PNAS Response – Winter 2024

The Authors

March 1, 2024

1 Review Team Comments

Editor Comments The paper is a very careful and elegant analysis of a new way of characterizing an easily measurable proxy for a structural network feature we know is important. I am recommending that we approve publication once the issues identified below have been addressed. Please accept my apologies for the delay in reaching a decision with respect to this manuscript. You can see the quite extensive and helpful comments of Reviewer 1 which have shaped my thinking, along with consultation with experts in areas of network science for which I am less expert. I want to thank you for the careful revision previously, and for your patience.

As Reviewer 1 nicely demonstrates with the clique removal example, inversity and assortativity are closely related but not identical ways of considering the interconnectedness of a static network topology. Is it possible that both of these measures ultimately boil down to the uniformity/skewness of density sampling for arbitrarily chosen vertex subsets, contrasting a focal individual against some scale of alternative induced subgraph density to characterize the evenness of the distribution of that ratio (i.e., focal individual vs induced subsample). If this is the case, then both are reasonable proxies for a marginally less accessible ”why” for the observed behaviors. Please look at work such as `https://doi.org/10.1073/pnas.1900219116` and cite as appropriate.

You set the local vs global (as a general point, this is really not global and it would be good to call it something else) measures as alternatives for assessing individual measure of responsibility for network-wide propagation of an infectious process and then saying that the global measure is entirely novel, even though it's just a relaxation of the ”local” measure. Further, assortativity is closely related to network modularity, and both assortativity and inversity are easily related to so called ”bridge nodes” (nodes that act as conduit connectors between modular communities). Modularity, bridge-node based methods, and related generalizations along these lines have been rather well studied already as a means of characterizing likely individuals for efficient spread/interruption, including consideration of local-only information. While most of these studies have been simulation based (Salathe and Jones 2010 and Gupta, Singh, and Cherifi 2016 both spring to mind), a few have also provided some analytic results. The implication here is to tone down the claims to novelty and connect more with cognate literatures.

The key issue from Reviewer 1 that needs to be addressed is: what kind of networks are you operating with? The Reviewer consideration of 3000 sociograms of different sorts which shows that while inversity is not the same as assortativity, it is very close (R2 0.97). begs the question of what is going on in figure S6 where you have cases where both are strongly positive - like about 0.5 where your series crosses over. As Reviewer 1 notes, these networks, whatever those networks look like, are not representative of the sorts of social networks we bump into regularly. To avoid building a metric around weird edge cases, please produce an inset sociogram or some such for figure s6 that makes clear what sort of structure generates the cases you find.

Please attend to the other issues raised by Reviewer 1 as well, which would take another couple of pages to summarize.

These changes aside, there is much to be enthusiastic about in this paper. Both reviewers agree that ”inversity” is a natural thing that has not yet been named or studied as its own thing. It is a simple metric that allows a slightly different understanding of how to evaluate mesoscale network structure using local information that seems quite valuable as a contribution on its own. It is always exciting to discover something, and I commend you.

Reviewer 1 Comments This revision has been informative and I appreciate that the authors have taken great pains to defend their measure. I have a great deal of respect for this authorship team; they are well-versed in the issues and clearly know what they are doing. One could probably not ask for a more qualified team to investigate this question. This is a case where asking for what I thought would be fairly trivial clarifications just seems to keep pulling up new questions. To some degree, that is a good thing: it means there is substantive, non- obvious meat here worth the attention of scholars digging into – so we should take it for what it is and let the field sort it out. We might recall, for example, that we spent years using the complex formulation for Burt's "effective size" measure (profitably), only to discover after Borgatti's 1997 paper that it is "merely" ego's degree minus the average of alter's degree (and similar simplifications for constraint). That simplification didn't undo the value of knowing what effective size was, nor undermine the theoretical logic that got us to it. My guess is that there's something similar here: you have discovered a property of the moments of friends-of-friends that taps into the imbalance of edgewise shared degree that is interesting and important, and it's captured well with the Pearson correlation of deg(i),1/deg(j). While I'm not convinced that something as mathematically idiosyncratic as the Pearson correlation, will be, in the end, the actual driving feature here – its just too arbitrary – I'm convinced that what's here is useful and not incorrect so worth moving forward on. That said...there are a couple of (please let them be..) minor points worth clarifying.

1. **Is inversity really scale invariant?** you state on line 211 that inversity is scale invariant, but x-axis of S6 is size – and since this is a network generation process defined by linking stars to cliques (by description, I can't figure out exactly what you did to replicate it), clearly the effect is strongly conditioned by the size of the network. Should that be troubling? Presumably this is a "real" feature of the underlying network – that the topology of this network depends on scale and the metric is merely reflecting that – but it's a bit disconcerting to see a nearly perfect relationship with size for a size-invariant measure, isn't it?

   Reply

   - By scale invariance, we mean a network property that remains the same *under unions of the same network*. Simply put, "doubling" a network by adding a copy of it would result in a larger network. By adding, we are taking the union of the set of all nodes and the union of the set of all edges across the networks. We note that while **mean degree is scale invariant** under this notion of scale invariance, **density is not scale invariant**. The number of potential invariant to unions of identical networks (graphs)node pairs increases when we combine two networks, so the density decreases when we add two identical networks.[1] **Inversity is then scale invariant under *this notion* of scale invariance.**

[1] This point is more readily apparent when wen consider the union of two identical cliques. Density is 1 for each clique, but not for the union of two cliques.

