

Comment on Geodesic Cycle Length Distributions in Delusional and Other Social Networks

John Levi Martin*

Introduction

I am delighted to see Stivala's piece on geodesic cycle length, which responds to and goes considerably beyond my 2017 JOSS. This article (1) regularizes the terminology I used; (2) replicates my analyses using exponential random graph models; and (3) applies these models to other data sets to examine the degree to which these models predict geodesic cycle length. All of these constitute a welcome (and impressively done) contribution. Yet, I also have a sense that some of the motivation of this paper is to establish the superiority of the ERGM approach, and to treat all others as, at best, fallbacks.¹ Given that part of my reason to write the first paper was precisely to try to help us avoid the monoculture that I see developing with the use of ERGMs, Stivala's contribution provides an excellent opportunity for social networkers to consider the implications and strengths of different models, and different ways of understanding our task as analysts.

Here, I will argue that the social networks community is increasingly moving towards an ill-considered ritualization of ERGMs, and in such a way to undermine the distinctiveness of network analysis/mathematical sociology, which had been the great hold-out against the "saming-of-everything" associated with the ideational sink of mainstream sociology. If we do

* University of Chicago, Chicago, Illinois, USA. jlmartin@uchicago.edu

¹ It should be noted that the version of this paper that I read as a reviewer, and on which I had suggested commenting, did not include the *dk*-series analyses, and lacked the final concluding paragraph of the current version. It thus had a stronger sense of defending the orthodoxy, and therefore seemed to be a very good place to begin a discussion of these issues.

not change course, we will import into our own field the contradictions that have, in the past generation, been recognized, but not solved, in mainstream statistical practice. I first address the contribution made by Stivala to the substantive problem at hand, then discuss the ritualism in current usage. I propose that close consideration demonstrates that the attributes of the ERGM model that Stivala suggests makes it superior to other techniques are more deleterious than advantageous. I try to demonstrate the collapse of the ERGM into the general linear modeling paradigm tends to lead us, and in this case, has led Stivala, to make the same interpretive errors that bedevil current social statistics more generally. I close by making a few suggestions as to where we can go in the future.

Patricia's network in comparative perspective

Regarding the substantive claims, I think that Stivala here shows some of the advantages of using the same model on many networks. He confirms my arguments about the geodesic cycles being surprisingly large, and surprisingly small, in Patricia's first two graphs (respectively). Stivala also notes that this is not quite true for the final graph, an analysis I had not done, as this graph had seemed to me to be the same as the second except for the addition of a heteroplanar component which complicates things. (I think that analyzing the two planes together leads to problematic results, but I still should have done the analyses that Stivala does and reported these results.) Further, Stivala takes a set of both real world and fictional networks and shows that in none of these does a straightforward, out-of-box-parameterization version of an ERGM fail to reproduce the largest geodesic cycle, or indeed the distribution of geodesic lengths. As Stivala says, this supports the argument that Patricia's mental model was such that the geodesic was an important consideration. Two different null models (the *dk*-series' most complex model and the out-of-the-box ERGM) fail to reproduce this observation.

Thus Stivala provides much stronger evidence than had I that there is something comparatively unusual in the networks made by Patricia. Whether this supports my argument as to the fundamentally *spatial* intuition that she, and other twentieth century Americans, had regarding the notion of social networks, remains to be seen. Yet (as noted above), it appears (especially from the structure of the first submission [see note 1]) that what motivated Stivala was not simply the desire to place my results in comparative perspective, but also to re-do them with an ERGM. Why might that be?

Ritualism

In my *Thinking Through Statistics* (Martin 2018, Chapter 8), I discussed the pressure on researchers, struggling to get ERGMs to converge, to change the model they are fitting, or to ignore certain cases. This is an indication of what in organizational research has been called "goal displacement" or "ritualism" – when the means to the end become the end itself. In the case of statistical practice, we would rather get the best estimates of the *wrong parameters* and estimate the *wrong model* than use a different technique that lacks the (ideal) asymptotic properties of the ERGM but actually is relevant to the question we have. We would, to take famous words from John Tukey, rather be "precisely wrong" as opposed to "approximately right."

Cases in which researchers change the model or drop data to achieve fit may be less common now than when I wrote *TTS*: certainly Stivala believes that “advances in model specifications such as the use of the ‘alternating’ or ‘geometrically weighted’ model terms..., curved ERGMs..., and alternative estimation algorithms...mean these degeneracy problems can generally be overcome (Schweinberger et al. 2019).”² Without quibbling over whether “generally” has any particular meaning here, I would never claim that the weaknesses of *any* model cannot be overcome, if by this one means that at the end of the day, one has fit a model. The question is often *what* model has been fit, and whether it answers the question with which we began.

Of course, sometimes science tells us that our first questions were unanswerable, and redirects us to ones that we *can* answer, and there is no shame in that. The issue with ritualism is that we *could* get traction on our question by using a different method, but do not, because of the allure of the method (or the capacity of reviewers to halt publication of useful results that require a different method).

I also do not deny that ritualism can help a field move towards increased reliability, the value of which is not to be underestimated. The opposite pole of ritualism – frenetic individualist innovation, never doing the same things twice – is just as incompatible with scientific advance as is ritualism. It may sound glorious to call for an end to fundamentalist orthodoxy, and to “let a thousand flowers bloom.” But when every piece is both a substantive claim and a methodological innovation (which did tend to characterize the earlier period of social networks research), something is wrong, even if a good time is had by all. What we should be prizing, then, are robust techniques that allow for comparable, theoretically relevant, answers across a wide variety of data sets. Stivala is confident that the ERGM is just such a technique. I am not so sure, and I *am* sure that it is far too early to attempt to push for a monoculture.³

This seems like a good time to consider the purported advantages. But before simply assuming that we know what to tally up as a plus and what a minus, it is worth being clear as what we want our models to *do*. Stivala seems to think that we want them to fit our data. I disagree, and think the best way of making this point is to remind ourselves as to how we got to the ERGM.

