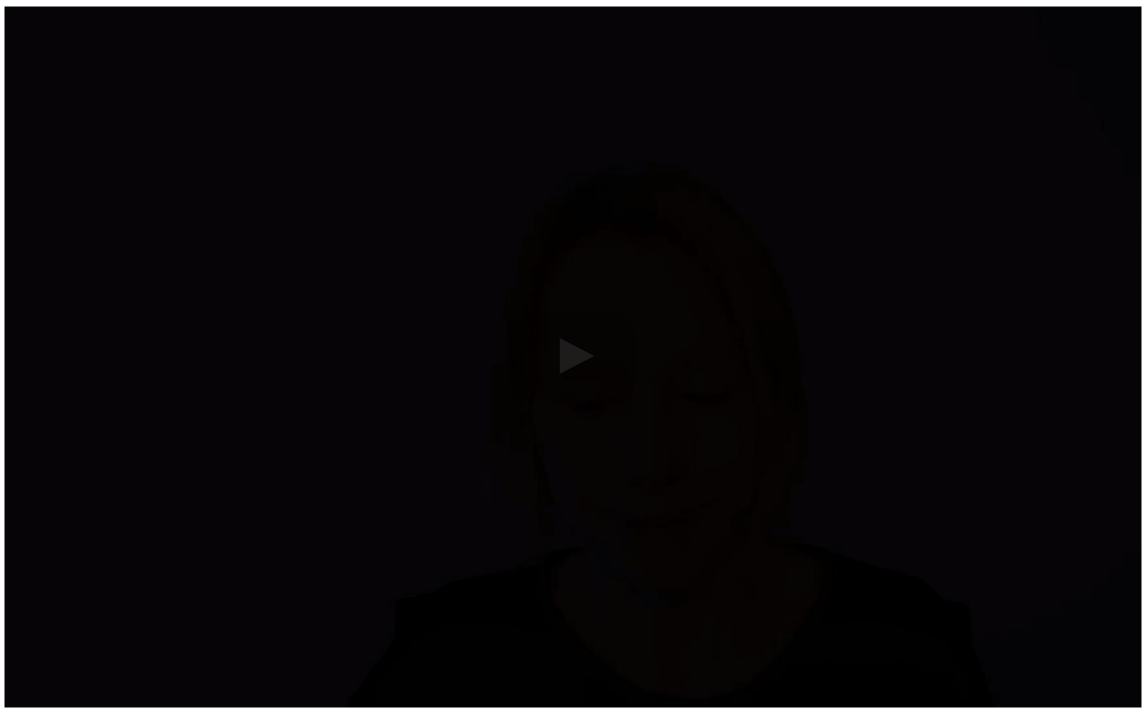
5. Long-range Climate Correlations

Bookmark this page

Exercises due May 21, 2021 19:59 EDT

Long-range Climate Correlations



you

can do is when you look at networks that you

build from data, do they actually change over time?

Do you have changes in their properties,

in their connectivity, and their patterns over time?

And that often gives us some interesting hints

on what may be happening or just notions that maybe

certain patterns are not as fixed

as we may easily believe them to be or simplify them to be.

14:31 / 14:31 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file
Download Text (.txt) file

Now, we will discuss long-range correlations. So far we have assumed that the correlation is determined by the geometric (or otherwise) distance between measurements. The correlation has been additionally assumed to increase with proximity, with the relationship described by a kernel function. Both the selection of the kernel and the selection of the parameters of the kernel are design problems.