Nevertheless, we understand that even though it may be correct, this might be confusing to the reader, since scale invariance can be conceptualized in different ways. Since this point is not central to our argument, *we have removed this claim from the paper*.

- With regard to your point about the change in inversity with regard to size in the example network made up of one star and many cliques, we have explained in more detail why we should expect inversity to vary. V: Need to add this point below to the supplement. In Figure S6, it is not just the size of the network that is changing as we move along the x-axis, but the network topology or structure is changing. Recall that the network is comprised of one star and 100 cliques. Specifically, we fix both the number of cliques and the distribution of sizes of cliques, and *vary only the size of the single star network*. Thus, there are two aspects of the network that are changing. First is the size of the network. Second is the proportion of the network that belongs to the star component. Thus, this figure is not an appropriate one to infer anything about scale-invariance. If we were adding copies of the network to itself, then indeed inversity would not change.

2. **What is actually driving this metric and is there a substantive difference with assortativity?** Assortativity answers a natural question naturally – we want to know how similar ego is to alter? The correlation is fine. Could just as well use a standardized difference score, but correlation puts it in a nice metric. Nothing depends on the exact parameterization for the sorts of questions people ask of it.

   For inversity you flip it: how similar is ego to the inverse of alter? And 1/deg(j) is non-linear, so (at least to me) not intuitive. The main difference between corr(deg(i),deg(j)) and corr(deg(i),1/deg(j)) is easily seen in a scatterplot. Consider a generalization of your figure 1, panel (a) where the network is a collection of cliques of different sizes. Here ranging from dyads (collections of degree=1) to (arbitrarily) 15 (degree=14). The baseline scatterplot and scores are as at right. The assortativity scatter in blue using the left y axis, inversity in red indexed on the right y axis. So assortativity is perfect, but because 1/deg(j) is curved in deg(j), the score is not perfect (it's -0.76). Now that lack-of-fit is not really informative in the natural correlation sense – we shouldn't expect a linear association measure to capture a non-linear functional form.

   Now let's just delete some of those middling cliques so that we no longer have the full range and see what happens: The font is a little small (sorry about that) but the pattern is obvious: assortativity stays constant, and inversity ranges arbitrarily close to -1.0 depending on the mix of cliques included. This is, of course, implied by eq 4 – as the selection of cliques is affecting the moments of the degree distribution, which is what drives the score (and it must also drive assortativity, I'd think?). So is this a feature or a bug? Is there any substantive difference between these cases? The dominant feature is a collection of cliques. Inversity varies across the cases because a collection of 1/d cases contribute more of that curve the more complete the range of d included in the network. When there are only 2 classes of numbers (isolated dyads and a big clique), there are only two points on the plane, and the correlation is perfect. To get positive inversity, you need hubs and pendants. Then the scatter will have a lot of points at degree=1 (all the pendants), each with alters having degree > 1 which implies an inverse-degree is $\ll 1$ generating a blur of points at x=1 and y <small>. The hubs then have $x \gg 1$ and at least some mass at $y = 1$ (inverse degree of a pendant=1), creating a positive association.

   And again we can play the same sort of game where we limit the collection of stars to get arbitrarily close to -1.0, though unlike the clique case here we also get variance in assortativity, so that seems a reasonable (if uninteresting) mapping, but again the issue is that the underlying feature here is that the graph needs many star-like substructures to generate a positive inversity. My instinct is something like this might be useful to spell out clearly in the supplement (perhaps more so than S3, at least for me). Now in all of these cases, the assortativity and inversity scores are strongly negatively correlated. They are not constant – in part because (as we see in the example for cliques) inversity varies based on the collection of sizes included – so in all those cases assortativity is constant (1.0) and inversity is not. Again, you claim this as a feature and (since we rarely see collections of cliques) I'm willing to give the benefit of the doubt on that. I've run the scores over about 3000 real-life networks from different domains that I happen to have on hand, which includes coauthorship, adolescent networks, village networks and some sundry others. I get this:

So inversity is not the same as assortativity, but darn that's close (the R2 of a simple regression model is 0.97). Color here is data source, so a lot of variation across types of networks. And you can see, for example, that there are cases where both have the same sign – right there around zero. The difference between the local mean and the global mean generally gets smaller the closer you get to the origin point there...so effectively this is a noise region. In figure S6 you seem to have cases where both are strongly positive – like about 0.5 where your series crosses over. Whatever those networks look like, they are not representative of the sorts of social networks I bump into regularly. So what? I think you (reasonably) want to push the idea that these are fundamentally different critters to signal value to your method. I would just temper that – in the sorts of settings where you might want to do an intervention (the 75 villages data is in this set, for example) – they don't seem that different, and you don't want to build an argument around edge-cases. (That said, you should probably produce an inset sociogram or some such for figure s6 that makes clear what sort of structure generates the at-least-to-me odd cases you find).