² However, I do note that Stivala saw fit to specify that “Only models that show acceptable convergence and goodness-of-fit according to statnet diagnostics are included in the results.” This does seem to imply that not all models *could* be adequately estimated.

³ We should remember the short life of “best practices” in this field: since the development of the first p^* models, we have had a relatively quick succession of enthusiasm for pseudo-likelihood estimates, for loglinear fits and indeed pseudo-likelihood comparisons of different models, for Gibbs-sampling type MCMC methods, for MCMC methods for specially devised parameters and, presumably, soon a reliance on Hamiltonian Monte Carlo methods (here see Stoehr, Benson, and Friehl 2018; I am not competent to evaluate the comparative advantages of these approaches for ERGMs but in statistics more widely there seems to be a wholesale move away from the methods currently employed for ERGMs).

History

In the late twentieth century, ideas coming from mathematical sociology had suggested some strong models for informal social structures, especially the notion of balance, still a topic of serious investigation today (Rawlings and Friedkin 2017). Since those structures were not observed in all their clarity in sociometric data, there was an attempt to make statistical tests for *substantive* models of social structure. This was, to my way of thinking, an eminently scientific way of proceeding: a clear analytic model was proposed that reality should approximate to the extent that disturbances were removed: it could be derived from (1) a few axioms regarding interaction supported by introspection and (2) logic. The *question* that researchers faced was whether the evidence tended to favor one or the other of these substantive models (most important here was the work of Paul Holland, Samuel Leinhardt, James Davis, and Eugene Johnsen [especially Holland and Leinhardt 1971; for a discussion see Martin 2009, Chapter 2]). The *problem* that researchers faced was one of “compared to what” – determining whether we see a structural tendency requires a baseline, which means a distribution of expected values of configurations given some null hypothesis. The first important *statistical model* (note that this is not the question itself, but the response to the problem generated by an attempt to answer the question) was the U|MAN triad distribution, which attempted to look for triadic structure beyond what would be expected given the types of dyads present. One limitation, as Holland and Leinhardt noted, was that individual level differences might be mistaken for structural effects. A hierarchy of attractiveness (tendency to be chosen as a friend) could, for example, be mis-cast as a general psychological tendency towards wanting to be friends with one’s friends’ friends.

Their next major effort (Holland and Leinhardt 1981), although mathematically related to the triad analysis, seemed to have the opposite problem. Individual attractiveness and expansiveness (tendency to make choices) were estimated, along with a tendency to reciprocity, but this required getting rid of almost all the structure, leading Breiger (1981) to use what was for him extremely strong language: he considered this “inappropriate.” The choice of the exponential function taken by Holland and Leinhardt (1981), despite the attendant difficulty of dealing with the normalizing constant, made a great deal of sense as a general approach. It seemed that there was an expectation among researchers that with grunt work, other probability distributions, presumably less individualistic, would be added to the p_1 , leading to a large family of interpretable models that generated probability distributions of graphs.

Shelby Haberman had noted the relation of the p_1 model to loglinear models, and Stanley Wasserman jumped on this, trying to push a general loglinear modeling framework as our “go to” for social network analysis (the reason Wasserman and Faust (1994) devote so much space to this). But Frank and Strauss (1981) had recognized the significance of Besag’s (1974) work on the Hammersley–Clifford theorem for statistical graph theory. This theorem demonstrates that the probability distribution of a “Markov” graph – one where two ties are conditionally independent if they are between four distinct nodes – can be factored, as it is a Gibbs distribution. This is, I want to emphasize, a *mathematical* result: the only necessary assumption is the identity of all homologous configurations, which made sense for graphs without covariates. If so, then any Markov graph has a probabilistic factorization. It is really that way – it is a non-arbitrary result. And it at first seemed to many that most social networks would be Markov.

Strauss and Ikeda (1990) then showed that Besag's (1975) pseudo-likelihood approach could be used for such graphs. As Stivala points out, there had been concerns about degeneracy from the start – models that worked well when anchored to a spatial substrate, like those in physics and in geographical science, when transferred to networks flapped about like a poorly tied down piece of plywood tied to a car roof. But the pseudo-likelihood estimates behaved better, and Wasserman and Pattison (1996; Pattison and Wasserman 1999) seized upon this as the way forwards as a general approach. Indeed, they christened it p^* because of the notion that this was an overarching parametric structure that could be used to generate *any* exponential distribution one could think of. I think they thought of this as wonderful – somewhat akin to when Nelder and Wedderburn (1972) developed generalized linear models.

But just as the very success of Nelder and Wedderburn's unification came at a great cost of allowing sociologists to have a single (and singularly false) vision of society in their minds (Abbott 1988), so the replacement of *sets* of *meaningful* distributions, even if nested, with a *single right hand vocabulary of effects*, has, I think, damaged the mental models social network researchers take to their data.

