OPEN

Networks from Flows - From Dynamics to Topology

Nora Molkenthin[1,2], Kira Rehfeld[1,2,4], Norbert Marwan[1] & Jürgen Kurths[1,2,3]

[1]Potsdam Institute for Climate Impact Research, P.O.Box 601203, 14412 Potsdam, Germany, [2]Department of Physics, Humboldt-Universität zu Berlin, Newtonstr. 15, 12489 Berlin, Germany, [3]Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB243UE, UK, [4]Alfred-Wegener-Institute for Polar and Marine Research, Telegrafenberg A43, 14473 Potsdam, Germany.

Complex network approaches have recently been applied to continuous spatial dynamical systems, like climate, successfully uncovering the system's interaction structure. However the relationship between the underlying atmospheric or oceanic flow's dynamics and the estimated network measures have remained largely unclear. We bridge this crucial gap in a bottom-up approach and define a continuous analytical analogue of Pearson correlation networks for advection-diffusion dynamics on a background flow. Analysing complex networks of prototypical flows and from time series data of the equatorial Pacific, we find that our analytical model reproduces the most salient features of these networks and thus provides a general foundation of climate networks. The relationships we obtain between velocity field and network measures show that line-like structures of high betweenness mark transition zones in the flow rather than, as previously thought, the propagation of dynamical information.

Complex networks allow to study underlying interaction structures of dynamical systems, where a detailed description of structure and dynamics may be impossible due to chaotic or otherwise complex behaviour[1–3]. In recent years complex networks have also found fruitful application in climate science, where the reduction of an inaccessibly complex continuous system to a discrete complex network helps to find large-scale interaction structures, that can not be found with conventional methods. Climate networks have provided important insights regarding various questions in climate sciences, ranging from the impact of the El Niño Southern Oscillation on global climate[4–6], to the dynamics of the Asian monsoons[7,8], ocean[9,10] and atmospheric dynamics[11,12]. Usually, these spatially embedded climate networks are constructed from nodes corresponding to (geographical) locations, and links corresponding to statistical interdependence between climate time series observed at the locations of the node[4,8,13–15]. The strength of statistical dependence is often[9,6,10,15] quantified using Pearson correlation[16]. Such statistical interdependences between climatic variables can occur due to a common driver (i.e. solar forcing) or be the result of a physical connection, which can be direct (i.e. ocean flows) or more indirect (i.e. teleconnections[17]). In our model, we focus on direct local and causal connections, where statistical interdependences imply the existence of dynamical exchanges of energy and matter mediated by the underlying flow field. In the real-world application on the Pacific ocean we choose a region and time window, such that common drivers and indirect connections are unlikely to play a dominant role.

While physical mechanisms were indicated in some studies[11,7,9], the dynamical processes behind the obtained networks, and the relationship between the underlying flow's dynamics and the network measures, have remained largely unclear. Donges et al.[9] observed in the climate network of global surface air temperature a striking resemblance between line-like structures of high betweenness and the locations of major ocean currents. They hypothesized that this "backbone of the climate network" resulted from atmospheric-oceanic coupling and the fact that "surface ocean currents play a major role in the energy and information transfer in the climate system". In this letter, we propose a fundamental analytical model to study transport mechanisms in synoptic currents with which we test this hypothesis and bridge the gap in reasoning between the system's dynamics and network results.

Instead of reconstructing the network from observations as in previous climate network studies, we compute the network directly and analytically from the underlying flow field and thus connect flow dynamics to the topology of networks. In spirit, this bottom-up approach from a model to the network, is similar to[18,19], but applies to very different systems. We provide a general framework that allows the inference of the topology of networks based only on the velocity field in the fluid. This way, the network approach can be used to compare the network imprint of the dynamics of the real-world climate system to that of a model system displaying an idealized flow with controlled dynamics.

