Geophysical Networks

Adway Mitra

ML for Earth System Sciences (AI60002)

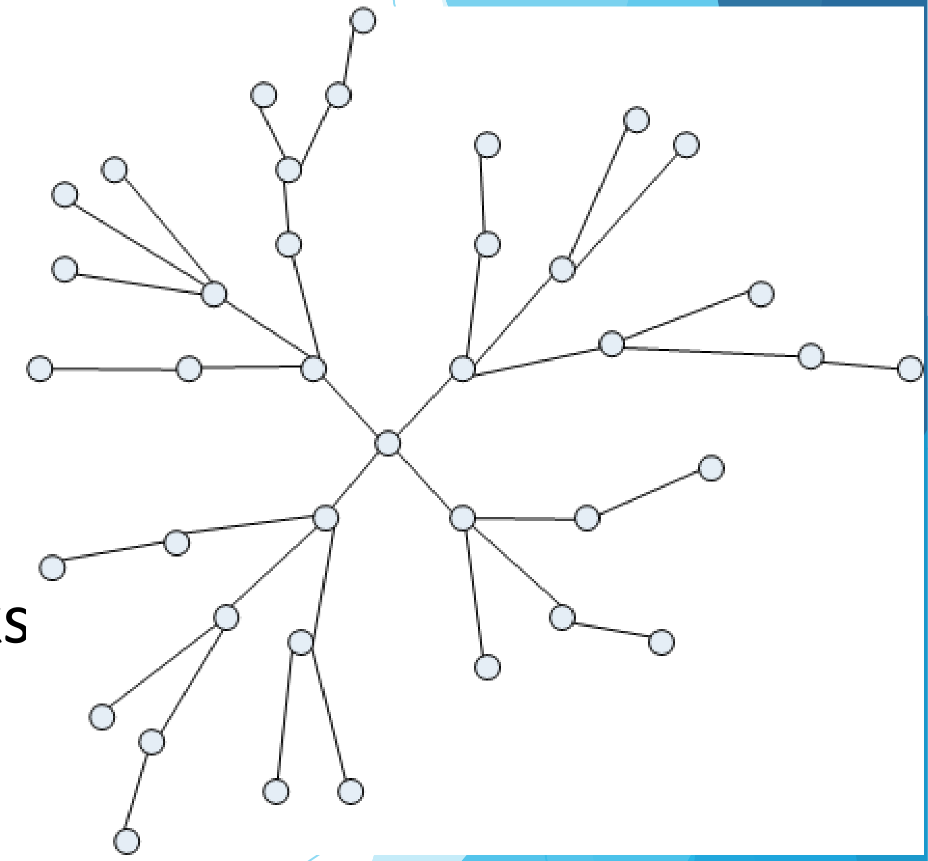
8th March 2021

Networks and Graphs

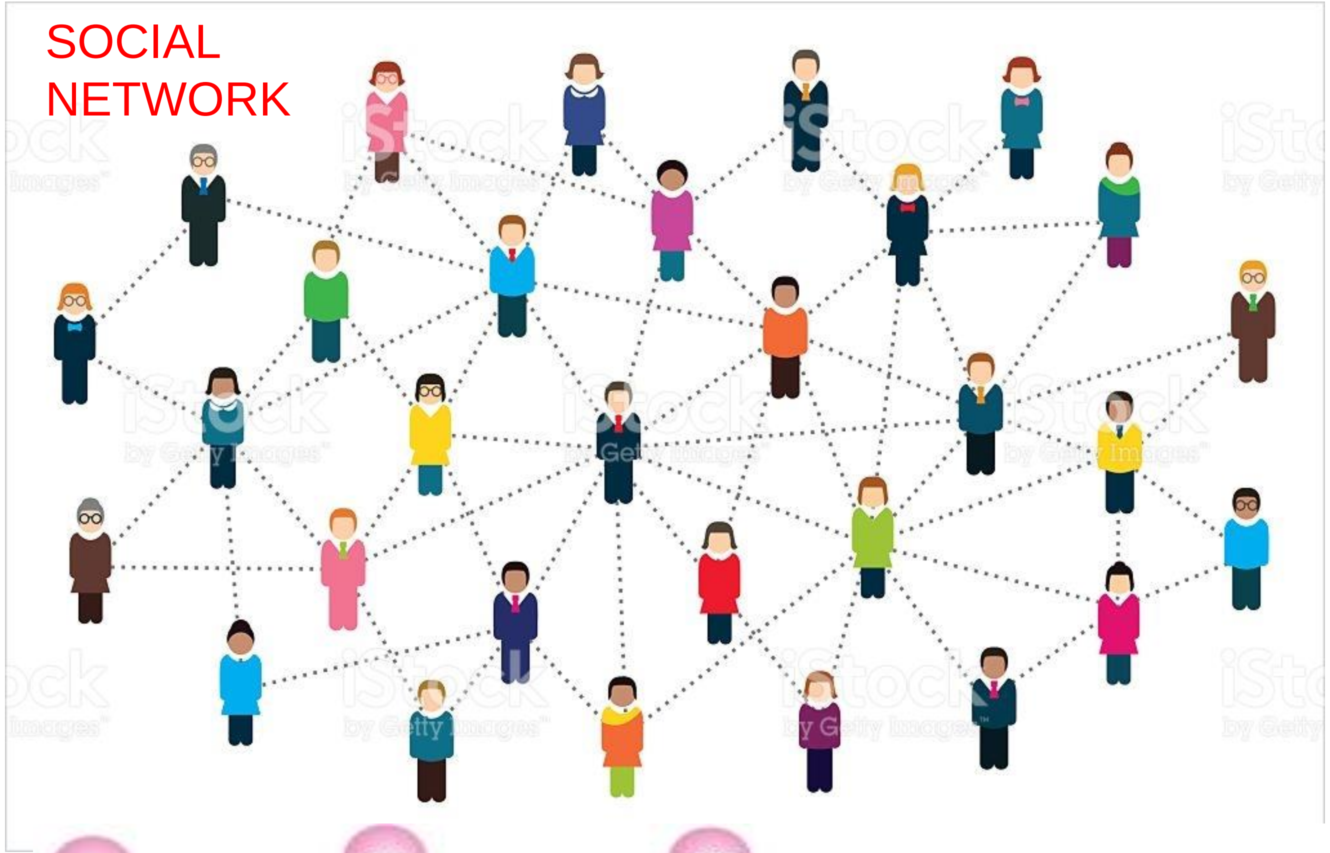
- V: set of nodes
- E: set of edges; an edge connects two nodes

- Each node represents an entity of some kind
- Edges represent interactions between them

