood

2010 releases [edit]						
A list of Bollywood films released in 2010 [5]						
January–March [edit]						
Opening	Title	Director	Cast	Genre [*]		
8	Dulha Mil Gaya	Muddassar Aziz	Fardeen Khan, Ishita Sharma, Sushmita Sen, Shahrukh Khan, Tara Sharma, Anushka Manchanda, Mohit Chadda, Vivek Vasani, Johnny Lever	Romantic comedy		
J A N	15	Pyaar Impossible!	Jugal Hansraj	Priyanka Chopra, Uday Chopra, Anupam Kher, Dino Morea	Romantic comedy	
	Chance Pe Dance	Ken Ghosh	Shahid Kapoor, Genelia D'Souza	Comedy, Romance		
	HejHo! Hum Lailo! Ba Rahe Hain	Dileep Shukla	Rajpal Yadav, Preeti Mehra	Comedy		
	The Waiting Room	Maneck Premji	Raj Singh Chaudhary, Radhika Apte	Thriller		
	22	Veer	Anil Sharma	Salman Khan, Zainab Khan, Mithun Chakraborty, Sohail Khan, Jackie Shroff	Epic film	
	Ishqya	Abshek Chaubey	Naseeruddin Shah, Vidya Balan, Arshad Warsi	Comedy, Crime, Romance		
	29	Rann	Ram Gopal Varma	Amitabh Bachchan, Sudeep, Ritesh Deshmukh, Paresh Rawal, Rajpal Yadav, Gul Panag	Drama	
		Road to Sangram	Amit Rai	Paresh Rawal, Om Puri, Javed Sheikh	Drama	
F E B	5	The Hangman	Vishal Bhardwaj	Om Puri, Shekhar Kapur, Gulshan Grover, Smita Jaykar, Tom Alter, Anita Kanwar, Yatin Kankar, Nazneen Ghani, Amrita Bedi	Drama	
	12	My Name Is Khan	Chandan Arora	Siddharth Narayan, Aditya Pancholi, Arulkirikaval, Anupam Kher, Seema Biswas	Action	
		Sukhmani	Karan Johar	Shahrukh Khan, Kajol, Jimmy Sheirgill	Drama	
	15	Aakhan Deception	Manjeet Mann	Gurdas Maan, Juhu Chawla, Divya Dutta	Action	
		Click	Deepak Bandhu	Anant Jog, Nagesh Bhonsle, Mushtaq Khan	Drama	
		Toh Baat Pakki!	Sangeeth Sivan	Shreyas Talpade, Smita Sirehia Utal	Action	
		Karthik Calling Karthik	Kedar Shinde	Tabu, Shammi Joshi, Vaibhavi Merchant	Honor	
	26	Teen Patti	Vijay Lalwan	Farhan Akhtar, Deepika Padukone, Priyanka Chopra, Aishwarya Rai Bachchan, Mahesh Manjrekar	Comedy	
			Leena Yadav			

Wikipedia Web Page

Bollywood Films - Microsoft Excel								
File	Home	Insert	Page Layout	Formulas	Data	Review	Developer	Acrobat
	Cut	Paste	Format Painter	Wrap Text	General	Conditional Formatting as Table Styles	Insert	Delete
	Font	Font	Font	Font	Number	Styles	Cells	Cells
E16	Mehul Kumar	Font	Font	Font	Number	Styles	Cells	Cells
1	MovieID	Title	SubGenre	Genre	Director	Cast		Opening Year
2	1	Mumbai Mirror	Action	Action	Ankush Bhatt	Sachin Joshi, Gihana Khan, Prakash Raj[30]		Jan 2013
3	2	Vishwaroop	Action	Action	Kamal Hassan	Kamal Hassan, Poja Kumar, Shekhar Kapur, Andrea Jeremiah, Rahul D	Feb 2013	
4	3	Zilla Ghaziabad	Action	Action	Anand Kumar	Arshad Warsi, Vivek Oberoi, Sanjay Dutt, Minisha Lamba, Ravi Krishan, F	Feb 2013	
5	4	4 Enemy	Action	Action	Ashu Trikhia	Mithun Chakraborty, Sunil Shetty, Mahakshay Chakraborty, Kay Kay M	Jun 2013	
6	5	Policegiri	Action	Action	K. S. Ravikumar	Sanjay Dutt, Prakash Raj, Om Purji, Prachi Desai[82]	Jul 2013	
7	6	Singh Saab The Great	Action	Action	Antil Sharma	Sunny Deol, Amritrao, John Lever, Urvashi Rautela[133]	Nov 2013	
8	7	Dhoom 3	Action	Action	Vijay Kumar Acharya	Aamir Khan, Abhishek Bachchan, Katrina Kaif, Uday Chopra[138]	Dec 2013	
9	8	T2	Action	Action	Prabhudeva	Ajay Devgn, Akshay Kumar, Kangana Ranaut, Zayed Khan, Sameera Reddy	Jan 2014	
10	9	Singham	Action	Action	Rohit Shetty	Asif Ali, Karan Johar, Aamir Khan, Akshay Kumar, Prakash Raj[43]	Jul 2011	
11	10	Shahri	Action	Action	Lalit Morarji	Eesha Koppikar[16]	Aug 2011	
12	11	Striker	Action	Action	Chandan Arora	SiddhARTH Narayan, Aditya Pancholi, Ankur Vikal, Anupam Kher, Seem	Feb 2010	
13	12	Aakkhar Decision	Action	Action	Deepak Bhandhu	Anant Jogi, Nageen Khoshe, Muftaq Khan	Feb 2010	
14	13	Right Yaa Wrong	Action	Action	Neeraj Pathak	Sunny Deol, Irrfan Khan, Esha Koppikar, Konkona Sen Sharma	Mar 2010	
15	14	Chase	Action	Action	Jagmohan Mundhra	Udit Goswami, Catherine Murino, Rajesh Khattar, Shweta Menon	Apr 2010	
16	15	Krantiveer - The Revolution	Action	Action	Mehul Kumar	Jahan Bloch, Sameer Afza	May 2010	
17	16	Red Alert: The War Within	Action	Action	Anant Mahadevan	Sunil Shetty, Vinod Khanna, Naseeruddin Shah, Sameera Reddy	Jul 2010	
18	17	Lamhaa	Action	Action	Rahul Roy	Sanjay Dutt, Bipasha Basu, Kunal Kapoor, Anupam Kher	Jul 2010	
19	18	Khuda Kasam	Action	Action	Abhishek Kapoor	Shreyas Talpade, Sonakshi Sinha, Aishwarya Rai Bachchan, Mahie	Oct 2010	
20	19	Team India	Action	Action	Manoj Bajpayee	Shreyas Talpade, Sonakshi Sinha, Aishwarya Rai Bachchan, Mahie	Oct 2010	
21	20	Khuda Kasam	Action	Action	C. K. Bosadia	Sunny Deol, Tabu, Farida Jalal, Govind Nihalani	Nov 2010	
22	21	Team: The Force	Action	Action	Alay Chandhok	Sohail Khan, Amrita Arora, Yash Tonk	May 2009	
23	22	Baabbar	Action	Action	Shreyas Talpade	Shreyas Talpade, Sonakshi Sinha, Urmila Matondkar, Urvashi Sharma	Oct 2009	
24	23	3 Acid Factory	Action	Action				
		Bollywood Films						

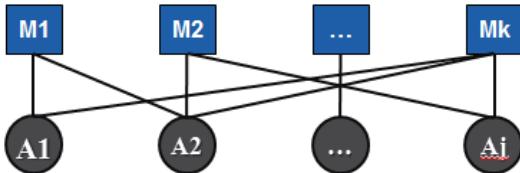
HTML- Page Source

```
*Network bollywood.net [2-Mode]
*Vertices 1643 627
1 "Mumbai Mirror" 0.0000 0.0000 0.5000
2 "Vishwaroop" 0.0000 0.0000 0.5000
...
628 "A. K. Hangal" 0.0000 0.0000 0.5000
629 "Aamir Ali" 0.0000 0.0000 0.5000
...
*Arcs
*Edges
1 896 1
1 1220 1
...
627 856 1
627 1053 1
```

Pajek .NET File

Social Network Analysis

Bipartite Network – Movies and Actors



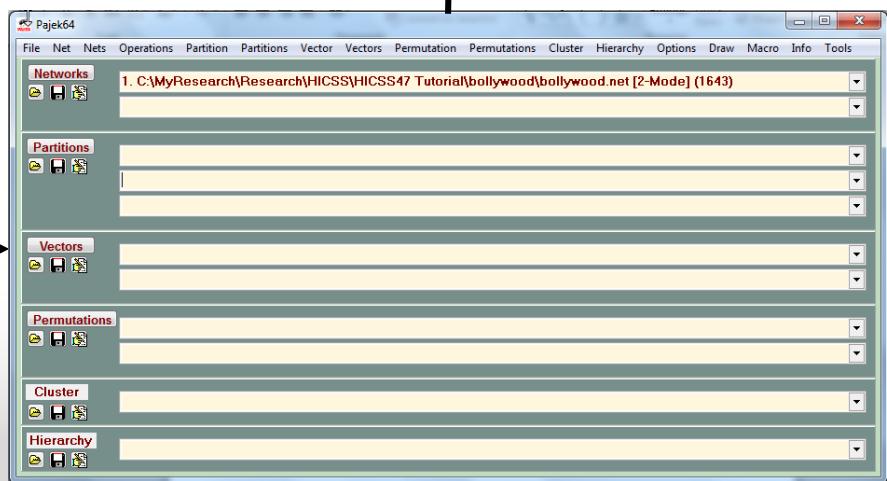
	M1	M2	...	Mk
A1	1	0		1
A2	1	1	...	1
...
Aj	0	1		1

	A1	A2	...	Aj
A1	-	2	...	1
A2	2	-	...	1
...
Aj	1	1	...	-

	M1	M2	...	Mk
M1	-	1	...	2
M2	1	-	...	2
...
Mk	2	2	...	-

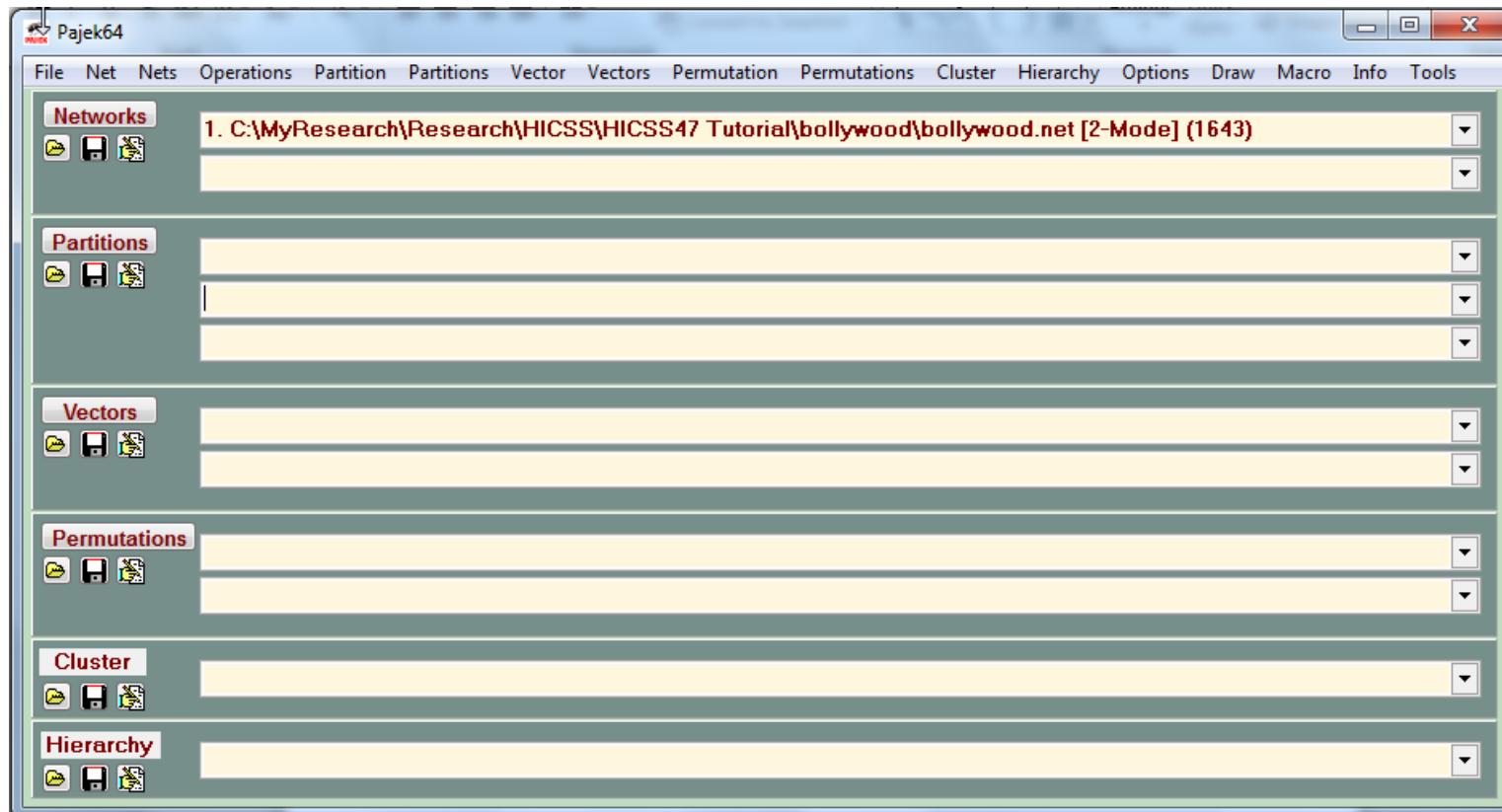
```
*Network bollywood.net [2-Mode]
*Vertices 1643 627
1 "Mumbai Mirror" 0.0000 0.0000 0.5000
2 "Vishwaroop" 0.0000 0.0000 0.5000
...
628 "A. K. Hangal" 0.0000 0.0000 0.5000
629 "Aamir Ali" 0.0000 0.0000 0.5000
...
*Arcs
*Edges
1 896 1
1 1220 1
...
627 856 1
627 1053 1
```