The ELM

Let me back up for a moment, to one of the earliest models for random graphs, the Erdős–Rényi (1959) model. It is quite common to hear or read that the Erdős–Rényi model is a probabilistic random graph model characterized by the number of nodes N and a number $0 \leq p \leq 1$, which is the probability of an edge being present. Such a characterization is not correct, as that would be the model proposed the same year by Gilbert (1959). Erdős–Rényi proposed a random graph characterized by N and by the total number of edges, E . While we can derive, for any Gilbert graph given N and p , a probability distribution for E , this makes no sense for an Erdős–Rényi graph, where E is fixed. Why do we mis-remember the Erdős–Rényi model as a *parametric* one? I think because we have moved to a conflation of parameters and probability in the sociological imagination.⁴ And I believe that the parametric hegemony may have some unhappy side-effects.

Harrison White, whose great mathematical vision, based on Lévi-Strauss's structuralism, was really the inspiration for much of network analysis worthy of the name, once wrote a minor paper, "Parameterize!" (2000). He noted that the great work of his that energized social network analysis – his studies of kinship systems, and then his transposition of these to informal networks via the notions of structural equivalence and role algebras – had nary a parameter in it. He considered this a fault, and was excited that, in his work on markets, he was going to be able to reduce the variation into a single exponentiated parameter. He hoped to do what science does – to look for relations between invariants. What sociology does is something very different, and it is (unlike most physical sciences) based on the Great Divide between the left hand side and the right hand side, and the notion that all parameters are fundamentally of the same (intellectual) nature. I wish that he had warned his readers that not *all* parameterization was a step in the right direction!

⁴ And we want to attribute it to Erdős because he was such a trip!

In terms of programming convenience, there is much to be said for the capacity to reconceive any formal data analysis as an application of a linear model of some very general form. But in terms of the direction of our capacity to generate important arguments about the world, especially structural models, it can prove deleterious. I propose that we think not so much about the GLM as the ELM – the *Exhausted* Linear Model. (Imagine it looking up bleary-eyed from its desk after an all-nighter: “What? I just finished fitting every multinomial loglinear model by treating it as a Poisson regression! Now you want me to do networks *also*?!”) The problem is that every problem seems to become the same – a kitchen sink (or garbage can) regression in which one tosses anything one thinks of on the right hand side, and keeps whatever fits (and whatever *can* be fit). In this light, let us now rethink the question of covariates.

The problem of covariates

Recall that when the excitement began for what became the ERGM, it did not have to do with the fitting algorithm used, nor with the notion that parameters were maximum likelihood estimates, but with the finding that a Markov graph could (assuming homogeneity of local effects) be factored in a *non-arbitrary* manner – indeed, as a Gibbs distribution which already had a very close connection to the exponential distribution. While Wasserman (or so I believe) moved quickly to vaunt the incorporation of other covariates, these did not necessarily have any privileged theoretical position and not all were happy with the idea that the models should be stretched to include parameters that broke the deductive link to the Hammersley–Clifford theorem. That is, if one had correctly parameterized a Markov graph (say as a series of stars and triangles), and one had a covariate such as sex, the only way to incorporate this while still having some connection to the mathematics underlying the approach would be to relax the homogeneity assumption and have different structural parameters for different groups of nodes (and not to throw in covariate-based terms at the individual or dyadic level, say). It is not impossible that there is a second stochastic process of sex homophily that happens to be independent of the structural processes (to the degree that any two predictors in an exponential model can be termed “independent”) but it is hardly obvious that this is frequently the case. If it is not the case, then all the ERGM can do is give us precise estimates of the wrong model.

Stivala takes for granted that it is a point in favor of the ERGM that it can take various other covariates into account. Indeed it often *has* to, otherwise it will not converge (“It’s not a bug, it’s a feature!”⁵). But are we sure that we *want* to? Few statisticians can find anything good to say about conventional sociological practice, which involves large strings of overlapping variables and results that rely on strong assumptions about their interrelations. It is even worse for the case

⁵ There is much to admire in the derivation of the geometrically weighted effects that have been used to damp the wild behavior that led to degeneracy. But even this is not always enough, and practitioners have found that for some data, they need to “ride the brakes” by including other parameters that introduce some heterogeneity to the edges. When refusing to get rid of my old car, I swear to my wife that the automatic choke on the carburetor works fine – you just have to stick a fork in it when it’s cold. These definitions of “work fine” display the admirable loyalty of the romantically attached, but are not the same “work fine” as others have a right to expect.

of the study of social networks.⁶ This is because if there actually are *structural* laws or principles out there –and if there are not, I do not think we have any reason to be particularly interested in spending much time fitting *any* such structural models – there is every reason to believe that they are partially separable from issues of *allocation*.

This point was made quite clearly by Lieberman (1985). The factors that shape an occupational structure, he insisted, are not the same as those that explain who, conditional on the existence of this structure, get what job. Yet, the ELM approach to mobility tended to lead people to blur these two and make implausible counterfactuals (if everyone goes to college, everyone will be a professional). So, too, if there are structural properties that we are attempting to isolate, we will mis-estimate these if we confuse individual covariates, that might help determine *who* occupies which positions in the structure, with the principles of the creation of that structure.

Indeed, in some interesting cases, individual covariates are as likely as not to be endogenous to structure. If the actual structure of a high school is the ranked cliques model, the highest ranking cliques may control access to certain extracurriculars which, if “taken into account” in the model (say, by entering a dummy variable for SHARED_EXTRACURRICULAR), might lead us to reject the ranked cliques model. The point is not that it never makes sense to enter dyadic or individual covariates, but that we must beware of falling back into the strange sociological conviction that the best model excludes no significant predictors (only, perhaps, relying on a claimed causal order to refrain from including post-treatment confounders). This conviction is based on an untenable assumption – and this is one of the few planks on which both those dedicated to causal analysis and those opposed to it can agree – that partialled coefficients on the right side can be treated as “effects,” whose precise metaphysical nature is left unexplored. The use of the ELM reinforces this, and I think we can even see this in Stivala’s analyses.

