# Discussion

### Summary

Using independent component analysis in place of distance- or density-based clustering on dynamic functional connectivity data is not a novel concept. It was initially suggested almost ten years ago (Yaesoubi, Miller, and Calhoun 2015) and saw some use even before then. Nonetheless, it has not yet been widely adopted, perhaps due to the novelty of the field of dynamic functional connectivity analysis. This article offers some insight into how such analyses might be used to extract differences in group connectivity dynamics.

### On Entropy Rate in Schizophrenia

Results indicate that schizophrenia patients consistently display elevated connectivity entropy rates compared to demographically matched healthy controls. This appears to contradict previous work which found reduced dynamical variance (Yu et al. 2015) and less dynamical activity (Miller et al. 2016) in patients than in controls. However, these articles do not measure entropy directly, and while signal variance generally correlates with entropy rate, this is not universally true. More recent work (Nataliia) does suggest that the functional connectivity of schizophrenia patients display elevated entropy compared to controls, a finding which matches the current work. How these two seemingly contradictory findings of reduced variance and decreased predictability might coexist is a question which warrants future study.

### On Estimating the Number of Sources in ICA

The divergence between previously proposed cluster counts and Marcěnko-Pasteur-based estimates highlights the unsolved nature of the problem of estimating the optimal number of independent components in blind source separation. Current methods are generally either somewhat arbitrary or designed for entirely different problems. The elbow method, for example, more often provides a rough guideline than a clear cutoff, which means researchers must rely on their intuition in choosing the optimal number of components. The present case, in which the cumulative sum of captured variance curved smoothly rather than displaying a clear “elbow”, is far from uncommon and leaves a wide range of dimensionalities possible. A means of selecting dimensionality in such cases would be extremely valuable addition to the field.

While the elbow criterion has, in this case, proven unable to determine the precise component count, it does suggest that the ideal number lies between 90 and 250 components. This is far more than the counts used in previous work, which range from the single to low double digits.

Most commonly, such numbers are established by estimating the variance contained in each principal component of the data and either selecting the percentage of variance desired, or heuristically selecting number based on the “elbow” of the variance plot. Such methods are, however, somewhat arbitrary, as they leave the choice of cutoff location to the researcher’s intuition rather than to a reproducible metric. This problem is not unique to blind source separation: clustering algorithms such as *k*-nn or Gaussian mixture modeling also require users to estimate the expected number of clusters before running the algorithm. As a result, many users have developed metrics to compare the goodness of clustering model order: examples include the silhouette score, the Calinski-Harabasz score, and the Davies-Bouldin score. Unfortunately, since blind source separation is a fundamentally different problem than distance- or density-based clustering, these scores may not capture the true number of sources in a blind source separation problem, such as that represented by dFNC dynamics.

The current article proposes solving this problem by leveraging the work of Ukrainian mathematicians Volodymyr Marcěnko and Leonid Pasteur, who conclusively demonstrated in 1967 that the eigenvalue distribution of a matrix of i.i.d. random variables has an analytically tractable upper limit . This law implies that any eigenvalue which exceeds this upper limit represents a nonrandom variable. Applying this law to correlation matrices, i.e. to FNC matrices, allows the user to quickly and efficiently estimate how many “clusters” of non-independent (correlated) activity exist by counting the number of eigenvalues which exceed this upper limit . The authors hope that this method may offer a solution to other exploratory studies in blind source separation, both within and outside of the field of dynamic functional connectivity.

### On Entropy Rate and its Implications

### Limitations and Future Steps