Check for updates

Zero Re-centered Projection: An Alternative Proposal for Modeling Empty Networks in ENA

David Williamson Shaffer, Brendan Eagan(✉), Mariah Knowles, Clara Porter, and Zhiqiang Cai

University of Wisconsin–Madison, Madison, WI 53711, USA
{beagan, zhiqiang.cai}@wisc.edu

Abstract. This paper examines the impact of having empty networks in an Epistemic Network Analysis model, that is, units whose networks contain no connections in a given model. These empty networks, also known as zero points, can negatively impact the interpretive validity of Epistemic Network Analysis spaces. In this study, we explore a change in the underlying mathematics and algorithm of Epistemic Network Analysis that we argue will make models easier to interpret accurately.

Keywords: Epistemic network analysis · ENA · Projections

1 Introduction

Many epistemic network analysis (ENA) models contain units (people, groups, or other items of interest) that have no connection between codes in the model. There is nothing particularly unusual or inherently problematic about having such *zero points*, or units in model that have no information on the variables of interest. However, in ENA models, empty networks can cause an interpretive problem in two ways: (1) the position of empty networks in an ENA model can be difficult to interpret, and as a result, (2) they can make interpretations of the overall model less reliable.

In this paper, we look at a change in the ENA mathematics and algorithm, and we argue that this change will make models easier to interpret accurately.

2 Background

2.1 Zero Values

In many analysis techniques, researchers can encounter units of analysis that have zero values for all of the variables in a model: for example, students who score a zero on a test—or perhaps a zero on every test—or an assembly line that has zero stoppages during the time of observation.

Depending on the statistical method being used, the nature of the question, and why points might have zero values in the first place, researchers deal with zero points in

B. Wasson and S. Zörgő (Eds.): ICQE 2021, CCIS 1522, pp. 66–79, 2022.
https://doi.org/10.1007/978-3-030-93859-8_5

different ways. For example, Fayaz explains that many statistical methods, including regression, involve dropping units that don't have any connections from the data if translations or transformations are not employed to handle the zero points [2]. Others approaches may actually introduce units with all zero values into their data to solve other problems. For example, attitude analyses sometimes purposely create a zero point provide reference or context for other points on the scale [3]. Scott [4] argues that in *social network analysis* (SNA), data points should be translated (that is, have a constant value added) such that units with no connections become outliers in a visualization. Some regression analyses place zero-valued units in the middle of the range of possible values for the variable or variables; others perform non-linear transformations (for example, log transforms) to account for zero values.

Because there are many possible ways of accounting for zero values, any analysis technique—and any analyses using that technique—has to provide a justification for the particular choice or choices made.

2.2 Zero Values in ENA

To date, the quantitative ethnography community—and users of ENA more widely—have not addressed this question (or provided such justifications) in detail.

To address this issue, we first ask: Do zero points actually present a problem for ENA analyses? After answering that question with a definitive *yes*, we then examine why, and provide a proposal for addressing the issue in a more theoretically sound way.

Wherefore Art Thou Romeo? Consider, for example, an analysis of two Shakespeare plays, *Romeo and Juliet* and *Hamlet* (taken from [1]).

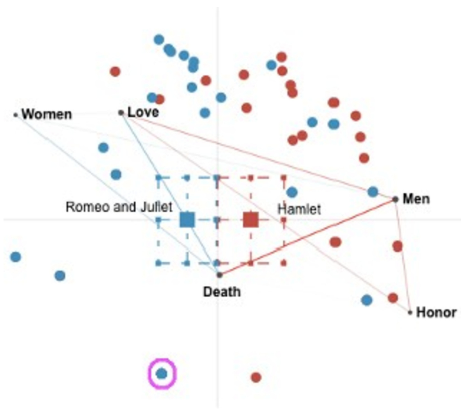

Fig. 1. ENA model comparing characters from *Hamlet* and *Romeo and Juliet*

In this example:

Data were the Lines of the plays
Conversations were each scene of each act of each play
Units of analysis were each character in each play
Codes were men, women, honor, love, and death

The stanza window was four lines [5], with a binary model. The data was normalized and the model used a means rotation [6] with Characters from *Hamlet* and Characters from *Romeo and Juliet* as the two groups.