Stivala argues that the ERGM has the advantage of being able to take nodal attributes into account, and does this in Models 2 and 3 of Table 3, the first suggesting that Christian alters have higher degree, and the second that (not surprisingly) those in the Sphere of the Blue Flame are more likely to be tied to one another than are random pairs of nodes. But if one examines Patricia’s maps, we see that this Sphere is only one of four large clumps of nodes (there are two components, the larger of which easily breaks into three pieces with one or two cuts to separate each). She has labeled this one, but what if she had labelled another (for example, that to the left of the Sphere of the Blue Flame?). We of course would find homophily here as well. What if she had labeled this “Sphere of Ju-Ju” after the most central actor? What if she in fact had labelled every “wheel” (every structure consisting of a hub and its spokes), and we were to take this into account? As we added more and more of these seemingly nodal attributes, our structural parameters would of course change. But in a particular way – we can imagine that, at the end of the day, we would no longer have any idea as to the nature of the structure, because we had mis-parameterized it as nodal attributes! It is just the nightmare that would cause Harrison White to faint in horror – all structure had been turned back into seemingly individual variables. And yet it

⁶. And I do not even want to go into the many cases of incorrectly interpreted ERGM models; even when they are correctly specified (and it is easy for complex models with covariates to be incorrectly specified), nested terms are notoriously tricky for even experienced analysts to deal with, and many articles get published that rest on misinterpreted parameters.

is difficult to prevent this sort of regress once one decides to envision one's job as fitting an ELM. We are ineluctably drawn to add parameters, and the only way to keep track of what we are doing is to treat each parameter as "an effect," thereby reifying it and projecting into our vision of the world that which is the most convenient interpretation for each that we can think of. And I think we see the way the ELMification of the ERGM pushes our interpretations in Stivala's discussion of Patricia's maps.

What do ERGMs tell us?

The sorts of interpretive slippages I will point to in Stivala's arguments are, I think, characteristic of the way in which sociology has had to make use of the ELM, and how increasingly we see social network researchers interpreting the ERGM. Widespread or not, however, these issues get to the heart of the choice before us, and so I want to look carefully at how, after all the work done, the results of the ERGM are interpreted. Stivala writes, "Given an observed network, we estimate parameters for local effects, such as closure (clustering), activity (greater tendency to have ties), homophily, and so on. The sign (positive for the effect occurring more than by chance, negative for less than by chance) and significance tell us about these processes, taking dependency into account. That is, the parameter tells us about the process occurring significantly more or less than by chance, given all the other effects in the model occurring simultaneously." This is the way we generally write about our models in sociology. We tend to take the ambiguity of the term "effects" (which can refer simply to a certain type of statistical predictor, but carries connotations of causality) as a cover for stretching a bit beyond what we really are doing.⁷

But in this case, I do not think that Stivala is correct to say that parameters in ERGMs should be interpreted as giving us *any* insight into process, except, perhaps, when they are at the limit convergent with SAOM parameters, and the SAOM model happens to be the true model.⁸ Even in a TERGM, these parameters must be understood as *descriptive* – they allow a stochastic model to reproduce a family of graphs such that the associated target statistics (e.g., number of homophilous ties) is similar to that in the observed graph. The notion that this statistic therefore measures the strength of a *process* such as a preference is quite implausible (e.g., perhaps we find that students of similar religious backgrounds tend to be friends, but this turns out to be completely explicable on the basis of residence). While this point about the potential slippage between parameter and underlying social process might be generally true, this is far more weighty in a model like the ERGM, in which most of our important parameters are one way of many of dividing up a complex set of interdependencies, as opposed to a single coefficient paralleling a vector of distinct observations. An ERGM using terms for degree and transitivity may do an excellent job at reproducing certain aspects of a network structure that was actually formed via a very different spatial process (nodes are located in an analytic space, with ties made on the basis of a stochastic function of the inverse of the distance). The good fit of whatever ERGM model is chosen, then, in no way indicates that the coefficients correspond to any

⁷ We lead like a runner on first base, until that reviewing pitcher trained at Harvard snaps the ball to the first baseman, at which point we frantically dive back to safety, hitting the dirt on our bellies!

⁸ Snijders (2001) shows this to be true in some conditions of detailed balance, though see Leifeld and Cranmer (2019) for more.

process.

A possible example here is raised by Stivala, in noting that Bearman, Moody, and Stovel (2004) find structures of romantic attachments to approximate spanning trees, but also to form structures with a large geodesic cycle. As Stivala notes, while “Bearman et al. (2004) propose a normative proscription against four-cycles (‘don’t date your old partner’s current partner’s old partner’),” there could be other reasons for such a structure. Stivala makes reference to one such alternative, an extremely-hard-to-fit ERGM that was able to produce similar data. But such structures will also arise if boys and girls are distributed in a space of likeness (say, a two-dimensional one), they have a relatively low degree, and there are strong tendencies for them to have relationships to those close in space.⁹

One might acknowledge the force of this critique but excuse such interpretive slippage (from parameter to effect, from effect to process) as a commonly cut corner in social statistics – we rarely explicitly remind the reader that our results are only interpretable if our assumed model is correct (one might argue), and so most of us bear this in mind as a mental reservation, making the omission of explicit mention innocuous. But such an omission cannot be seen as innocuous if it is made as part of an argument for the superiority of a certain model!

