Reply to adams: Multi-dimensional edge inference

We completely agree with adams that, in social network analysis, the particular research question should drive the definition of what constitutes a tie (1). However, we believe that even studies of inherently social phenomena, such as the spread of influence (2) or supposed "social contagions" (3), can benefit strongly from a focus on objective behavioral data (4). For instance, the conventional wisdom is that social influence only travels along self-perceived ties. However, in truth, it remains unknown how much is being hidden from us by recency and saliency cognitive filters (5), and significant social influence may, in fact, travel across unperceived ties. Behavioral data are not prone to such filters and thus, when used properly, may shed considerable light on such important questions.

We also agree that an appropriate combination of self-reported edges and behavioral data, as suggested by adams, has the potential to address previously unanswered questions about the underlying dynamics of a group of interacting individuals. However, we urge caution in combining these very different kinds of data: a network with multiple types of edges, such as self-reported ones with those inferred by a factor analysis of behavioral data, can obscure important nuances that should be leveraged through parallel analyses rather than flattened into a single monolithic network. Indeed, there are times when it may be more appropriate to use "raw" behavioral data instead of the nonparametric output of a factor analysis. For instance, Fig. 1 shows how the self-reported friendship network compares with networks corresponding to communication, proximity on Saturday night, and travel, showing some marked differences. Understanding how behavioral data may influence, conflict with, or derive from the construction of the self-reported network is an interesting future line of inquiry.

It should, however, be possible to develop significantly better edge inference techniques by combining information at the local level of the dyad (including edge and node attributes) with information about ties elsewhere in the network. For instance, it was recently shown that the nested or hierarchical structure of entire social networks can predict missing dyadic links with high accuracy (6). Mathematically, this general approach has the form:

$$P\left(xy_{friendship}|xy_{\mathrm{behavior}}, \Delta T_{E_{/xy}}\right)$$

where $P(xy_{friendship})$ is the probability of a friendship between x and y, $xy_{behavior}$ is the local behavioral variables associated with the dyad, and $\Delta T_{E_{/xy}}$ represents how the global topological summary statistics associated with the network would change with the addition/removal of the edge between x and y.

To conclude, we point out that purely behavioral data, of course, are not a panacea, and their proper interpretation can be difficult without appropriate cognitive models. Instead, the use of behavioral data in social network analysis provides a highly complementary and very powerful tool for understanding social phenomena of all kinds (4). It also provides an objective way to identify, wrestle with, and mitigate the cognitive biases humans express, which often muddy the waters for scientific understanding of human phenomena. We strongly believe that the appropriate collection and analysis of such behavioral data have crucial roles to play in social network analysis.

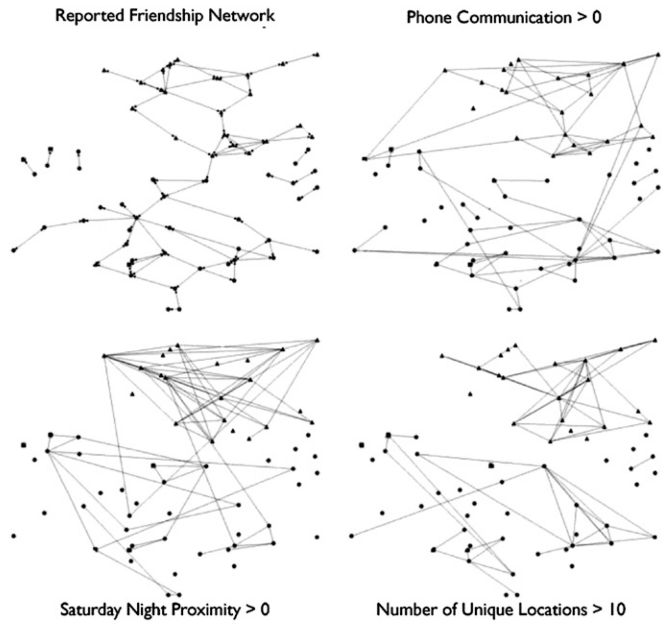

Fig. 1. Networks representing reported friendship, phone communication, proximity on Saturday night, and travel (a dyad is connected if proximate while being associated with more than 30 unique cellular towers). Nodes reflect the two groups of colleagues—the first year business school students and the Media Laboratory students working together in the same building on campus.

Nathan Eagle[a,b,1]***, Aaron Clauset***[a]***, Alex (Sandy) Pentland***[b]***, and David Lazer***[c,d]

[a]*Santa Fe Institute, Santa Fe, NM 87501;* [b]*Massachusetts Institute of Technology Media Laboratory, Cambridge, MA 02139;* [c]*Northeastern University, Boston, MA 02115; and* [d]*Harvard University, Cambridge, MA 02138*

1. adams j (2010) Distant friends, close strangers? Inferring friendships from behavior. *Proc Natl Acad Sci USA* 107:E29–E30.
2. Watts DJ, Dodds PS (2007) Influentuals, networks, and public opinion formation. *J Consum Res* 34:441–458.
3. Cohen-Cole E, Fletcher JM (2008) Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. *BMJ* 337:a2533.
4. Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. *Proc Natl Acad Sci USA* 106:15274–15278.
5. Freeman L, Romney A, Freeman S (1987) Cognitive structure and informant accuracy. *Am Anthropol* 89:310–325.
6. Clauset A, Moore C, Newman MEJ (2008) Hierarchical structure and the prediction of missing links in networks. *Nature* 453:98–101.

Author contributions: N.E., A.S.P., and D.L. designed research; N.E. performed research; N.E. analyzed data; and N.E. and A.C. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: nathan@mit.edu.