In climate, advection and diffusion in atmospheric and oceanic flows are amongst the most important mechanisms by which dynamics are mediated. Temperature dispersion in such flows is governed by the advection diffusion equation (ADE). As an abstraction of local random fluctuations in the temperature field on such a flow, we compare the decay of temperature δ-peaks over time throughout the flow. We use them to define a cross-correlation analogue (CCA), given in equation (3), as the continuous normed scalar product between the temperature development of a tracer peak, evaluated at locations x_1 and x_2. We apply this method to derive networks i) analytically for a homogeneous and a circular flow, ii) from two more complex paradigmatic flows (using an appropriate approximation for the solution of the differential equation) and iii) from observed surface temperature data from the equatorial Pacific (using Pearson correlation) and investigate the relationships between the properties of the velocity field of the underlying flow and the observed network.

Results

Application to paradigmatic flows. We constructed flow networks analytically for homogeneous and circular flows and found that the links are longer in flow direction than perpendicular to it (in the supplementary material (SM)). For more complex flows $\vec{v}(\vec{x})$ we have to use an approximate solution to the ADE, as a direct analytic derivation is not possible any more. We assume that these stationary flows vary slowly over space, $|\nabla \vec{v}(\vec{x})| \ll \chi$ (we have used throughout this paper $\chi = 1$), and all derivatives of the velocity field are ignored in the following. This is necessary to ensure the applicability of the approximation. In d dimensions this gives us the approximated temperature field

$$T_{appr}(\vec{x},t;\vec{x}_0) = \frac{e^{-\frac{|\vec{x}-\vec{x}_0-\vec{v}(\vec{x})t|^2}{4\chi t}}}{\sqrt{4\pi\chi t}^d},\qquad(1)$$

where χ is the diffusivity of the fluid. To evaluate the validity of this assumption for a given velocity field $\vec{v}(\vec{x})$, we compute a diagnostic residual R from:

$$R(\vec{x},t) = \frac{\partial T_{appr}}{\partial t} - \chi \Delta T_{appr} + \nabla \cdot (\vec{v}(\vec{x}) T_{appr}).\qquad(2)$$

$R(\vec{x},t)$ is zero for a perfect solution, which is the case if $|\nabla \vec{v}(\vec{x})| = 0$. If the maximum of this function is small compared to the other terms in equation 2, $R(\vec{x},t) \ll \frac{\partial T_{appr}}{\partial t}$, the approximation is considered to be good. For the velocity functions used in this paper this is indeed the case. The spatial integration for the norm is a simple Gaussian integration (SM).

Then the correlation function takes the form (derivation in the Methods section):

$$C(\vec{x}_1,\vec{x}_2) = \frac{\int_{t_0}^{t_1} \iint_{\mathbb{R}^2} \frac{1}{(4\pi\chi t)^2} \frac{1}{t(t+t_l)} e^{-\frac{B}{4\chi t}} d\vec{x}_0 dt}{\frac{1}{8\pi\chi t}(\log(t_1)-\log(t_0))},\qquad(3)$$

where

$$B = |\vec{x}_1 - \vec{x}_0 - \vec{v}(\vec{x}_1)t|^2 + \frac{t}{t+t_l}|\vec{x}_2 - \vec{x}_0 - \vec{v}(\vec{x}_2)(t+t_l)|^2$$

and t_l is defined in equation (9). We now compute the correlations in a grid and connect any pair of sites with a correlation larger than α with a "link". We determine the threshold α such that the link density ρ is constant, $\rho = \frac{L_{net}}{L_{full}}$, where L_{net} denotes the number of links in the flow network and L_{full} denotes the number of links in the fully connected graph with the same nodes. We choose a value for ρ such that
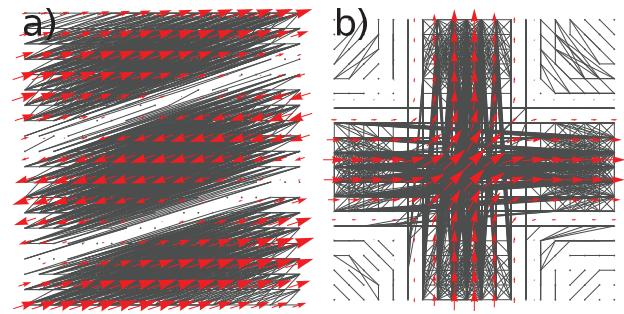

Figure 1 | The correlation network (black) computed from the given velocity field (red arrows) for two flow fields for: (a) Counter-currents, (b) Crossing currents, for better visibility, a low link density of 2 percent was chosen. The networks display longer links in flow direction and a higher link density in regions with higher velocity.

the network has almost no isolated nodes and is sufficiently far from being fully connected. We find this to be the case in a large range of values for ρ and, out of those, we choose $\rho = 0.2$. Our results are robust for a large range of link densities (see section S3 in the SM).