Pajek .NET File



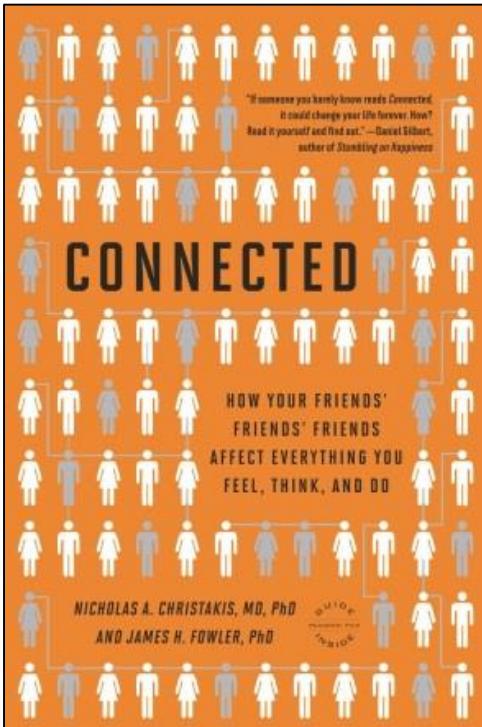
Social Network Analysis

Analyzing *Bollywood*



Social Networks

Rules of Networks

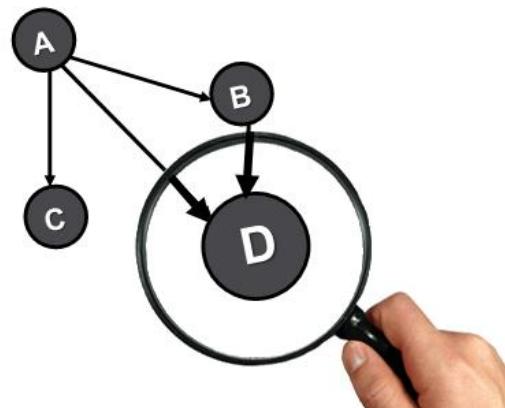


- RULE 1: WE SHAPE OUR NETWORK
- RULE 2: OUR NETWORK SHAPES US
- RULE 3: OUR FRIENDS AFFECT US
- RULE 4: OUR FRIENDS' FRIENDS' FRIENDS AFFECT US
- RULE 5: THE NETWORK HAS A LIFE OF ITS OWN

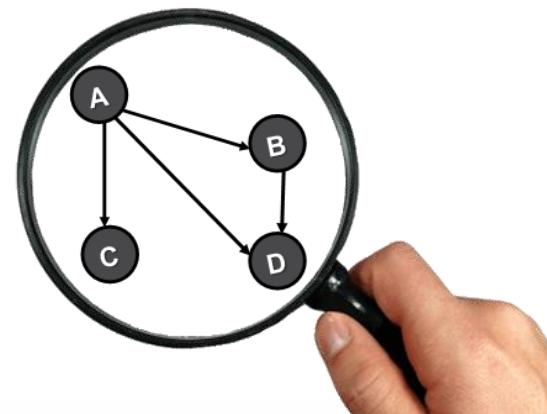
Social Network Analysis

Bifurcated Methodology

Local
Measures



Global
Measures

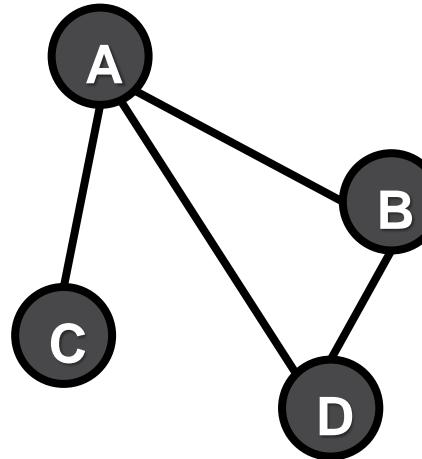


Social Network Analysis

Walks and Paths

- Two points may be directly connected or indirectly connected thru a sequence of lines called a *walk*.
- If points and lines are distinct it's a *path*.
- The *length* of a path is the number of edges (or lines) that make it up.
- The shortest length is called the *distance* or *geodesic*.

Undirected



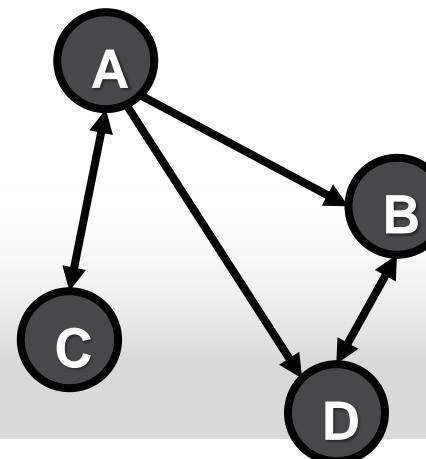
Sample Walks

((C,A),(A,D))
((C,A),(A,B),(B,D))
((C,A),(A,D),(D,A),(A,D))

Sample Paths

((C,A),(A,D))
((C,A),(A,B),(B,D))

Directed



Sample Walks

((C,A),(A,D))
((C,A),(A,B),(B,D))

Sample Paths

((C,A),(A,D))
((C,A),(A,B),(B,D))

Social Network Analysis

Centrality – Who is most influential?

Measure	Definition	Interpretation	Reasoning
Degree	Number of edges or links. In degree- links in, Out-degree - links out	How connected is a node? How many people can this person reach directly?	Higher probability of receiving and transmitting information flows in the network. Nodes considered to have influence over larger number of nodes and or are capable of communicating quickly with the nodes in their neighborhood.
Betweenness	Number of times node or vertex lies on shortest path between 2 nodes divided by number of all the shortest paths	How important is a node in terms of connecting other nodes? How likely is this person to be the most direct route between two people in the network?	Degree to which node controls flow of information in the network. Those with high betweenness function as brokers. Useful where a network is vulnerable.
Closeness	1 over the average distance between a node and every other node in the network	How easily can a node reach other nodes? How fast can this person reach everyone in the network?	Measure of reach. Importance based on how close a node is located with respect to every other node in the network. Nodes able to reach most or be reached by most all other nodes in the network through geodesic paths.
Eigenvector	Proportional to the sum of the eigenvector centralities of all the nodes directly connected to it.	How important, central, or influential are a node's neighbors? How well is this person connected to other well-connected people?	Evaluates a player's popularity. Identifies centers of large cliques. Node with more connections to higher scoring nodes is more important.

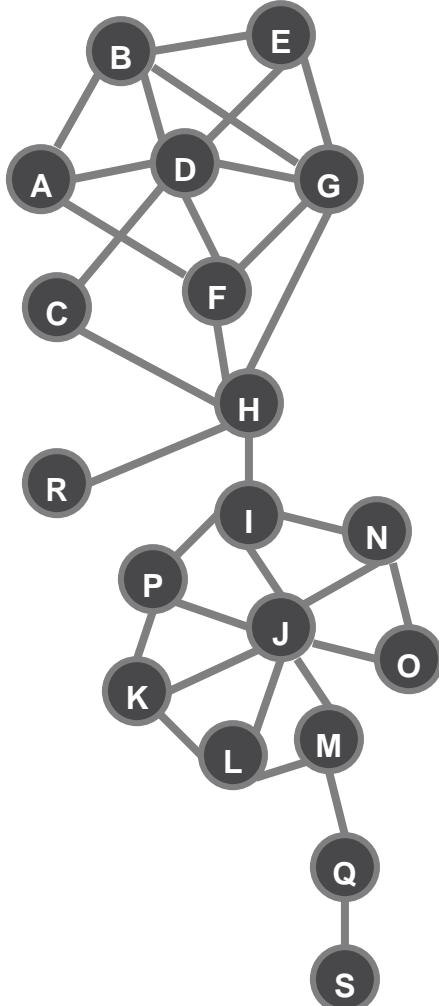
Social Network Analysis

Cohesion – How well connected?

Cohesion	Definition	Interpretation	Reasoning
Density	Ratio of the number of edges in the network over the total number of possible edges between all pairs of nodes	How well connected is the overall network?	Perfectly connected network is called a "clique" and has a density of 1.
Clustering	A node's <i>clustering coefficient</i> is the density of its 1.5 degree egocentric network (ratio of connecting among ego's alters). For entire network it is the average of all the coefficients for the individual nodes.	What proportion of ego's alters are connected? More technically, how many nodes form triangular subgraphs with their adjacent nodes?	Measures certain aspects of "cliquishness." Proportion of you friends that are also friends with each other. Another way to measure is to determine (in a undirected) graph the ratio of the number of times that two links emanating from the same node are also linked.
Average Path Length (Distance)	Average number of edges or links between any two nodes (along the shortest path)	On average, how far apart are any two nodes?	This is synonymous with the "degrees of separation" in a network.
Diameter	Longest (shortest path) between any two nodes	At most, how long will it take to reach any node in the network? Sparse networks usually have greater diameters.	Measure of the reach of the network
Centralization	Normalize ratio of the sum of the variances of the centrality of each node from the most central node to the maximum sum possible	Indicates how unequal the distribution of centrality is in a network.	Measures how much variance there is in the distribution of centrality in a network. The measure applies to all forms of centrality.

Social Network Analysis

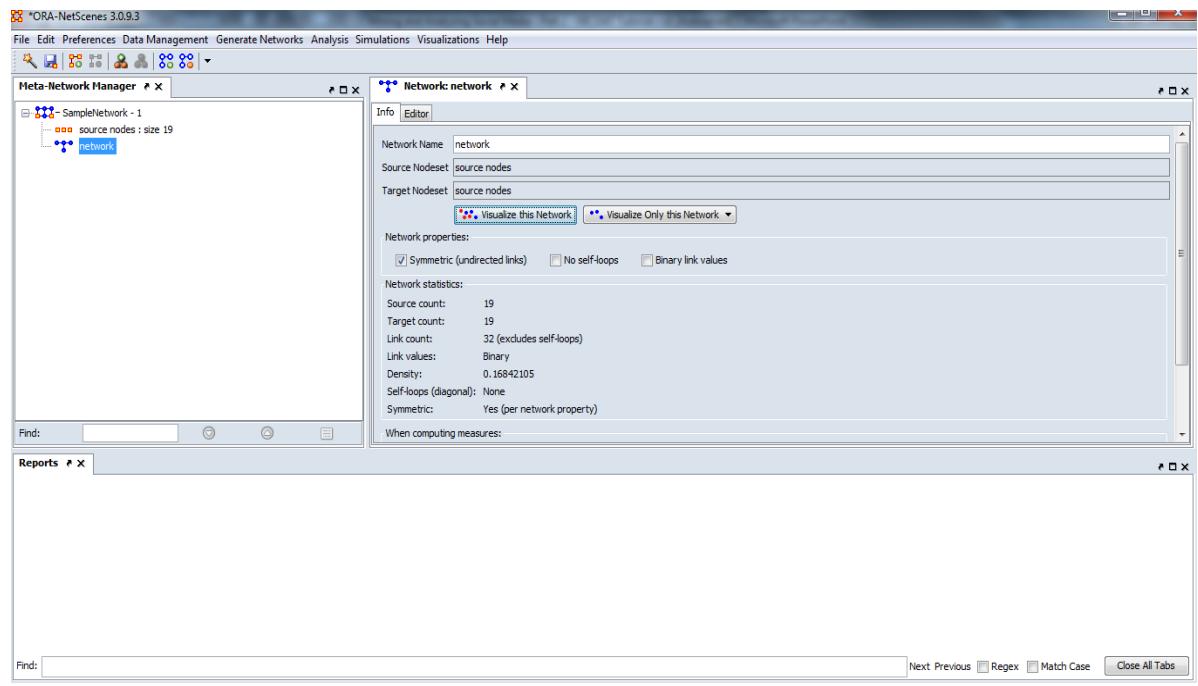
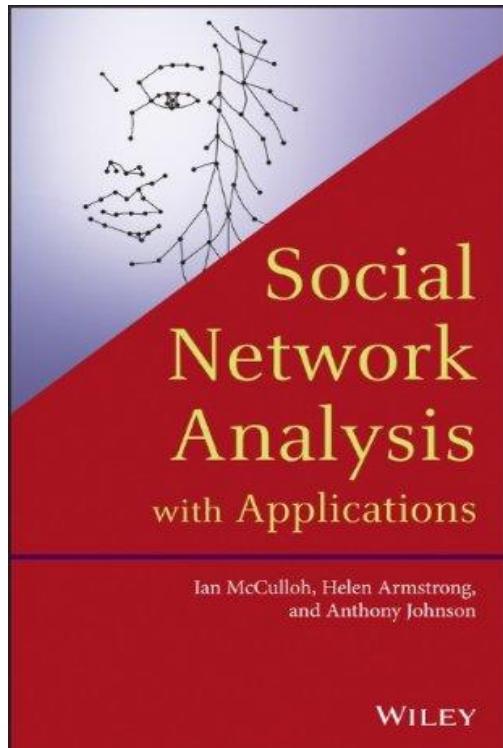
Generic Example for



