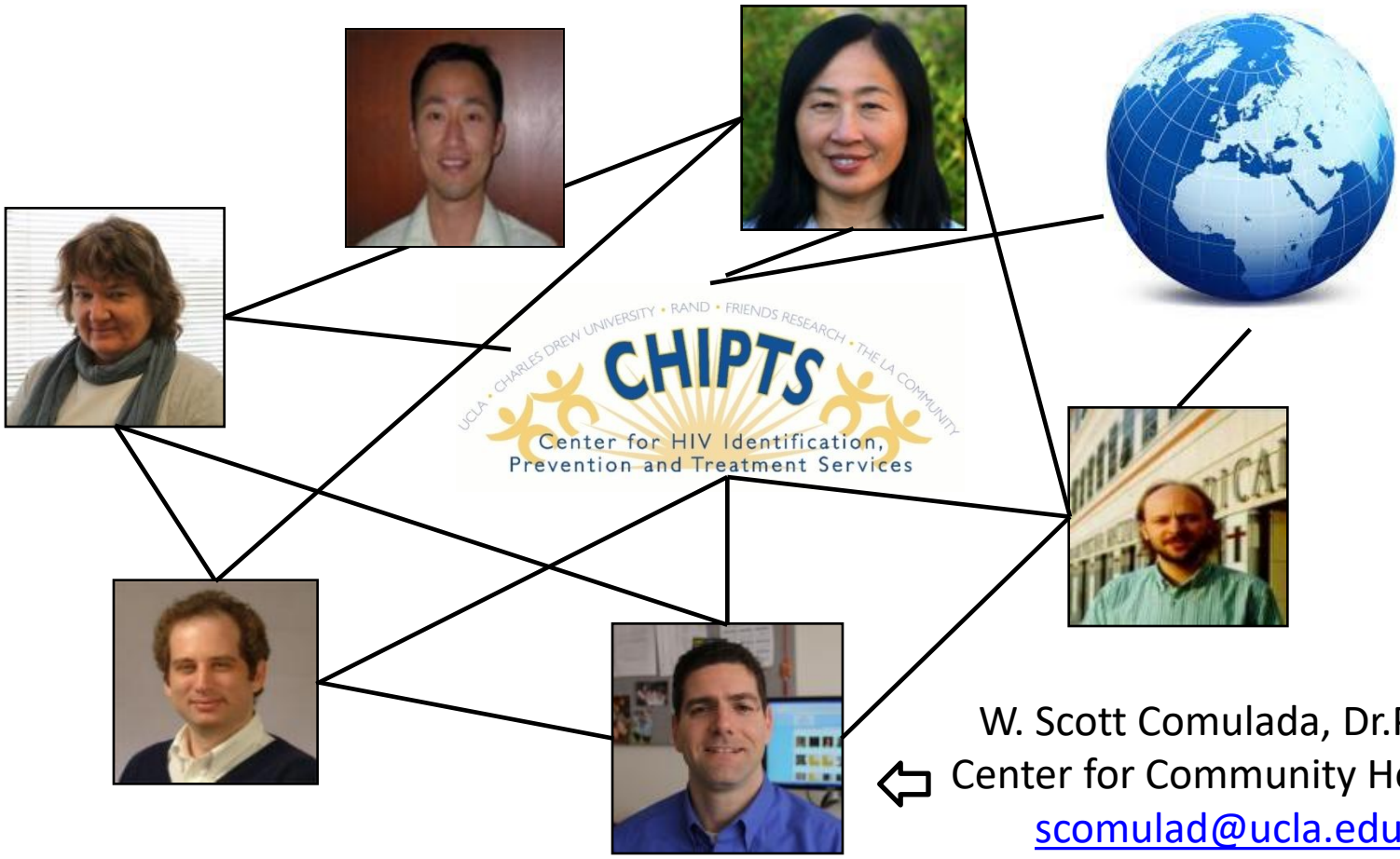
Social Network Analysis in HIV Research



W. Scott Comulada, Dr.P.H.
Center for Community Health
scomulad@ucla.edu

September 13, 2011

Background

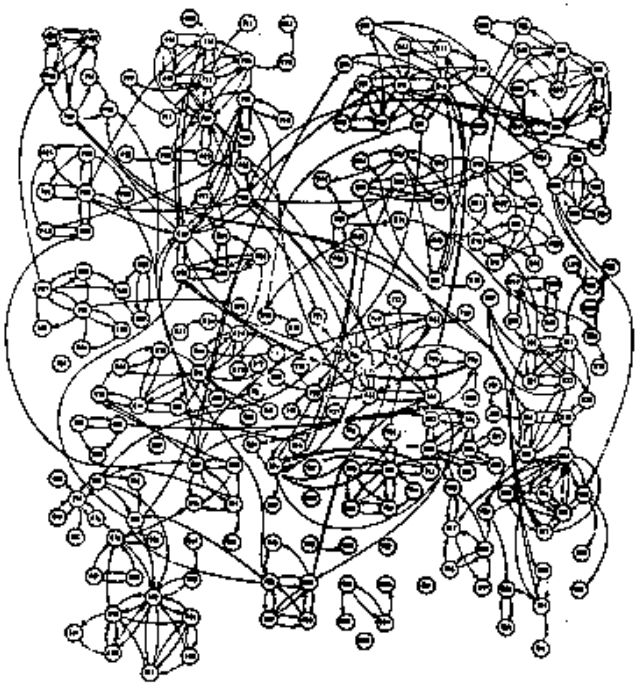
What do we mean by ‘social network analysis’?

Focus on …

relationships between individuals

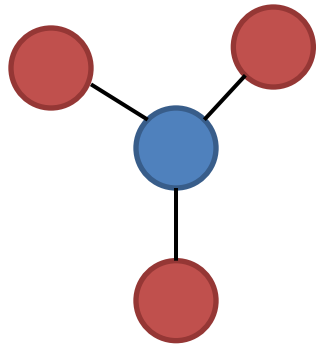
rather than characteristics of individual

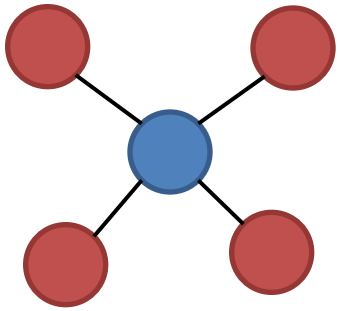
Background

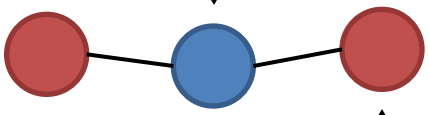


Original studies collected sociometric data
A few large networks in closed environments
e.g. schools, offices, etc.

Background







Ego
Study participant

Alter
Network member

Subsequent studies also collected egocentric data

Many smaller (often disconnected) networks

Applicable to HIV research on marginalized populations

Why is the social network framework important?

Strikes balance between

Ecological fallacy – inferences on individuals from aggregate-level data

And

Reductionist fallacy – inferences at aggregate level from individual-level data

Why is the social network framework important?

Example in condom use practices with sex partners

Ecological fallacy – infer practices from aggregate partner data

Reductionist fallacy – infer practices only from partner-level data

Why is the social network framework important?

Successful HIV interventions need to address structural factors (Coates et al., 2008)
* Includes the social environment *

Solution: Social network-based interventions
a.k.a. Positive-peer deviant models, popular opinion leader models

<u>HIV prevention examples</u>

Injection drug users (Heckathorn, 1999)
South African infants (Rotheram-Borus et al., in press)

Egocentric study designs

Basic egocentric studies (assessment only on egos)

- Same recruitment strategies as non-S.N. studies, e.g. convenience samples

Hybrid egocentric-sociometric studies (assessment on egos and some alters)

- Snowball sampling (Goodman, 1961) – friends recruit friends
- Respondent driven sampling (Heckathorn, 1997, 2002) weights snowball sample to account for non-random recruitment

Egocentric designs in HIV research

General challenges

- Often conducted with marginalized, hard-to-reach populations
  - E.g. Drug users
- We observe incomplete networks
- Hard to generalize results

Egocentric designs in HIV research

Longitudinal challenges in collecting repeated alter observations

- Networks not stable over time, e.g. unstable relationships with drug using network members
- Difficulty in linking alter data that is repeated
  - IRB issues in collecting enough information to indentify alters