The resulting model is shown in Fig. 1.

Perhaps not surprisingly, there are a number of characters in the plays who do not refer at any point to men, women, honor, love, or death (or, more specifically, do not make connections between these codes). The location of these characters in the ENA space is indicated by the pink circle.

This location of the zero points is difficult to interpret. In what sense are characters who say nothing (or nothing relevant to the model) more like characters in *Romeo and Juliet* than *Hamlet*? Moreover, an ordinary reading of the ENA space would indicate that these characters would be making more connections to death than to love—being low in the ENA space where death is rather than high in the space where love is. But this makes little sense, as by definition these characters make no connections between any codes!

Where Truth is Hid, Though it Were Hid, Indeed, /Within the Center. We thus propose that from an *interpretive* point of view, a more logical position for the zero points would be to place them at the origin of the graph, as shown in the right diagram of Fig. 2. Now these characters are neither more similar to *Romeo and Juliet* or *Hamlet*, and they are not associated with any of the codes in the model more than any others—which makes sense as they are associated equally to all with a strength of zero.

Notice that when this happens, the model changes: that is, the plotted points, node positions, means, and confidence intervals are all different.

We refer to this method as *zero re-centered projection*, and in the following sections, we will attempt explain why the current modeling approach leads to this interpretive problem, how zero re-centered projection addresses the issue, and why resolving the problem in this way causes the model to change.

2.3 Interpretation of ENA Objects

In ENA, information about each unit of analysis, i is represented in four distinct but coordinated ways:

1. as an *adjacency matrix*, A_i, showing the cumulative connections for each pair of codes for that unit;
2. as a (usually normalized) *adjacency vector*, $\boldsymbol{v}_i$, that represents the location of the adjacency matrix, A_i in a high dimensional space, where each dimension represents the (normalized) connections between a pair of codes in the adjacency matrix;

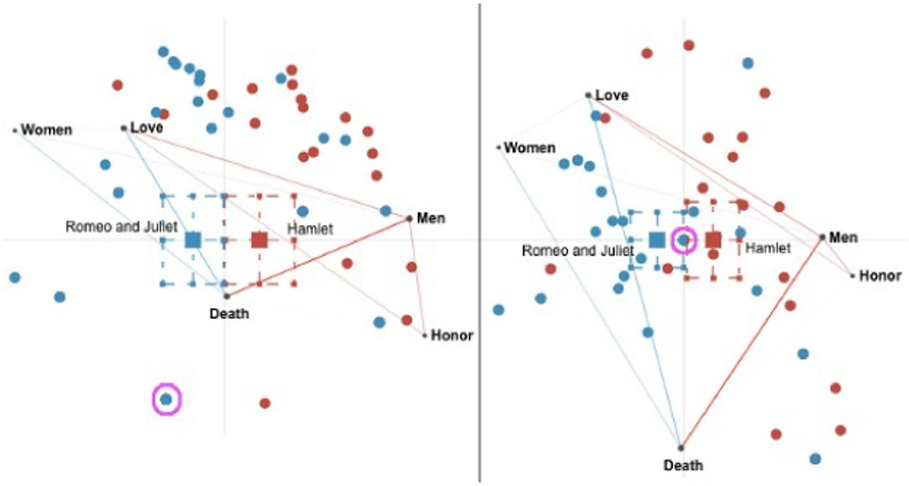

Fig. 2. Zero re-centered ENA model comparing characters from *Hamlet* and *Romeo and Juliet* (right) compared to the same model using the current ENA projection (left)

3. as a *plotted point*, p_i that represents the location of the adjacency vector $\boldsymbol{v}_i$ under a projection from the space of adjacency vectors; and
4. as a *network graph*, G_i, whose edges represent the values in the adjacency vector, A_i, and whose centroid approximates the location of the plotted point

A key feature of ENA is that the locations of a plotted point, p_i, can be *interpreted* in terms of its network graphs, G_i, in the following sense:

> *Points* that are located up/down/right/left in the space, have *network graphs* that have strong connections in the up/down/right/left part of the space.