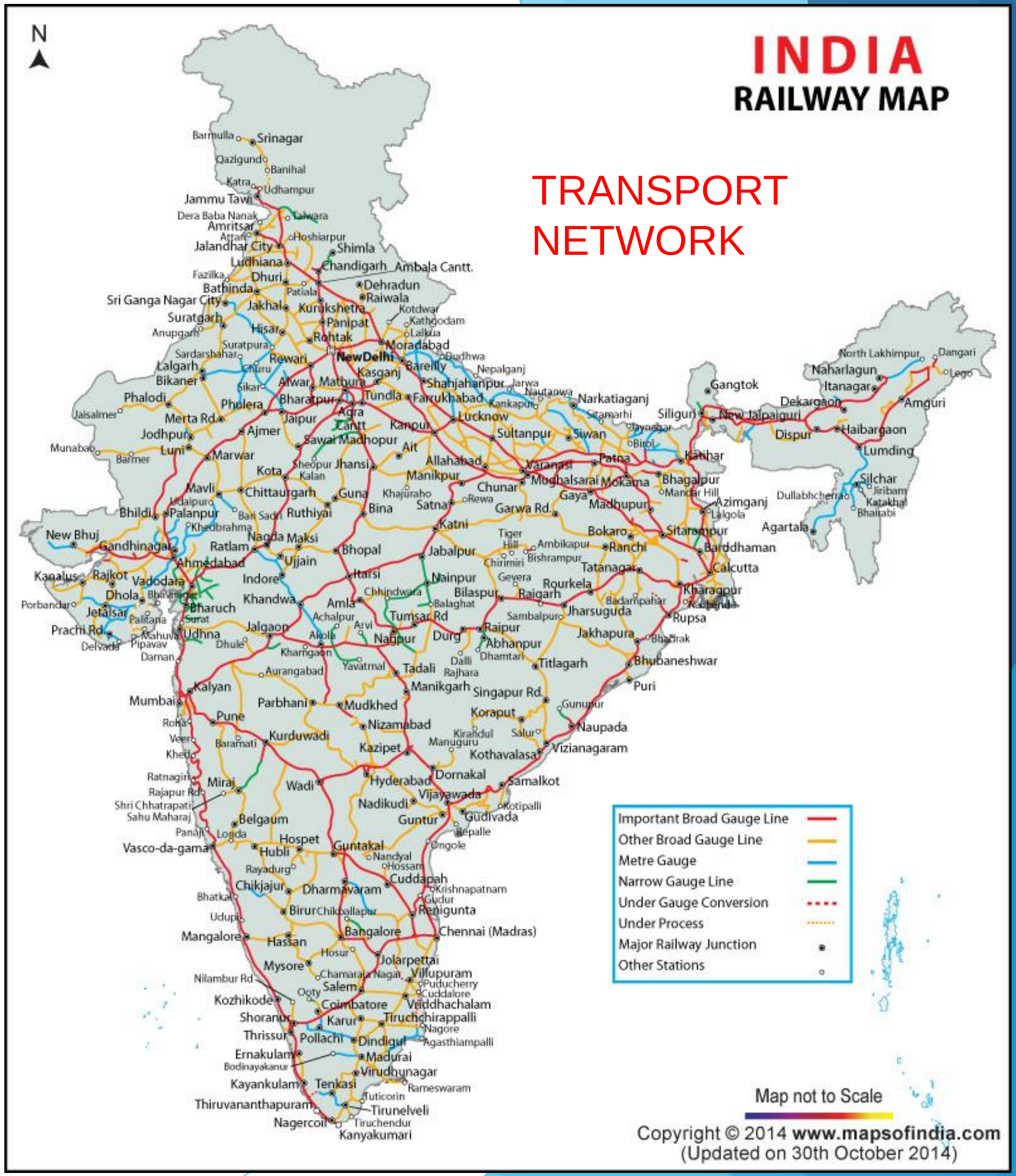
- Examples: computer networks, social networks biological networks, road/transport networks



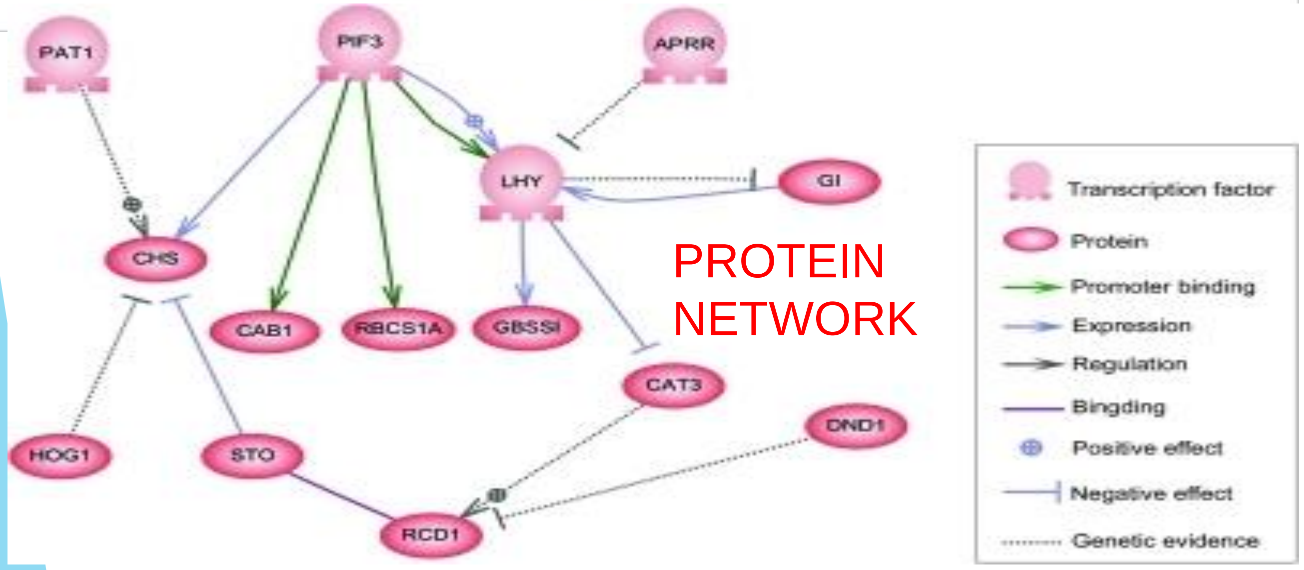
SOCIAL NETWORK



TRANSPORT NETWORK



PROTEIN NETWORK



Why use networks?