- For each of the nodes what is it's
 - Degree Centrality
 - Betweenness Centrality
 - Closeness Centrality
 - Eigenvector Centrality
- For the entire network what is it's
 - Degree Centralization
 - Betweenness Centralization
 - Closeness Centralization
 - Eigenvector Centralization

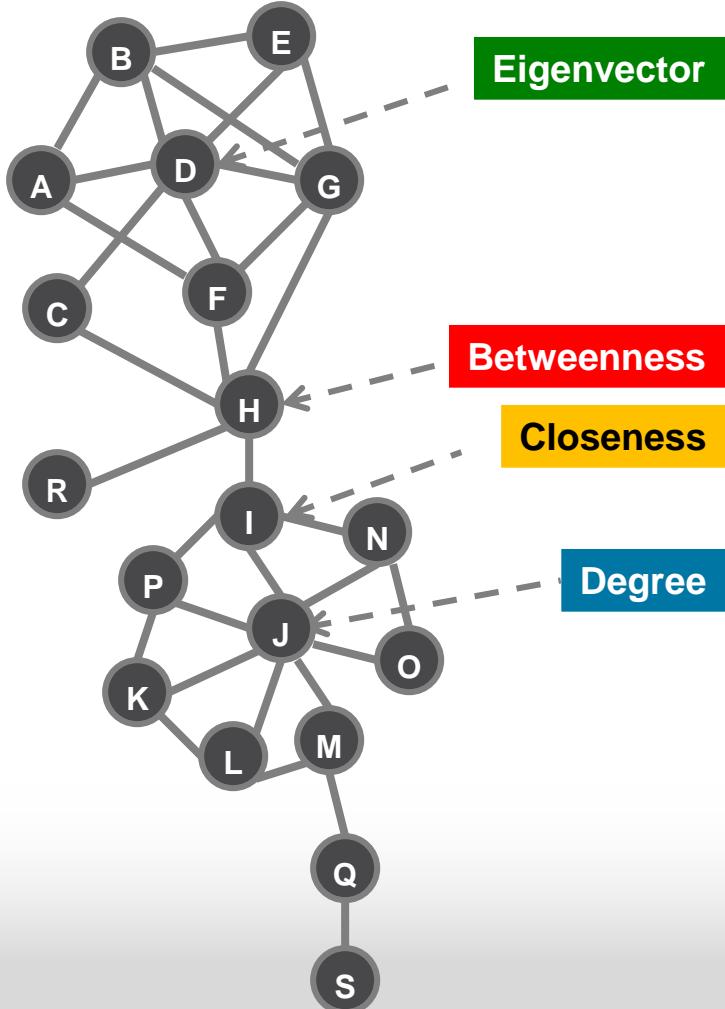
Social Network Analysis

Visualizing and Analyzing with ORA



Social Network Analysis

Centrality – Who is most important?

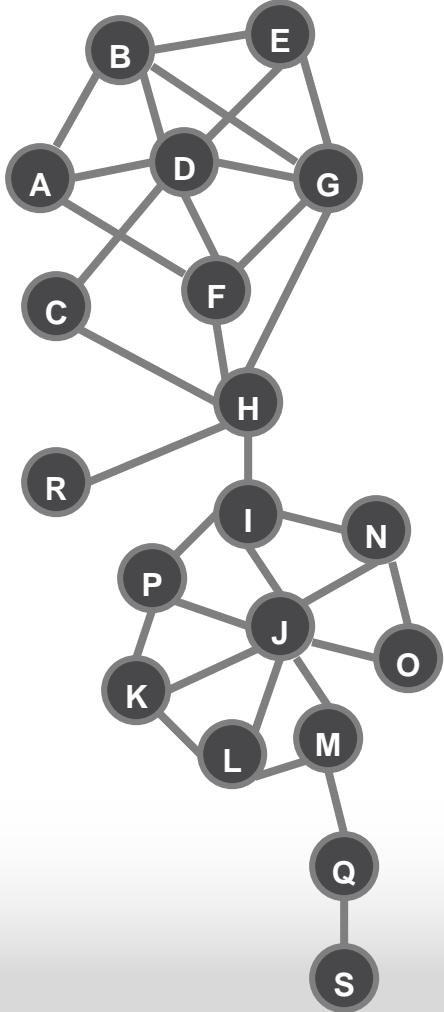


Node	Degree	Normed Degree	Betweenness	Closeness	Eigen Vector
A	3	0.17	0.00	0.29	0.29
B	4	0.22	0.01	0.30	0.36
C	2	0.11	0.03	0.35	0.18
D	6	0.33	0.04	0.31	0.46
E	3	0.17	0.00	0.29	0.30
F	4	0.22	0.11	0.36	0.35
G	5	0.28	0.19	0.37	0.43
H	5	0.28	0.58	0.45	0.28
I	4	0.22	0.53	0.46	0.13
J	7	0.39	0.43	0.43	0.12
K	3	0.17	0.00	0.32	0.06
L	3	0.17	0.01	0.33	0.05
M	3	0.17	0.21	0.33	0.04
N	3	0.17	0.03	0.38	0.07
O	2	0.11	0.00	0.31	0.05
P	3	0.17	0.03	0.38	0.08
Q	2	0.11	0.11	0.26	0.01
R	1	0.06	0.00	0.32	0.07
S	1	0.06	0.00	0.21	0.00

Correlations	Degree	Betweenness	Closeness	Eigenvector
Degree	-	0.57	0.59	0.59
Betweenness		-	0.79	0.07
Closeness			-	0.13
Eigenvector				-

Social Network Analysis

Cohesion – How well connected?



Measure	Value
Network Size	19
Average Degree	3.37
Degree Centralization	0.22
Betweenness Centralization	0.48
Closeness Centralization	0.27
Eigenvector Centralization	0.56
Clustering Coefficient	0.43
Density	0.19
Average Distance	3.06
Diameter	8
Number of Unreachable Nodes	0

Node	Clustering
A	0.67
B	0.67
C	0.00
D	0.40
E	1.00
F	0.50
G	0.50
H	0.10
I	0.33
J	0.29
K	0.67
L	0.67
M	0.33
N	0.67
O	1.00
P	0.67
Q	0.00
R	NA
S	NA

Social Network Analysis

The World of Facebook

Active Users
1.2B  18%

Mobile Users
.9B  45%

Logon Daily
.7B  25%

Likes Daily
4.5B  67%

Pieces of Content Shared Daily
4.8B  94%

New Profiles
5/second

Photo Uploads
300M/day

Of Total Web Page Views
1/5 on FB

Avg Time Per Visit
20 minutes

Sex Ratio
53%F/47%M

Europe
.2B Users

Age 24-34
30% of users

Social Network Analysis

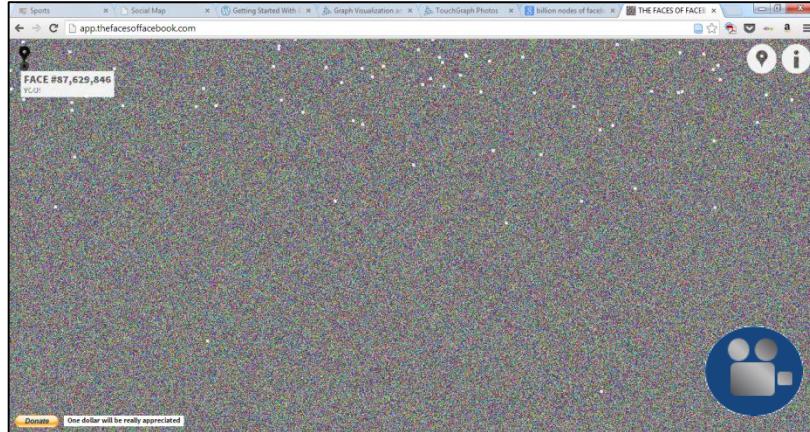
Facebook in the World



[facebookstories.com/stories/1574/
#color=language-official&story=1&country=SA](http://facebookstories.com/stories/1574/#color=language-official&story=1&country=SA)