So, for example, the characters with red plotted points in Fig. 1 are more toward the right side of the diagrams because they have network graphs with strong connections between codes on the right side of the diagrams—which is what the red lines on the network graphs indicate. Characters with blue plotted points in Fig. 1 are more toward the left side of the diagrams because they have network graphs with strong connections between codes on the left side of the diagrams—which is what the blue lines on the network graphs indicate. Thus, we interpret the location of the points in terms of the kinds of connections that their networks contain. In Fig. 1, we can thus interpret the differences between the plays in the sense that both plays are about death, but characters in *Romeo and Juliet* make more connections between death, women, and love, whereas characters in *Hamlet* associate death more with themes of men and honor.

2.4 Coordination of ENA Objects

These interpretations are made possible because the different representations (adjacency matrices A_i, adjacency vectors $\boldsymbol{v}_i$, plotted points p_i, and network graphs G_i) are *coordinated* in the following very specific sense:

> The *nodes* of the network graphs are positioned so that for any unit, i in the model, the *centroid* c_i of its network graph G_i approximates the position of the associated plotted point p_i.

This relationship between plotted points, p_i, and centroids, c_i, is illustrated in Fig. 3. In the figure, the centroid of G_1 is at the center of the graph, because all of the edges have the same weight. As weight is added (G_2) or subtracted (G_3) from edges in the graph, the centroids move toward the heavier edges and away from the lighter edges.

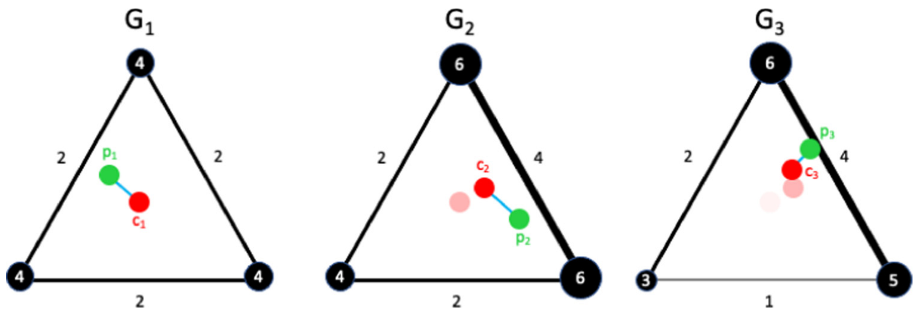

Fig. 3. Three network graphs, showing the relationship among node weights (white numbers in black circles), connection strength (numbers along graph edges), centroids ($c1$–$c3$ in red), and plotted points ($p1$–$p3$ in green). Distance between the plotted point and centroid for each graph is shown with a green line. Larger distances between centroids and plotted points make interpretations of plotted points in terms of network graphs less reliable. (Color figure online)

The centroid for a network graph is computed as follows. First, compute the total weight, w_j, for each node, N_j as the sum of the weights of the edges connected to it. (This is shown in Fig. 3) Then, for the x coordinate of c_i, X_{c_i} is computed as the weighted average of the x coordinates, x_j, of the nodes n_j:

$$X_{c_i} = \frac{x_j w_j}{w_j}$$

Y_{c_i} is computed similarly.

Thus, if the plotted point, p_i for some adjacency matrix, A_i, is close to its centroid, c_i, then the edge weights in its corresponding network graph, G_i, will allow us to interpret the position of its plotted point in terms of the connections between nodes.

