



Networking Network Scholars: Generating Best Practices for Archiving Network Data

Thomas M. Carsey
carsey@unc.edu

Director, Odum Institute for Research in Social Science
UNC-Chapel Hill

SAMSI CMSS Opening Workshop
August 19, 2013



The “Grand Challenge” Facing Social Network Analysis

- NSF’s “Rebuilding the Mosaic” challenges include:
 - Understanding social networks
 - Building data and technology infrastructure
 - Integration and synthesis across data, methods, and disciplines
 - <http://www.nsf.gov/pubs/2011/nsf11086/nsf11086.pdf>



The “Grand Challenge” Facing Social Network Analysis

- NSF’s “Rebuilding the Mosaic” challenges include:
 - Understanding social networks
 - Building data and technology infrastructure
 - Integration and synthesis across data, methods, and disciplines
 - <http://www.nsf.gov/pubs/2011/nsf11086/nsf11086.pdf>
- My goal is to help to develop best practices for collecting, storing, and sharing social network data.
 - Supported by the National Science Foundation (OCI-1255826)



The Primary Obstacles

- Network files are large and unwieldy.
 - Relational database rather than a flat file



The Primary Obstacles

- Network files are large and unwieldy.
 - Relational database rather than a flat file
- This makes many networks examples of “Big Data”



- Network files are large and unwieldy.
 - Relational database rather than a flat file
- This makes many networks examples of “Big Data”
- Column in *The Washington Post* by Samuel Arbesman titled “Five myths about big data” (8-16-2013) referenced the following tweet offered as a definition of “Big Data”

Big Data, n.: the belief that any sufficiently large pile of shit contains a pony with probability approaching 1 (by James Grimmelmann)



But What Kind of Pony?





The Primary Obstacles (cont.)

- Networks are dynamic.
- It is hard to know if networks are complete.
- Social networks present privacy concerns.
 - Sampling methods (e.g. RDS, web scraping) raise questions of selection, completeness, and privacy.
- Social network analysis is multidisciplinary.



The Primary Obstacles (cont.)

- Networks are dynamic.
- It is hard to know if networks are complete.
- Social networks present privacy concerns.
 - Sampling methods (e.g. RDS, web scraping) raise questions of selection, completeness, and privacy.
- Social network analysis is multidisciplinary.
- There is no common software solution.
 - Common tools include UCINet, R, Pajek, Gephi, Ora, AutoMap, SIENA, and many others
(<http://www.casos.cs.cmu.edu/tools/tools.php>)



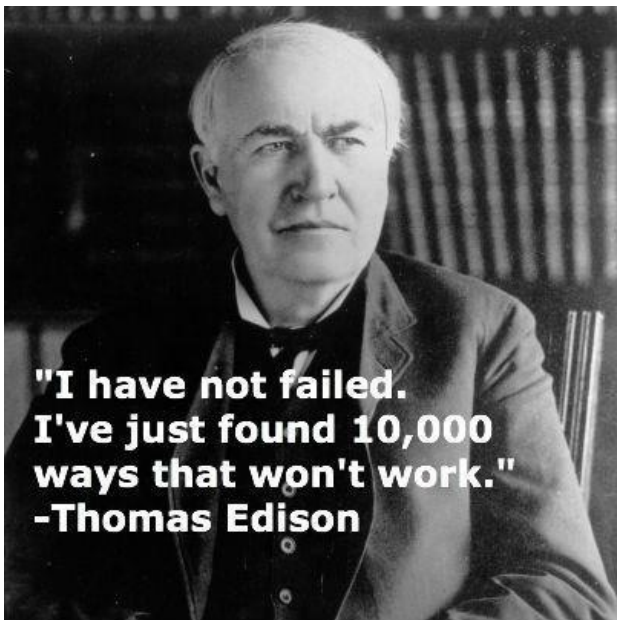
What Has Been the Impact?

- Common archives of social network data have not emerged.
- Researchers expend a lot of time and effort developing unique workflow adaptations to existing software environments.
- Faculty spend substantial class time and office hours on data management that could be spent on theory, methods, and/or applications.



What Has Been the Impact?

- Common archives of social network data have not emerged.
- Researchers expend a lot of time and effort developing unique workflow adaptations to existing software environments.
- Faculty spend substantial class time and office hours on data management that could be spent on theory, methods, and/or applications.
- Ironically, if you measure ties between network scholars based on how easy it is to share data, the network of network scholars is full of small clusters and isolates.



**"I have not failed.
I've just found 10,000
ways that won't work."
-Thomas Edison**



- Project Goals:
 - Create a shared understanding of the challenges.
 - Design an archive infrastructure prototype.
 - Maybe produce an R package something like “StatTransfer” for network files/objects
 - Begin education and outreach efforts.



- Project Goals:
 - Create a shared understanding of the challenges.
 - Design an archive infrastructure prototype.
 - Maybe produce an R package something like “StatTransfer” for network files/objects
 - Begin education and outreach efforts.
- Why am I leading this?
 - I have an interest in network analysis, but I am not an expert. This fits with a need to lower barriers to entry into the field.
 - I am Director of the Odum Institute at UNC which operates a large data archive, and I’m co-Director of our Social Network Analysis at Carolina (SNAC) program.



So let's start the conversation now!

Go ahead SAMSI — I'm listening.

(and e-mail me any time: carsey@unc.edu)