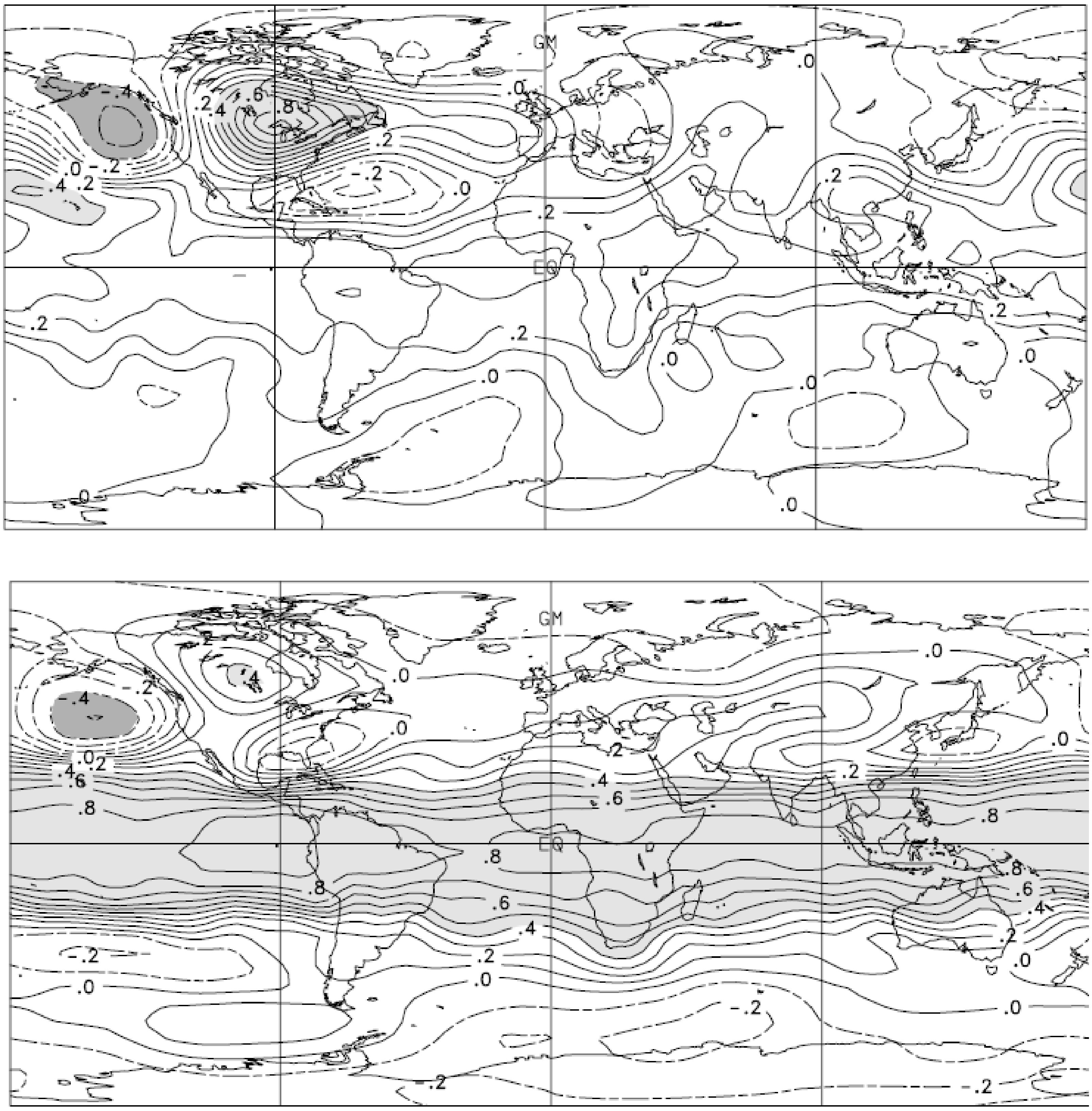
We have also discussed multiple ways to find a model for the data we have at hand. Such an approach is reasonable in many scenarios. However, these are not the only possible way to relate various random variables and observations.

The below figure shows an example of long-range correlations. In the top image, we observe two gray shaded areas, one darker than the other.

Suppose we pick a point inside the light gray area and compute the correlation between that point and all other regions on the map. Other points that are positively correlated with it are then shown as light grays, and points that are negatively correlated are shown as dark grays. The selected point is not correlated with much of the other points on the map, and so these are not shown as shaded.

It is curious that there is some positive correlation not only with the points that are immediately located next to

It is curious that there is some positive correlation not only with the points that are immediately located next to the selected point, but also with other non-connected areas. For example, in the extreme top left, there is a small area of positive correlations and another one with negative correlations.



43: Long-range spatial correlations.

The bottom image in the above figure shows a similar example, but in this case, we have picked a point on the equator. We can observe that this point is positively correlated with all other points along the equator; and, there is an area in the top left for which there is an apparent negative correlation present.

These correlations – that are not dependent on physical closeness – cannot be modeled with the kernel function approach we discussed before.