We compute networks for the analytically homogeneous case (SM) and, using numerical integration, for two basic, low-gradient velocity fields given in Fig. 1, where i) one is composed of three narrow parallel flows, with alternating directions, and ii) the other flow is made up of two narrow flows intersecting in the middle. The resulting networks and underlying flows are illustrated in Fig. 1. Please note that the image resolution is equal to the grid resolution in all network figures. In areas of the flow with a higher velocity, the resulting networks show a higher density and length of links than in slower regions. We analyze these networks using the network measures degree k_i (equation (14)) and betweenness centrality b_i (equation (15)), in order to find relationships between them and the underlying velocity field. The network measures are given in Figs. 2 and 3.

We mainly find that high absolute velocity coincides with high node degree. For low velocities, degree and flow speed are approximately proportional, for higher speeds a saturation occurs due to the finite size of the grid (see Fig. S.2 in SM). High values for shortest path betweenness occur in the transition zones between in our case opposing flow directions (Fig. 2), or regions of distinctly different flow velocities (Fig. 3). In both cases, the regions of highest betweenness outline the underlying velocity field. The position of the high betweenness zone depends on the value of the threshold (link density), a lower threshold increases the size of the well-connected region and pushes the transition zone further out.

Other network measures such as local clustering coefficient or local assortativity[3] yield structures similar to that of the node degree (results not shown).

Application to ocean data. In the next step, we compute correlation networks from sea surface temperatures (SST) in the tropical Pacific and compare them with measured ocean currents velocity field (see data description in Methods). Flow velocity, gradient, and the obtained network measures are given in Fig. 4. To suppress turbulent effects, we use only the longitudinal component of the gradient. As for the paradigmatic flows we investigated earlier, also here we find a reasonable agreement between the absolute values of the velocity field and the degree in the correlation network (Fig. 4 (a) and (b), also Fig. S.3 in SM). Again, the degree is maximal where the current's velocity is, and the betweenness shows large values in regions with large values of the longitudinal velocity field's absolute gradient (Fig. 4 (c) and (d)), hereby confirming the results obtained for the paradigmatic flows.
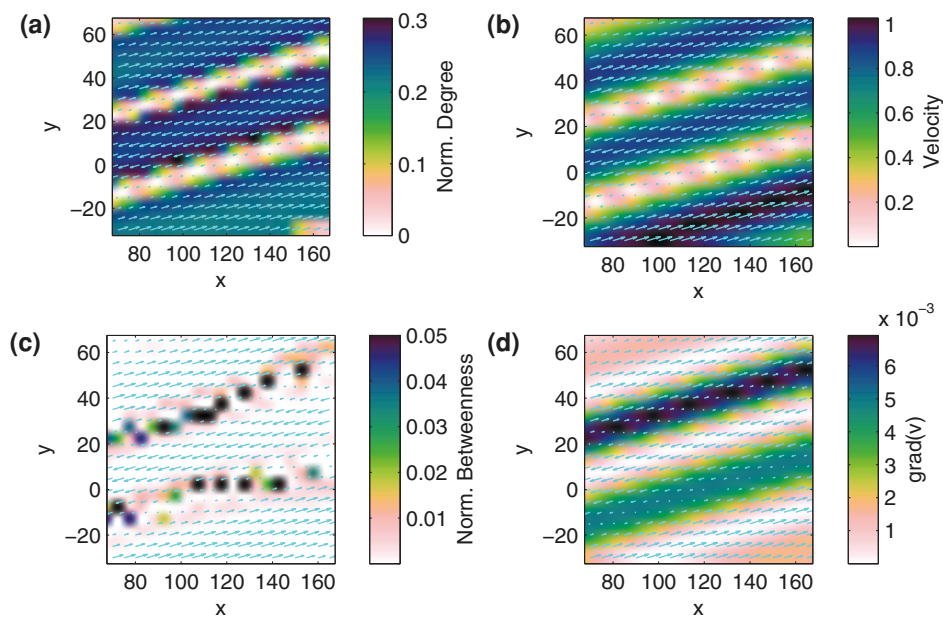
Figure 2 | **Flow field and network measures for the counter-currents in Fig. (1a).** (a) The normed degree, relates to (b) the absolute value of the flow's local velocity; (c) The maxima of the normed betweenness are co-located with (d) the maxima of the absolute value of the gradient gradient of the absolute current velocity. See equations (14) and (15) for definitions of the network measures.