In turn, if *in general* plotted points and centroids are aligned in this way, then the positions of the nodes make it possible to interpret the *dimensions* of the plotted points in terms of the *positions* of the nodes. This is referred to as the *co-registration* of plotted points, p_i, and network graphs, G_i, and it is a critical—indeed, crucial—step in interpreting ENA models.

2.5 More Mischance/On Plots and Errors Happen

Of course, plotted points and centroids are never completely coincident, as illustrated by the green lines in Fig. 3. This discrepancy can be quantified for each dimension by correlating the coordinates of the plotted points, p_i, and centroids, c_i. Thus, if X_c and X_p are vectors of the x coordinates of the centroids and plotted points respectively, we compute the Pearson correlation as:

$$\text{X axis goodness of fit} = r_x = \frac{k \frac{Xc_k - \overline{X}_c}{s_c} \quad \frac{X_{p_k} - \overline{X}_p}{s_p}}{n - 1}$$

where n is the total number of points and s_c and s_p are the standard errors of X_c and X_p respectively. Goodness of fit for the y axis is computed similarly. Other methods of correlation can be used as well, and ENA automatically computes both Pearson and Spearman correlations.

Correlation is an appropriate measure of goodness of fit because network graphs have no inherent scale relative to the values of plotted points. Thus, correlation provides a scale-invariant (or scaleless) measure of the degree of coregistration of a set of plotted points, p_i, and network graphs, G_i.

As a result, the goodness of fit quantifies *the degree to which the dimensions of the plotted points can be accurately interpreted by the network graphs*—and thus whether the interpretations of differences between plotted points are valid.

2.6 A Mote It is to Trouble the Mind's Eye

And this brings us to the heart of the matter. In general, it is important to have high correlations between plotted points and centroids to warrant reliable interpretations of the coordinates of the plotted points—and generally desirable to have correlations $r \geq 0.9$. However, has shown in Table 1 in the analysis of *Hamlet* and *Romeo and Juliet* above, the original correlations (using the usual ENA positioning of zero points) are below this desirable threshold, and well below it in the case of the y axis.

Table 1. Correlations between plotted points and network centroids.

Axis of interest	Original model correlation	Zero re-centered correlation
X	0.88	0.93
Y	0.75	0.94

On the other hand, Table 1 shows that if we position the zero points at the origin, as in the right diagram of Fig. 1, the goodness of fit measure for this data is higher, and more specifically above 0.9 on both dimensions.

Thus, the problem with positioning zero points in the usual manner is two-fold:

1. The location of zero points, or the points corresponding to empty networks, is difficult to interpret; and
2. The interpretation of the positions of *all of the other plotted points* in the model becomes less reliable

In fact, it was precisely the problem of understanding why some models had low goodness of fit (sometimes far worse than the example above) that led us to investigate the mathematical implications of zero points in ENA models.

We turn to that issue in the next section of the paper.

3 Zero's Fault

3.1 Projecting Zero

To understand this fault or flaw—that is, why empty networks can cause interpretive problems—recall that each adjacency matrix, A_i, in the model shows the cumulative connections for each pair of codes for that unit. Each adjacency matrix, A_i is represented as an adjacency vector, $\boldsymbol{v}_i$, that represents the location of the adjacency matrix, A_i in a high dimensional space, where each dimension represents the (normalized) connections between a pair of codes in the adjacency matrix.

The set of adjacency vectors, $\boldsymbol{v}_i$, is then projected into a 2-dimensional plane by centering the vectors on their mean, $\overline{\boldsymbol{v}}$ and multiplying by a rotation matrix, R. Different projections (for example, means rotation or singular value decomposition) use different rotation matrices, but those differences are not important here.