The way we will analyze these kinds of long-range correlations is by the construction of graphical networks. We will describe this approach next.

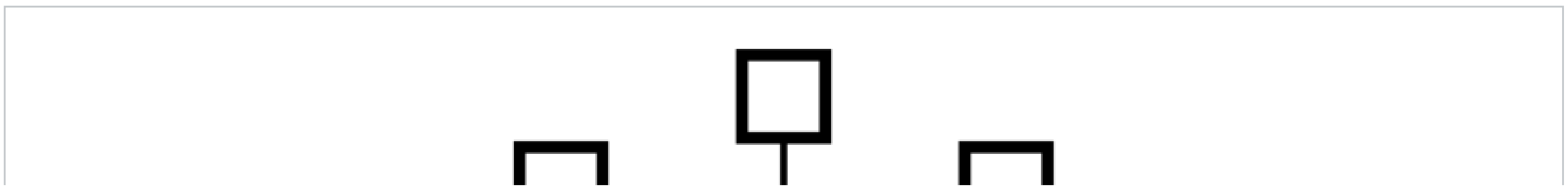
First, recall the definition of a network or a graph as $\mathcal{G} = (V, E)$, where V is the set of nodes $(v_1, \cdots, v_n)$, and $E \subseteq V \times V$ is a set of edges such that $(v_i, v_j) \in E$ implies that nodes v_i and v_j are connected.

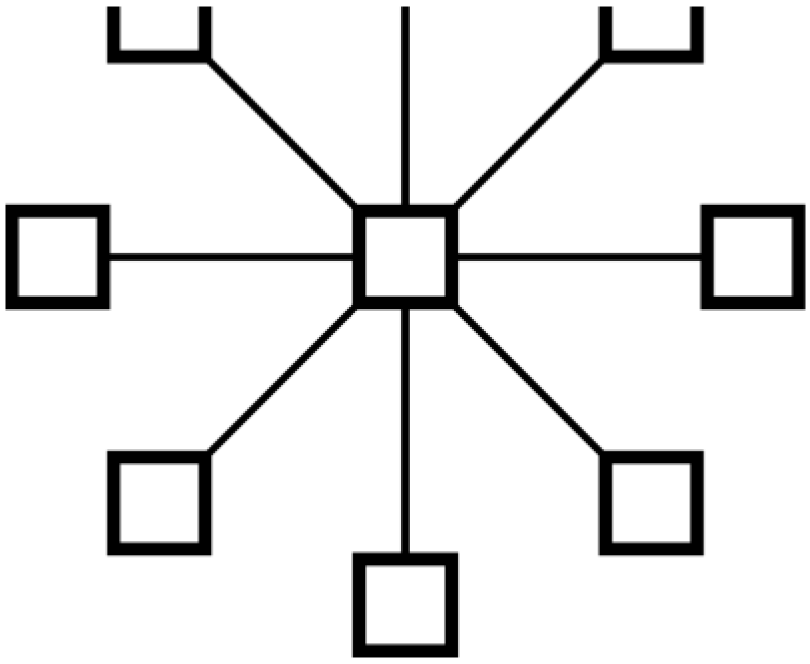
Properties of Graph

4 points possible (graded)

The degree is defined as the number of neighboring nodes, that a node is connected to.

Consider the graph below:





44: Star graph.

What is the maximum degree of this graph? That is: what is the degree of the node(s) with the largest degree?

- 1
- 9
- 8
- 2
- 10
- 0

What is the multiplicity of the maximum degree? That is: how many nodes share this maximum degree?

- 1
- 9
- 8
- 2
- 10
- 0

What is the minimum degree of this graph?