Social Network Analysis

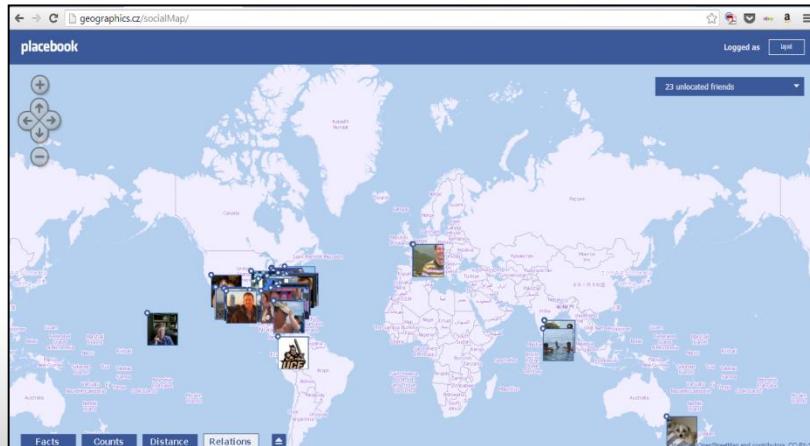
Your World of Facebook



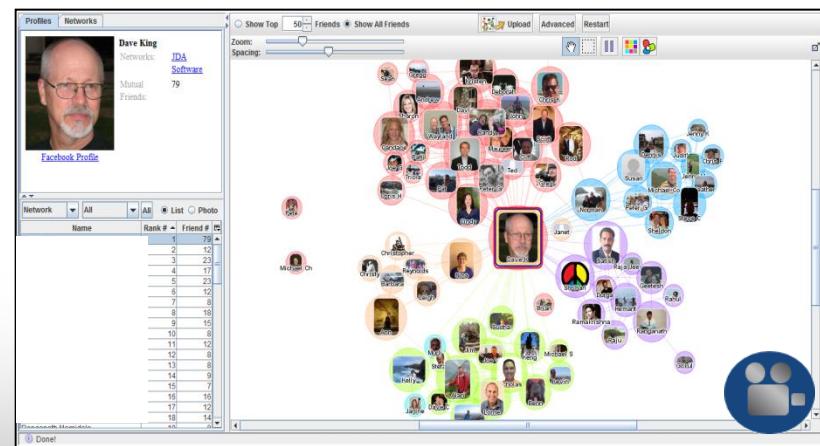
app.thefacesoffacebook.com



apps.facebook.com/challenger_meurs/
?fb_source=appcenter&fb_appcenter=1



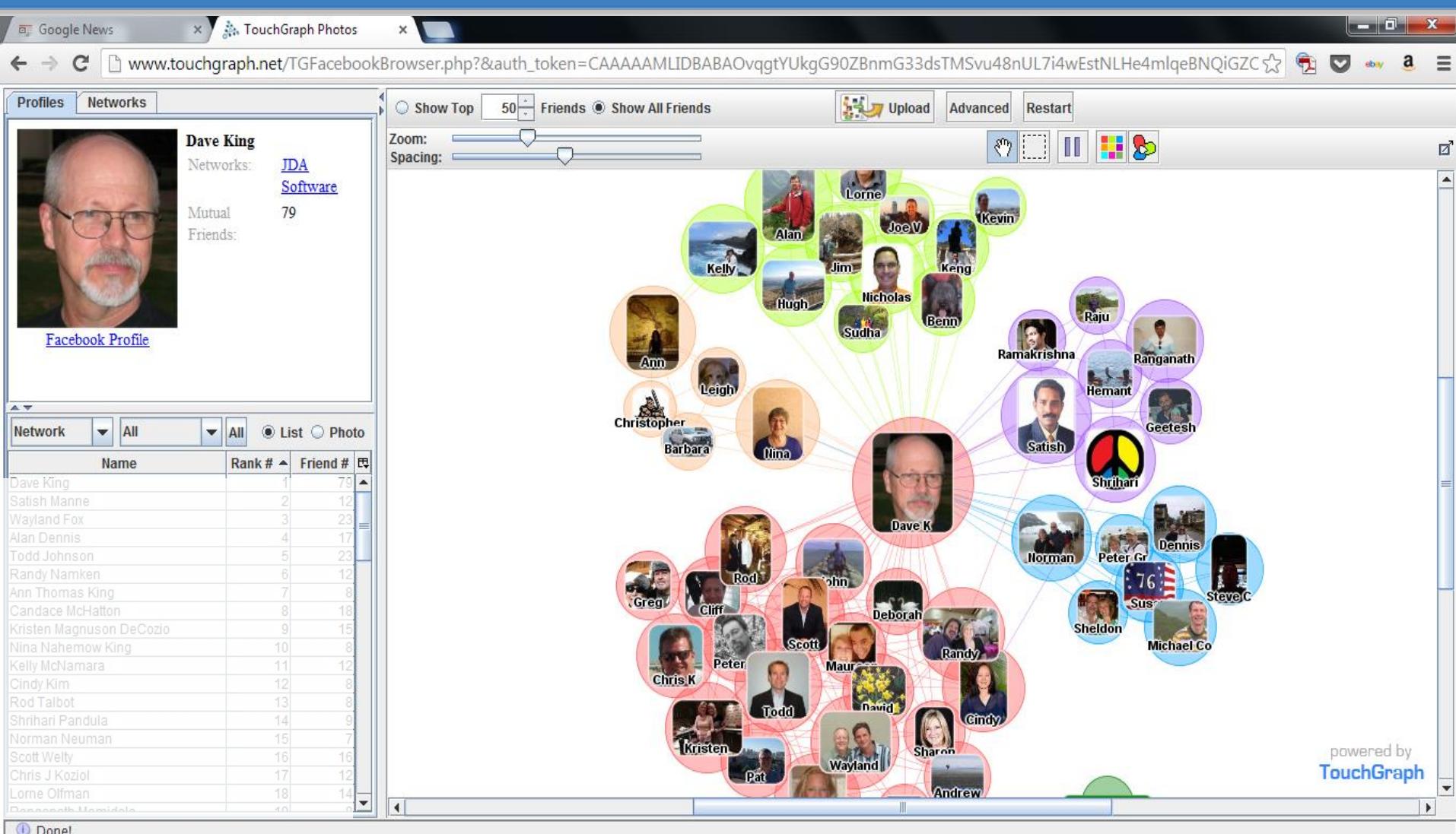
Geographical.cz/socialMap/



apps.facebook.com/touchgraph/

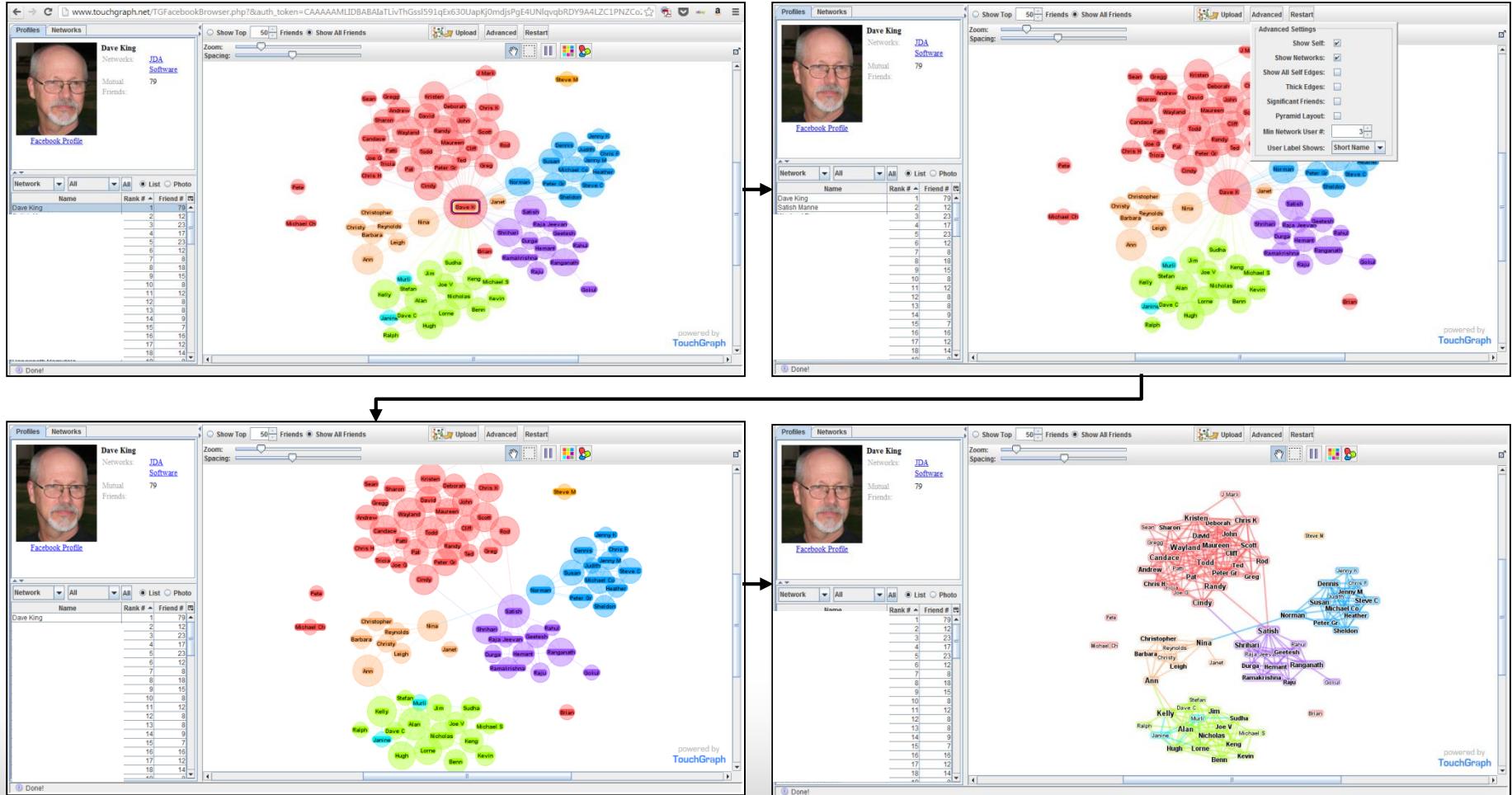
Social Network Analysis

Egocentric Network – Facebook Example



Social Network Analysis

Egocentric Network – Facebook Example



Social Network Analysis

Egocentric Network – Facebook Example

Netvizz v0.93

The screenshot shows the Netvizz v0.93 web application. At the top, there's a header with the URL <https://apps.facebook.com/netvizz/>. Below the header, the title "netvizz" is displayed. The main content area has a heading "netvizz v0.93". It contains several sections of text and checkboxes:

- your personal friend network:** Describes how it creates a network file with all friendship connections.
- Step 1 - Select user data to include in the file (sex, interface language, and account age ranking are standard):** Includes a checkbox for "friends' like and post count (public and visible to logged user, first 1000 only), includes counts for received likes and comments on posts, adds an additional #6 seconds of waiting time per friend".
- Step 2 - create a gdf file from your personal network by clicking [here](#):**
- file fields:** Lists sex, locale, age rank, like count, post count, post like count, post comment count, post engagement count.
- Attention:** Data depends on your friends' privacy settings and the filtering choices you made for your newsfeed.
- your like network:** Describes how it creates a bipartite network (gdf file) from friends' likes.
- groups:** Describes how it extracts social networks from groups.

.gdf (GUESS) file format

The diagram shows a large grey arrow pointing from the Netvizz interface on the left to the GDF file format on the right. The GDF file format is presented as a block of text with several data entries:

```
nodedef>name VARCHAR,label VARCHAR,sex VARCHAR,locale VARCHAR,agerank INT,  
like_count INT,post_count INT,post_like_count INT,  
post_comment_count INT,post_engagement_count INT  
4945386,HW,male,en_US,67,6,3,11,,1  
13307576,LO,male,en_US,66,19,21,54,,0  
512057631,AA,male,en_US,65,20,37,32,,1  
534718919,CB,male,en_US,64,38,239,1384,,2  
539959548,RA,male,en_US,63,69,448,1737,,1  
543650513,SR,female,en_US,62,0,1,2,,3  
558159260,NR,male,en_US,61,15,262,148,,5  
...  
100001597623987,SM,male,en_US,4,3,22,99,,1  
100003001526236,SP,male,en_US,3,0,3,2,,1  
100003603794852,CF,male,en_US,2,5,36,24,,0  
100004022119721,JG,female,en_US,1,0,0,,,0  
edgedef>node1 VARCHAR,node2 VARCHAR  
4945386,13307576  
4945386,558159260  
4945386,601254884  
4945386,626874213  
4945386,662011579  
...  
100000599719815,100001132688907  
100000634883450,100001132688907  
100001132688907,100001378371279  
100001305975715,100001597623987  
100001305975715,100003001526236  
100001393609640,100003603794852  
100001597623987,100003001526236
```