Thus, we compute each p_i as:

$$p_i = R \times (\boldsymbol{v}_i - \overline{\boldsymbol{v}})$$

This is useful because it projects the mean of the vectors $\boldsymbol{v}_i$ to the origin of the ENA space:

$$R \times (\overline{\boldsymbol{v}} - \overline{\boldsymbol{v}}) = R \times 0 = 0$$

However, because all of the connections in an empty network, A_0, are zero, all of the values of $\boldsymbol{v}_0$ will be zero. And as a result:

$$p_0 = R \times (0 - \overline{\boldsymbol{v}})$$

That is, empty networks appear as the projection of $-\overline{\boldsymbol{v}}$ in the original space.

Thus, empty networks are not projected to the origin of the ENA space unless the mean of the vectors is zero: that is, $\overline{\boldsymbol{v}} = 0$. But that is essentially impossible, because all of the connection counts in the adjacency matrices A_i are either zero or positive. As a result, there are no negative values in any adjacency vector, $\boldsymbol{v}_i$. So unless all $\boldsymbol{v}_i = \boldsymbol{v}_0$, $\overline{\boldsymbol{v}} = 0$.

The empty networks will only be projected to the origin of the ENA space if all of the networks are empty!

3.2 Zero's Centroid

The fact that empty networks are not projected to the origin by itself creates the interpretive problem described above: how should we interpret the position of any plotted point, p_0, from an empty network. But, perhaps more important, it also causes the overall goodness of fit for the model to drop.

To see why, recall that we compute the centroid, c_i of a network graph, G_i as:

$$X_{c_i} = \frac{x_j w_j}{w_j}$$

However, because in an empty network, A_0, all of the node weights, $w_j = 0$, the location of the centroid, c_0 will be:

$$X_{c_0} = \frac{x_j w_j}{w_j} = \frac{x_j \times 0}{w_j} = 0$$

That is, the centroid of an empty network will always be at the origin.

In other words, while the projected point, p_0, for an empty network will *never* be at the origin, its centroid, c_0 will *always* be at the origin.

As a result, using ENA's usual projection method:

1. The presence of a an empty network will always reduce the goodness of fit—that is, the interpretive reliability—of an ENA model; and, more to the point,
2. The presence of many empty networks in a data set has the potential to significantly reduce the goodness of fit of a model.

4 Zero Re-centering

The result of this unavoidable mathematical mismatch between p_0 and c_0—that is, the projected location of empty networks and the centroid of empty networks—is that the location of empty networks using ENA's current projection method will always be difficult to interpret, and models with many empty networks will be difficult interpret reliably overall.

In other words, we argue that ENA's projection method should be revised so as to deal with zero points in a way that will not compromise interpretive validity.

Given that the centroid of empty networks, c_0, will always be at the origin, it seems unavoidable that any solution to this problem would involve a projection where the projected point for empty networks, p_0, should also be at the origin. In this section of the paper, we examine two alternatives: un-centered projection and zero re-centered projection.

4.1 Un-centered Projection

Recall that ENA's current projection is of the form:

$$p_i = R \times (\boldsymbol{v}_i - \overline{\boldsymbol{v}})$$

That is, the set of adjacency vectors, $\boldsymbol{v}_i$, is projected into a 2-dimensional plane by centering the vectors on their mean, $\overline{\boldsymbol{v}}$ and multiplying by a rotation matrix, R.

If instead we compute:

$$p_i = R \times (\boldsymbol{v}_i)$$

that is, we do not center the vectors on their mean, then instead of projecting the *mean* to the origin of the ENA space, the empty networks are placed at the origin, as shown in Fig. 4.

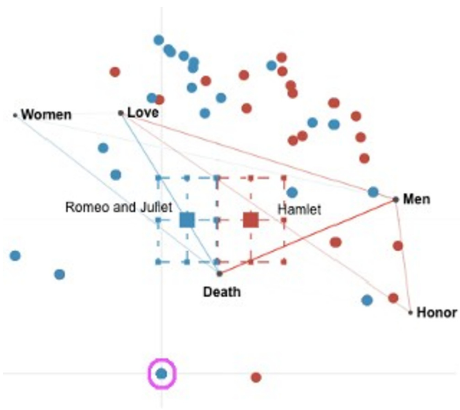

Fig. 4. ENA model of *Hamlet* and *Romeo and Juliet* using an un-centered projection

Because ENA uses a linear projection, this ENA model is essentially equivalent to the original (Fig. 1), but shifted so that the zero points are at the origin of the ENA space rather then the mean.