- 1
- 9
- 8
- 2
- 10

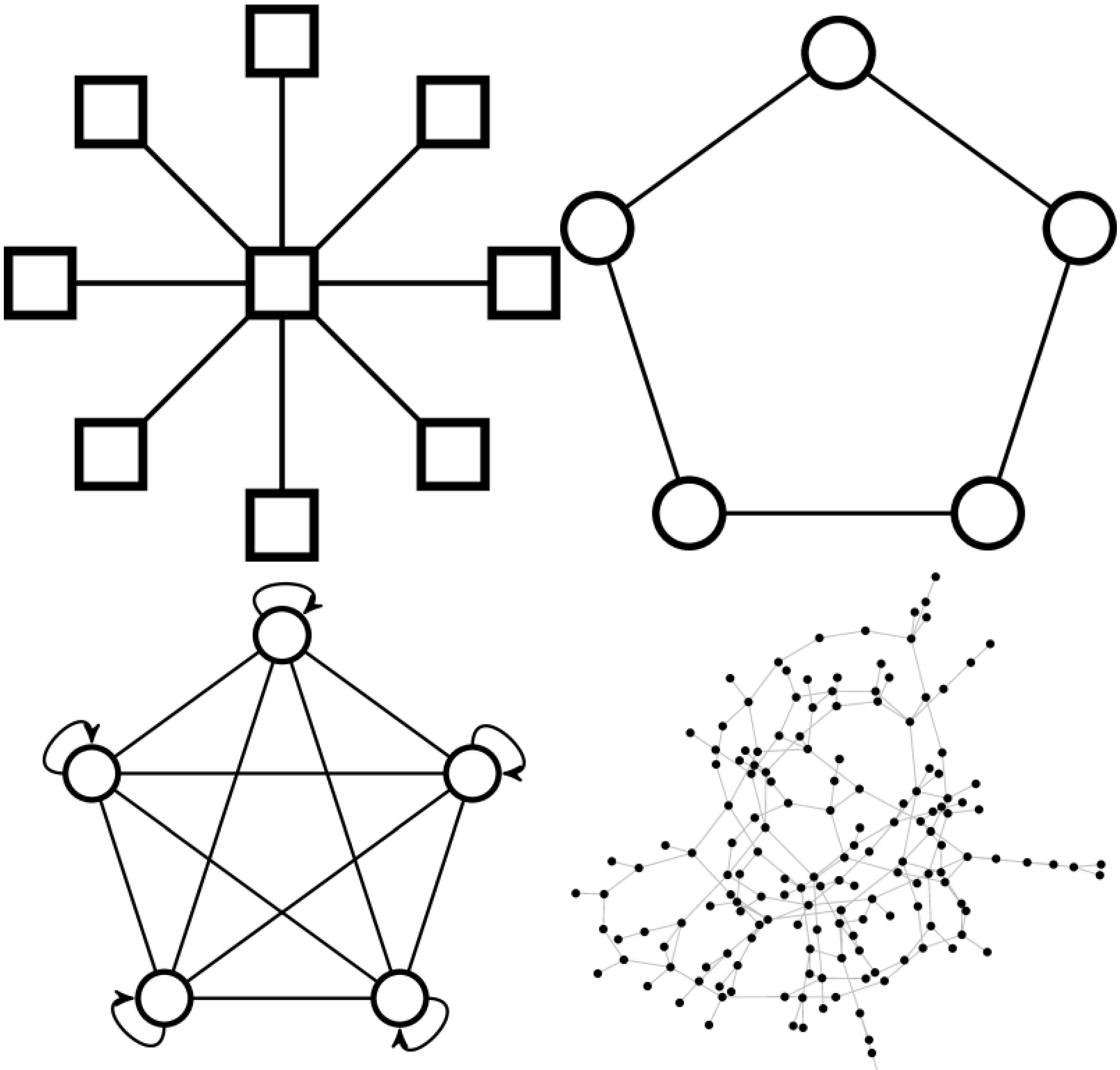
- 10
- 0