One simple way to generate SNA data from Facebook

Social Network

Egocentric Network – Facebook Example

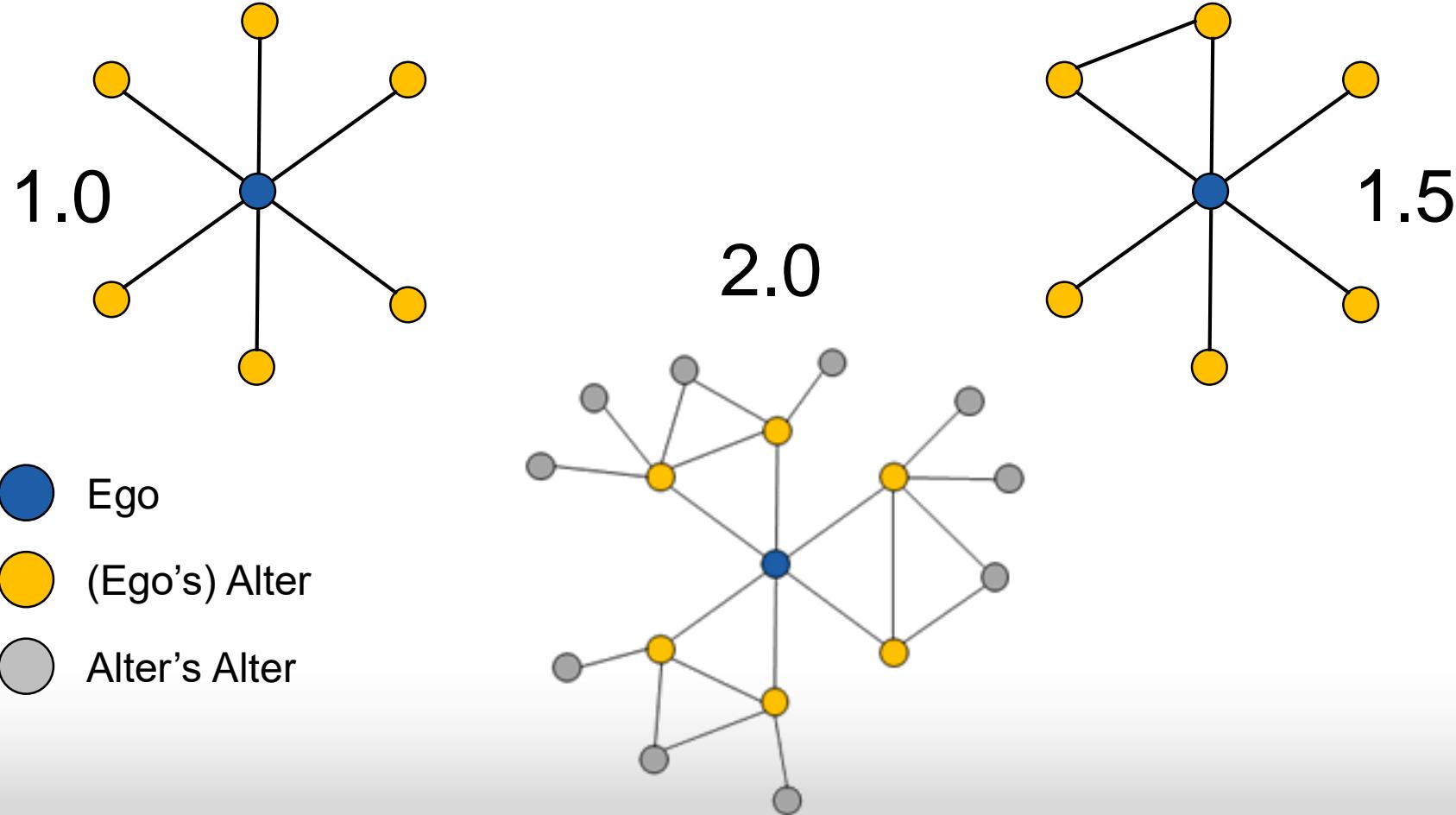
Cleansed-Augmented Guess .GDF File used for analysis

```
nodedef>name VARCHAR,label VARCHAR,ShortName VARCHAR,Group VARCHAR,GroupNum INT,  
Sex VARCHAR,SexNum INT,Age VARCHAR,AgeNum INT,  
Country VARCHAR,CountryNum INT,Relationship VARCHAR,RelNum INT  
1,Alan,Alan,Academic,1,Male,2,Middle,3,United States,1,Married,1  
2,Andrew,Andrew,JDA,5,Male,2,Young,1,United States,1,Single,2  
3,Ann,Ann,Family,4,Female,1,Middle,2,United States,1,Married,1  
4,Barbara,Barbara,Family,4,Female,1,Middle,2,United States,1,Married,1  
5,Benn,Benn,Academic,1,Male,2,Older,3,United States,1,Married,1  
...  
76,Ted,Ted,JDA,5,Male,2,Older,3,United States,1,Married,1  
77,Todd,Todd,JDA,5,Male,2,Middle,2,United States,1,Married,1  
78,Tricia,Tricia,JDA,5,Female,1,Middle,2,United States,1,Married,1  
79,Wayland,Wayland,JDA,5,Male,2,Middle,2,United States,1,Single,2  
80,DaveK,DaveK,Ego,7,Male,2,Older,3,United States,1,Married,1  
edgedef>node1 VARCHAR,node2 VARCHAR  
1,29  
1,37  
1,38  
2,7  
2,13  
...  
80,75  
80,76  
80,77  
80,78  
80,79
```

guess.wikispot.org/The_GUESS_.gdf_format

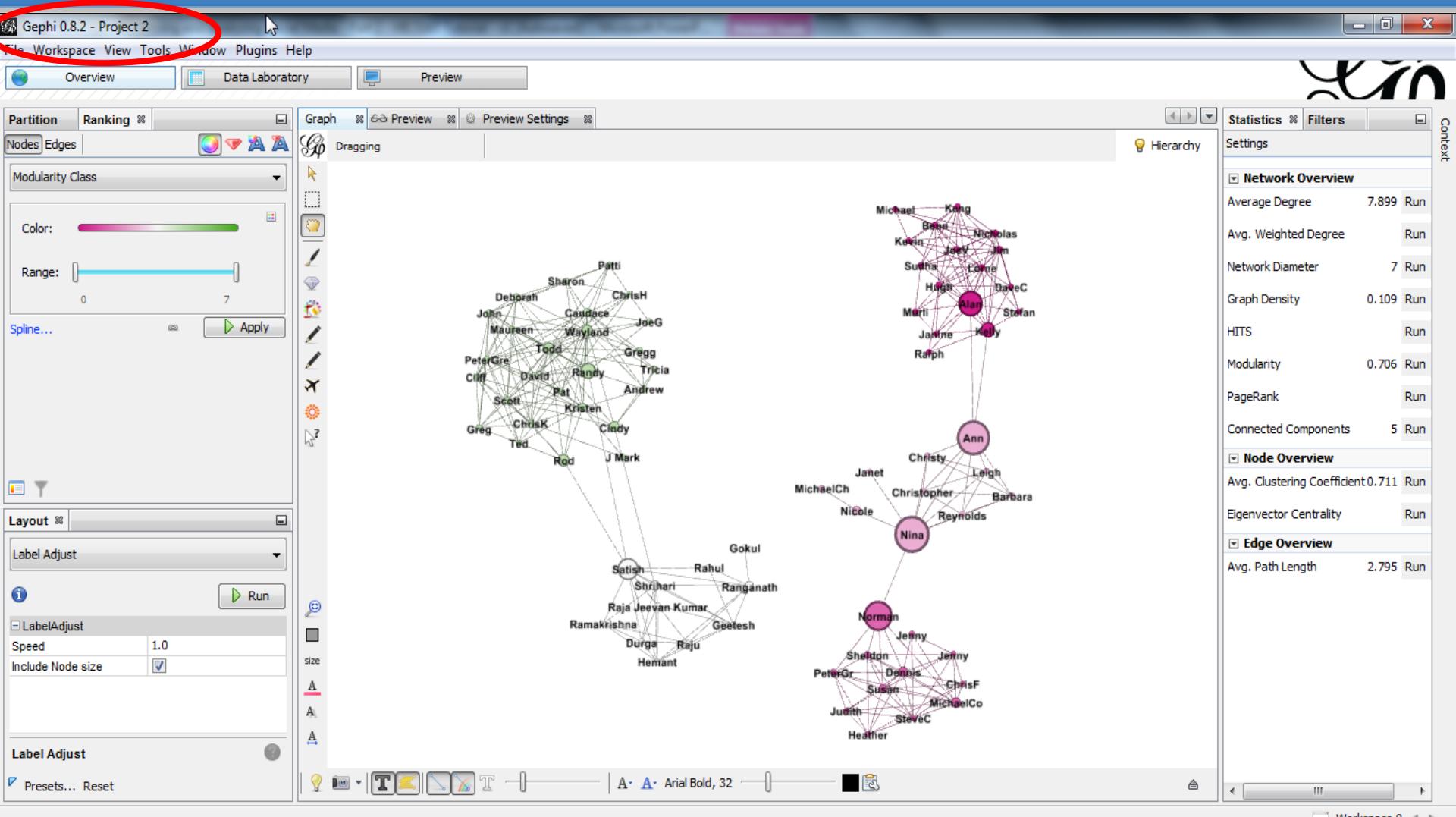
Social Network Analysis

Egocentric Networks – by Degree



Social Network Analysis

Egocentric Network – Facebook Example

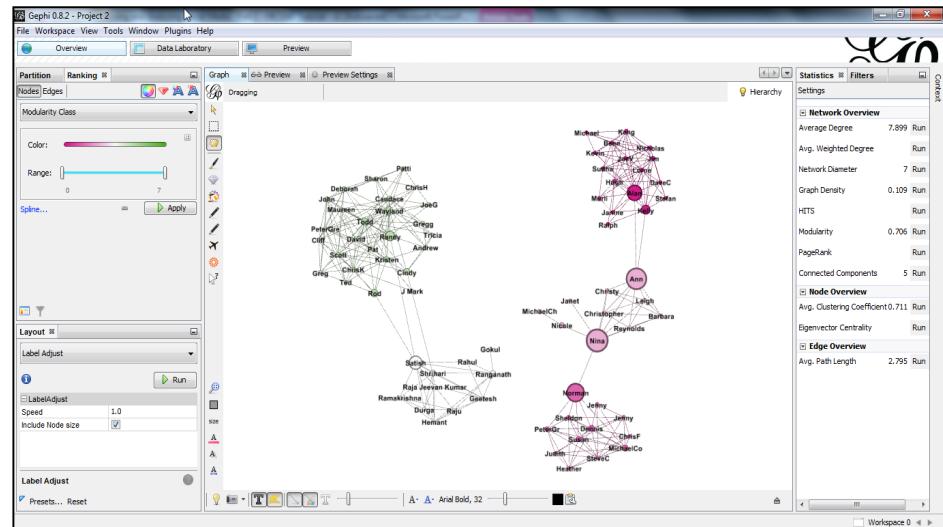


Social Network Analysis

Egocentric Network – Facebook Example

The image shows four overlapping windows from a Gephi tutorial:

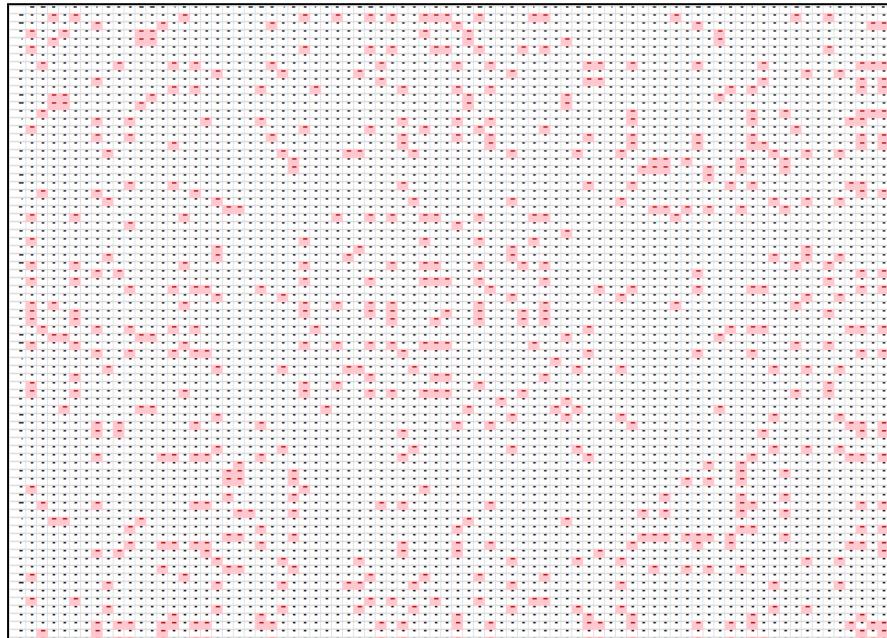
- Gephi Tutorial Visualization:** Shows a menu with "Introduction" selected.
- Gephi Tutorial Layouts:** Shows a menu with "Introduction" selected.
- Gephi Tutorial Quick Start:** Shows a menu with "Introduction" selected.
- Getting Started With The Gephi Network Visualisation App – My Facebook Network, Part I:** A main window with the URL blog.ouseful.info/2010/04/16/getting-started-with-gephi-network-visualisation-app-my-facebook-network-part-i/ at the bottom.