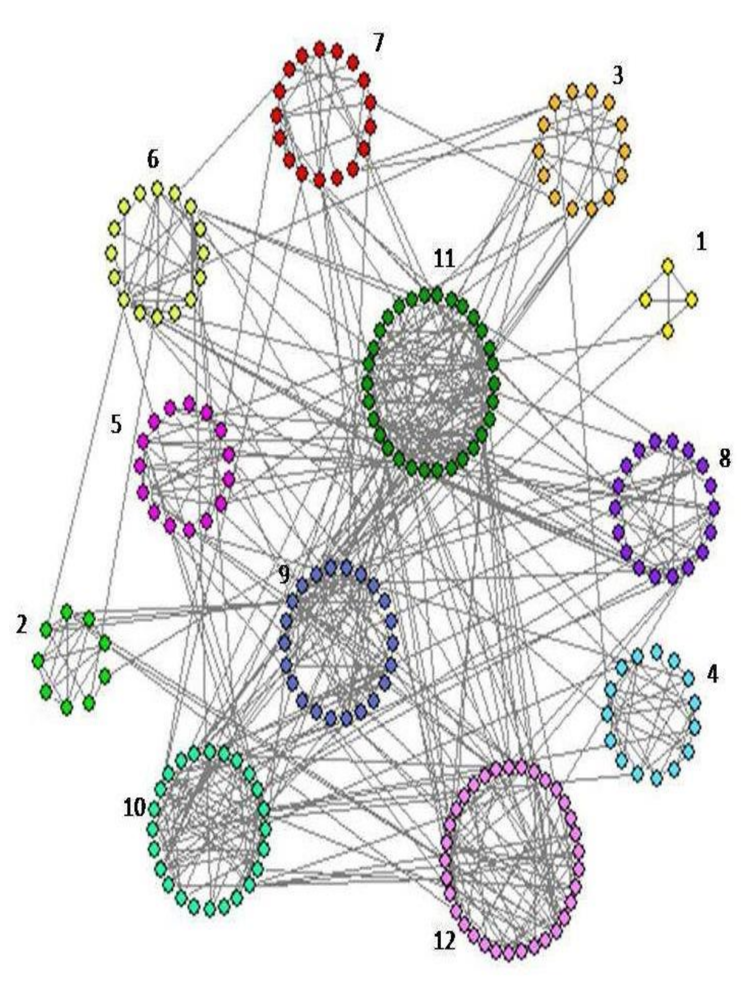
- Study the nature of interactions among entities and their dynamics
- Identify “communities” where members interact strongly,

  e.g. common interest groups in social networks

Model dynamic processes,

e.g. flow of information in social network

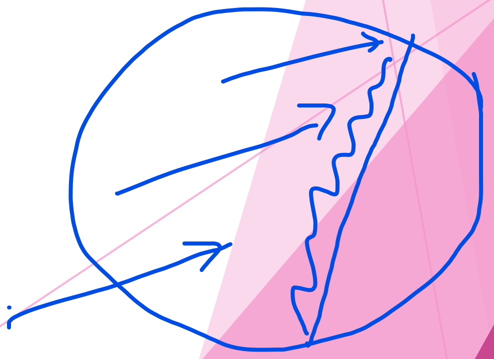
flow of infectious disease



Geophysical Networks

- Geophysical networks to visualize and analyse spatio-temporal geophysical data
- To identify regions whose geophysical conditions are strongly related
- To identify teleconnections
- To identify causal relationships between geophysical events
- To identify relationships among different geophysical variables

- But how to construct a geophysical network?

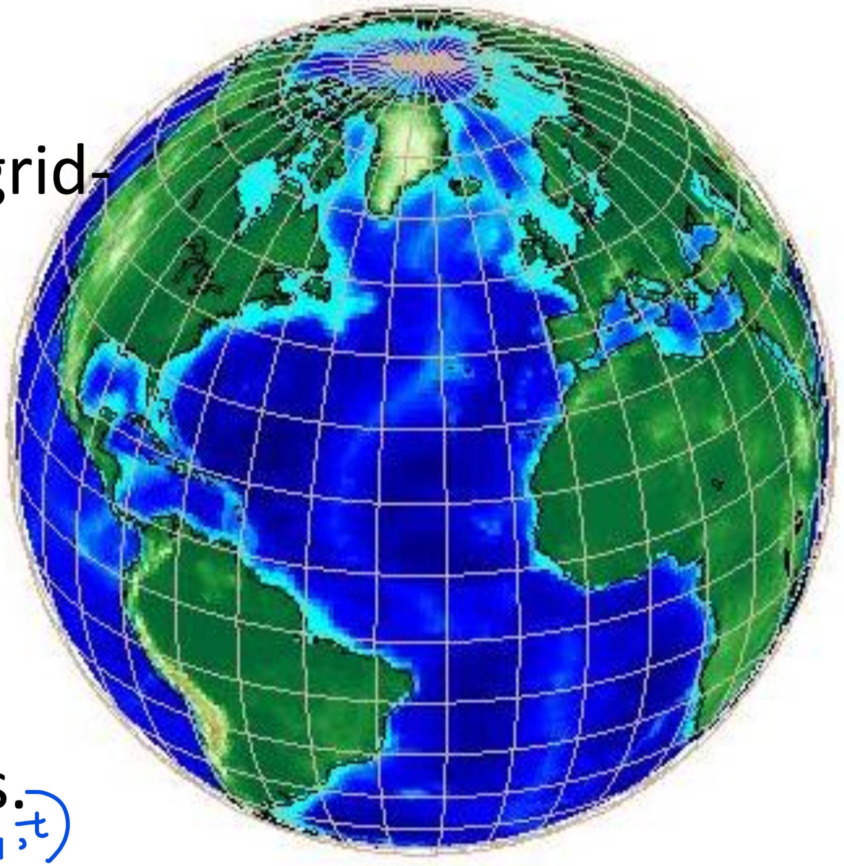


Networks and Graphs

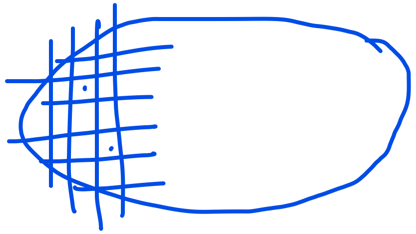
Design Issues –

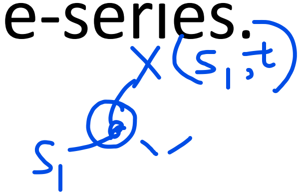
- What does each node represent?
- Which nodes are connected by edges?
- Are edges directed?
- What does each edge represent?
- Is the graph fixed or does it evolve over time?

Geophysical Networks: Defining Nodes



- In most geophysical networks, each node represents a spatial location e.g. one node per grid-point.
- Each node has an attached variable, e.g. rainfall received by the corresponding location in any time-step (hour/day/month/year).
- At each time-step, we get an "instantiation" of the network; for each node we get a time-series.