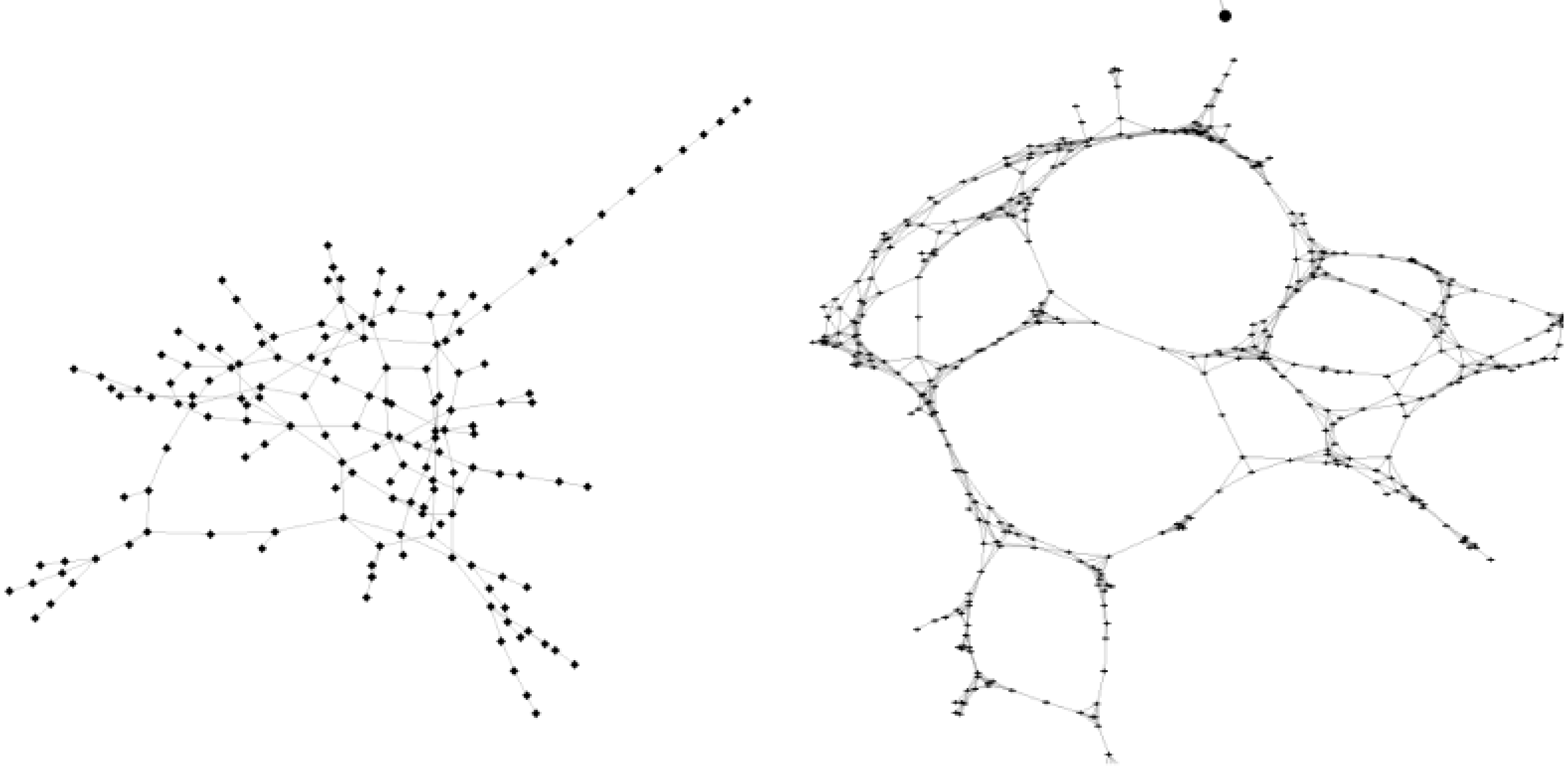
What is the multiplicity of the minimum degree?