Reply

Thank you for taking such a detailed and insightful investigation of how inversity varies (and assortativity does not) as we examine collections of cliques of different sizes. We have aimed to replicate this analysis in the Appendix 2 of this response letter in Figure 1. We begin with the collections of all cliques from size 2 (i.e., a dyad of 2 nodes) to size 15. We then vary the network by removing cliques as you have done. For each of the plots, we also include the network or sociogram corresponding to the plot. Briefly, we find the following:

(a) We begin by noting that for any sub-collection of cliques in this size range, *assortativity is always equal to 1*. However, inversity is not equal to -1 for any of the collections of cliques.

(b) When we have all cliques ranging from size 2 to size 15, we find that inversity $\rho = -0.77$. As we start removing cliques, we find that inversity becomes less negative and goes closer to zero. When removing only one clique, in panel (B), we observe inversity $\rho = -0.764$. Similarly, further removing cliques results in less negative inversity values (closer to 0). However, we note that this pattern is **not monotonic**. When we remove all cliques except the smallest (size 2) and the largest (size 15), then we find that inversity $\rho = -1$. We see this non-monotonicity as interesting and worthy of careful examination in a follow-up study.

(c) We would like to note that the values of inversity we have computed differ from those you have computed. We think the discrepancy might be because there are multiple edges in the network corresponding to a single point on the plot. For example, in the top left panel of Figure 1, corresponding to the plot point $(d_i, d_j) = (14, 14)$, there are 15 edges. However, at plot point $(d_i, d_j) = (1, 1)$, there is only one edge. If we ignore the number of such edges, we are able to obtain the inversity values you had indicated in your review report. However, to accurately compute inversity, we do need to account for each edge, since *inversity is defined across edges of the network*, i.e. $\rho = cor(d_i, 1/d_j)$, where the correlation is taken across the set of edges.

(d) In Appendix XYZ of this response letter, in Figure 2, we try to replicate the analysis you have conducted by varying the number of stars. We again include the sociograms corresponding to the plots. Do we want to highlight the differences between our plots and what the reviewer has produced? Not sure what is the purpose here.

(e) Comment: Reviewer 1 has directly asked us to include these figures in the supplement.

(f) Differences made evident by formula. As apparent from the formula $\rho = cor(d_i, 1/d_j)$, inversity is particularly sensitive to edges where one node has low degree and the other one has relatively high degree. In particular, we observe that inversity is most sensitive when you have nodes with degree of one tied to hubs or high degree nodes. This is also illustrated in Figure 1 of the paper. Assortativity is more sensitive to variance in the degree of the nodes that are connected.

(g) TO DO: Keep the size of the star constant. Multiply the size of the cliques, or even the range of the cliques. See how assortativity and inversity respond. DONE. Need to write about it.

(h) Need to caveat the graph of stars and cliques and say that this is a stylized example, and may not be representative of real world networks. However, we have the India village networks where 3 out of 44 villages have positive values of inversity. However, the reviewer's figure shows this too.

(i) We need to make or repeat the point that the magnitude of differences might be high. Even small errors in estimating inversity can result in much greater differences in local means. Pick some villages in individual network india villages that has big differences in local mean / proxy local mean, but smaller differences in -assortativity / inversity. Show these villages in red and show how the local mean is different.

(j) We need to respond directly to R1's claim of real world networks where $R^2 = 0.97$.

3. Some details on the text I noted as reading.

   (a) the last sentence of the abstract is redundant with the stuff above, strike one or the other.

   - **Reply:** Agree. We have updated the abstract to avoid this redundancy.

   (b) I would switch early from "friends" to "neighbors" and use that consistently throughout, fine to reference "friendship paradox" but this holds for non-affective relations (it holds generally for graphs), so go with the more general features (I just saw a Scientific Reports paper, for example, that made hay out of the fact that it holds for negative ties. Go figure). So, instead of "friends of friends" use "two-step neighbors" or "neighbors' neighbors." *Note:* there's a natural confusion here in addition to the two forms you layout. I think most people will not think of i naturally as being a neighbor's neighbor. Of course they are, but friends-of-friends are not usually "me." This is clear in figure S1, but perhaps not in the text at this point.

   - **Reply:** We agree that "neighbors" would be a more general and natural usage, given the wider applicability. We have used neighbor beyond the first paragraph of the main paper, but have retained the well-known term "friendship paradox," rather than use a new term for this phenomenon. We have also added a note about a node being a neighbor's neighbor to make the point above more obvious. Please see the footnote at the bottom of page 2 of the main paper.

(c) Line 141 .. always holds for *non-regular* networks (I think that's what you mean by "weakly" but I think being clear about what class of graphs is better.

- **Reply:** We agree, and have made the change suggested.

(d) I'm not sure what the phrase "prove that the sign ... helps us determine" how do you prove something "helps"?

- In my sample, if global $>$ local, then inversity is always negative. If local$>$global, there is also a small subset where inversity is negative. So perhaps "helps" means "is not quite certain"?
- **Reply:** We have simplified it to: "prove that the sign ... determines" since the value of inversity directly indicates which intervention is better. If local $>$ global, then inversity ρ **would always be positive**, based on equation (4), reproduced below. Similarly, if global $>$ local, inversity $\rho < 0$ always. Note that the function Ψ is always positive.

$$\mu_L = \mu_G + \rho\, \Psi(\kappa_{-1}, \kappa_1, \kappa_2, \kappa_3) \tag{1}$$

(e) Line 217, I think you're missing an "is" somewhere there.