Geophysical Networks: Defining Edges

- Once the set of nodes has been chosen, we need to select how we define connections between pairs of nodes (edges).
- Edges can be weighted or unweighted; directed or undirected.
- Edge weight: a measure of “similarity” between the pair of nodes.

- Consider each pair of nodes in a network
- Measure their “similarity”
- Put an edge if similarity is above a threshold.

But what is a measure of “similarity”?

Geophysical Network Edges: Correlation

S_i , S_j

0.4 . −0.3

- Consider two nodes (two grid points): i, j.
- Their corresponding climatic variables: Vi, Vj.
- → two time series: {Vi(t)}, {Vj(t)}.
- Define edge weight W(i,j) = Pearson correlation coefficient between the two time-series.

$\boxed{-1 \leq \rho_{ij} \leq 1}$

- **Seminal paper by** Tsonis and Roebber "**climate networks**" **(2004)**
- Identify all **pairs** of grid points with correlation > 0.5.

→ "Correlation Network"

→ First type of climate network ever defined.

Tsonis, A. A., & Roebber, P. J. (2004). "The architecture of the climate network." *Physica A: Statistical Mechanics and its Applications*, *333*, 497-504.

Geophysical Network Edges: Mutual Information

- Another criterion: Mutual Information between the variables
- Measures the *statistical dependence* between them

$$M_{ij} = \sum_{\mu\nu} p_{ij}(\mu,\nu) \log \frac{p_{ij}(\mu,\nu)}{p_i(\mu)p_j(\nu)},$$

$$\frac{p(X_i = \mu, X_j = \nu)}{p(X_i = \mu)\, p(X_j = \nu)}$$

$$\begin{bmatrix} X_i(t) = \mu \\ X_j(t) = \nu \end{bmatrix}$$

- The excess amount of information generated by falsely assuming the two time-series at nodes i and j to be independent.
- Able to detect nonlinear relationships.

$$\begin{bmatrix} X_i(t) = \mu \\ X_j(t+\delta) = \nu \end{bmatrix}$$

Donges, J. F.; Zou, Y.; Marwan, N.; Kurths, J. (2009). "Complex Networks in Climate Dynamics". *The European Physical Journal Special Topics*. Springer-Verlag. **174** (1): 157–179

Synchronization and Lag

- Often, time-series at two locations are not perfectly synchronized.
- Influence of one region may take time to reach another region.
- Especially true for spatially distant locations.
- Lag networks: compare {Vi(t)}, {Vj(t+Δ)}, where Δ is a suitable “lag”.

Δ_{ij}

G. Tirabassi and C. Masoller, 2013. On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields. Europhysics Letters Vol 102 (2013).

- How to identify the “best” lag for any pair of nodes to maximize their correlation? Sequence Alignment/Dynamic Time Warping?

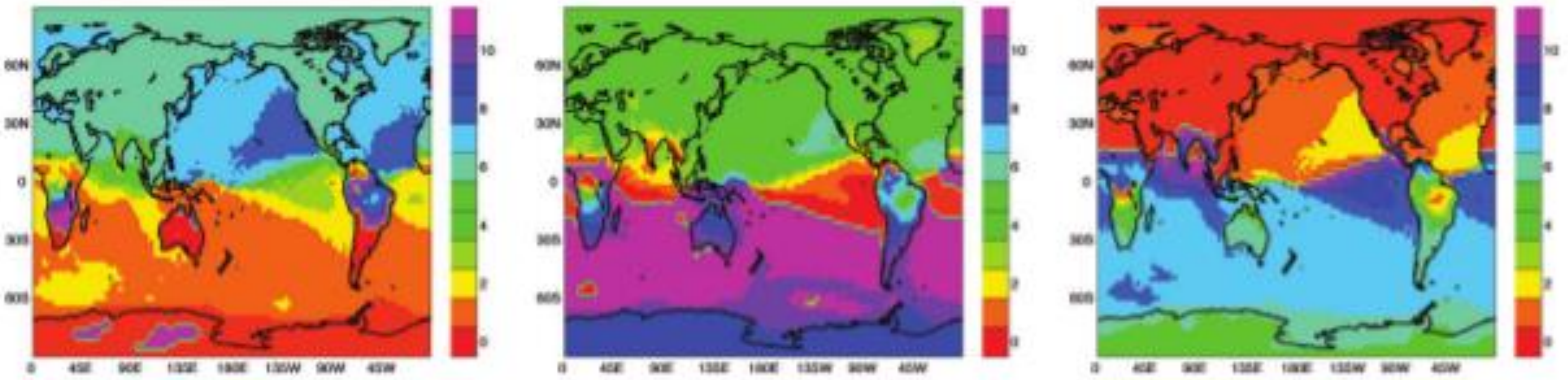


Fig. 2: (Color online) Lag-times of a node in Mongolia (left), in Australia (center) and in the El Niño basin (right).

PC: Tirabassi et al, EPL 102 (2013)

Geophysical Network Edges: Event Synchronization

- Define "events" for each time-series.
- E.g. annual rainfall at a location exceeding a threshold.
- Event a in time-series Vi, event b in time-series Vj are synchronized if |a-b|< threshold.
- How often are events of two time-series synchronized?
- Very relevant for extreme event analysis!

3/11, 1

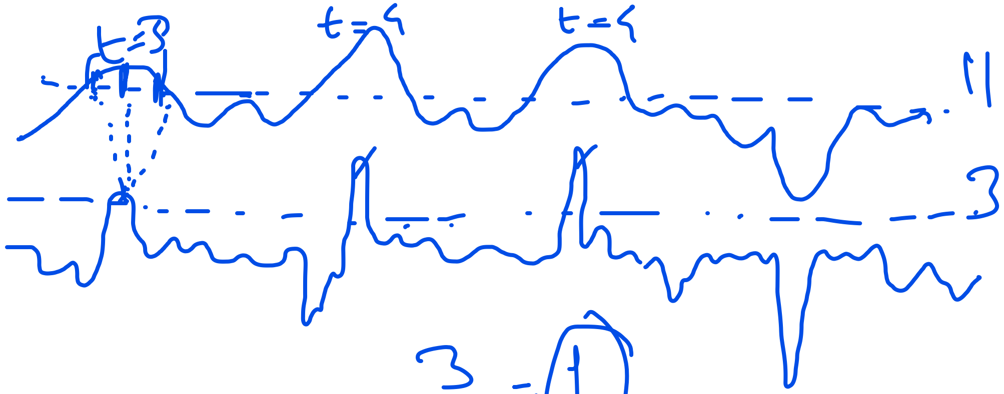
$\boxed{\Delta = 0}$

δ_i X_i

δ_j X_j

$\boxed{\Delta = 2}$

9/11, !