One might also accept my argument in principle, but demand that evidence be shown of a concrete misinterpretation. Such evidence may rarely be forthcoming, if all we have is a single model with parameters crying out for tendentious interpretation. This is where the virtues of comparing results across *types* of models, which Stivala carries out here, will appear. I believe that in this case, the results demonstrate such a misinterpretation in Stivala’s analysis of the 1992 data, precisely because Stivala tends to go from the fact that (were the model correct) the ERGM gives defensible estimates of parameters to the assumption that, by definition, the ERGM model is true. “We can conclude,” Stivala writes, “from these models that there is a significant tendency [towards] preferential attachment [by]¹⁰ degree (the GWDEGREE parameter is positive and significant). This is not what we might have expected from the *dk*-series models as discussed in Martin (2017) where this network is described as having degree heterophily or a hub-spoke structure (Martin 2017:12).”

One will note here the unwarranted leap from the GWDEGREE parameter, which really describes the *ceteris paribus* degree *distribution*, to a particular *process* that might generate similar distributions. Indeed, some reflection suggests the difficulty in seriously claiming that a process of preferential attachment occurred within Patricia’s head, unless one were to propose that earlier alters would tend to have more connections because of their age, leading to a skewed distribution (though as noted in Martin 2017, it is clear that this sort of process did *not* take place). Further, the fact that the network *does* have a hub-spoke structure, as suggested by the *dk*-series analysis, and obscured by the ERGM, is easily seen if one were to view Patricia’s *own* visualizations (reproduced in Martin 2017: Figures 2 and 3), as opposed to those constructed by applying a conventional graph algorithm to the data (Stivala 2020: Figures 3 and 4). *Patricia* organized the data unmistakably in hub-spoke structures. This example, then, wonderfully

⁹ Demonstration of this is left as an enjoyable exercise to the reader.

¹⁰ I have corrected what I believe to be typographical errors in this sentence.

illustrates the dangers of conceptual monoculture: had we only the data, Pajek and statnet, and were we permitted only to employ ERGMs, we would not understand the actual structure of Patricia's network.

I have the feeling that many social networkers go from the admirable properties of the *estimates* of MCMC ERGM fits to the truth of the *model*. Thus, model *assumptions* are treated as a kind of magic which makes the world itself re-arrange itself meekly to allow for the application of the model. Indeed, this notion that assumptions *allow for* an unproblematic mapping from parameter to process (as opposed to creating *problems for* drawing conclusions) is explicitly defended by Stivala: "The social circuit dependence assumption means that statistical significance of parameters tells us something about the corresponding local processes generating the observed global structure of the network." One often reads similar arguments regarding ERGMs, but this is quite backwards – the *assumption* cannot tell us anything about the world! What one should say is rather "The social circuit dependence assumption means that if the processes generating the data include dependencies that are here ruled out *a priori*, the resulting model parameters do not have any clear meaning." (I have also noted this same slippage – treating assumptions as efficacious in creating the desired world – in reviews given by devotees of the very elegant Stochastic Actor Oriented Models of Snijders, as some insist that authors of submitted papers *should* use this model because, in their minds, this will construct a virtual world which is more tractable than the reality in question. Snijders himself takes the exact opposite approach, emphasizing the scope conditions of the method.¹¹) We make assumptions because we *have* to in order to obtain useful estimates, not because they change reality to conform to our model.

Even if the model is correct, its coefficients are not necessarily indicative of any particular *process*. Indeed, I do not think that many ERGM modelers would have been able to rattle off a clear behavioral-process analogue to the once-canonical alternating two paths statistic! Indeed, Robins et al. (2007), if I remember correctly, confessed to being somewhat mystified by this statistic. For better or worse, all our parameters are, in and of themselves, model-dependent *descriptive* devices. They could be used to identify the degree of certain processes in known-to-be-true models-of-process, but as far as I know, we do not have any such known-to-be-true models-of-process. If we are going to make use of them, it has got to be in another way.

How should we interpret ERGMs?

What do we want from ERGMs? One common answer among non-network-researchers is that we want them to adjust our models for dyadic data to deal with the statistical non-independence of observations. For example, someone is interested in high school friendship formation, and has a model including observed covariates, but a conventional logistic regression on these covariates, even if the model is correct, will not reach the maximum likelihood estimates of the parameters because of the violation of conventional sampling axioms. Since, however, we invariably do *not*

¹¹ I know this because I once urged a student at a conference to use SIENA for his/her project, as it would be so elegant, pointing out that Snijders was literally quite close at hand if there were some issues in adapting the model to the data structure. "I already asked him," the student reported sadly, "and he said he didn't think the assumptions of the model were satisfied by my data, and so he wouldn't do it."

have the right model, and because there are many other techniques for dealing with statistical non-independence, it seem quite implausible to try to channel all efforts into using models and estimation routines whose properties are still debated and which sometimes refuse to yield estimates.¹² I do not know of one sociological argument involving observed covariates (such as GPA homophily) that hangs on whether MCMCMLE or PLE estimates are used for structural parameters, for example. But even if there were one, there is no reason to think that the results from the former are better than those of the latter, given that we must expect that we do not happen to have all the covariates in all the right functional forms. If using ERGMs does not make very much sense if we are merely attempting to estimate *other* parameters, let us assume that we are interested in using ERGMs to learn about structure itself. In that case, how should we interpret the results?