Egocentric designs in HIV research

Two examples: Network-based interventions to reduce HIV-transmission behaviors conducted in drug-using neighborhoods in Baltimore Maryland

- SHIELD study (recruitment, 97-99; Latkin et al., 2003a)
  - Linkage through alter first names and demographics
- STEP study (recruitment, 04-06; Tobin et al., 2010)
  - Linkage by explicitly asking egos which alters were discussed during previous assessment

The promise of social media

The promise of social media

Assessment of “off” and online networks potentially capture different networks and measure different constructs

- Alters reported on paper and Facebook (Vernon, 2011)
- Sociometric networks reported on paper and through e-mail (Quintane & Kleinbaum, 2011)

The promise of social media

- Fits in with what has been long known: Observed and self-reported behavioral interactions do not relate very well (Bernard & Killworth, 1977)
- Pen paper: perceptions
  - Sometimes more important for behavior change
  - Perceived versus actual alter behavior more strongly related to ego behaviors (Valente, 2005)
- Electronic: actual contacts
  - More important for disease spread, diffusion of ideas

The promise of social media

Supplement traditional pen-and-paper name generating questions with online media

- e.g. SIM card reader to allow for alter identification from phone contact list (Schneider et al., 2011)

The promise of social media

- Allows for sociometric approach necessary to understand risk network structure

e.g. 1996 Swedish survey of sexual behavior (n = 2810)

- Important finding: Distribution for number of partners not normal / exponential, closer to scale-free distribution
- Explanation: Increased skill in acquiring new partners as number of previous partners grows, “rich get richer”
- Implications for disease propogation
- Survey was pen-and-paper! Think of possibilities that social media brings.

Reference: Liljeros et al., 2001

Social media tools

- Recruitment
  - Social networking websites, e.g. Facebook
- Data collection
  - Computer tablets, e.g. Apple iPad: “SNA Observer”, S.N. application (Hansberger, 2011)
  - Mobile phones

The promise of social media

Potential barrier:

IRB

Early adopting CHIPTS researchers are paving the way…

in recruitment (Facebook; PI: Sean Young) and

in assessment (mobile phones; Dallas Swendeman; Nithya Ramanathan)

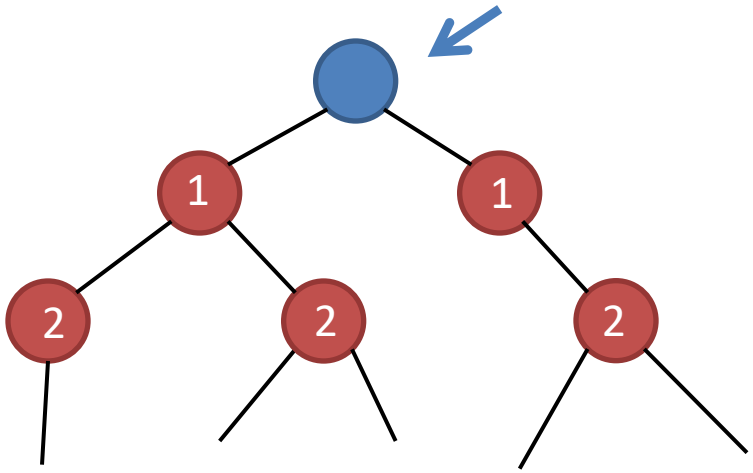
Online recruitment strategies

(See Gjoka et al. for overview)

Graph traversal techniques

- Breadth-First Search
  - Visit all nodes within given distance of starting node

e.g. within 2 degrees



Problem: Favors larger networks

Online recruitment strategies

Random walks ("Biased" walks since nodes are not resampled)

- Next node is chosen with equal probability from among adjacent nodes that have not been sampled

Problem: Favors larger networks

Assumption: Large population, bias is minimal (Tomas, 2011)

Online recruitment strategies

Metropolis-Hastings Random Walk

- Corrects for bias from favoring larger networks
- Example selection procedure (k = # of connections):

Step 1: Select candidate node (B) with equal probability from nodes adjacent to A