- **Reply:** We have fixed this, by adding an "and" to the sentence.

(f) Line 249 – this sort of place replace friends with neighbors.

- **Reply:** Agree, we have done what you suggest.

(g) Is there a better way to say "depends on the structure of connections (who is connected to whom)" which you repeat throughout? You are trying to distinguish between things that are just a function of the degree distributions and those that actually depend on the edgelist. I'd use "topology" or "structure". Perhaps just note it first and don't repeat the "who with whom" parts.

- **Reply:** We agree *network topology* might be a useful way to indicate this and reduce this repetition. We only use the longer form when introducing this term.

(h) Line 351; this first point seems redundant with 291.

- **Reply:** We think it would help the reader to highlight the contrast, at the risk of mild repetition. Please see the revised sentence based on this suggestion.

(i) Again, stuff around line 392 is built on, seemingly, pretty weird networks. In most real-life cases, they are very similar. Ditto with the rank ordering – my sense is that when this is true, the networks either don't look much like social networks we typically collect or are effectively equal.

- **Reply:** We have added a caveat in the main paper that it would be an important thing for future research to examine the relationship between inversity and assortativity or other related variables. In the present revision, we have tried to make it absolutely clear and transparent exactly what kinds of networks generate the results observed, where both inversity and assortativity have the same positive sign. We have revised the figure (also included in this response letter as Figure 5) to include inset sociograms, and have simplified the network generating process so that the set of cliques is always from size 2 to size 20. We observe that the networks in inset (B) and (C) have networks that have both positive inversity and assortativity. The network in (B) is a star of size 30 with cliques of size 2 to 20 (all such cliques are included and each is included exactly once). For this network

B, $\rho(B) = 0.027$ and $\rho_a(B) = 0.553$. Similarly, for network (C), we have $\rho(C) = 0.296$ and $\rho_a(C) = 0.228$, where (C) has a star of size 39, and the cliques are the same as (B). We agree that including the sociograms gives the reader more insight into understanding the kinds of network structures that produce the results obtained.

(j) Line 504: you can't, in practice, exclude isolates if you don't know who is not connected.

- **Reply:** We agree, excluding isolates was actually not required in the implementation of the seeding strategies. We draw a sample from the set of nodes, which *can include isolates*, and have modified the text in the paper to reflect this.

(k) Do you need a fixed proportion? Again, thinking practically, most people won't be very good at estimating their number of friends. So asking them to nominate a proportion of their peers is going to be hard.

- **Reply:** We have altered this to indicate a specified probability of nominating each friend, rather than a fixed proportion. We note that many of these steps can be automated with programs that can undertake these steps which reduce the need for human memory. However, such methods may not be suitable for manual data gathering. We agree with your point that the ego-based strategy would be easier to implement than the alter-based strategy, and we would like to explore easier and more effective methods to *implement* these strategies in follow-up research.

Reviewer 2 Comments Having reviewed the response to reviewers and the updated manuscript, I think the authors have done a good job in responding to the issues raised (and, particularly, why assortativity and inversity are related but distinct measures). I think the manuscript makes a very interesting contribution to the field, and the inversity measure is likely to trigger new work; I support publication.

Response: *We thank Reviewer 2 for the supportive comments at this stage. We look forward to seeing it in print.*

---

2 Inversity and Assortativity in Cliques and Stars

We have tried to recreate some of the plots Reviewer 1 had included in their comments. These plots are detailed in Figure 1 below.

3 Assortativity and Inversity in Real Networks – India Villages

4 Revised Plot of How Inversity and Assortativity vary with Stars and Cliques

4.1 Variation with size of Star Network

We have revised the plot of how inversity and assortativity change as the size of the star network is varied, keeping the cliques constant. To make this deterministic, we have fixed the cliques to include all cliques from size 2 to size 20. This plot also includes inset sociograms (A)-(D) at specific points of interest.

In each network, we include all cliques from size 2 to size 20, where the clique size refers to the number of nodes. Observe that the cliques (in red) are kept exactly the same, and only the size of the star network (in blue) is varied from 2 to 61.

4.2 Variation with Cliques

We now retain a single star network (of size 30), and vary the number of cliques, with sizes from 2 to 20.

