

Symmetry Is Beautiful

Uri Shwed^a and Peter S. Bearman^bAmerican Sociological Review
77(6) 1064–1069
© American Sociological
Association 2012
DOI: 10.1177/0003122412463018
<http://asr.sagepub.com>

The Temporal Structure of Scientific Consensus Formation (Shwed and Bearman 2010) developed a procedure—which did not require expert judgment—to evaluate the level of contestation in scientific literatures. Examining different cases of consensus and contestation, we showed that science may progress in a spiral pattern that quickly generates new questions in new domains from recent answers (e.g., climate change research), stagnate around old questions in a cyclical pattern (e.g., smoking research in the 1950s), or entrench in a flat pattern responding to irrelevant external critiques (e.g., research on autism and vaccines). Bruggeman, Traag, and Uitermark (hereafter BTU) argue that without distinguishing between positive and negative citations, our procedure is of little value. They arrive at this conclusion by misunderstanding the role of network communities in our procedure. They then propose as an improvement a self-defeating solution that relies on distinguishing between positive and negative ties, requires expert evaluation, and destroys the possibility of an unbiased evaluation. Avoiding that trap was the point of our article.

While much of their comment considers instances of contestation in politics that have nothing to do with science, the relevant findings they offer with respect to our analyses support our original findings. Given this, it is not surprising that BTU's own calculations show that their concern with the valence of citations has no bearing on the evaluation of contestation levels. By accepting the comment for publication, *ASR*'s editors prolong a period of flat argumentation in the sociology of science. In this reply we reiterate the original article's contribution and show how

BTU's findings strengthen ours. We conclude with remarks on the importance of symmetry, a guiding principle of our analysis that BTU completely abandoned.

If it seems that the original article, the comment, and the reply are talking past each other, it is because they are. In the original article we made arguments about scientific contestation, modularity as a global property of graphs, symmetry, comparability, and empirical results. In the comment, BTU discuss the allocation of nodes to communities, non-scientific discourse, and doing what the 457 previous papers using modularity on topics unrelated to ours did. Although of great interest to many people, these are different foci than those we considered.

Our article exploited the modularity community detection algorithm not for community detection, but as a general property of networks. This enabled us to link black-boxing (an idea) with modularity (a measure) to provide a useful, symmetrical tool for sociologists of science to evaluate the state of contestation around politically laden issues. Specifically, we used maximal directed modularity as a way to quantitatively operationalize the imagery of black-boxing (Latour) or bottled ships (Collins). We showed that the measure fits the contestation and consensus trends in disparate science domains described

^aBen Gurion University of the Negev^bColumbia University**Corresponding Author:**Uri Shwed, Department of Sociology and
Anthropology, Ben Gurion University of the
Negev, PO Box 653, Beer-Sheva 84105 Israel
E-mail: shwed@bgu.ac.il

qualitatively (for validation) by field-specific experts. We developed our procedure by observing a simple homology between what Latour and Collins suggest changes in the process of consensus making, and what modularity quantifies—observed differences in the salience of internal parts. Validation of our procedure, however, is purely empirical: we did not claim or assume that modularity indexes contestation; we tested that idea and found that it did.

The main benefit of our measure is that it allows comparative research into science without field-specific expertise. A main requirement for a useful measure is what Latour defined as generalized symmetry—rejection of any pre-categorization or pre-demarkation of the field of inquiry. We achieved symmetry by offering a tool that requires no parameter specification or analyst's judgment. To attain symmetry, the analyst must not impose interpretations on the data. Oddly, this is precisely what BTU criticize us for. We have no interest in defending claims that we never made and are wrong. However, in the next section we focus on the analyses provided by the authors, and outline why their critique is irrelevant to the ideas in our article and actually strengthens our results.

SYMMETRY AND NEGATIVE CITATIONS

BTU's main critique is that we ignore the content of ties and assume or interpret all ties as positive. BTU propose a modularity algorithm for signed networks; that is, their algorithm does not simply maximize modularity, it does so only on ties that are signed positive, while minimizing modularity on ties that are signed negative.

This could be a welcomed addition, for instances where the signing process itself may be taken for granted. Science, of course, is no such instance. This is why BTU fail to provide signed networks that are relevant for the discussion of scientific consensus. In their own words, "To distinguish negative from positive references, we would have to acquaint ourselves

with the vernacular of cancer researchers and read thousands of articles, which is beyond our capabilities here" (Bruggeman, Traag, and Uittermark 2012:1053). Had it been within BTU's capabilities, all they would have provided us with is yet another expert-based, asymmetrical interpretation of the literature based on content analysis. The idea of our article is that one ought to refuse to do so—regardless of capabilities—because doing so endangers the symmetry and objectivity that our analysis strives for.