One possibility would be to return to the notion that we are attempting to generate known distributions of graphs – the parameters are merely a flexible way of doing what was being done in, say, the U|MAN analysis via combinatorics. (This is the interpretation of Fuhse 2018.) We want to preserve a theoretical understanding of the nature of the predictions, but the precise value of the parameters is uninteresting. But if this is the case, then a method like that of the *dk*-series – which would be a bit like the Erdős–Rényi approach in contrast to the Gilbert – seems preferable, as it has less interpretive ambiguity. And ad hoc approaches, in which a stochastic generating process is used to create graphs (somewhat like that in the Bearman et al. 2004 article cited by Stivala) may also be preferable, even if they lack any known statistical properties.

We see some residue of this way of thinking about ERGMs in the concern with *fit*. Such a concern once was widespread in the sociological use of statistics, and models with low R^2 were considered to be “bad models.” We no longer think this way – the degree of residual variance can be high in a model with important (and not merely statistically significant) predictors (and, as Mike Hout used to say, “Who wants to live in an $R^2=0.9$ world, anyway?”). But in ERGM modeling, “fit” is used (quite properly) in a different way – if we cannot reproduce the graph statistics, then we are not actually generating the family of graphs that we might be claiming is the class which includes the observed.

It is, therefore, quite understandable that fit becomes of great concern to those using ERGMs. Still, fit is only a means to an end of making meaningful statements about the world – it is not an end in itself. We know that bad models (e.g., those that condition on post-treatment confounders, for those doing causal modeling) can have better fits than good models. Yet I see with the current use of ERGMs a tendency to slip into a working consensus that one’s job is to fit data, and the better the fit, the better job has been done – even if this does not advance our understanding. I think there are two problems with this. One is that it can lead researchers to prefer ad hoc elaborated models that fit any particular case (or that *can* be fit) to the use of better structured and interpretable models fit to many cases. The most important substantive application of

¹² To the extent that such models are really oriented to a rigorous test of whether *non-structural* parameters are significant, the ERGM model is only a half-way step. Bayesian methods are used to fit the graphs, but, so far as I know, model uncertainty is never taken into account. Recognizing that there are many different equally plausible ways that the dependence in a graph could be parameterized would probably send our t -ratios plummeting downwards like a lead balloon!

ERGMs of which I am aware – one in which we learn something new that we did not know before, as opposed to replicating well-known facts about structure – is Faust and Skvoretz’s (2002; also Skvoretz and Faust 2002) use of a base model across many different social networks. While things like this had been tried with triad analysis, the use of a single ERGM model with a restricted set of parameters facilitates the development of knowledge that can cut across a single type of tie. These are not necessarily the best *fitting* models to any particular set of data, but they are the most *useful* ones for learning.

For this reason, I think that the out-of-the-box models that Stivala uses are the *right* ones to use, even though in most cases they assuredly do not correspond to *processes*, for our first crack at any network chosen at random. In general, it seems that we have tended to find that to adequately position any example on the most important axes of structural variation certain types of parameters are generally useful. For symmetric networks, we want something about degree distribution, and something about transitivity. For an asymmetric network, we would also want something tapping reciprocity, thereby getting at the moments of nodes, dyads, and triads that tend to contain the key to positioning any structure in a general space of possibilities (in *Social Structures* [2009] these moments of 1,2, and 3 are termed *differentiation*, *dependence*, and *involution*).

The second problem with the emphasis on fit is that it actually flies in the face of what I at any rate hold to be the most successful procedure for using statistics to build social scientific knowledge, namely, falsification. The key evidence supporting my argument about the root social networks schema being spatial was not the capacity of a spatial network model to fit the data – this capacity should be obvious upon inspection of the raw data. Rather, it was the *failure* of reasonable models to predict the absence of large geodesics in the 1992 graph. Why not rest my argument on the capacity of a spatial model to fit the data? Because a spatial model will fit the data from many non-spatial processes (indeed, the capacity of spatial models to fit almost *any* data if you have enough dimensions is the reason why these are widely employed in information-science contexts in which knowing the actual generating process is unimportant). We can only really tell a spatial from a relational model (in Watts’s 1999 sense) when a spatial model cannot account for ties crossing far reaches of space, or when a relational model makes predictions that are not borne out in practice (as we shall shortly see!).

Let us see whether fit is a good guide for determining when a model is helping us by considering the difference between the ERGM results for Patricia’s 1992 and 1993 networks. Stivala compares the results of the ERGM and the *dk*-series with an eye to which fits better. “For the 1990 and 1992 networks, the ERGM does not fit the maximum geodesic cycle length or geodesic cycle distribution any better (or worse) than the *dk*-series 2:5*k* distribution, despite the ERGM making use of nodal attributes for the 1992 network, which the *dk*-series cannot. However for the 1993 network, the ERGM has a better fit to the geodesic cycle length distribution than the *dk*-series 2:5*k* distribution (Figure 7). Note that all four ERGM models fit the maximum geodesic cycle length as well or better than the *dk*-series 2:5*k* distribution, including ERGM Model 1, which does not include any nodal attributes and has only three estimated parameters (Table 4).”

The implication seems to be that we should be comparatively happier with the ERGM (compared to the *dk*-series) for 1993 than for 1992, as it is here that the ERGM, with few parameters, fits the

data better. Yet, I would maintain that the opposite is true: for 1992, we learn something from the ERGM precisely *because* of its failure to fit, while in the 1993 data, we learn nothing. In the 1992 network, the incapacity of the ERGM to predict the small geodesics alerts us to the fact (if indeed we are correct in seeing it as one) that Patricia’s world is in some sense a large spatial world. In the 1993 data, the ERGM hits it right, not because the model is different, but because “this world” is different from the previous. Now Patricia has one geodesic of five (Polly-Beth/Annie-Juju-Wendy-Jill) and one of eight (Polly-Beth-Ashes-Allison-Vanessa-Christine-Julie-Patricia*); I use Patricia* to refer to the node named Patricia in patient Patricia’s graphs).