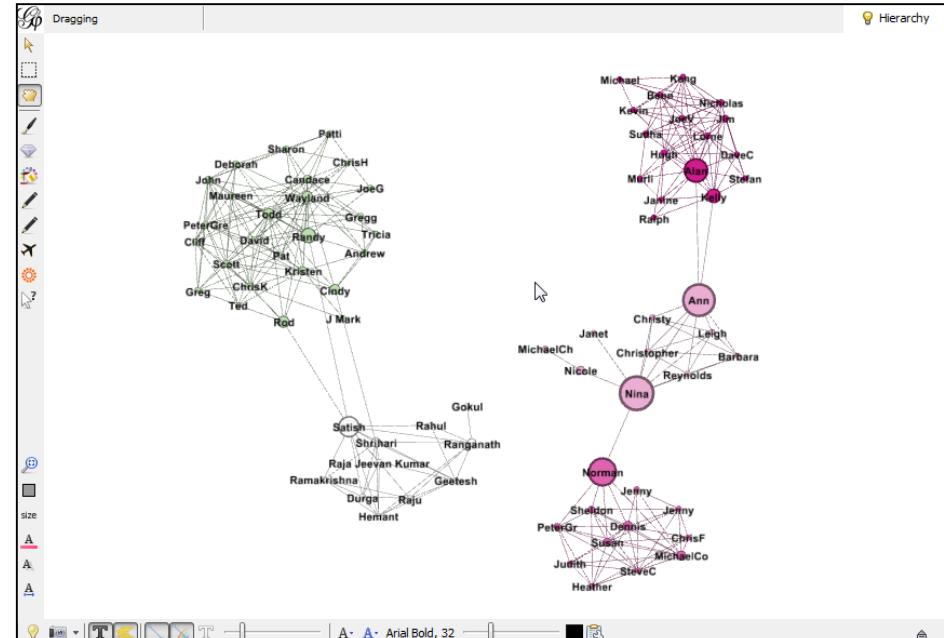
Social Network Analysis

Egocentric Network – Facebook Example

Matrix Representation



Graph Representation

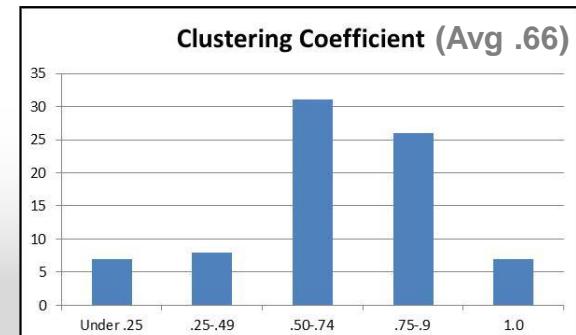
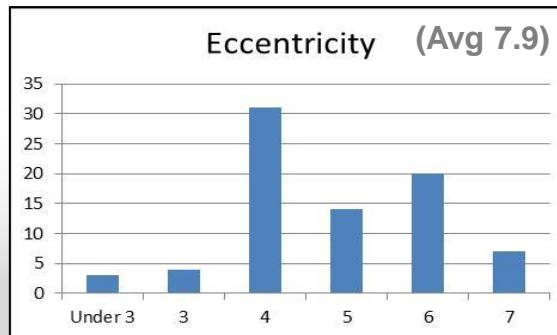
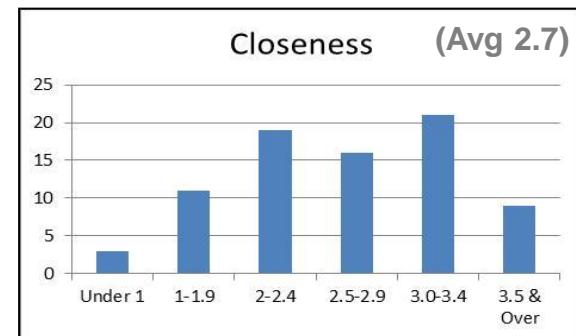
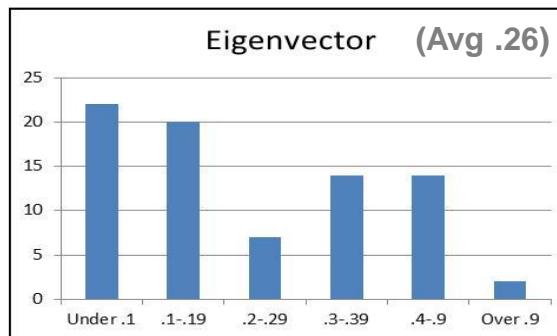
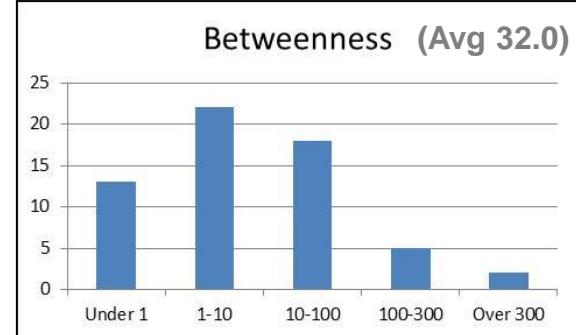
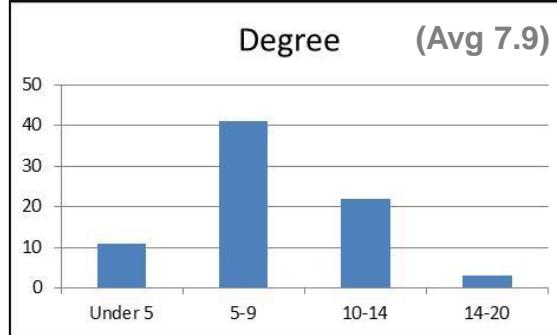


$$PL = N(N-1)/2 = 3081 \quad \text{Ego Density} = L/PL = 312/3081 = .10$$

Social Network Analysis

Egocentric Network – Facebook Example

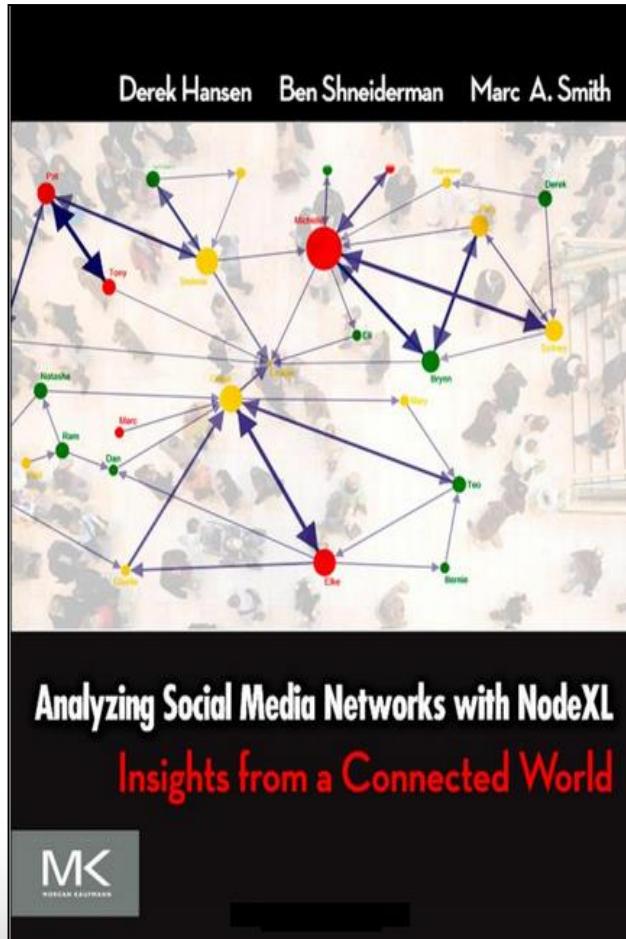
Age	Number	%
Under 30	17	22%
30-50	44	56%
Over 50	18	23%
Grand Total	79	100%
Sex	Number	%
Female	22	28%
Male	57	72%
Grand Total	79	100%
Group	Number	%
Academic	17	22%
Comshare	12	15%
Execucom	2	3%
Family	9	11%
JDA	28	35%
JDA-I	11	14%
Grand Total	79	100%
Country	Freq	%
United States	61	77%
India	12	15%
Other	6	8%
Grand Total	79	100%



Corr	Deg	Betw	Close	Eigen	Eccen	Clust
Deg	-	0.24	-0.12	0.80	0.14	0.01
Betw		-	-0.15	0.01	-0.13	-0.23
Close			-	-0.37	0.92	0.46
Eigen				-	-0.08	-0.09
Eccen					-	0.48
Clust						-

Social Network Analysis

Another Analytical Alternative - NodeXL



The screenshot shows the NodeXL software running within Microsoft Excel. The ribbon at the top includes tabs for File, Home, Insert, Page Layout, Formulas, Data, Review, View, Developer, Acrobat, NodeXL, and Design. The main area displays a network graph with nodes and edges. A tooltip for 'Vertex 1 Name' is visible, stating 'Enter the name of the edge's first vertex.' To the right, a 'Document Actions' sidebar provides information about NodeXL, including its purpose as a network graphs tool for the Social Media Research Foundation, links to discussion lists, and a 'Donate' button.

Social Network Analysis

Facebook Analysis

The Anatomy of the Facebook Social Graph

Johua Ugander^{1,2*}, Brian Karrer^{1,2*}, Lars Backstrom¹, Cameron Marlow^{1†}
1 Facebook, Palo Alto, CA, USA
2 Cornell University, Ithaca, NY, USA
3 University of Michigan, Ann Arbor, MI, USA
* These authors contributed equally to this work.
† Corresponding author: cameron@fb.com

Abstract

We study the structure of the social graph of active Facebook users, the largest social network ever analyzed. We compute numerous features of the graph including the number of users and friendships, the degree distribution, path lengths, clustering, and mixing patterns. Our results center around three main observations. First, we characterize the global structure of the graph, determining that the social network is nearly fully connected, with 99.91% of individuals belonging to a single large connected component, and we confirm the ‘six degrees of separation’ phenomenon on a global scale. Second, by studying the average local clustering coefficient and degeneracy of graph neighborhoods, we show that while the Facebook graph as a whole is clearly sparse, the graph neighborhoods of users contain surprisingly dense structure. Third, we characterize assortativity patterns present in the graph by studying the basic demographic and network properties of users. We observe clear degree assortativity and characterize the extent to which ‘your friends have more friends than you’. Furthermore, we observe a strong effect of age on friendship preferences as well as a globally modular community structure driven by nationality, but we do not find any strong gender homophily. We compare our results with those from smaller social networks and find mostly, but not entirely, agreement on common structural network characteristics.

Introduction

The emergence of online social networking services over the past decade has revolutionized how social scientists study the structure of human relationships [1]. As individuals bring their social relations online, the focal point of the internet is evolving from being a network of documents to being a network of people, and previously invisible social structures are being captured at tremendous scale and with unprecedented detail. In this work, we characterize the structure of the world’s largest online social network, Facebook, ~~in order to enhance understanding in the management of social networks~~.

In its simplest form, a social network contains individuals as vertices and edges as relationships between vertices [2]. This abstract view of human relationships, while certainly limited, has been very useful for characterizing social relationships, with structural measures of this network abstraction finding active application to the study of everything from bargaining power [3] to psychological health [4]. Moreover, social networks have been observed to display a broad range of unifying structural properties, including homophily, clustering, the small-world effect, heterogeneous distributions of friends, and community structure [5, 6].

Quantitative analysis of these relationships requires individuals to explicitly detail their social networks. Historically, studies of social networks were limited to hundreds of individuals as data on social relationships was collected through painstakingly difficult means. Online social networks allow us to increase the scale and accuracy of such studies dramatically because new social network data, mostly from online sources, map out our social relationships at a nearly global scale. Prior studies of online social networks include research on Twitter, Flickr, Yahoo! 360, Cyworld, Myspace, Orkut, and LiveJournal among others [7–11].

The trend within this line of research is to measure larger and larger representations of social networks,

The emergence of online social networking services over the past decade has revolutionized how social scientists study the structure of human ... previously invisible social structures are being captured at tremendous scale and with unprecedented detail.

Active *	Global	US
Members	721M	149M
Friends	68.7B	15.9B
Aver. Friends	190	214
Total Pop	6.9B	260M

Accessed within 28 days of May '11
At least one friend
Over 13 years of age

Social Network Analysis

The Anatomy of the Facebook Social Graph

Johan Ugander^{1,2*}, Brian Karrer^{1,3*}, Lars Backstrom¹, Cameron Marlow^{1†}

¹ Facebook, Palo Alto, CA, USA

² Cornell University, Ithaca, NY, USA

³ University of Michigan, Ann Arbor, MI, USA

* These authors contributed equally to this work.

† Corresponding author: cameron@fb.com

Abstract

We study the structure of the social graph of active Facebook users, the largest social network ever analyzed. We compute numerous features of the graph including the number of users and friendships, the degree distribution, path lengths, clustering, and mixing patterns. Our results center around three main observations. First, we characterize the global structure of the graph, determining that the social network is nearly fully connected, with 99.91% of individuals belonging to a single large connected component, and we confirm the ‘six degrees of separation’ phenomenon on a global scale. Second, by studying the average local clustering coefficient and degeneracy of graph neighborhoods, we show that while the Facebook graph as a whole is clearly sparse, the graph neighborhoods of users contain surprisingly dense structure. Third, we characterize the assortativity patterns present in the graph by studying the basic demographic and network properties of users. We observe clear degree assortativity and characterize the extent to which ‘your friends have more friends than you’. Furthermore, we observe a strong effect of age on friendship preferences as well as a globally modular community structure driven by nationality, but we do not find any strong gender homophily. We compare our results with those from smaller social networks and find mostly, but not entirely, agreement on common structural network characteristics.

Introduction

The emergence of online social networking services over the past decade has revolutionized how social scientists study the structure of human relationships [1]. As individuals bring their social relations online, the focal point of the interest is evolving from being a network of documents and people to being a network of people, and previously invisible social structures are being captured at tremendous scales with unprecedented detail. In this work, we characterize the structure of the world’s largest online social network, Facebook, in an effort to advance the state of the art in the empirical study of social networks.

In its simplest form, a social network contains individuals as vertices and edges as relationships between vertices [2]. This abstract view of human relationships, while certainly limited, has been very useful for characterizing social relationships, with structural measures of this network abstraction finding active application to the study of everything from bargaining power [3] to psychological health [4]. Moreover, social networks have been observed to display a broad range of unifying structural properties, including homophily, clustering, the small-world effect, heterogeneous distributions of friends, and community structure [5, 6].