But what are negative citations in science? Consider, for example, this exchange. Judging from BTU's tone, their critique is clearly a negative citation of our article, focusing on what they see as flaws. And yet, they expose our analysis to more readers, prolong its tenure in *ASR*'s pages, and broaden its linkage to the scientific communities interested in community detection algorithms, while substantively strengthening our results. The border between negative and positive citations is fuzzy; negative citations imply recognition and contribute to the progress of science as much as positive citations (Cole and Cole 1981:24–27). To argue that the symmetry constraint (which prohibits us from signing ties) is flawed, BTU need to demonstrate that when asking the same questions asymmetrically they reach significantly different conclusions regarding scientific consensus. This is not something their comment is oriented toward.

In the original article, we neither argued nor assumed that citations are positive. We noted that two different theories of citations (by Merton and Latour) imply that most—although not all—citations are positive. From this we proposed that changes in contestation levels (which also likely mean changes in the ratio of negative to positive citations) are observable through scaled modularity. Our original article tested this and found it is the case. Scaled modularity distinguishes between periods of contestation and consensus in part because the ratio of negative citations to positive citations is different in such periods and the structural properties of negative citations are different from positive citations.

BTU's own results demonstrate this. However, acknowledging the difference between positive and negative citations at the micro level of article construction does not mean the macro measures we employ must feed on such data. Signing citations and coding them as negative or positive requires domain specific expertise and human judgment—two issues we minimized in favor of a symmetrical analysis. That was the point of our article.

COMMUNITY ALLOCATION VERSUS SCALED MODULARITY

BTU charge us with assuming that negative citations have no implications for the communities detected, and they provide an elegant analysis to demonstrate that this assumption does not hold. We did not make this assumption. BTU are confused between the common use of community detection algorithms—for detecting communities—and our use of the same procedure for an evaluation of a network macro property. That was one of the innovations of the original article.

BTU show that when they introduce even a small ratio of negative citations into the equation they get different results, which is reasonable. But this finding is not really relevant for the argument we made. We made no claims about the actual community structure, which is the only thing they report. Huge dissimilarities in the allocation of specific nodes to specific communities can yield similar (macro) modularity scores. As a matter of fact, in many cases, they do (Good, Montjoye, and Clauset 2010). In addition, it is not surprising that arbitrarily signing some citations as positive and others as negative—while taking that arbitrary value into the calculation—generates different results.

BTU show (in their Figure 1) that as more citations are arbitrarily defined as negative, different draws yield increasingly different community allocations. Our Figure 1 simply overlays theirs with the corresponding line representing the trend in scaled modularity for

the cases and periods from our original article. BTU provide no explanations for the clear drop in the variance of different draws, apparent in 1991 to 1992 in the smoking case and in 1994 to 1995 in the skin cancer case. This trend corresponds closely to our findings;¹ in 1992, the final EPA report on hazards of passive smoking terminated that debate, and in 1994 our analysis of the skin cancer case crosses the .1 value of scaled modularity, which we carefully suggested as a possible threshold for consensus. Thus, it seems that (yes) arbitrarily defined negative citations generate different allocations of nodes to communities (which our procedure does not care about), but that when the literature is consensual, different draws are more coherent among themselves. In other words, when employing an algorithm that is sensitive to some random quality, consensual black-boxed networks are more robust than networks with salient internal divisions.

Their Figure 1 thus replicates our findings. In fact, BTU (2012:1054) summarize their analysis saying: “During periods of epistemic rivalry, when the percentage of negative ties is higher, the difference will usually be larger. . . . the actual pattern of negative ties is unknown to us and remains an empirical question.” Our article answered this empirical question for seven cases and a century of research, and showed that periods of epistemic rivalry correspond with scaled (unsigned) modularity.

SPECIALIZATION AND THE “SERIOUS FLAW” OF SCALING MODULARITY

BTU go to great length to argue that our scaling of the modularity property is “a serious technical flaw” (BTU:1051). They call it renormalization rather than scaling. They argue it is wrong to scale modularity, which is “by definition normalized” (BTU:1052). Modularity (as part of the process of community detection) is normalized for the number of ties (references); we scale it for the (logged) number of nodes (articles). Those are very different things. The normalization predefined