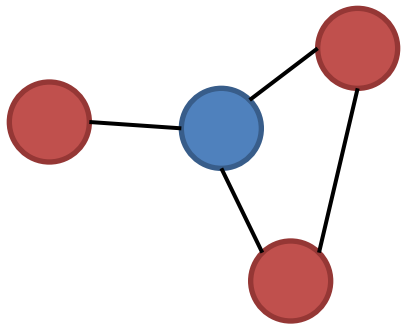
Step 2: Generate random uniform number U~ Unif[0,1]

Step 3: If $U < k_A / k_B$, jump to B; Otherwise, stay at A

Step 4: Repeat

Common egocentric variable types

- Characteristics of egos and alters
- Ego-alter and alter-alter ties
- Network structure, commonly
  - Network size = # of alters
  - Density = # of alter ties / # of possible alter ties
    - If n alters, are n (n-1) / 2 possible alter ties



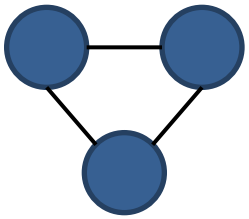
e.g., Density = 1 / 3

Three network structure variables related to risk taking in "risky" networks (Trotter II et al., 1995)

*e.g., Sociometric drug using networks *

- ↑ Density, ↑ risk

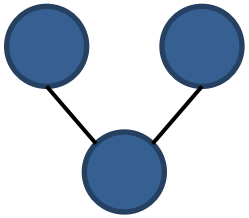
Transitive



- ↑ Transitivity, ↑ risk

  (Proportion of transitive triplets)

Broken



- ↑ Centrality (Bonacich eigenvector), ↑ risk

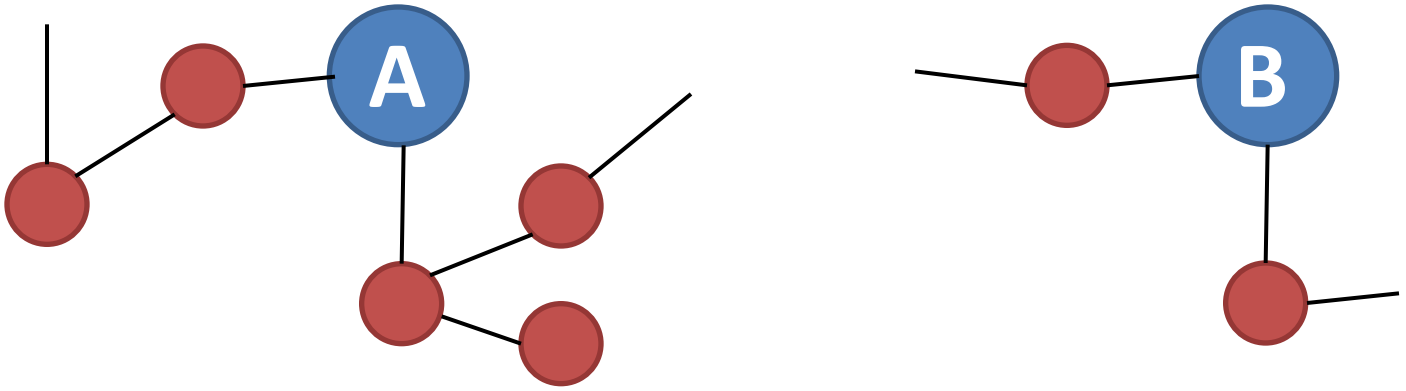
More on centrality

- Basic centrality measure: Degree centrality

  # of connections or "degrees" each network member has with other members

Problem: May not capture "influence"

e.g., A and B have same degree centrality



A may be more central but less influencial than B (members connected to B are more isolated)

More on centrality

Alternative measure to capture centrality and influence: Bonacich eigenvector centrality

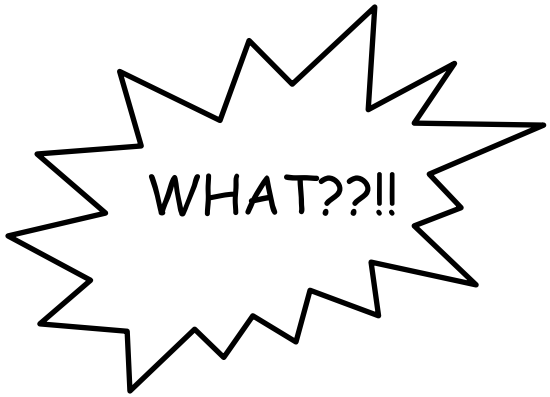
Calculation incorporates weights for connectedness of neighbors