The model fits this statistic – but what does that tell us about the *processes* that gave rise to the graph? If I am following correctly, Stivala would have us believe that the model gives us evidence of processes of preferential attachment and triadic closure. Fortunately, we can compare this 1993 network to that of 1992 to understand what produced these two larger geodesics. The 5-geodesic was formed by the entrance of a new node, Jill, who tied together Wendy and Polly, completing a new cycle. The 8-geodesic comes from a sadder pattern. There already had been a path connecting Patricia*-Julie-Christina-Vanessa-Allison-Ashes-Beth-Polly. However, Polly and Patricia* were not connected, while Patricia* and Vanessa had a direct connection (which, had it remained, would have short-circuited Julie and Christina from any geodesic). The changes here do *not* support a hypothesis of a tendency to triadic closure. Instead, what is happening is that Patricia* (the original, at one time “true,” personality) is losing ground and becoming increasingly unpopular. In 1992, she is tied to Charlotte, but in 1993, Charlotte drops Patricia* and affiliates with Polly, who is a rising hub (her degree goes from 3 to 2 to 5 across Patricia’s three graphs). Patricia* (perhaps in desperation) also affiliates with rising Polly, while Patricia*’s tie to Vanessa breaks, leading Patricia* to go from degree 4 to degree 2. Inspection of *Patricia’s* graph for 1992 shows that the new Polly-Charlotte, Patricia*-Polly ties are not triadically implied. They are, however, spatially close in Patricia’s rendering.

Our knowledge, then, comes less from fit than from *failures*, or at least from the *comparison* of models. The Orsini et al. (2015) approach formalizes this as successive comparison of nested models, as had the initial Goodman loglinear approach. The first way of thinking about the p^* model was that it was going to be able to do something similar (and see the wonderful piece by Pattison and Snijders (2013) trying to chart the relation of different models). While one can of course add or subtract terms to the ERGM like any ELM, the fact that ritualized tweaks are incorporated into most modeling, and that strong conventions have been established like the GWDEGREE and GWESP used by Stivala, suggests that we will not learn as much from ERGMs as we would if we could see where they failed.

Let me give as an example of how we best learn from ERGMs using a paper by Gondal and McLean (2013), studying data on loans among renaissance Florentines. They begin with the simplest random graph models, which fit terribly, but they show that much of that comes because these ignore reciprocity. Already, without a good model in hand, we have learned something (and something non-trivial – we might imagine that data on loans could tend *away* from reciprocity). They then move to something more like the “out of the box” ERGM model, first including five terms to capture the skew in both in-degree and out-degree. This model fits rather well, produces interpretable and significant parameters – but it does quite poorly at reconstructing the observed number of transitive triads. Gondal and McLean then move

cautiously, guided both by their substantive understanding and by the refusal of some ERGMs to converge, to a more complex model.

Had Gondal and McLean *started* with their final, complex and necessarily tangled, model, and simply declared that it fit, and that the parameters could be conveniently interpreted as indicating processes, I could not think it safe to accept their interpretations. But because they show the successive failures of cumulatively more inclusive models, we can learn something. We do not know whether there is an actual tendency for preferential attachment in the sense that those who are successful at getting loans become better able to get more loans (the “credit rating system”) or whether this simply captures a distribution in *need*. But we know that a theory that denies individual heterogeneity here is almost certainly untrue. And we learn this not from the final, well fitting model, but from the failures of the simpler models. Thus, I argue that what Stivala considers to be the recent success – that one can always get an ERGM to cross the finish line, if you coax it enough – may be no cause for celebration.

For the future

I am grateful to Stivala not only for taking seriously (as do I myself) the analysis of these somewhat strange data, but, by placing this analysis in comparative perspective, strengthening the conclusion. And I am grateful to Stivala for connecting this idea to proper mathematical vocabulary. But I am also grateful to Stivala for giving us the opportunity to reflect on current practice, and on where we are going. It would be a shame if we were to commit ourselves to a monoculturalism that aligns us with the thoughtways of the ELM, precisely what structural sociology was trying to escape. But I do not mean to argue that the problem with the current use of the ERGM-as-ELM is that it is supporting a rather fundamentalist orientation among some adherents (and I certainly do not think that this characterizes all users, let alone the pivotal developers of the method). Rather, I think what we need to do is to re-awaken our interest in the ERGM-not-as-ELM.

Whether or not any particular model cast as an exponential function and fit using an MCMC method is an advance, we should recognize that the core approach that lies at the heart of the ERGM is a beautiful and generative idea. Indeed, it turns out that the same fundamental vision underlies the random graph model, and the Gibbs sampling used to identify it, as well as being deeply connected to the pseudolikelihood as well. This is the Boltzmann equation, and the notion that there are consistent analogies that can be made between graph configurations and physical systems with variable energy levels. It is this sort of interest in pursuing elegant and rigorous mathematical derivations that separated the field of mathematical sociology from statistics-as-generally-understood (which concerned itself largely with issues of inference).¹³ The field of

¹³ If we are not going to try to rigorously make use of the potentials of the exponential distribution, we might do better to use a completely different approach. Twenty years ago, anyone who used an OLS for dichotomous data was considered to be criminally incompetent; today the rechristened “Linear Probability Model” is found superior in interpretability for many purposes (Angrist and Pischke 2009). It is quite possible that in 20 years, many researchers will have switched to OLS models with random effects for network questions that are not about structure and where the networks have density $.2 < d < .8$.

social networks research needs heads better than mine to grapple with the capacity of applying such thermodynamic models in a consistent way to social life, and to resist the temptation to run after data fitting. As Lieberman (1985) says, physics would have gotten nowhere if the Galileo had tried to explain *variation* and left it at that. We are perhaps a field whose bread and butter is variance, but this still must be a tool in the pursuit of *invariants* if we are to do anything worthy of the name “science.”