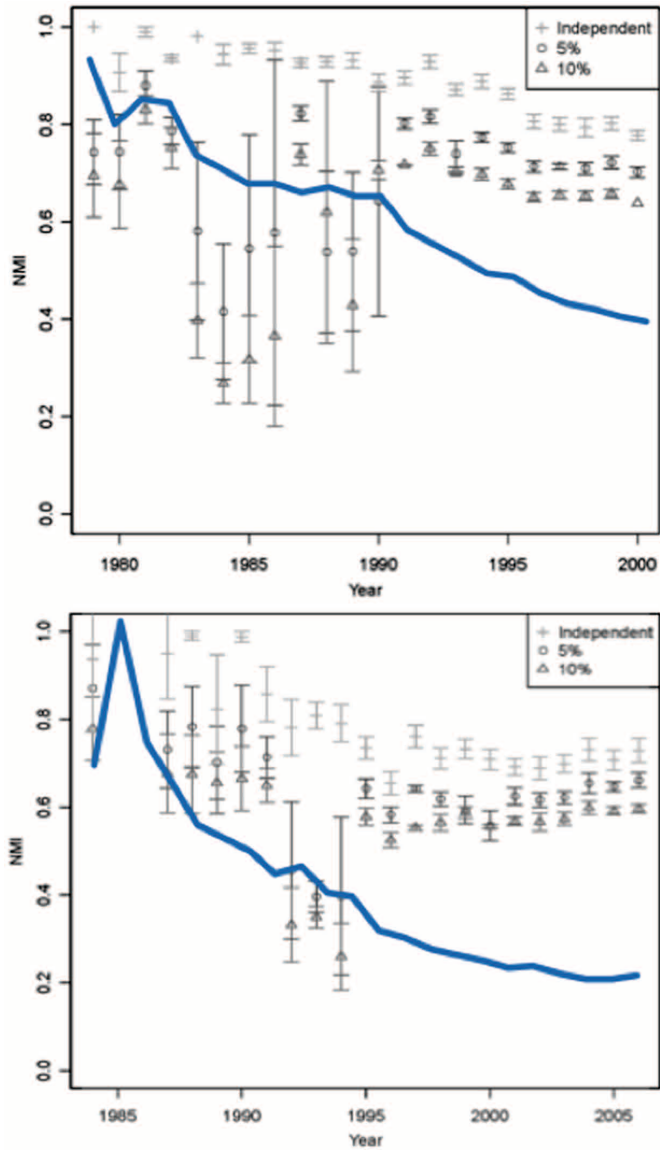


Figure 1. An Overlay of BTU’s Figure 1 with Modularity Trends from Shwed and Bearman (2010)

Note: The top panel refers to the literature about smoking hazards and the bottom panel to the carcinogenicity of solar radiation. In both cases, modularity trends suggested by Shwed and Bearman (2010) are indicative of a reduction in the variance of results from BTU’s procedure, following an independent random signing of parts of the original networks. This means that as a literature becomes consensual (indexed by the decreasing modularity trend line), even BTU’s random manipulation of 10 percent of the network ceases to have an effect. Consensual literatures, black-boxed, are more robust.

in the calculation of modularity pertains to connectivity, the number of ties in the network, and not to network size. It is the required minimal normalization to allocate nodes into communities. At the risk of being didactic, this assignment process was not what our article was about.

We presented scaling to control for benign contestation—the natural process of scientific specialization, as well as the development of invisible colleges. These processes are a function of the literature’s size, regardless of its connectivity. BTU mention that out of 457 expert papers cited by Fortunato (2010) none used our scaling. This is no coincidence, of course, but it is not condemning evidence either. It establishes that we were the first to focus on the modularity of a graph as a diagnostic concept (rather than on the communities it creates) in the specific context of citation networks where a big part of community creation is a product of size.

Thus we find BTU’s Figure 2 unremarkable: it demonstrates our point that much of what goes on in science is benign contestation. The figure also shows that without our scaling of modularity, it remains fairly constant throughout the period. BTU (2012:1052) manage to support our substantive argument for scaling, while ignoring the scaling itself because we “did not infer renormalization within the theoretical framework on which modularity is built.” They show the growing specialization over size, they show that raw modularity remains stable over that time, and they ignore the empirical fact that throughout this period our scaling tracks important processes such as the end of the dispute about safer cigarettes and the forming of consensus around passive smoking.

SCIENTIFIC CONSENSUS VERSUS DUTCH POLITICS

BTU consider as a new case a public debate in Dutch newspapers. They bridge the immense gap between what they comment on—our discussion of scientific consensus formation—and a political debate in the Dutch public sphere by arguing that their method is “very general” (BTU 2012:1051).² They provide

three colorful figures that reiterate the (undisputed) point that information on the nature of a tie may change community allocation in cross-sectional networks. They show—with data far removed from the original context—that community allocation would have been different if tie valence were taken into account. The analysis is not longitudinal and BTU do not report modularity—so we cannot tell whether this has any bearing on our approach—although the context is so different it is hard to determine what similar or different modularity trends would mean for our model of distinguishing benign contestation from epistemic rivalries in scientific literatures.