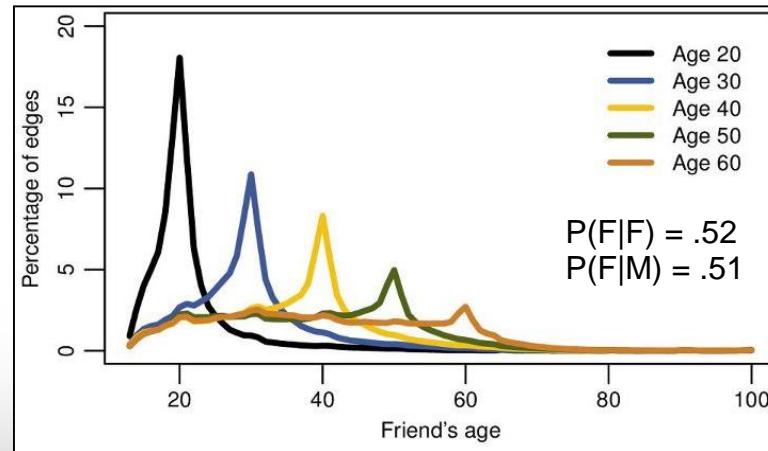
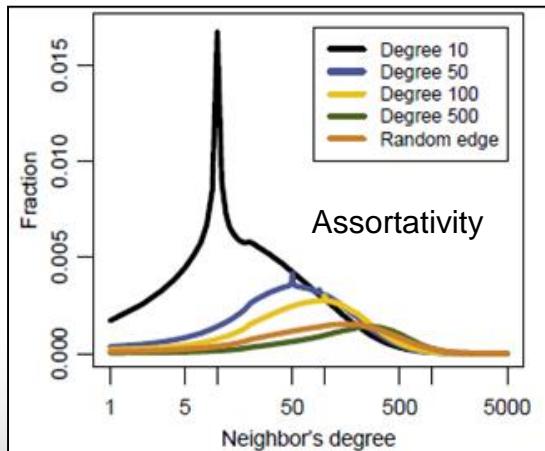
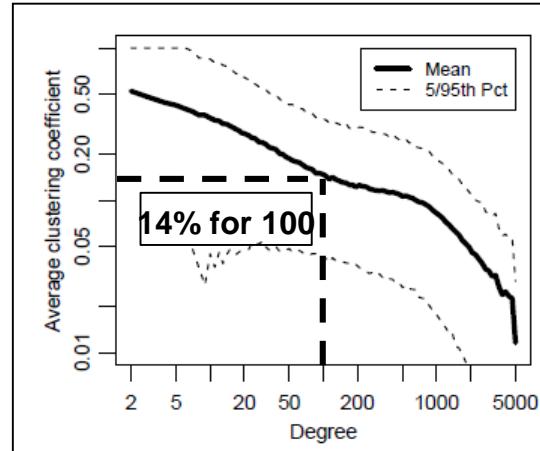
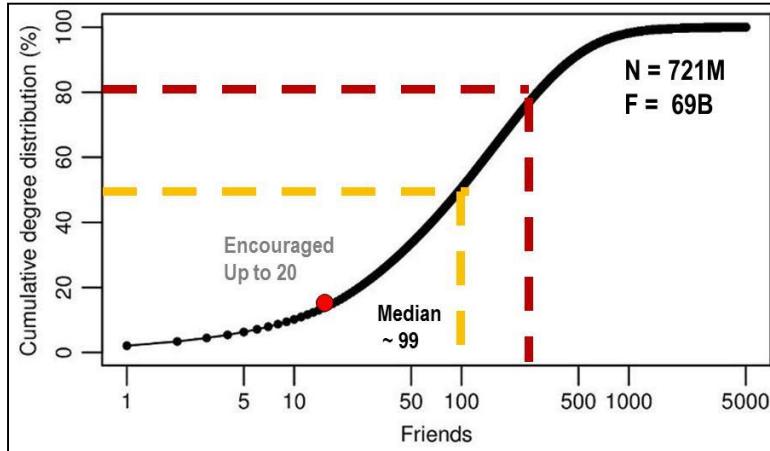
Quantitative analysis of these relationships requires individuals to explicitly detail their social networks. Historically, studies of social networks were limited to hundreds of individuals as data on social relationships was collected through painstakingly difficult means. Online social networks allow us to increase the scale and accuracy of such studies dramatically because now social network data, mostly from online sources, map out our social relationships at a nearly global scale. Prior studies of online social networks include research on Twitter, Flickr, Yahoo! 360, Cyworld, Myspace, Orkut, and LiveJournal among others [7–11].

The trend within this line of research is to measure larger and larger representations of social networks,

- Nearly fully connected, with 99.91% of individuals belonging to a single large connected component
- Confirm ‘six degrees of separation’ phenomenon on a global scale. Second, by studying
- While the Facebook graph as a whole is clearly sparse, the graph neighborhoods of users contain surprisingly dense
- Clear degree assortativity.
- Strong impact of age on friendship preferences as well as a globally modular community structure driven by nationality, but no gender homophily.

Social Network Analysis

Facebook Analysis



Social Network Analysis

Facebook Analysis

Four Degrees of Separation

Lars Backstrom*, Paolo Boldi† Marco Rosa† Johan Ugander* Sebastiano Vigna†

December 8, 2011

Abstract

Frigyes Karinthy, in his 1929 short story “Lánchezemek” (“Chains”) suggested that any two persons are distanced by at most six friendship links.¹ Stanley Milgram in his famous experiment [20, 23] challenged people to route postcards to a fixed recipient by passing them only through direct acquaintances. The average number of intermediaries on the path of the postcards lay between 4.4 and 5.7, depending on the sample of people chosen.

We report the results of the first world-scale social network graph-distance computations, using the entire Facebook network of active users (≈ 721 million users, ≈ 69 billion friendship links). The average distance we observe is 4.74, corresponding to 3.74 intermediaries or “degrees of separation”, showing that the world is even smaller than we expected, and prompting the title of this paper. More generally, we study the distance distribution of Facebook and of some interesting geographic subgraphs, looking also at their evolution over time.

The networks we are able to explore are almost two orders of magnitude larger than those analysed in the previous literature. We report detailed statistical metadata showing that our measurements (which rely on probabilistic algorithms) are very accurate.

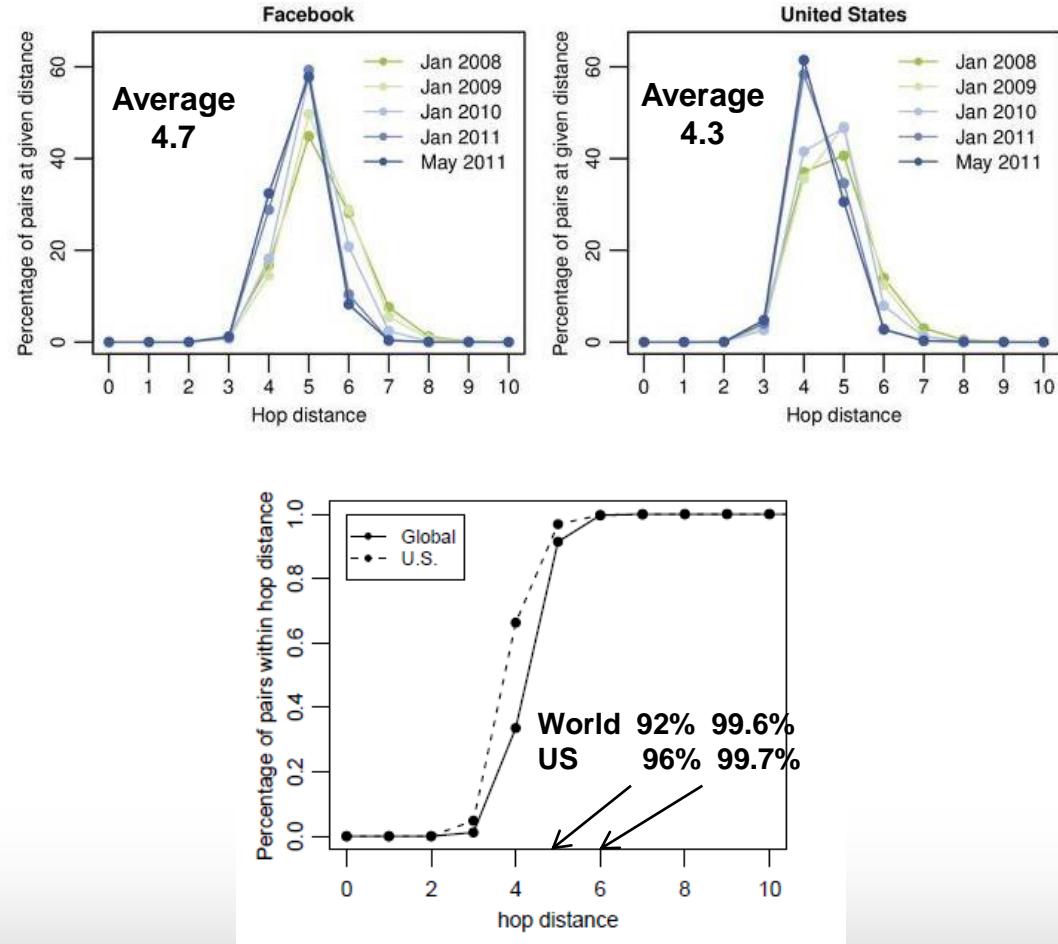
1 Introduction

At the 20th World Wide Web Conference, in Hyderabad, India, one of the authors (Sebastiano) presented a new tool for

*Facebook.
†IDSI, Università degli Studi di Milano, Italy. Paolo Boldi, Marco Rosa, and Sebastiano Vigna have been partially supported by a Yahoo! faculty grant and by MIUR PRIN “Query log e web crawling”.

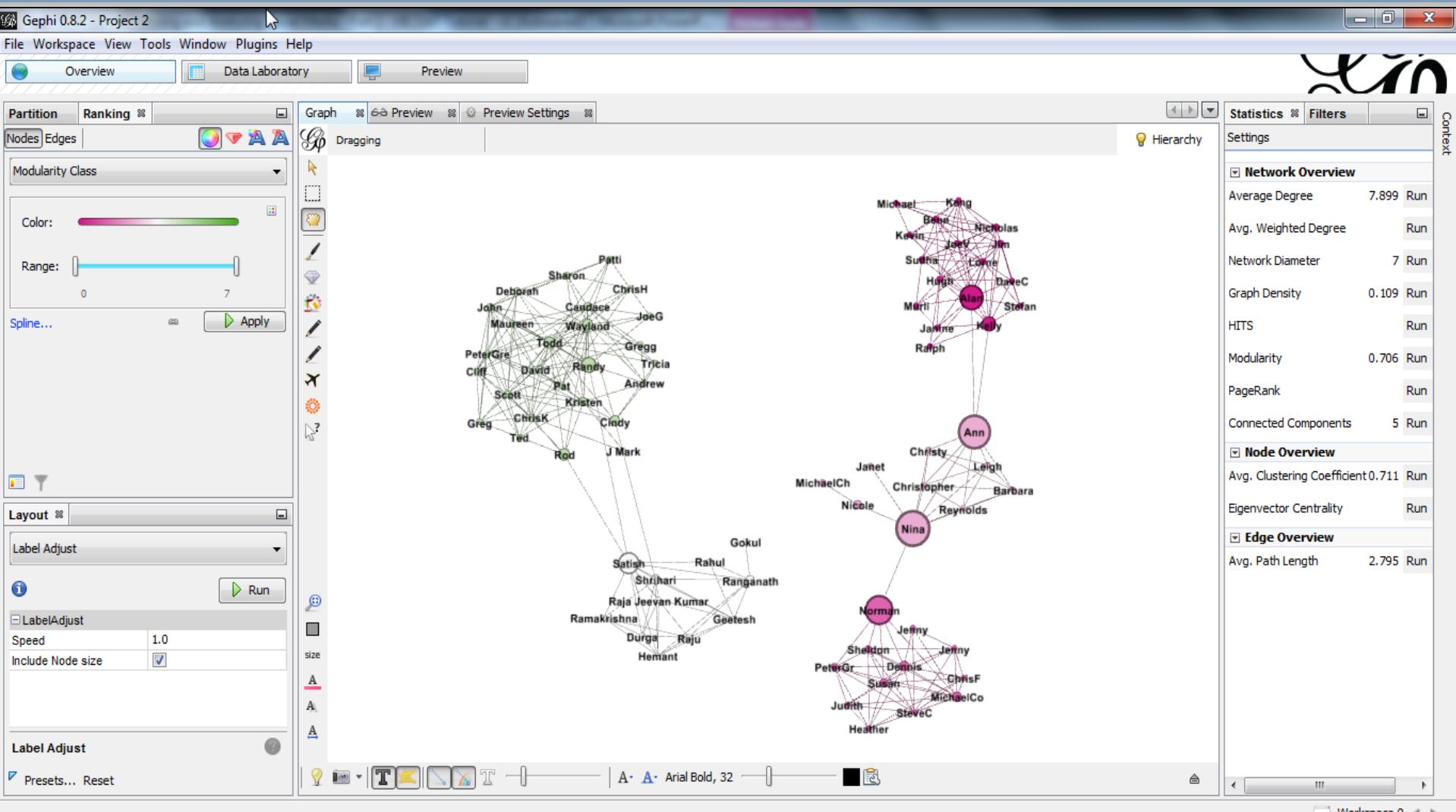
¹The exact wording of the story is slightly ambiguous: “He bet us that, using no more than five individuals, one of whom is a personal acquaintance, he could contact the selected individual [...]”. It is not completely clear whether the selected individual is part of the five, so this corresponds to the definition of “six degrees of separation” in the story, but the “six degrees of separation” phrase stuck after John Guan’s 1990 eponymous play. Following Milgram’s definition and Guan’s interpretation (see further on), we will assume that “degrees of separation” is the same as “distance minus one”, where “distance” is the usual path length (the number of arcs in the path).