This approach addresses the problem that empty networks create for goodness of fit because using an un-centered projection, $p_0 = c_0 = 0$. However, the positions of the nodes now show that the y axis is capturing almost exclusively the difference between the characters with empty networks and all of the other characters: most of the plotted points are at the top of the graph (high y values), with many zero points at the origin and a few points in between.

This highlights a critical feature of many ENA models: namely, that this model uses *normalized adjacency vectors*.

The question of normalization arises because two units of analysis (characters in the ENA models here) can have the same *patterns* of discourse, but one might have more data than the others. To see why, imagine that two characters in a play had exactly the same coded lines, but one repeated those same coded lines 50 times each. Both adjacency vectors would go in the same direction from the origin in the space of adjacency vectors, $\boldsymbol{v}_i$, but the character with 50 times as many lines would have a vector 50 times as long.

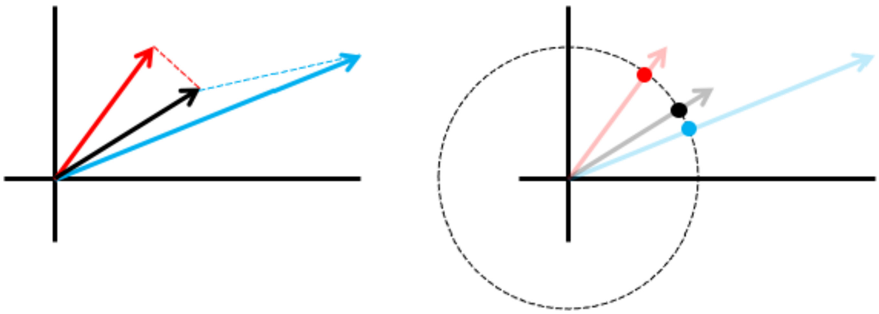

Fig. 5. Illustration of the process of normalizing ENA data

Figure 5 illustrates this idea. The black vector points in a similar direction to the blue vector, but is closer in length to the red vector. Without normalization (left image), the black vector appears more similar (closer) to the red vector than the blue. When the vectors are normalized to have the same length (right image), the black vector is more similar to the blue vector.

Thus, normalization accounts for differences in the total number of coded lines of data for each unit in a model.

The process of normalization has been described elsewhere [1], but briefly, normalization transforms the original adjacency vectors, $\boldsymbol{v}_i$, into normalized vectors $\boldsymbol{v}_i^* = \frac{\boldsymbol{v}_i}{|\boldsymbol{v}_i|}$, which are then used in the projection:

$$p_i = R \times \left(\boldsymbol{v}_i^* - \overline{\boldsymbol{v}}^*\right)$$

That is, each adjacency vector is divided by its length, such that all of the vectors have length $\left|\boldsymbol{v}_i^*\right| = 1$. Or almost all the vectors. Because zero vectors have no length, their normalized value is:

$$\boldsymbol{v}_0^* = \frac{\boldsymbol{v}_0}{|\boldsymbol{v}_0|} = \frac{0}{0}$$

which is undefined.

That is, zero points in a normalized plot have no inherent position in the normalized vector space. The current ENA projection (and thus an un-centered projection) place zero points at the origin of the space of normalized vectors, $\boldsymbol{v}_i^*$. But that position—in an ENA model with normalized data—is, in fact, arbitrary.