- 1
- 9
- 8
- 2
- 10
- 0

Submit You have used 0 of 2 attempts

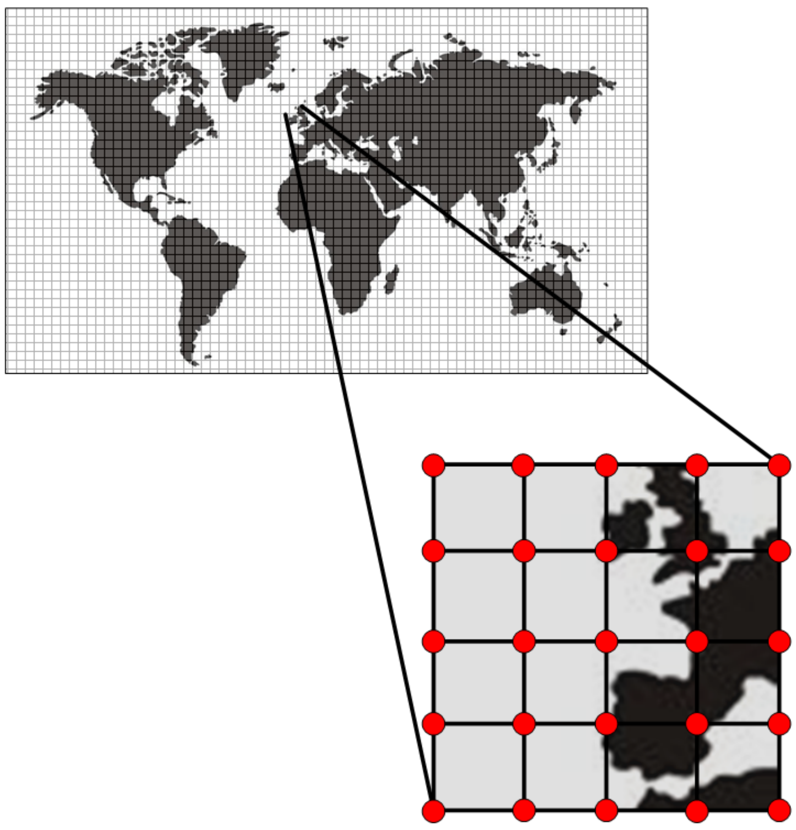
The below figure shows six examples of graphs, with various numbers of nodes and edge configurations. This is sometimes referred to as the graph topology or shape.





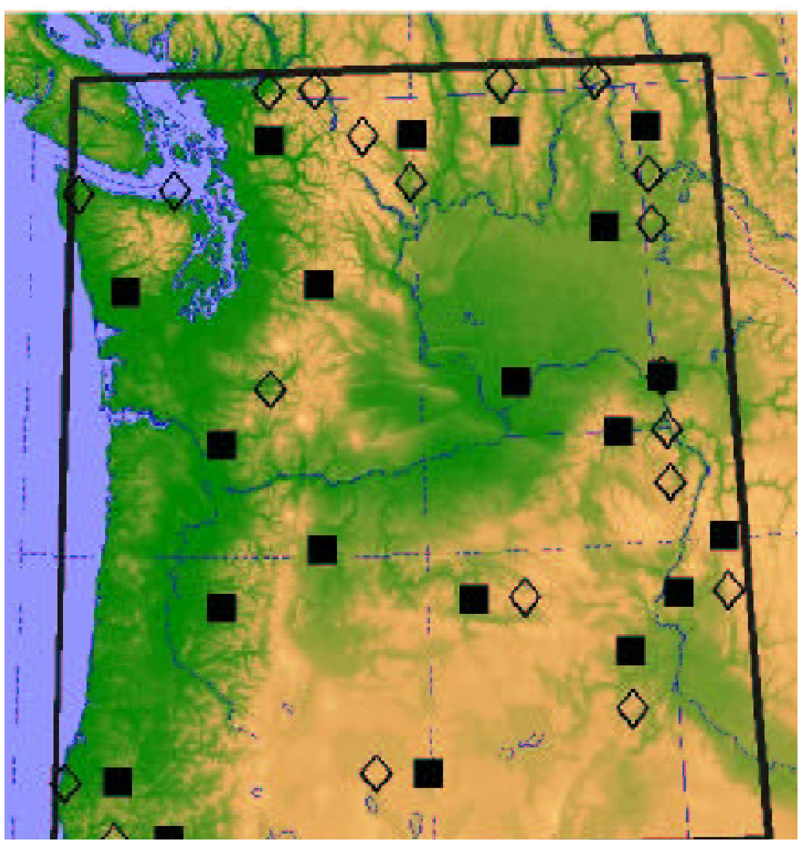
45: Several examples of graphs.

Now, we are going to proceed to connect our understanding of the network models with environmental data. To do so, we will assume that the physical locations we are interested in have some mesh structure on it, which allows us to define a set of nodes or particular locations. For example, the below figure shows a map with a grid overlaid, where each vertex in the grid might represent a particular location. We will define a network in which the set of notes V will correspond to all the vertices in the grid, represented as the red dots below.



46: A map with a grid on top.

Of course, a grid is not the only option to define the set of nodes. One can also define the nodes in some regions of interest as in the figure below.

