

Brief Report

Knowing Your Social Network Data and Measures and Big Data: A Summary of Jeffrey C. Johnson's Keynote XXXIV Sunbelt Social Network Conference

Tracy Van Holt

*East Carolina University
Greenville, North Carolina*

Author

Tracy Van Holt, is an assistant professor in the Department of Geography at East Carolina University in Greenville, North Carolina.

Please send all correspondence to vanholtt@ecu.edu.

Electronic data from the internet and cellphones have made an unprecedented amount of information about human behavior and social structure available. The analysis of Big Data (BD) is increasingly playing a role in social network research. But while BD is appealing, there are trade-offs. With what would now be considered small data (SD), it is easier to test for reliability and validity because there is more control in the research design and data collection stage. With BD, which are often mined, it is harder to know if these data are meaningful because there is little control over the questions asked of respondents and there is no on-the-ground understanding of the kind that comes from direct interviewing or from participant observation. Jeff's keynote summarized how ethnographic knowing, aids the development of theory, of measures, and of data collection in the study of social networks and how ethnographic knowing can be applied to BD.

While writing a blog post about Jordan White (a singer-songwriter who was a "rising star" with a big following on Twitter), Emily Guendelsberger, found that the BD she was analyzing was filled with false leads. Based on a hunch, Guendelsberger spent a day tracking Jordan's Twitter followers and found that many of the 200,000 followers that Jordan had amassed were an artifact of mass behavior—an illusion. BD can be filled with false leads and requires ethnographic knowing.

For his doctoral research, Jeff immersed himself

in an Alaskan Salmon Fishery as a fishing camp boat carpenter and did SD ethnographic research on scarcity and resource management. As a carpenter, Jeff saw how informal social roles affected group structure, dynamics, and outcomes. One person was willing to be the butt of jokes and pranks—a role labelled "court jester" (Johnson and Miller 1983). Academics at that time mostly had a negative view of social deviants like the court jester. Indeed, the court jester was the worst fisher in the camp and he had no kinship relations to the group. However, Jeff observed that the court jester fostered group cohesion when the fishers in the camp were stressed over a strike. He was rewarded for his role: this ostensibly lower status actor hung out with and interacted with higher status actors in the camp and he was rewarded for his informal role with over-the-limit fish transfers. Had Johnson not known his informants well, the court jester role would have appeared less influential in group dynamics.

Building on this work, at the Antarctic Research Station Jeff and colleagues James Boster and Lawrence Palinkas tested the hypothesis that networks that evolve into core-periphery structures—those with higher global coherence--will function better than groups that evolve into clique structures (Johnson et al. 2003a). Each year, for three years, they did ethnographic research on small groups—~30 people each—that lived in winter-over isolation at the station for eight months. The groups contained the researchers (the beakers) and the people

who ran the station (the trades) from different genders and social classes. Over the course of the Austral winter, crew were given a series of ratings and pile sorting tasks on social interactions to map the networks. A globally coherent group structure, that is, a core periphery structure with no subgroups were identified in year 1. In year 2, a core group of people initially emerged and some cliques developed over the year, though the structure still maintained a high global coherence. Distinct clique structures evolved in year 3 and this network had low global coherence (Johnson et al. 2003a). The research team created an index of core-periphery structure [the coefficient of relative variation (CRV)] by incorporating the first factors from a factor analysis of the correlation matrices that represented the interaction among station members (Johnson et al. 2003a). Low CRV scores reflected lower variation in the first factor and a higher core-periphery structure, or higher global coherence (Johnson et al. 2003a). When he included multi-year data from research groups from stations in Poland, Russia, and China where each station-year represented a unit of analysis, higher morale was highly correlated with higher global coherence (Johnson et al. 2003b).

The team also tested whether formal and informal leadership roles (instrumental and expressive leadership) and the role of social deviants explained the evolution of the social network structure (Johnson et al. 2003a). Positive deviants (like the court jester) foster cohesion in times of stress, while negative deviants (like people who challenge the leadership) can be disruptive. Roles were measured by a sentence completion task, with questions like: “_____ is a natural leader in getting things done around the station” and “_____ is one of the entertainers or comedians.” They examined the consensus on perceptions of the formal and informal leadership roles and deviants at the end of winter and compared that to global coherence. The networks in years 1 and 2, with more globally coherent structure showed a trend toward high agreement on and overlap of both formal and informal leadership roles compared with the network in year 3, with low coherence. Globally coherent structures (years 1 and 2) were also associated with positive deviants and lower coherence networks had fewer positive deviants and more negative deviants (year 3) (Johnson et al. 2003a).

Jeff’s keynote address raised the question: Can we gain similar understanding of human behavior in BD analysis, where it is difficult to know the informants or to get ethnographically informed social measures of interest (e.g. trust)? Much of BD work is being driven by data and are exploratory in nature rather than being driven by theory. BD analyses search for patterns or key structures

or nodes. But is there theoretical meaning to the metrics used? Are there theoretical and empirical meaningfulness of these patterns? How are these patterns associated with outcomes? Knowing in BD is the big challenge. If Jeff just had the networks from the South Pole station without the ethnography he could not have understood the social structure or developed a theory to account for the structure. The equivalent in BD is to find original sources and check for validity and reliability. Johnson, Van Holt, and their collaborators have been applying ethnographic knowing to BD. They compared the trade-offs in recall and accuracy between human-coded and machine-coded datasets (Van Holt et al. 2013) and they tested whether environmental factors were linked to conflict as theory predicts by analyzing eight years of articles from the Sudan Tribune; they show that ethnic groups with more news reports of severe conflict had more news links to livestock (Van Holt et al. 2012).

Johnson will be working on the sequel to this talk and the search for meaning in mined data in his new position as a Preeminent Professor of Anthropology at the University of Florida’s initiative on informatics and big data.

References

- Johnson J.C. and M.L. Miller. 1983. Deviant Social Positions in Small Groups: the Relation Between Role and Individual- *Social Networks* 5:51-69, 1983.
- Johnson, J.C., Boster, J.S. and L. A. Palinkas. 2003a. Social roles and the evolution of networks in extreme and isolate environments. *Journal of Mathematical Sociology* 27 (89-121).
- Johnson, J.C., Palinkas, L.A, and J.S. Boster. 2003b. Informal social roles and the evolution and stability of social networks. In “Dynamic Social Network Modeling and Analysis”. The National Academies Press: Washington, D.C.
- Van Holt, T., Johnson, J.C., Brinkley, J., Carley, K., and J. Diesner. 2013. Rapid Ethnographic Assessment for Cultural Mapping. *Poetics* 4(4): 366–383
- Van Holt, T., Johnson, J.C., Brinkley, J., Carley, K.M., and J. Caspersen. 2012. Structure of ethnic violence in Sudan: an automated content, meta-network, and geospatial analytical approach. *Computational and Mathematical Organization Theory* 18:340-355.