Subject: Re: he Spread of Evidence-Poor Medicine via Flawed Social-Network Analysis
From: "T.A.B.SNIJDERS" <[log in to unmask]>
Reply-To: [log in to unmask]
Date: Fri, 17 Jun 2011 12:52:12 +0200
Content-Type: text/plain
Parts/Attachments: text/plain (187 lines)

***** To join INSNA, visit http://www.insna.org *****

Dear all,

I would like to add to this thread (although the title is not charming) some of my thoughts about the issue of influence in networks and how to investigate it, carrying on after Sinan's contribution. I'm sorry that it has grown into an overly long discussion piece.
Summary: to study social influence we need not only experiments but also observational studies, and the possibilities are not as bleak as suggested by Lyons.

What struck me most in the paper by Lyons with the non-charming title are the following two points. The argument for social influence proposed by Christakis and Fowler (C&F) that earlier I used to find most impressive, i.e., the greater effect of incoming than of outgoing ties, was countered: the difference is not significant and there are other interpretations of such a difference, if it exists; and the model used for analysis is itself not coherent. This implies that C&F's claims of having found evidence for social influence on several outcome variables, which they already had toned down to some extent after earlier criticism, have to be still further attenuated. However, they do deserve a lot of credit for having put this topic on the agenda in an imaginative and innovative way. Science advances through trial and error

and through discussion. Bravo for the imagination and braveness of Nick Christakis and James Fowler.

How people influence each other is a central issue in social network analysis, as Sinan Aral writes in his contribution to this thread. Our everyday experience is that social influence is a strong and basic aspect of our social life. Economists have found it necessary to find proof of this through experimental means, arguing (Manski) that other proofs are impossible. Sociologists tend to take its existence for granted and are inclined to study the "how" rather than the "whether". The arguments for the confoundedness of influence and homophilous selection of social influence (Shalizi & Thomas Section 2.1) seem irrefutable. Studying social influence experimentally, so that homophily can be ruled out by design, therefore is very important and Sinan Aral has listed in his message a couple of great contributions made by him and others in this domain. _However, I believe that we should not restrict ourselves here to experiments._ Humans (but I do not wish to exclude animals or corporate actors) are purposive, wish to influence and to be influenced, and much of what we do is related to achieve positions in networks that enable us to influence and to be influenced in ways that seem desirable to us. Selecting our ties to others, changing our behaviour, and attempting to have an influence on what others do, all are inseparable parts of our daily life, and also of our attempts to be who we wish to be. This cannot be studied by experimental assignment of ties or of exchanges alone: such a restriction would amount to throwing away the child (purposeful selection of ties) with the bathwater (strict requirements of causal inference).

The logical consequence of this is that we are stuck with imperfect methods. Lyons argues as though only perfect methods are acceptable, and while applauding such lofty ideals I still believe that we should accept imperfection, in life as in science. Progress is made by discussion and improvement of imperfections, not by their eradication.

A weakness and limitation of the methods used by C&F for analysing social influence in the Framingham data was that, to say it briefly, these were methods and not generative models. Their methods had the aim to be sensitive to outcomes that would be unlikely if there were no influence at all (a sensitivity refuted by Lyons), but they did not propose credible models expressing the operation of influence and that could be used, e.g., to simulate influence processes. The telltale sign that their methods did not use generative models is that in their models for analysis the egos are independent, after conditioning on current and lagged covariates; whereas the definition of social influence is that individuals are not independent.

Together with colleagues I have developed models for the simultaneous operation of social influence and tie selection (homophilous or otherwise). The best reference currently is "Dynamic networks and behavior: separating selection from influence" by Christian Steglich, Tom Snijders, and Michael Pearson in /Sociological Methodology/, 40 (2010), 329-392; the methods are implemented in the Siena software (there is an extensive website www.stats.ox.ac.uk/siena/

<http://www.stats.ox.ac.uk/siena/>). These models and the methods based on them indeed are not perfect, but I think they help to get a better understanding of influence and selection processes, and we are working on their weaknesses. They assume the availability of data on networks and individual behaviour or other outcomes observed in a panel design, provided that the network is not too big (a couple of hundred actors, currently being extended to a couple of thousand).