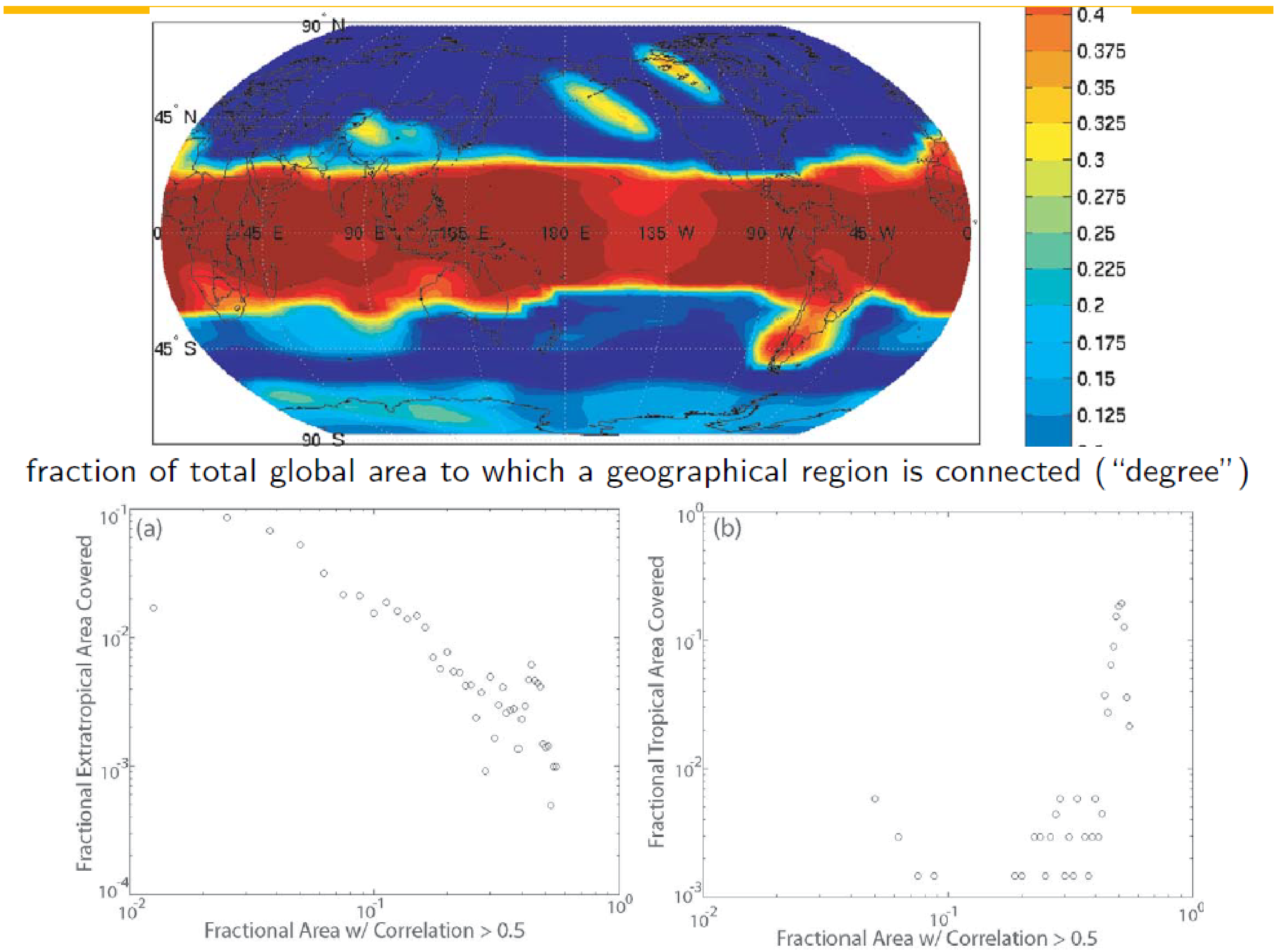


47: Definition of some locations of interest on a map.

The next question is: how do we define the set of edges? One needs to define some criteria by which two nodes, defined above as regions of interest, are connected via an edge.

Assume that for each region or location of interest, for which we have defined a node, we have available a time series of data for the variable we are interested in, such as temperatures or pressures. A time-series indicates some sequence of measurements over a defined period annotated by the time of measurement. We will define an edge between two nodes by identifying some predefined relationships between the two time-series defined at the two nodes if they have “something to do with each other". For example, one can compute the correlation between the two time-series and add an edge between the two points if they are sufficiently correlated, whereby “sufficiently" we mean a design parameter, such as a threshold.

