Role of RPCA
Rotated principal component analysis (RPCA) or Empirical Orthogonal Functions (EOF) with varimax rotation have a long history in climatology [1, 2]. It can play a role within complex network analysis and namely community detection that both appeared in climate research relatively recently [3, 7]. Determination of optimal community structure is well known hard problem and there are several methods excelling in specific situations [4] and several ways of measuring quality of resulting community structure such as modularity [5].

A question under study is how RPCA results can be used either solely as community or as dimensionality reduction preprocessing for further community analysis. We use data from National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP-NCAR) Reanalysis [8], more specifically SAT and SLP.

Community structure
Roughly speaking, distribution of edges has tendency to concentrate within group of nodes. Let’s have a graph $G = (V, E)$ and subset of vertices $C \subseteq V$. For $C$ to be a community $E(C) / |C|^2 \geq E(V) / |V|^2$ and $E(C) / |C|^2 < E(V) / |V|^2$.

$E(X)$ denotes edges induced by vertex set $X \subseteq V$ and $E(C)$ number of outgoing edges from vertices of $C$.

Several variants were tested, algorithm walktrap [9] is shown. All algorithms are from package python-graph accessible at http://igraph.sourceforge.net/

Used measure of quality of community partition
Quality of partition by modularity [5]. For graph $G = (V, E)$, $|E| = m$, set of communities $C / |C|/m$, define [4] $\delta(C, C)$ indicator that vertices $i$ and $j$ have same community.

$Q = \frac{1}{m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{m} \right) \delta(C_i, C_j)$

$k_i$ denotes degree of vertex $i$ (number of edges adjacent to vertex $i$), and $A_{ij}$ elements of adjacency matrix.

Interpreting RPCA as community (scheme)
We proceed as follows:
1. Construct RPCA components using preprocessed input data.
2. Construct weighted network using the same input without cos. transform.
3. Compute threshold for binarization from density of network constructed as union of complete graphs each corresponding to particular component.
5. Construct community structure as nonoverlapping clusters derived from RPCA components using maximization criterion (i.e. cluster is given by maximal membership in component).

Dimensionality reductions

1. RPCA components
Whol globe SLP 58 SLP, 67 SAT →
2. GEOGRID Spherical Geodesic Grid [10]
162 or 642 grid points, 162 \times 162
3. LATLON sub-sampling of lat-lon grid [3]
Original 73x144 is subsampled to 10 \times 10 ' bins that gives 642 points

RPCA as community
RPCA results $Q_{RPCA}$ and algorithms fastgreedy $Q_{FG}$ and walktrap $Q_{WT}$

SAT: $Q_{RPCA} = 0.4323, Q_{FG} = 0.6549, Q_{WT} = 0.7283$

SLP: $Q_{RPCA} = 0.4674 while$ $Q_{FG} = 0.5418, Q_{WT} = 0.6500$

Discussion (further results in second column)
RPCA seems to preserve some of community structure information – GEOGRID with slightly larger dimension (number of nodes) gives comparable results in sense of modularity. GEOGRID and LATLON with much larger dimension gives larger modularity. Further research with equal graph sizes and data preprocessing is required.

References