Figure 1: Networks with varying number of cliques

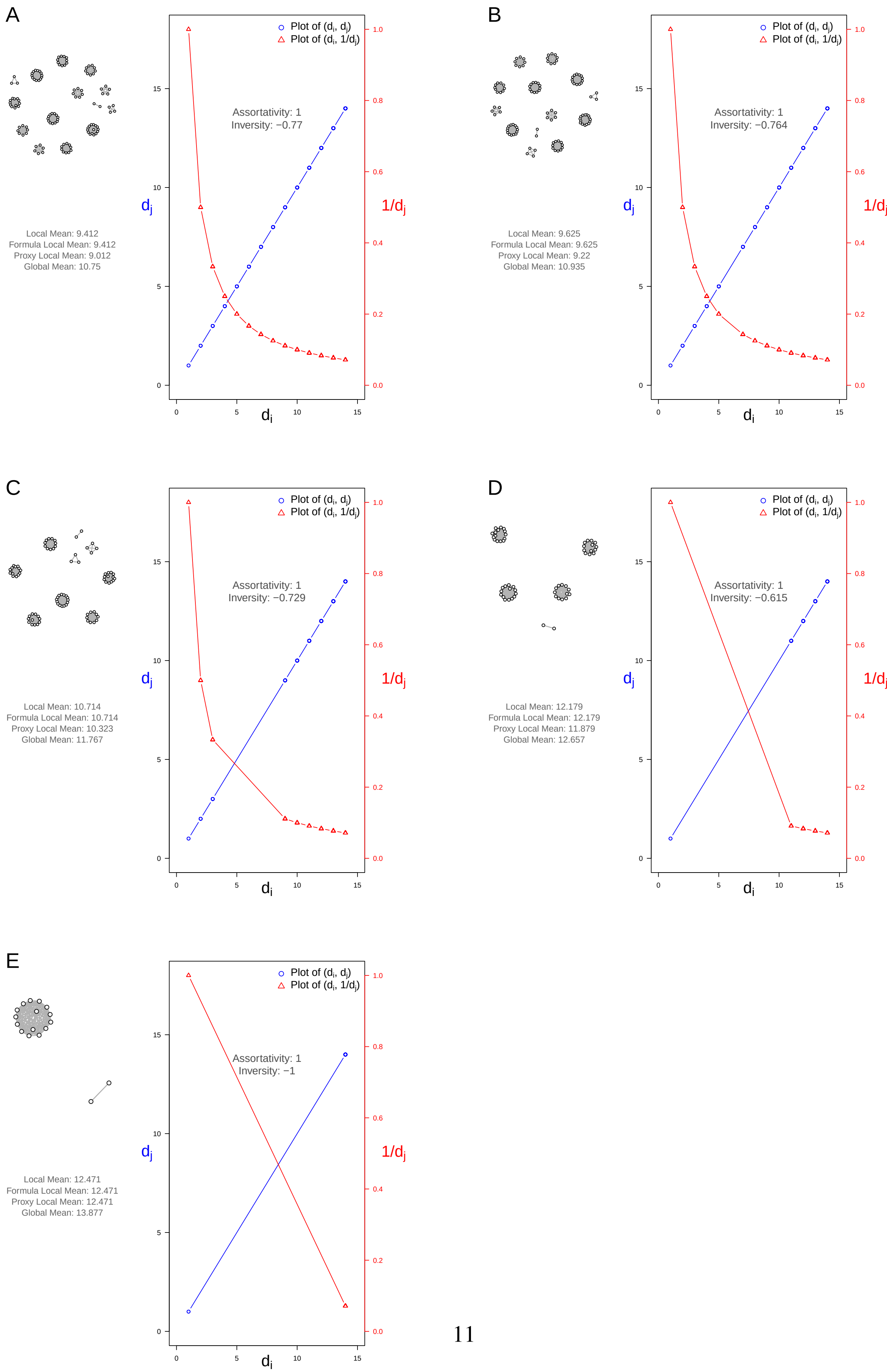


Figure 2: Networks with varying number of stars