Discussion

In this paper we have established a connection between data networks and the underlying physical system. The approach can easily be generalized beyond 2D static flows and to flow systems outside of climate science, as temperature can be replaced by any quantity described by the heat equation such as density or chemical concentrations. In multivariate settings, reaction, advection and diffusion processes could be studied simultaneously. Given sufficient computing resources, non-stationary flows $\vec{v}(\vec{x}, t)$ could be treated similarly, using a time offset for integration range and peak appearance, as the ADE can still be solved analytically for time dependent velocity fields. This could give new insights in the dynamics of evolving flows, highly valuable not only in the analysis of changing climates.

The line-like structures in the betweenness fields of global climate networks[9] were previously attributed to "information flow" in underlying ocean currents. We found that regions of high betweenness outline the flow rather than tracing it. Our results therefore suggest some corrections concerning the former interpretation and suggest that a high betweenness occurs in transition zones between regions of different magnitude or direction of the underlying velocity. This qualitative observation can be seen when comparing the betweennes with the absolute gradient. Physically, this could be due to the fact that advection dominates in fast flowing regions, which results in a higher parallel but lower perpendicular link density compared to the stagnant case. At the same time, we observe a correlation between regions with a high node degree and high average current velocity. Considering the advective-diffusive nature of these surface
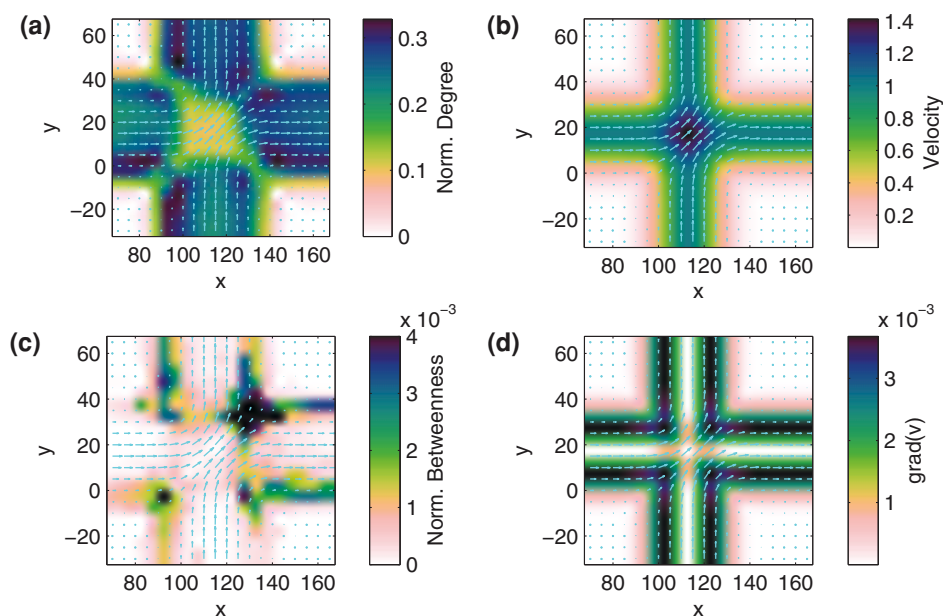