We should be more enthusiastic about the equivalent of an Ideal Gas Law that can set the direction for plausibly cumulative social science than Actual Gas Fits that only predict the past using parameters that we all agree to pretend are meaningful. There is nothing outlandish in proposing that we pursue such *covariateless* models in the sense of not *controlling* for error but collecting data in pure conditions. Only a science that is “out of control” has a chance of getting anywhere. If you think that it is implausible that we could develop anything like an ideal gas law for human beings, because we humans are so complex and individual, do not despair! I believe I have heard that Neon atoms think the same about themselves.

References

- Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6:160-85.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.
- Bearman, P.S., J. Moody, and K. Stovel. 2004. "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks." *American Journal of Sociology* 110(1):44-91.
- Besag, J.E. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society, Series B* 36:192-236.
- Besag, J.E. 1975. "Statistical Analysis of Non-Lattice Data." *Journal of the Royal Statistical Society, Series D* 24: 179-95.
- Breiger, Ronald L. 1981. "Comment on Holland and Leinhardt's 'An Exponential Family of Probability Distributions for Directed Graphs.'" *Journal of the American Statistical Association* 76:51-3.
- Erdős, P., and A. Rényi. 1959. "On Random Graphs. I" *Publicationes Mathematicae* 6:290-7.
- Faust, Katherine, and John Skvoretz. 2002. "Comparing Networks across Space and Time, Size and Species." *Sociological Methodology* 32:267-99.
- Frank, Ove and David Strauss. 1981. "Markov Graphs." *Journal of the American Statistical Association* 81:832-42.
- Fuhse, Jan A. 2018. *Soziale Netzwerke; Konzepte und Forschungsmethoden*, 2nd ed. Konstanz: utb (UVK).
- Gilbert, E.N. 1959. "Random Graphs," *Annals of Mathematical Statistics* 30/4:1141-4.
- Gondal, Neha, and Paul D. McLean. 2013. "What Makes a Network Go Round? Exploring the Structure of a Strong Component with Exponential Random Graph Models." *Social Networks* 35(4):499-513.
- Holland, Paul W., and Samuel Leinhardt. 1970. "A Method for Detecting Structure in Sociometric Data." *American Journal of Sociology* 76:492-513.
- Holland, Paul W., and Samuel Leinhardt. 1981. "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association* 76:33-50.
- Leifeld, Philip, and Skyler J. Cranmer. 2019. "A Theoretical and Empirical Comparison of the Temporal Exponential Random Graph Model and the Stochastic Actor-Oriented Model." *Network Science* 7(1):20-51.
- Liebertson, Stanley. 1985. *Making It Count*. Berkeley: University of California Press.
- Martin, John Levi. 2009. *Social Structures*. Princeton, NJ: Princeton University Press.
- Martin, John Levi. 2017. "The Structure of Node and Edge Generation in a Delusional Social Network." *Journal of Social Structure* 18(1):1-21.
- Martin, John Levi. 2018. *Thinking Through Statistics*. Chicago: University of Chicago Press.
- Nelder, J., and R. Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society, Series A* 135(3):370-84.
- Orsini, Chiara et al. [11 co-authors]. 2015. "Quantifying Randomness in Real Networks." *Nature Communications* 6:8627. <https://doi.org/10.1038/ncomms9627>
- Pattison, Philippa, and Tom Snijders. 2013. "Modeling Social Networks: Next Steps." In *Exponential Random Graph Models for Social Networks*, edited by Dean Lusher, Johan Koskinen and Garry Robins, 287-301. Cambridge: Cambridge University Press.

- Pattison, Philippa, and Stanley Wasserman. 1999. "Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations." *British Journal of Mathematical and Statistical Psychology* 52:169-93.
- Rawlings, Craig M., and Noah E. Friedkin. 2017. "The Structural Balance Theory of Sentiment Networks: Elaboration and Test." *American Journal of Sociology* 123(3):510-48.
- Robins, G., T. Snijders, P. Wang, M. Handcock, and P. Pattison. (2007). "Recent Developments in Exponential Random Graph (p^*) Models for Social Networks." *Social Networks* 29(2):192-215.
- Skvoretz, John, and Katherine Faust. 2002. "Relations, Species, and Network Structure." *Journal of Social Structure* 3(2). <https://www.cmu.edu/joss/content/articles/volume3/SkvoretzFaust.html>
- Snijders, Tom A.B. 2001. "The Statistical Evaluation of Social network Dynamics." *Sociological Methodology* 31:361-95.
- Stoehr, Julien, Alan Benson, and Nial Friel. 2018. "Noisy Hamiltonian Monte Carlo for Doubly-Intractable Distributions." arXiv:1706.10096v2.
- Strauss, David, and Michael Ikeda. 1990. "Pseudolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85:204-12.
- Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Wasserman, Stanley, and Philippa Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p^* ." *Psychometrika* 61:401-25.
- Watts, Duncan J. 1999. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton: Princeton University Press.
- White, Harrison C. 2000. "Parameterize!: Notes on Mathematical Modeling for Sociology." *Sociological Theory* 18(3):505-9.