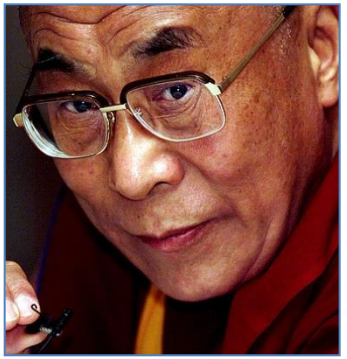
Bottom line: Centrality / influence are not always easy to measure or interpret

For example……

Is Bieber more influential than…

Obama, the Dalai Lama, or





your mamma?

At least online, according to *Klout* (for Obama and the Dalai Lama). Having a lot of connections, i.e. twitter followers, only gets you so far in the rating…

Social network analysis

- Historically, more focused on sociometric data
- Cross-sectional, Exponential random graph models (e.g., Wasserman et al., 1996)
- Longitudinal, Stochastic actor-based models (e.g., Snijders, 1996), “Rsiena” downloadable from http://cran.r-project.org/ for R

Common egocentric analyses

*Same tools used for non-S.N. analysis *

- Basic statistical tests, e.g. chi-square, t-test
  - Relationship between # of drug users / sex partners / kin in network and trading sex (Latkin et al., 2003b)
- Regression
  - Impact of network intervention (Latkin et al., 2003a)
  - Relationship between ego-alter ties and other characteristics (dyadic analyses)
    - e.g., Relationship between syringe sharing and order of alter nomination (Valente and Vlahov, 2001)

Dyadic analyses

- Issue: Correlations between ego-alter ties within networks
- Cross-sectional: Hierarchical models (Valente, 2010, chapter 4)
  - Include random effect for ties within each egocentric network
  - Analyze through random effect regression in favorite software package, e.g. binary ties through logistic regression in PROC GLIMMIX in SAS

Longitudinal dyadic analyses

Very recent!

(e.g., Lubbers et al., 2010)

Bivariate longitudinal models for egocentric networks

Example from Bohnert et al., 2009

- Bivariate outcomes: Ego drug use and composite measure for network-level use
- Tested competing theories:
  1. Drug users choose drug using networks (social selection; homophily; Lazarsfeld & Merton, 1954)
  2. Drug users influence drug use (social influence)
- Structural equation models in MPLUS