Figure 3 | **Flows and network measures for the crossing currents, see captio of Fig. 2.**
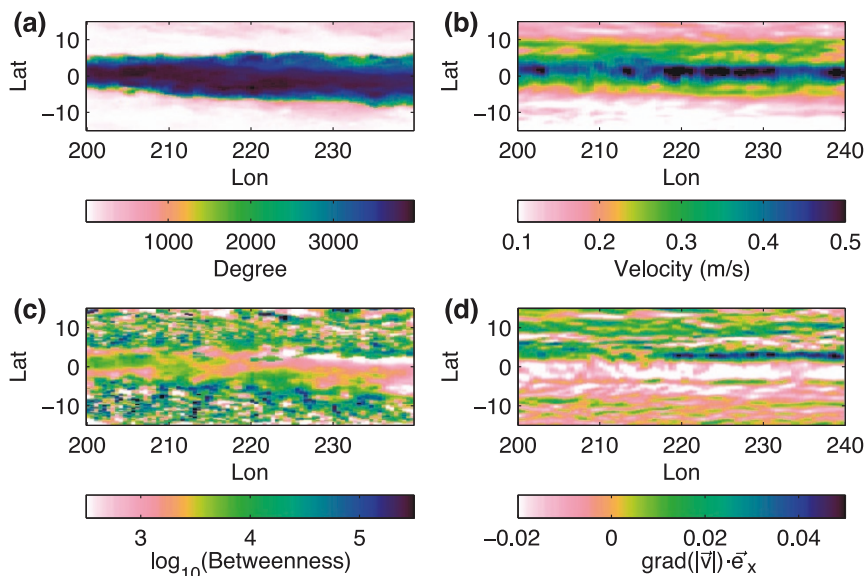
Figure 4 | Network measures of the correlation network of the equatorial counter-currents from 1997 daily anomaly SST data in comparison with flow velocity and gradient. The region of highest degrees coincides with the region of highest flow velocity, while the regions of highest betweenness coincide with the highest velocity gradient.

currents, a physical explanation could be that a fast flow transports the signal farther.

We find that both, the degree and the betweenness increase marginally along the flow direction. This can be understood as the signals from the slow flowing region first travel through diffusion, once they hit the fast region their main peak will travel downstream (the trajectory approximately follows the red arrow in Fig. 5). This leads to points downstream in the fast flowing area to have connections even to points in the slow region upstream from them, leading to increased degree and betweenness there.

In future research, such idealized case studies may be highly useful to study the influence of spatial embedding, and to test hypotheses
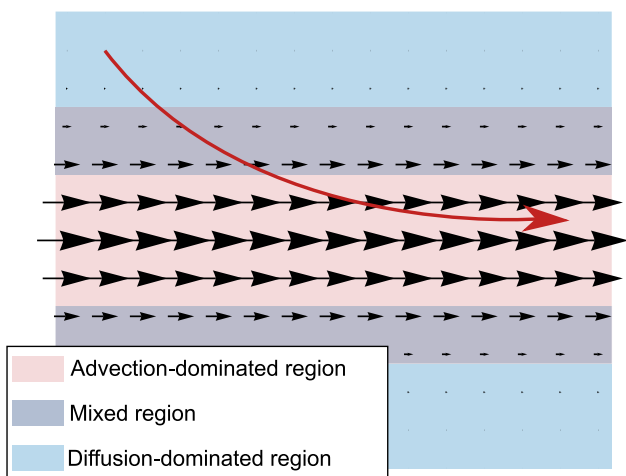

Figure 5 | Schematic illustration of flow properties, that result in distinctive network properties: While advection dominates the transport of temperature fluctuations in regions of fast propagation, localized diffusion dominates in stagnant regions. Signals that leave the stagnant area by diffusion through the mixed region are subsequently transmitted along the flow. This leads to the asymmetry seen in the betweenness, where the betweenness values rise in flow direction.

concerning the dynamics of observed correlation networks. Given sufficiently low-gradient flow data, this method can be used to construct correlation networks from observed oceanic or atmospheric flows.

We have shown how correlation networks can be constructed directly from flow fields and given an example of how to use these networks to interpret network measures. We thereby provide a foundation for climate network analysis and bridge the gap between the dynamics of underlying flows and climate network interpretation.