The below figure shows a map with a uniform grid used to defined the set of nodes or regions of interest. Moreover, the correlations of the time measurements at each location was used to build a network where an edge was added when two points generated measurements that were correlated more than certain prescribed threshold. The top image in the figure below shows areas with a higher degree on the corresponding nodes in the network as red points, and areas with lower degrees are shown as blue areas. One can observe that in general, the area on the equator is a band with high degree nodes, whereas areas outside the equator have lower degrees with some exceptions. This implies that areas near the equator have a much higher correlation with other areas than off-the-equator zones.



48: Degree distributions of a network build from spatial correlations.

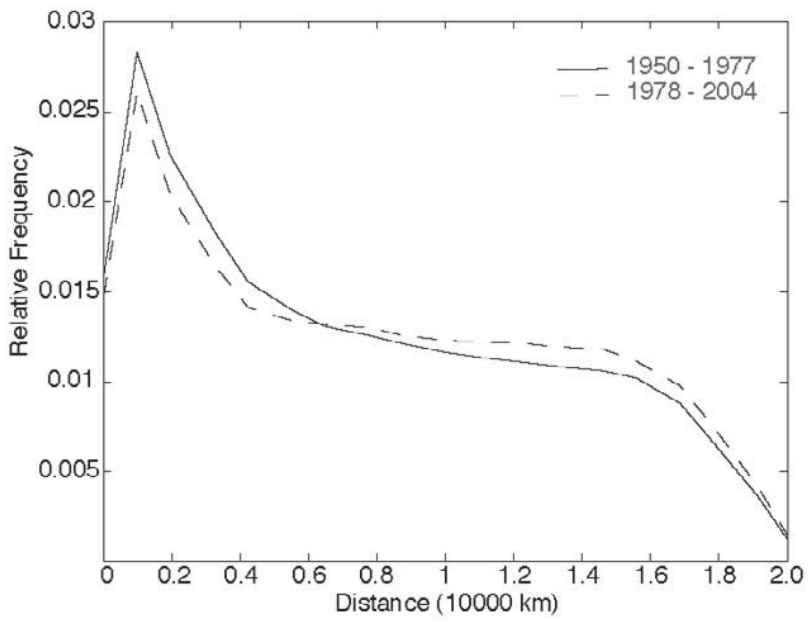
The two bottom plots in the above figure show the degree distribution of the two zones where the degree is high and low, respectively, namely: the red equator area and the blue off-the-equator area. The x-axis is a measure related to the degree, and the y-axis is the frequency of that measure. The measure is essentially the degree divided by the total number of nodes.

The bottom right image shows the degree distribution of the tropical (equatorial) area, where we can observe that there is some uniformity on the values of the degrees. All possible degrees appear at approximately the same frequency. If all nodes have the same number of neighbors or connections in a graph, such a graph is called a regular graph. On the other hand, the extratropical area shows that many nodes have a low degree, while just a few have a large degree — similar to a scale-free or power-law like network.

Note that it is important to define the threshold by which we have defined an edge to represent a “sufficiently" strong correlation. In the simulations in the above figure, the threshold in correlation was chosen to be 0.5.

However, such a parameter is data-dependent and needs to be selected appropriately.

Another analysis we can do on networks built from correlations on time-series data is to explore how the connectivity of the generated networks change with time. The below figure shows the edge length of the networks generated with the same data source as in the above figure with the difference that now, the data was split into two time periods.



49: Variations in the edge length over time.

We can see that for the later time period, the edges, and thus the correlations, became longer in distance. One could also generate a similar time comparison for the degree distribution.

Discussion

Hide Discussion

Topic: Module 5: Environmental Data and Gaussian Processes:Sensing and Analyzing global patterns of dependence / 5. Long-range Climate Correlations

Add a Post

< Previous Next >

edX

[About](#)
[Affiliates](#)
[edX for Business](#)
[Open edX](#)
[Careers](#)
[News](#)

Legal

[Terms of Service & Honor Code](#)
[Privacy Policy](#)
[Accessibility Policy](#)
[Trademark Policy](#)
[Sitemap](#)

Connect

Blog
Contact Us
Help Center
Media Kit
Donate

GET IT ON Google play
Download on the App Store