$\frac{3}{3} = 1$

3/11

$X_i(t) > \delta_i$

$X_j(t') > \delta_j$

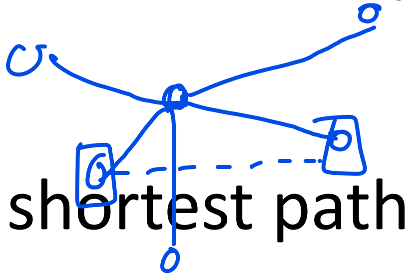
where $|t - t'| \leq \Delta$

Events are synchronized

Geophysical Networks: Properties

Some networks are so large/dense that one can only print properties of the networks, not the connections themselves.

- **Degree distribution** - number of edges per node. as a fn. of dist
- **Local/global clustering coefficient** – probability that two randomly chosen neighbors of any node are themselves neighbors.
- **Centrality** – mean inverse shortest path to each node from a given node
- Number and size distribution of **connected components.**
- **Area-weighted Connectivity** – earth area covered by neighbors of each node.
- **Diameter, path length** distribution.
- **Small-world property** – distribution of shortest path length between pairs of nodes.



Geophysical Networks: Degree Distribution

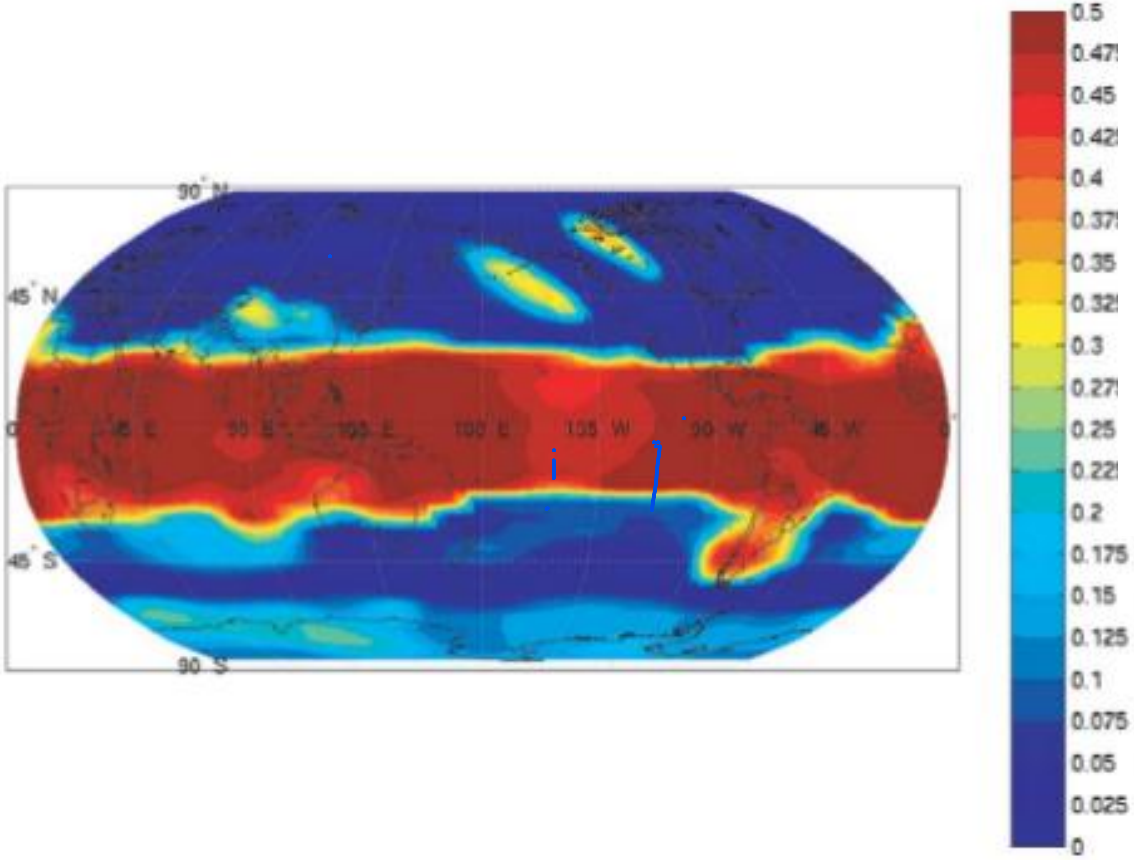
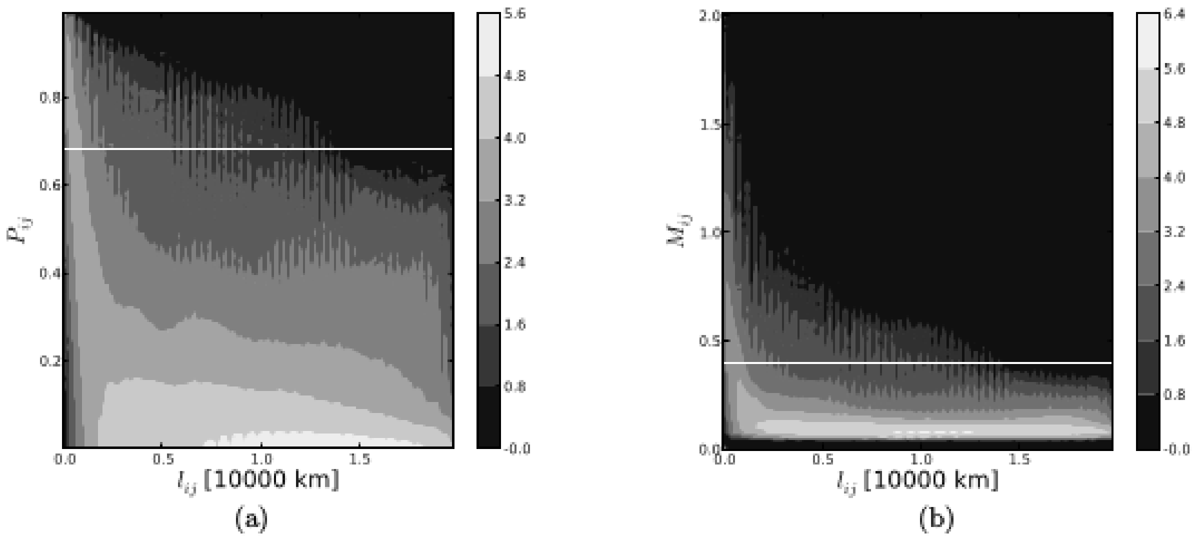