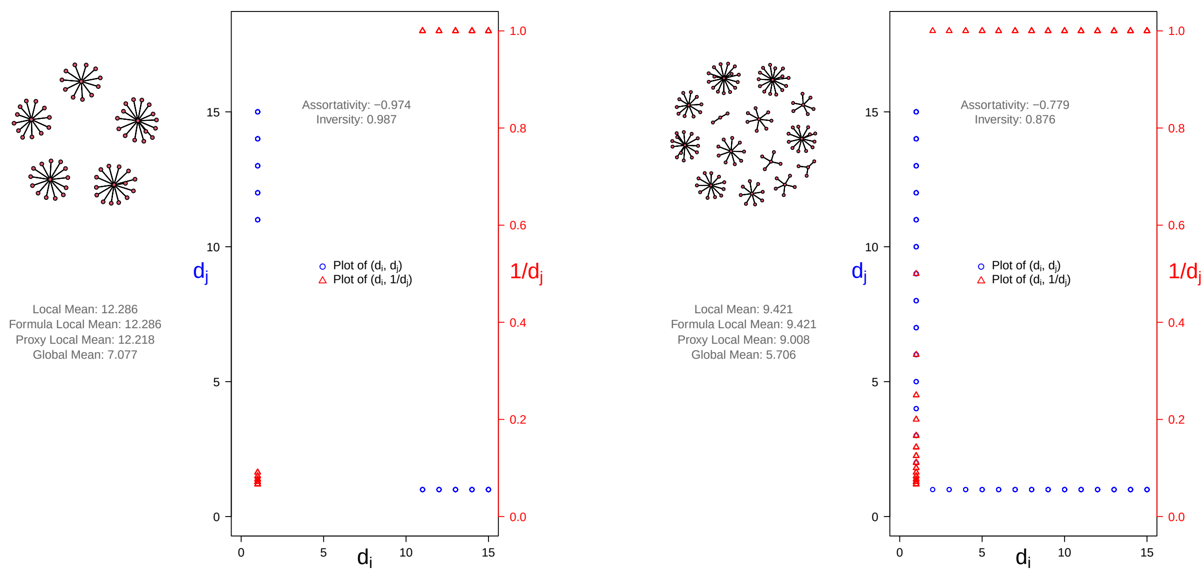
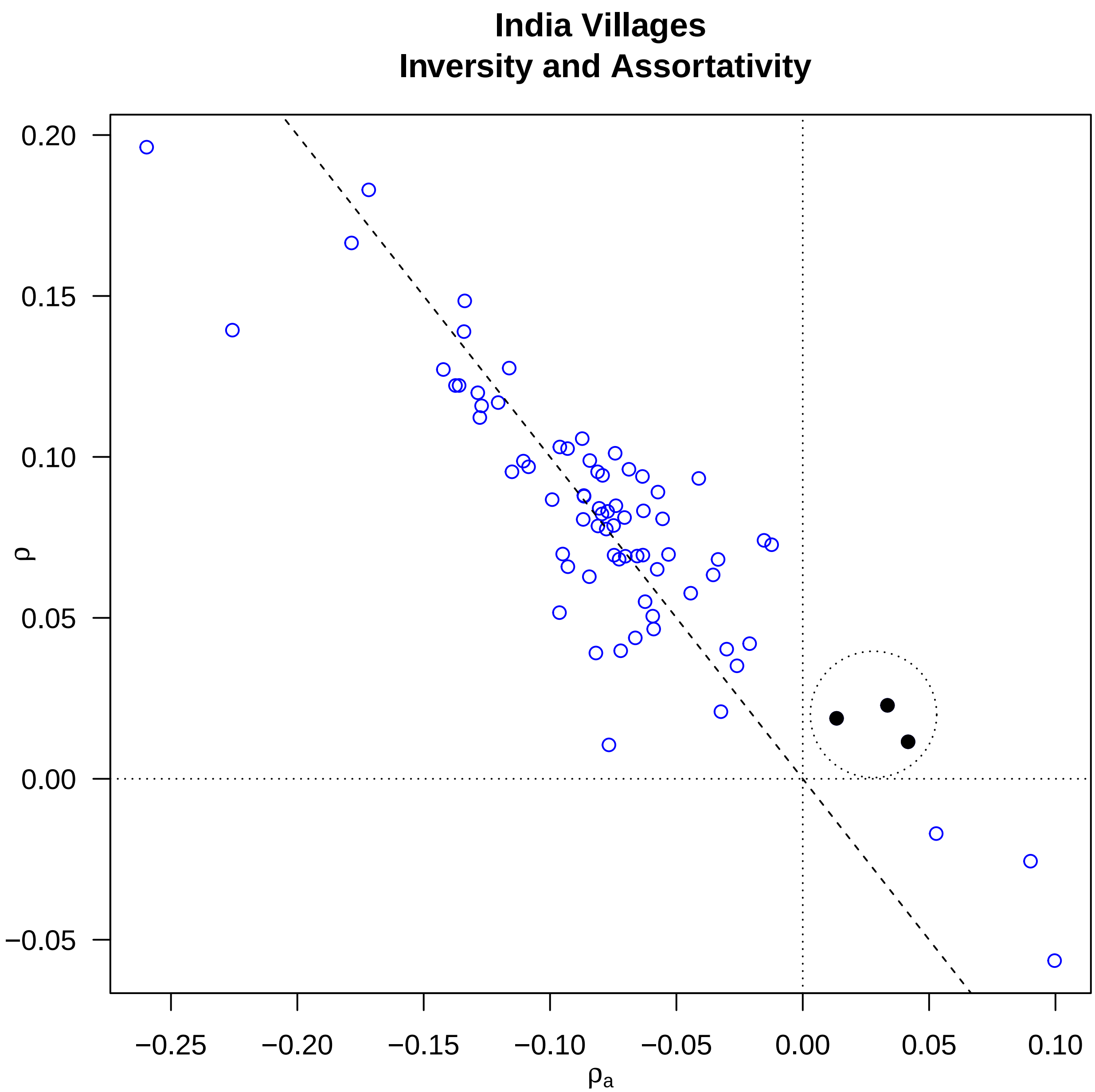
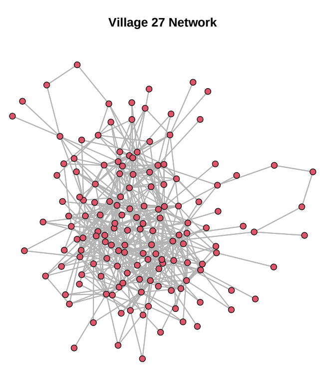