Bivariate longitudinal models for egocentric networks

Can also examine bivariate outcomes at alter level

- Same data, ego-alter ties on trust and drug use with alters

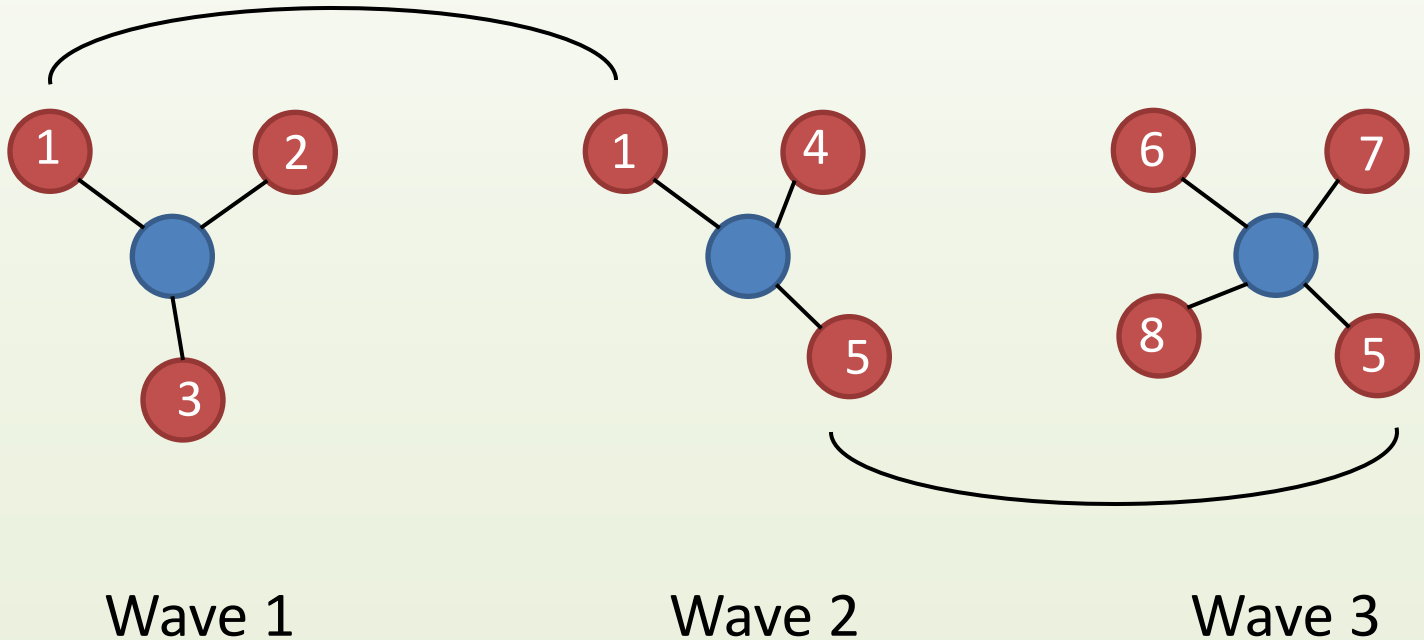
Problem: Additional random effects

Solution: Bayesian approach in WinBUGS

General issues with hierarchical egocentric analyses

Sparse nesting

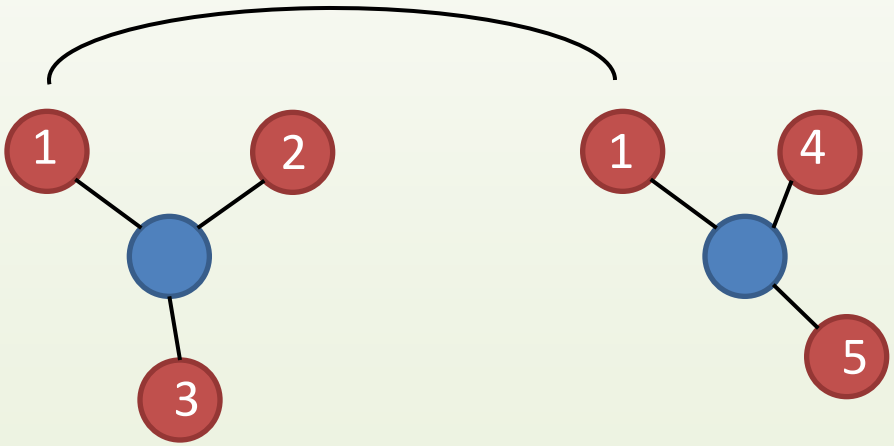
* Most alters reported on at one time point *



General issues with hierarchical egocentric analyses

Sparse nesting

Network members shared across networks



Network 1

Network 2

General issues with hierarchical egocentric analyses

Network-level random effects give more weight to larger networks

- Assumption: measurements are representative of population, regardless of network size
- Possibly reduces bias if reporting on smaller networks more likely to be incomplete
- Conversely, what if reporting on larger networks more biased? Egos less likely to know or become fatigued

References

Bernard, H.R., Killworth, P.D., 1977. Informant accuracy in social network data ii. Human Communication Research, 4, 3-18.

Bohnert, A.S.B., Bradshaw, C.P., Latkin, C.A., 2009. A social network perspective on heroin and cocaine use among adults: Evidence of bidirectional influences. Addiction 104, 1210-1218.

Coates, T.J., Richter, L., Caceres, C., 2008. Behavioral strategies to reduce HIV transmission: how to make them work better. Lancet, 372, 669 – 684.

Goodman, L.A., 1961. Snowball sampling. Annals of Mathematical Statistics, 32**,** 148–170.

Gjoka, M., Kurant, M., Butts, C.T., Markopoulou, A. Walking in Facebook: A Case Study of Unbiased Sampling of OSNs. Downloaded August 24, 2011: http://mkurant.com/publications/papers/Walking_in_Facebook_Infocom_10.pdf.