4.2 Zero Re-centered Projection

Because the placement of the zero points is undefined, we thus propose that rather than placing the zero points at the origin, we place the zero points at the mean of the non-zero points in the vector space. That is:

$$\boldsymbol{v}_0^* = \frac{\boldsymbol{v}_i^* \forall \boldsymbol{v}_i^* \boldsymbol{I} = 0}{|\boldsymbol{v}_i^* \forall \boldsymbol{v}_i^* \; 0|}$$

The resulting projection will appear somewhat different from the current ENA projection because now $\boldsymbol{v}^* \boldsymbol{I}_0 = 0$—that is, the zero points are in a different location, so the projection will change.

However, as shown in Fig. 2, this produces a model where the zero points are easy to interpret sensibly: they have no connections, and thus sit at the center of the ENA space. They are neither similar to nor different from either play in the model.

5 For ’tis a Question Left Us Yet to Prove

To test whether these alternative projections address questions of goodness of fit, we conducted a simulation study using the data from *Hamlet* and *Romeo and Juliet*.

We hyopthesized that:

1. Models with the un-centered projection and the zero re-centered projection would both outperform the current ENA projection method in terms of goodness-of-fit; and
2. The difference in goodness of fit between the current ENA projection and the alternative methods would be larger as the number of zero points in the model rose.

The original data set had a total of 71 characters, 21 of whom had empty networks using the model described above. To conduct the simulation, we first chose a set of target numbers of empty networks, $k = \{0, 1, \ldots, 61\}$. We kept a minimum of 10 non-zero networks because if there are fewer non-zero points than codes, nodes cannot be reliably placed.

For each value of k, we randomly sampled the original data 1000 times. For each sample, we randomly chose (with replacement) 71-k characters with non-zero adjacency matrices. We then added k units with zero adjacency matrices to make a new set of 71 units.

For each of the 1000 data sets, we constructed six ENA models. We used *singular value decomposition* (SVD) and a *means rotation* (MR). Then, for each rotation method, we constructed three models: one using ENA's current projection method, one using un-centered projection, and one using zero re-centered projection.

For each model over the 1000 sampled data sets, we computed the goodness of fit on the x axis using a Pearson correlation coefficient, and found the mean correlation and standard error using the Fisher z-transformation.

Figure 6 shows the results, with the percentage of zero points $\left(\frac{100}{71}k\right)$ on the x axis and correlation on the y axis.

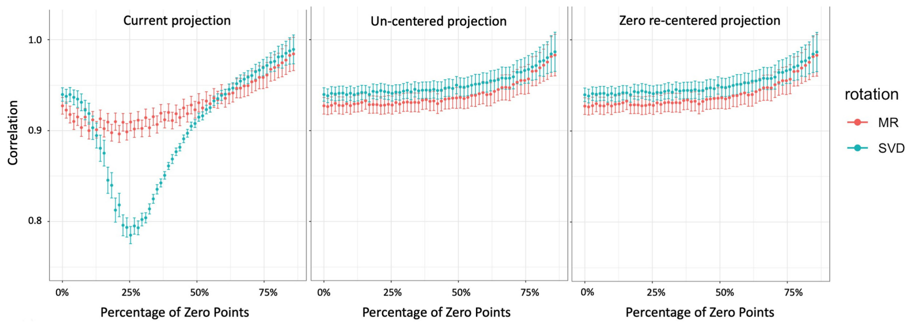

Fig. 6. Correlations for MR (red) and SVD (teal) for the current projection (left), un-centered projection (center), and zero re-centered projection (right). (Color figure online)

As expected, both the un-centered projection and the zero re-centered projection outperformed the current ENA projection. However, the simulation study suggests that the difference was only significant for the means rotation; SVD performed similarly for both models.

The results did not confirm our second hypothesis. The difference in performance did increase as the percentage of zero points went from 0% to 25%. However, the difference *decreased* as the percentage of zero points rose further, with differences becoming statistically insignificant after 50% of the points in the model were zero points.