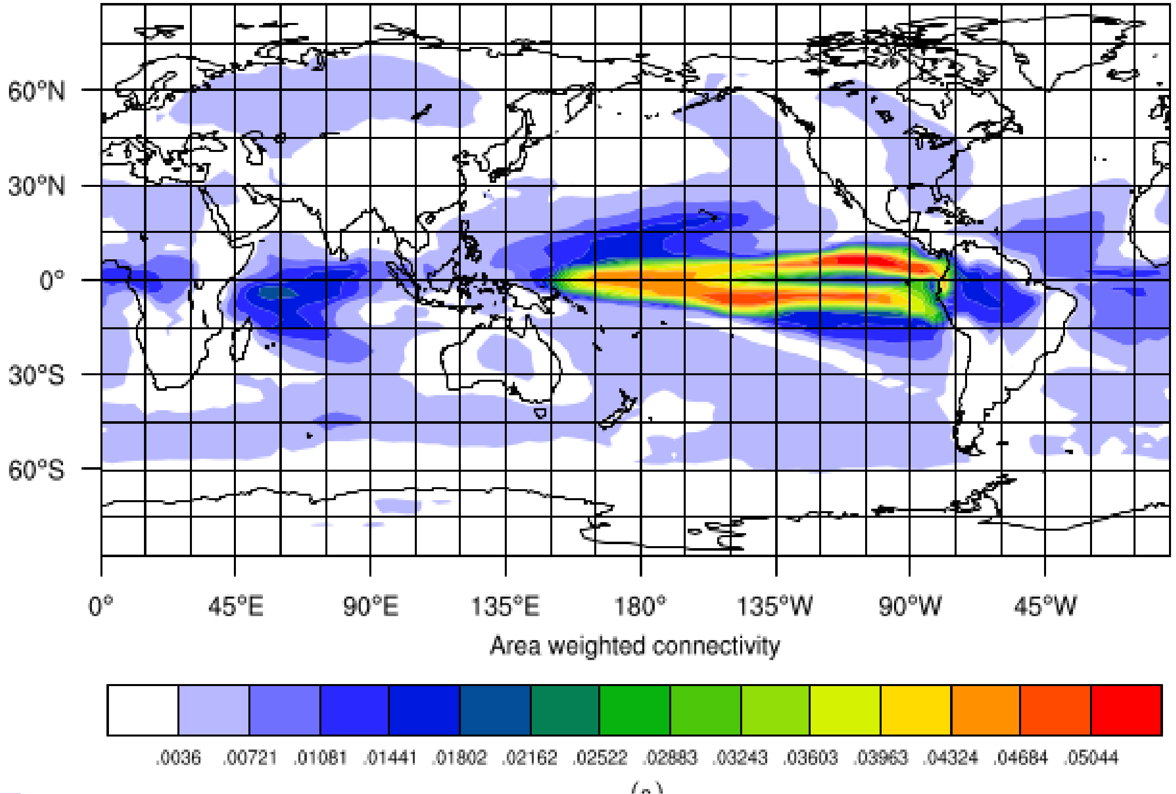
Fig. 6. Total number of links (connections) at each geographic location. The uniformity observed in the Tropics indicates that each node possesses the same number of connections. This is not the case in the extratropics where certain nodes possess more links than the rest.



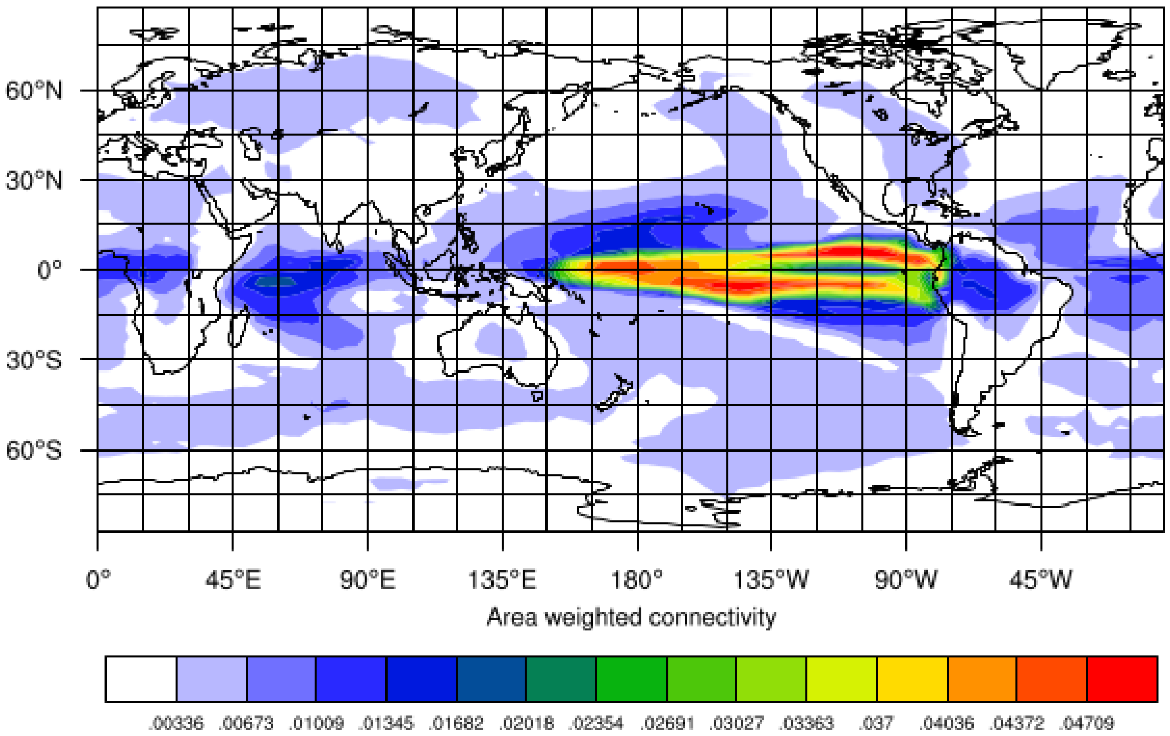
PC: Donges et al, EPJST 174 (2009)

PC: Tsonis et al,
BAMS May 2015

Geophysical Networks: Area Weighted Connectivity



Correlation



Mutual Information

PC: Donges et al, EPJST 174 (2009)

Teleconnections

- How does climatic influence pass over long distances?

- Zhou, Dong; Gozolchiani, Avi; Ashkenazy, Yosef; Havlin, Shlomo (2015). "Teleconnection Paths via Climate Network Direct Link Detection". *Physical Review Letters*. **115** (26).