Figure 3: India Village Household Networks – Assortativity and Inversity





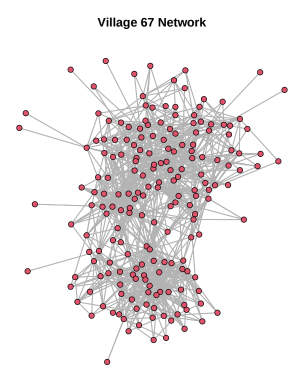
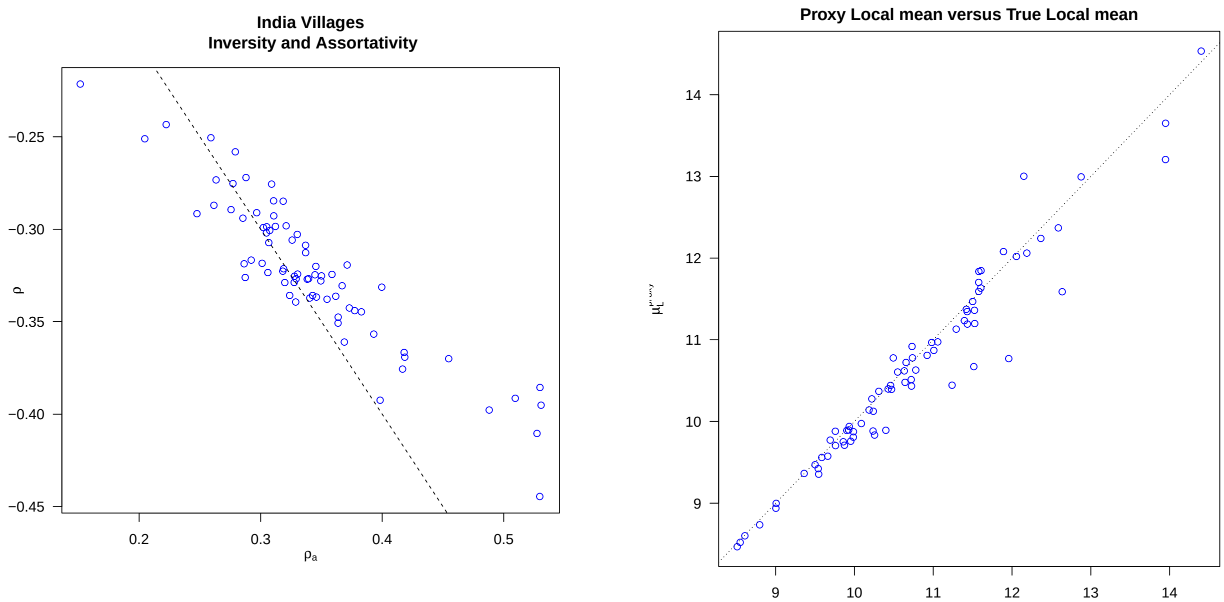