Hansberger, J.T., 2011. The SNA Observer: A tool for collecting real-time longitudinal sna data. Poster session, INSNA conference, Florida, February 2011.

Heckathorn, D.D., 1997. Respondent driven sampling: A new approach to the study of hidden populations. Social Problems 44, 174-199.

Heckathorn, D.D., Broadhead, R.S., & Anthony, D.L., 1999. AIDS and social networks: HIV prevention through network mobilization. Sociological Focus 32, 159-179.

Heckathorn, D.D., 2002. Respondent-Driven Sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. Social Problems 49, 11-34.

Latkin, C.A., Sherman, S., Knowlton, A., 2003a. HIV prevention among drug users: outcome of a network-oriented peer outreach intervention. Health Psychology 22, 332-339.

Latkin, C.A., Hua, W., Forman, V.L., 2003b. The relationship between social network characteristics and exchanging sex for drugs or money among drug users in Baltimore, MD, USA. International Journal of STD & AIDS 14, 770-775.

Lazarsfeld, P., Merton, R.K., 1954. Friendship as a Social Process: A Substantive and Methodological Analysis. In: Berger, M., Abel, T., Page, C.H., editors. Freedom and Control in Modern Society. Van Nostrand: New York, pp. 18–66.

Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., Aberg, Y., 2001. The web of human sexual contacts. Nature, 411, 907-908.

References (continued)

Lubbers, M.J., Molina, J.L., Lerner, J., Brandes, U., Avila, J., McCarty, C., 2010. Longitudinal analysis of personal networks. The case of Argentinean migrants in Spain. Social Networks 32, 91-104.

Quintane, E., Kleinbaum, A.M., 2011. Matter over mind? E-mail data and the measurement of social networks. Connections, 31, 20-43.

Rotheram-Borus, M.J., LeRoux, I.M., Tomlinson, M., Mbewub, N., Comulada, W.S., Le Roux, K., Stewart, J., O'Connor, M.J., Hartley, M., Desmond, K., Greco, E., Worthman, C.M., Idemundia, F. Philani Plus (+): a randomized controlled trial of mentor mother home visiting program to improve infants' outcomes. Prevention Science, in press.

Schneider, J.A., Kapur, A., Oruganti, G., Schumm, P., Laumann, E.O., 2011. A novel hybrid egocentric-archival network characterization approach using cell phones to identify bridging actors in a high risk HIV/sti network in India: The Secunderabadi Men's Study (sms). Oral presentation, INSNA conference, Florida, February 2011.

Snijders, T.A.B., 1996. Stochastic actor-oriented dynamic network analysis. Journal of Mathematical Sociology 21, 149–172.

Tobin, K.E., Kuramoto, S.J., Davey-Rothwell, M.A., Latkin, C.A., 2010. The STEP into Action study: a peer-based, personal risk network-focused HIV prevention intervention with injection drug users in Baltimore, Maryland. Addiction 106, 366-375.

Tomas, A. Model[l]ing the effect of differential recruitment on the bias of estimators for respondent-driven sampling. Dowloaded August 31, 2011. http://www.stats.ox.ac.uk/~tomas/html_links/DiffRecruitBandB.pdf.

Trotter II, R.T., Baldwin, J.A., Bowen, A.M., 1995. Network structurre and proxy network measures of HIV, drug and incarceration risks for active drug users. Connections 18, 88-103.

Valente, T.W., 2005. Network models and methods for studying the diffusion of innovations. In: Carrington, P.J., Scott, J., Wasserman, S. (Eds.), Models and Methods in Social Network Analysis. Cambridge University Press, New York, NY., p. 106.

Valente, T.W., 2010. Social Networks and Health: Models, Methods, and Applications. New York: Oxford University Press.

Valente, T.W., Vlahov, D., 2001. Selective risk taking among needle exchange participants in Baltimore: Implications for supplemental interventions. American Journal of Public Health 91, 406-411.

Vernon, M.C., Danon, L., House, T.A., Read, J.M., Keeling, M.J., 2011. Online and offline ego-centered network data collection. Oral presentation, INSNA conference, Florida, February 2011.

Wasserman, S., Pattison, P., 1996. Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and p*. Psychometrica 61, 401-425.