And yet, the Dutch newspaper analysis warrants the observation that BTU introduce a radically different context in order to show the utility of their extension of modularity, only to end up analyzing a subset of a subset of a subset of a subset of their original data.

Let us compare the data selection involved in our analysis with theirs. Both analyses define an issue with boolean search keywords. Our analysis selects on journals indexed by the Web of Science, they select on Lexis-Nexus. BTU add four more levels of data selection. First, they rely on only two newspapers, neither of which are among the top Dutch newspapers with respect to circulation. Second, among all the articles they retrieved, BTU select only articles over 1,000 words. From these they manually generated a population of actors and references between them, labeling the references as positive, negative, or neutral. They then discarded the data labeled neutral (which actually means ambiguous; BTU:1058). Finally, after dispensing with an unreported ratio of the data twice, introducing coder bias and discarding more than half of the data, they restrict their analysis to only the largest connected component—discarding an unreported proportion of whatever data they had left. That is not the symmetry we strived for.

CONCLUDING REMARKS

BTU present three distinct analyses to challenge our methods and results. The first analysis shows that community allocation changes

when tie type is taken into account. This has no relevance for our analysis, which makes no use of community allocation outcomes in the first place. In addition, community allocations may vary while modularity remains robust (Good et al. 2010). Their second analysis aptly demonstrates our idea (see Shwed and Bearman 2010:823–24) that a big part of scientific contestation is what we label benign contestation—the normal process of specialization. BTU’s final analysis provides support for all of the claims we made with respect to what is wrong with expert judgment and why symmetry is a goal to strive for.

Our finding is that scaled modularity of citation networks indexes contestation. Scientific consensus sometimes increases and sometimes decreases, on specific issues in specific times. BTU (2012:1058) write that they “believe that [our] claims and findings do not stand up to scrutiny.” And yet, the trends we find in scaled modularity, without any data manipulation, fit broad expert reports on each of our seven cases—two of which (in the cases of autism and non-ionizing radiation) were made after our analysis. Nowhere in their comment do BTU supply any shred of relevant evidence that contradicts any of our findings, theories, or methods. That is the nature of belief.

BTU (2012:1052) also argue that our scaling is not “within the theoretical framework on which modularity is built.” We plead guilty. We started with theoretical puzzles and insights from the sociology of science, found a tool (modularity) that seemed to relate to the theory, and went out to the field to test the utility of our tool. In the process, empirical reality (the fact that benign contestation is clearly size dependent) led us to modify our tool by scaling for size. We never attempted to be “within the theoretical framework” of our tool, because we use tools and theories to help us understand data. Theories and tools are something to work with, not something to be entrapped within.

The main benefit of our method is that it enforces a robust symmetry. We leave no room for discretion, coder bias, or manipulation of context and scope in data selection.

This is why our method is useful for evaluating the state of contestation on scientific issues laden with political interests. The symmetry constraint actually makes it possible for us to say that we can identify when scientists agree and disagree about important issues. We show that our approach works across such diverse cases as gravitational waves, anthropogenic climate change, and cancer and smoking. The method proposed in the comment is a technical expression of qualitative analyses that people have been doing for years. Our idea in the original article was to show why that analysis strategy is not useful. That our commentators did that for us is very kind indeed.

Notes

1. BTU report they replicated our data collection procedures, but they start their analysis some 27 years later than ours in the smoking case.
2. It appears not general enough to be applied to the subject of our article, which is why BTU use arbitrary randomization instead of real data, and real data is used in a radically different context.

References

- Bruggeman, Jeroen, V. A. Traag, and Justus Uitermark. 2012. “Detecting Communities through Network Data.” *American Sociological Review* 77:1050–1063.
- Cole, Jonathan R. and Stephen Cole. 1981. *Social Stratification in Science*. Chicago: University of Chicago Press.
- Fortunato, Santo. 2010. “Community Detection in Graphs.” *Physics Reports* 486:75–174.
- Good, Benjamin H., Yves-Alexandre de Montjoye, and Aaron Clauset. 2010. “The Performance of Modularity Maximization in Practical Contexts.” *Physical Review E* 81, 046106.
- Shwed, Uri and Peter S. Bearman. 2010. “The Temporal Structure of Scientific Consensus Formation.” *American Sociological Review* 75:817–40.

Uri Shwed is an Assistant Professor of Sociology at Ben Gurion University of the Negev. He chairs the masters program in critical studies of society. His research focuses on the sociology of science, stratification, networks, and methodology.

Peter S. Bearman is the Cole Professor of Social Science at Columbia University. He is currently working on the autism epidemic, qualitative research design, large-scale civil conflict, and the problem of continuity.