- Considers lag-corrected correlations between pairs of nodes
- Separates “indirect” effects from each node’s time-series and extracts “pure” time-series for each node
- Constructs “direct” correlation network based these “pure” time-series between pairs of distant nodes
- Also identifies optimal path of influence between pairs of strongly correlated distant nodes

Teleconnections

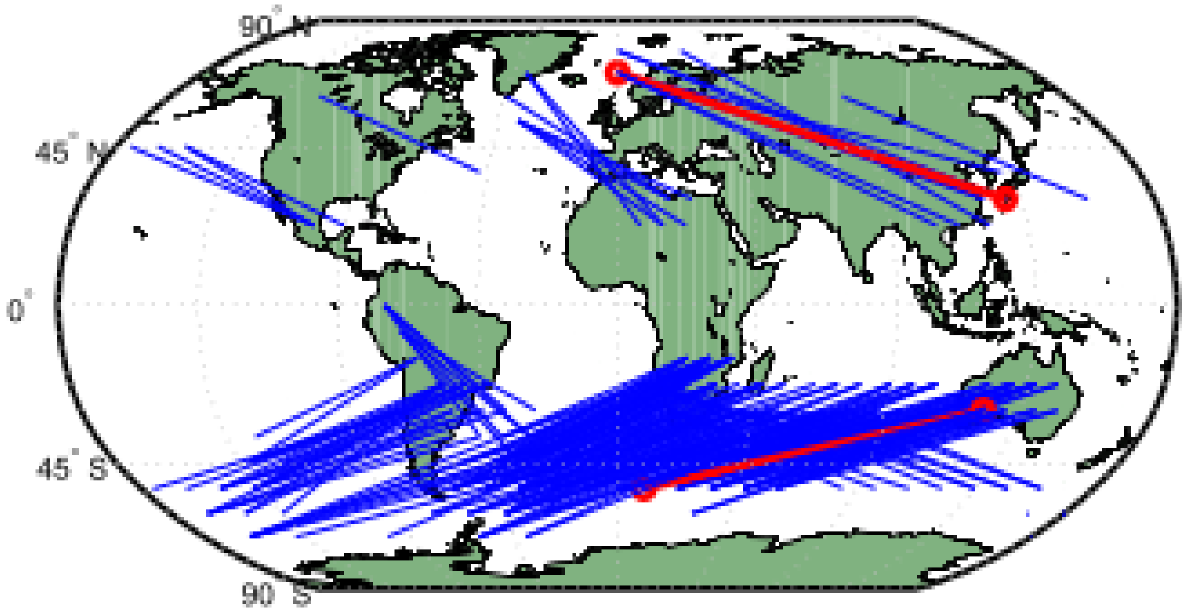
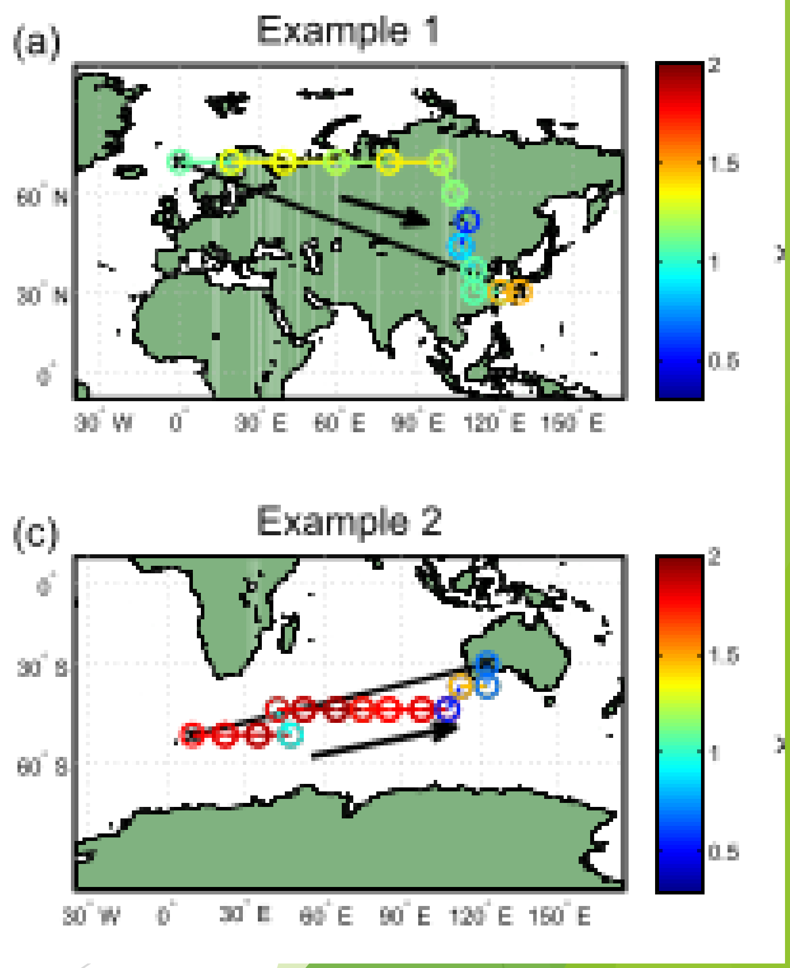


FIG. 3 (color online). 226 strong and long observed positive links which have (i) distance larger than 5000 km, (ii) $W_{i,j}^{obs,pos} \geq 9$, and (iii) latitude difference above 20°. The red circles and lines indicate the two examples considered in the text and in Fig. 4.



PC: Zhou et al, PRL 115 (2015)

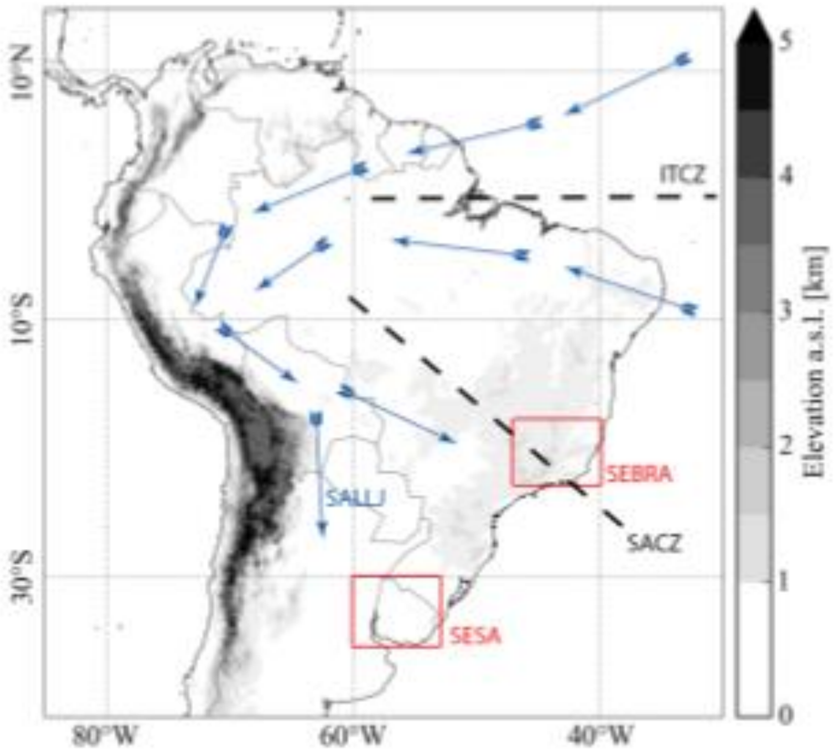
Extreme Event Synchronization

- Extreme Events at individual locations identified.
- How synchronized are such extreme events in two locations?