In this research we have been making claims of the kind that we aim to "disentangle influence and selection", and given the results by Shalizi & Thomas about the confoundedness of these two, there is the question what this means and whether this aim is reasonable at all. A brief summary of my position is the following. We can never exclude the possibility that what seems to be social influence with respect to a variable Z is the consequence of earlier homophilous choice on an unobserved variable Z' that later on leads to changes in the variable Z. This is a simple formulation of some of the more general and mathematical results obtained by Shalizi and Thomas (section 2.1). "Disentangling" selection and influence is possible only under the assumption that the available observed networks and individual variables contain all the variables that play a role in the causal process, and if moreover a number of distributional assumptions are made (cf. the remark made by Shalizi and Thomas where they refer to Steglich et al., unfortunately to a preprint and not to the recently published version). The sensitivity to the distributional assumptions is a serious question, and this is a topic that should and will be investigated. The assumption that all relevant variables are observed is always questionable, but statistical inference very often is done under such assumptions. We must strive after observational designs where this is, to the best of our knowledge, a reasonable approximation; and we can make progress on this front by what we always do as social scientists: try to find out better what drives these processes, come closer to determining the type of network ties and the individual variables that "really" matter and how they affect one another. As the great statistician R.A. Fisher said when asked how to make observational studies more likely to yield causal answers (cited by Cox and Wermuth, 2004): "Make your theories elaborate". Instead of "true" causality, we can obtain results about time ordering: are individuals similar first, and then become tied (~ homophily) or are they tied first, and become similar later (~ influence)? Such results, for richer and more and more relevant variables, can give important scientific advances about selection and influence, based on observational studies combining rich data collection, insightful theorizing, and good modelling.

Lyons in his discussion section criticizes statistical modelling, and here I find his formulations a facile attack on statistical modelling of observational studies. This section does not do justice to the difficulties of the topic and the possibilities to make reasonable advances. He writes "Yet viewing observational data through the lens of statistical modelling produces new biases, generally unknown and mostly unacknowledged, lurking in mathematical thickets. …. Observational studies often lead to publications whose causal conclusions contradict one another or are contradicted by experiments … this is a natural

consequence of poor methodology." These are words of a knight riding in shining armour high above the fray, not of somebody who honours the muddy boots of the practical researcher. Lyons' discussion section ignores that observational studies are inevitable for many scientific aims, difficult indeed, but possible as I have tried to argue above. It also ignores that a lot of methodologically careful observational studies have been done, as well as that collectively we learn from our mistakes as long as we keep our eyes open and are not intimidated by authority. Most concretely, this discussion ignores that some assumptions are more important for practical applicability than others. For example, in linear regression, the assumption of a continuous distribution is practically totally irrelevant but theoretically extremely convenient; the assumption of normal distributions is unimportant; the assumption of constant residual variances is important; and the assumption of independent residuals is extremely important. Such distinctions are argued by robustness studies: we are worried by deviations from assumptions only if they invalidate expressions of uncertainty such as standard errors or posterior standard deviations, type I and type II error rates, etc. Methods can make invalid assumptions and still give good answers.

/Reference additional to those mentioned earlier in the thread:
/D.R. Cox and N. Wermuth. Causality: a statistical view. /International Statistical Review /*72 (2004)*, 285-305.

Cheers,

Tom

--

==========

Tom A.B. Snijders

Professor of Statistics in the Social Sciences

University of Oxford

Professor of Statistics and Methodology

Department of Sociology

University of Groningen

http://www.stats.ox.ac.uk/~snijders/
<**http://www.stats.ox.ac.uk/%7Esnijders/**>

==========

Back to: **Top of Message** | **Previous Page** | **Main SOCNET Page** **Permalink**