Figure 4: India Village Individual Networks – Assortativity and Inversity



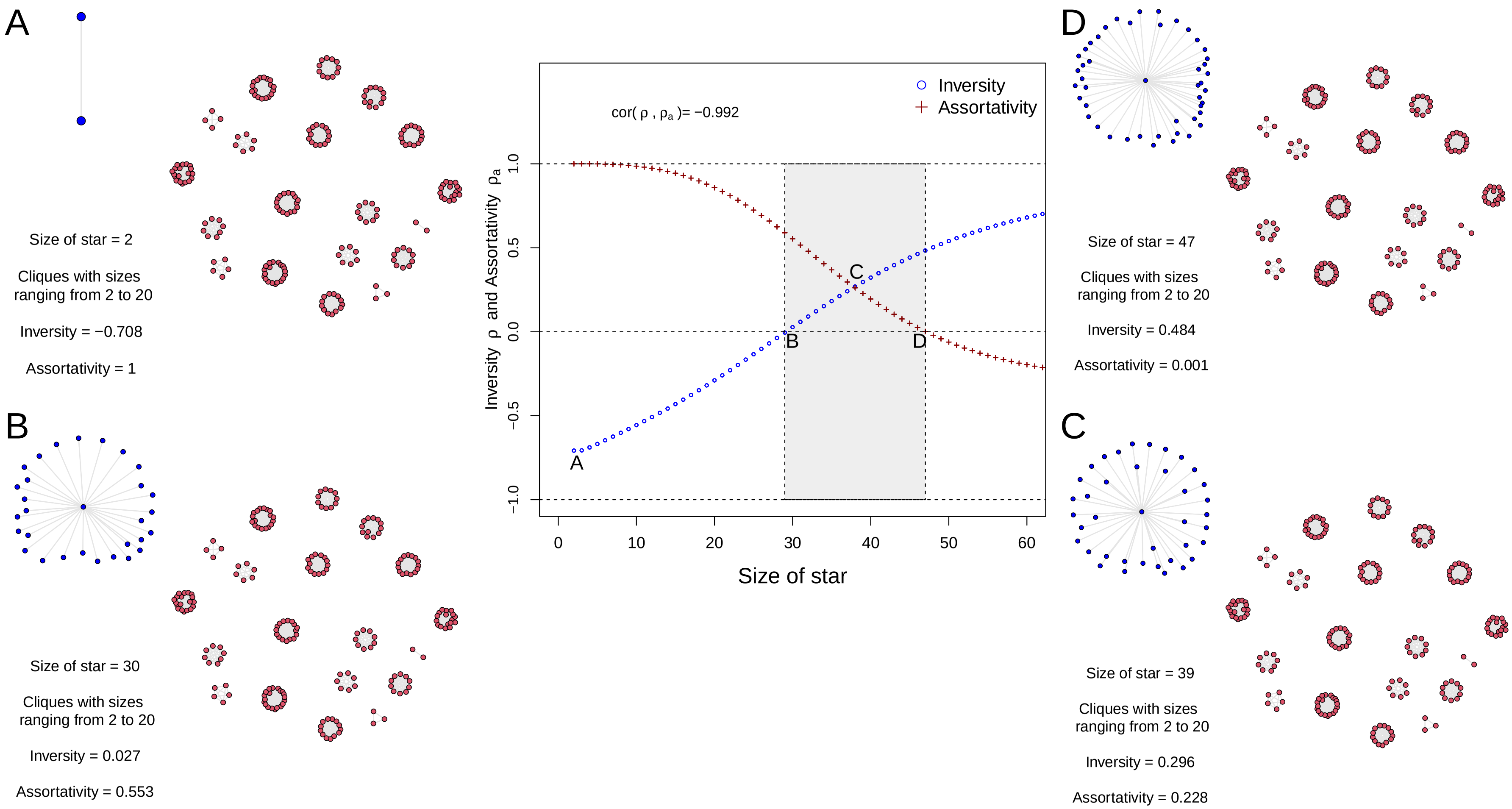


Figure 5: Variation of Inversity with Star Size

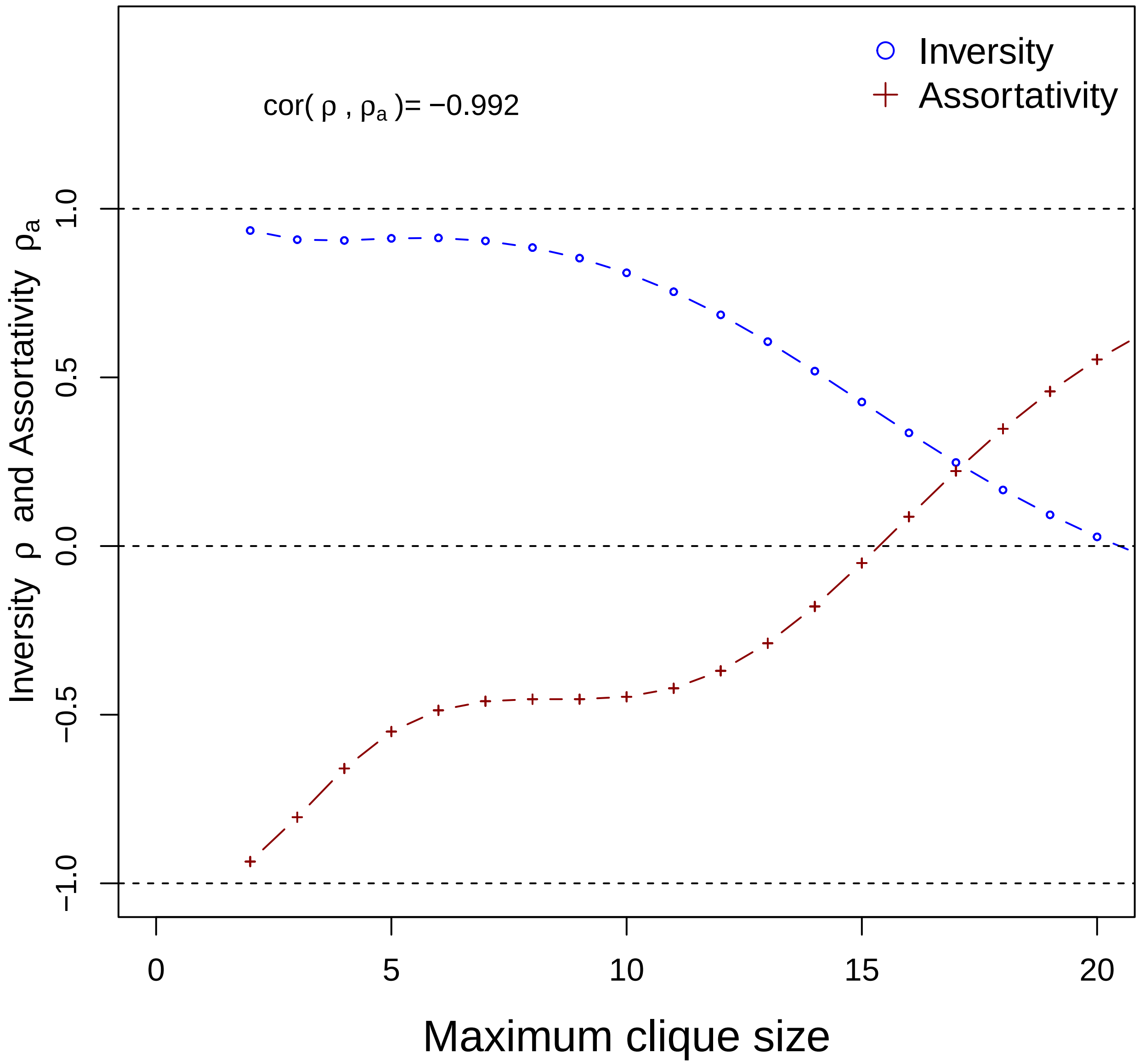


Figure 6: Variation of Inversity and Assortativity with Cliques