- N. Boers, A. Rheinwalt, B. Bookhagen, Henrique M. J. Barbosa, N. Marwan, J. Marengo and J. Kurths, 2014. “The South American rainfall dipole: A complex network analysis of extreme events”. Geophysical Research Letters, 2014

- Analyze rainfall extremes (above 90-th percentile) of a dipole over South America in Brazil and Argentina.
- Separate networks for the two phases of the dipole.
- Strength of each edge equal to number of synchronized rainfall. extremes (upto 3 days) at the connected nodes.
- Degree of each node studied during both phases of dipole.

Extreme Event Synchronization



PC: Boers et al, GRL 41 (2014)

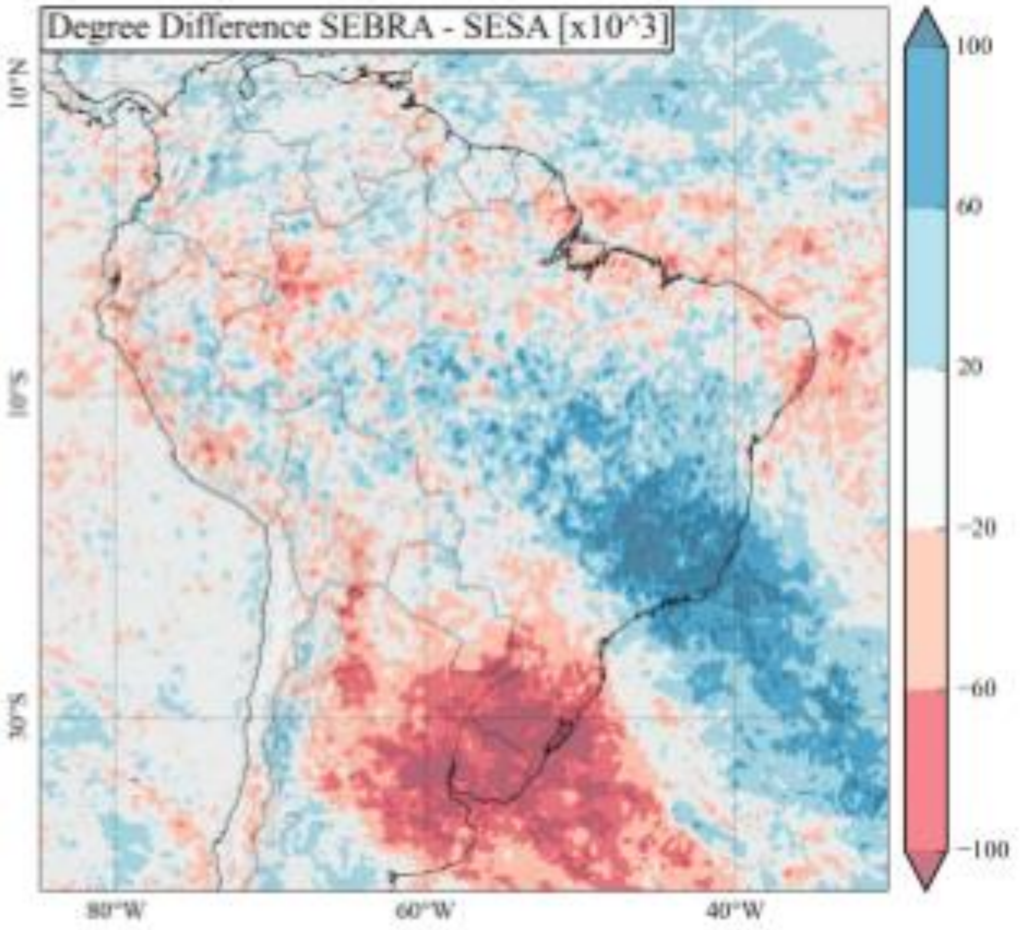


Figure 4. Difference between degree fields for the SEBRA and for the SESA phase. Note the oscillation between positive and negative values extending over the entire continent beyond the dipole between the SESA and SEBRA regions.

Other Geophysical network types exist

Other networks:

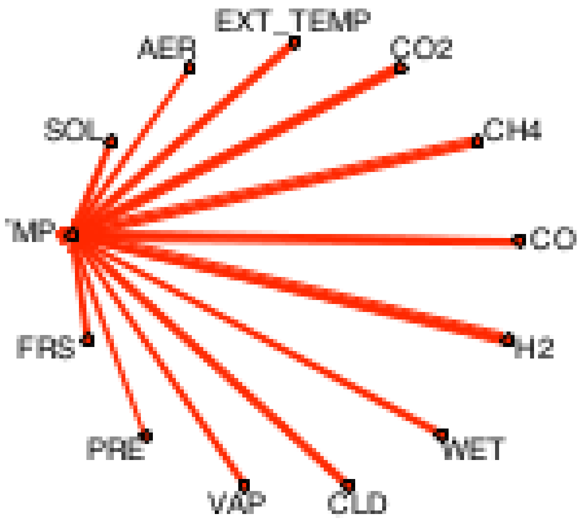
- Phase synchronization networks.
- Variety of causality-based networks.

Optional: We can discuss causal networks in more detail next time, if there is interest.

Geophysical Networks: Nodes

- Sometimes, each node can represent (location, time)
- E.g. Rainfall received by a location in the month of July of a year.

- Each node may also represent a climatic variable.
- E.g. one node for temperature, one for rainfall, one for humidity, etc.
- One instantiation of the network per location and time-step; one time-series per location.



PC: Lozano et al, KDD 2009

Time-evolving Network

- Divide the total data duration into time-slices.
- Compute the edge-weights and build the network for each time-slice
- Edge between any node-pair may be present in some slices, absent in others
- How do the network properties change across the slices?