²The spid (shortest-paths index of dispersion) is the variance-to-mean ratio of the distance distribution.



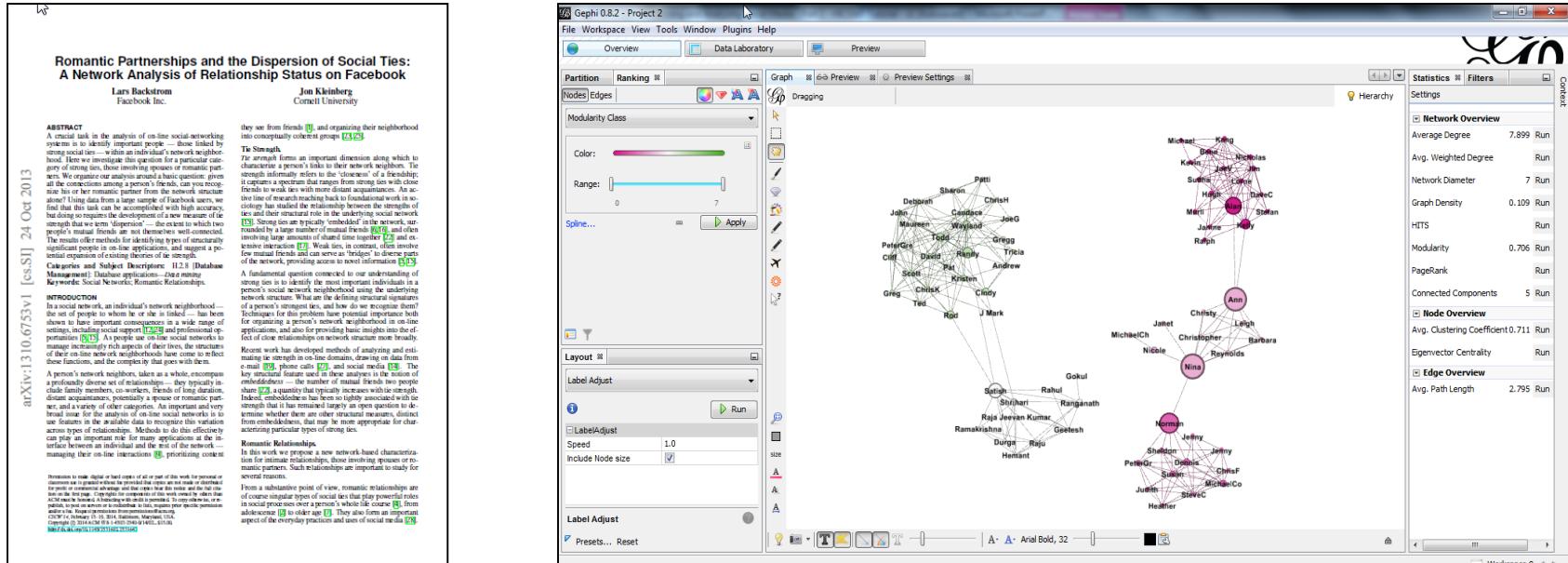
Social Network Analysis

Looking for Cohesive Subgroups



Sociology

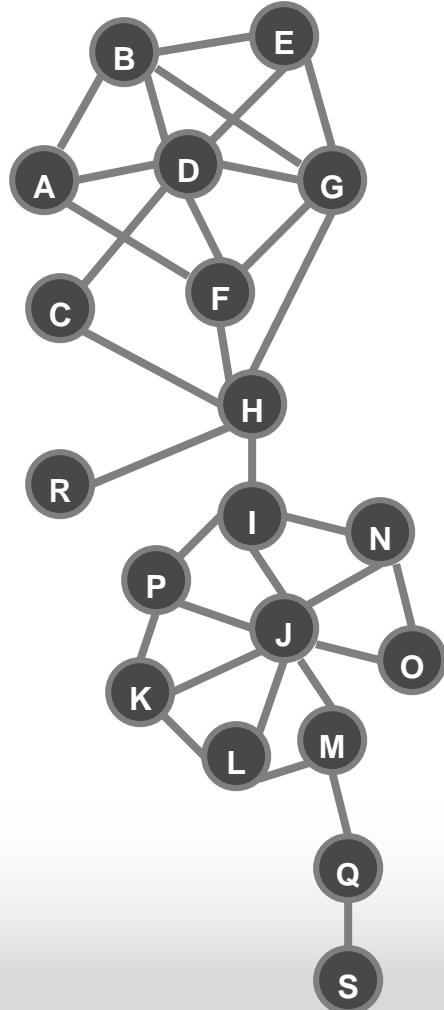
Cohesive Subgroups



Many individuals have large clusters of friends corresponding to well-defined foci of interaction in their lives, such as their cluster of co-workers or the cluster of people with whom they attended college. Since many people within these clusters know each other, the clusters contain links of very high embeddedness, even though they do not necessarily correspond to particularly strong ties.

Social Network Analysis

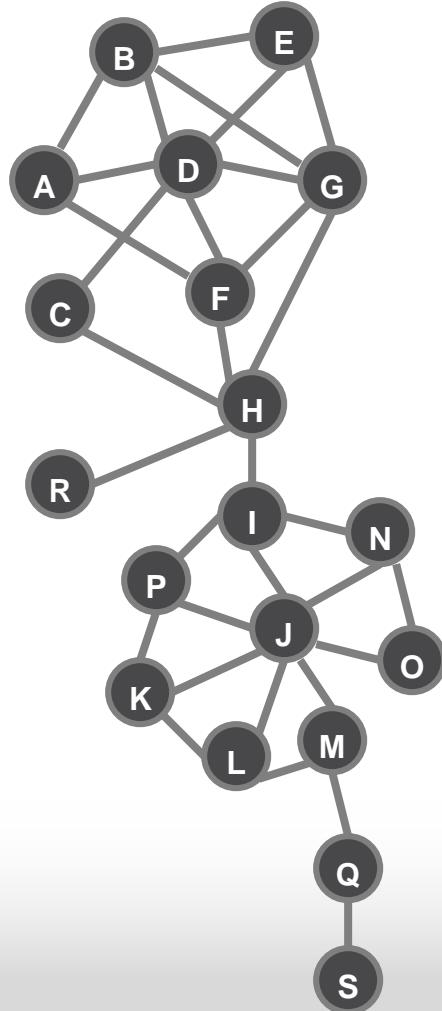
Cohesive Subgroups



- In a **complete** graph all the nodes are adjacent to one another
- A network is **connected** if every two nodes in the network are connected by some path in the network
 - A directed graph is **strongly connected** if and only if for every pair of vertices x, y , there is a directed path from x to y and a directed path from y to x between them. It is **weakly connected** if there is at least an undirected path for every pair of vertices.
- A graph G has **components** G_1 and G_2 if no (undirected) path exists between any node in G_1 to any node of G_2 (component is an isolated subgraph)
- A link is a **bridge** in a network if its removal increases the number of components in the network

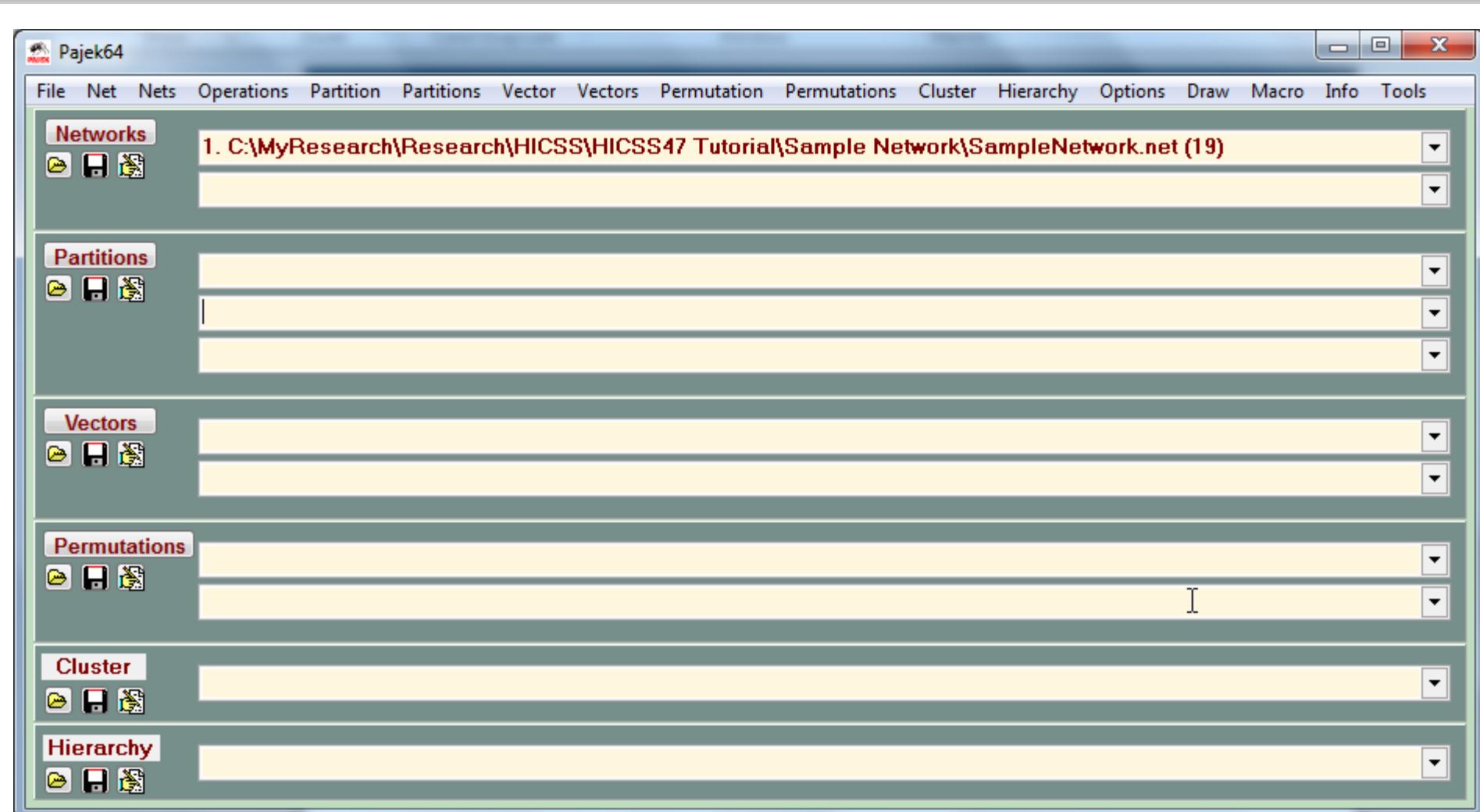
Social Network Analysis

Cohesive Subgroups



- A **clique** is a maximal complete subnetwork containing three vertices or more.
- **n-clique** is a maximal subnetwork in which the maximum path length for connection is “n” (e.g. 2-clique has nodes connected by paths of 1 or 2 lengths).
- A **k-core** is a maximal subnetwork in which each vertex has at least degree k within the subnetwork (e.g. 3-core)
- **Girvan-Newman Algorithm** breaks complex networks into “communities” by removing edges (betweenness) resulting in fragmentation

Social Network Analysis



Social Network Analysis

It's a Small World after all, or is it?

6.0 vs. 4.7