This is likely because as the number of zero points increases, the mean of the vectors, $\bar{v}$, gets closer to zero, so the difference between plotted points and centroids for zero points decreases: that is, $p_0 - c_0 \rightarrow 0$. Thus, as the number of zero points increases past some point (25% of the data in this case), their impact on goodness of fit becomes smaller.

Further work will be required to determine why these effects were not seen in a model using SVD rather than a means rotation.

6 Discussion

This work has clear limitations, not least of which is that we only conducted an empirical study using one data set. Thus, although the mathematical argument may be sound, further research will be needed to determine what levels of zero points create poor goodness of fit for the current ENA projection relative to either alternative projection. Further work should also address the question of why models using SVD did not show the same differences as models using a means rotation—including a more thorough analysis of whether this finding holds in a wider range of data sets and models, and if so, whether the difference can be explained mathematically.

Perhaps more important, this study only considered ENA models using normalized data. In models where data is not normalized, the position of zero points is not undefined. Thus, further work will be needed to determine whether zero points deflate goodness of fit statistics when using the current ENA projection in un-normalized models.

Despite these limitations, it appears that the zero re-centered projection provides interpretive clarity without sacrificing mathematical rigor. It places the zero points at the center of the ENA space, consistent with the fact that they correspond to empty networks. The un-centered projection accomplishes this as well, but despite the fact that the zero points have an undefined position in the space of vectors, $\boldsymbol{v}_i$, it allows the zero points to influence the location of the mean of the projection.

A parsimonious reader might ask: *Why not just remove the zero points entirely?* This would surely solve the problem they create for goodness of fit. However, it would also potentially lead to inflated Type I error rates: that is, it could lead a researcher to conclude that the differences between two groups are statistically significant when in fact they are not.

To see why, imagine any data set with a statistically significant difference between two groups. If we add one hundred million zero points to each group, the difference would almost certainly no longer be significant. Zero points impact statistical measures on the plotted points other than goodness-of-fit, so they cannot simply be removed to improve goodness-of-fit without potentially impacting other important model statistics.

Similarly, an empirically-minded reader might ask: *If this is such a big problem, why hasn't anyone said anything about it before?* Again a fair question, and one that we think is answered in two ways. First, although researchers always should be concerned with goodness-of-fit in models, that is not always the case, even in a community as concerned with interpretive validity as QE researchers. That is, people may just not have noticed that there was a problem with goodness-of-fit in their models. Second, as our experiment shows, these problems only arise when the number of zero points is in a particular range of proportions to the total number of data points—and the number of data sets for which this is the case may be small.

There may be situations in which a zero re-centered model is not the best choice when using normalized data, although what such reasons might be is not evident from the current study. In fact, we were not able to think of any such situations—although of course that doesn't mean that they don't exist.

We thus recommend that researchers use the zero re-centered model when working with normalized data unless they have specific reasons not to.

Acknowledgements. This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

1. Bowman, D., et al.: The mathematical foundations of epistemic network analysis. In: Ruis, A. R., Lee, S.B. (eds.) ICQE 2021. CCIS, vol. 1312, pp. 91–105. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67788-6_7
2. Fayaz, M.: Dealing with Missing Data. http://www.realstatistics.com/descriptive-statistics/missing-data/
3. Guttman, L., Suchman, E.A.: Intensity and a zero point for attitude analysis. Am. Sociol. Rev. **12**(1), 57 (1947). https://doi.org/10.2307/2086491
4. Scott, J.: Social Network Analysis. CPI Group (2017)
5. Siebert-Evenstone, A.L., Irgens, G.A., Collier, W., Swiecki, Z., Ruis, A.R., Shaffer, D.W.: In search of conversational grain size: modeling semantic structure using moving Stanza windows. J. Learn. Analyt. **4**(3), 123–139 (2017)
6. Zachari Swiecki, A.R., Ruis, C.F., Shaffer, D.W.: Assessing individual contributions to collaborative problem solving: a network analysis approach. Comput. Hum. Behav. **104**, 105876 (2020)