Methods

Definition of continuous cross-correlation analogue. For the model system we assume stationary two-dimensional flows in a square area in a two-dimensional boundaryless fluid of constant diffusivity χ described by the velocity field $\vec{v}(\vec{x})$. The ADE states how the change of temperature over time is governed by the spatial temperature change and the velocity:

$$\frac{\partial T}{\partial t} = \chi \Delta T - \nabla \cdot (\vec{v}(\vec{x}) T), \tag{4}$$

and is obtained by inserting the advective and diffusive flux

$$\vec{j} = \vec{j}_{diff} + \vec{j}_{adv} = -\chi \nabla T + \vec{v} T \tag{5}$$

into the sourceless continuity equation for temperature

$$\frac{\partial T}{\partial t} = -\nabla \cdot \vec{j}. \tag{6}$$

Here, $T(\vec{x},t)$ is the value of the temperature in position $\vec{x}$ at time t. We use a δ-peak as a tracer of the flow, analogous to local temperature fluctuations. It is inserted at an arbitrary point $\vec{x}_0$ in the fluid as the initial condition, so, in other words, we solve the Cauchy problem of equation (4) with the initial condition

$$T(\vec{x},0;\vec{x}_0) = \delta(\vec{x} - \vec{x}_0). \tag{7}$$

Analogous to the commonly used Pearson correlation[16], we define the continuous cross-correlation analogue (CCA) as the normed scalar product of solutions of the Cauchy problem of the ADE at two points $\vec{x}_1$ and $\vec{x}_2$

$$C(\vec{x}_1,\vec{x}_2) = \left\langle \frac{T(\vec{x}_1,t;\vec{x}_0)}{\|T(\vec{x}_1,t;\vec{x}_0)\|}, \frac{T(\vec{x}_2,t+t_l;\vec{x}_0)}{\|T(\vec{x}_2,t+t_l;\vec{x}_0)\|} \right\rangle. \tag{8}$$

The time lag t_l is the difference in travel time of the peak from $\vec{x}_0$ to $\vec{x}_1$ and $\vec{x}_2$, the norm is defined in the SM in equation (S.1)

$$t_l = t_{max}(\vec{x}_1,\vec{x}_0) - t_{max}(\vec{x}_2,\vec{x}_0), \tag{9}$$

where $t_{max}(\vec{x},\vec{x}_0)$ is the time when the temperature at $\vec{x}$ reaches its maximum, with the initial peak starting at $\vec{x}_0$. The scalar product is then defined as the integral over time and peak position $\vec{x}_0$. The time integration is analogous to the sum over time steps, the

integration over the peak position is an integral over realizations of the peak, corresponding to stochastics in the time series, where peaks appear at random in arbitrary places. So we define the CCA in this context as:

$$C(\vec{x}_1, \vec{x}_2) = \frac{\int_{t_0}^{t_1} \int_{\mathbb{R}^2} T(\vec{x}_1, t; \vec{x}_0) T(\vec{x}_2, t + t_l; \vec{x}_0) d\vec{x}_0 dt}{\sqrt{\|T(x_1, t; \vec{x}_0)\|} \sqrt{\|T(x_2, t + t_l; \vec{x}_0)\|}}, \quad (10)$$

where $\vec{x}_0$ is the position of the peak. The lower limit of the integration, t_0 is chosen small but non-zero (here $t_0 < 10^{-2}$) as the correlation function is not defined for $t = 0$. The upper limit is chosen such that all temperature profiles have decayed to a value very close to zero (here: $t_1 = 5000$).

Network construction. We evaluate the CCA on a regular grid between all pairs of grid-points. This provides the correlation matrix C_{ij} from which the adjacency matrix A is constructed by choosing a fixed significance threshold α (see "results" section on page three). This can be expressed with the Heaviside θ function and Kronecker δ as

$$A_{ij} = \Theta\left(C_{ij} - \alpha\right) - \delta_{ij}. \quad (11)$$

For any given flow field, we first have to solve the ADE (equation (4)) and use the result to compute the correlation matrix using equation (10). The Cauchy problem of the ADE in d dimensions can be solved as

$$T(\vec{x}, t; \vec{x}_0) = \frac{e^{-\frac{|\vec{x} - \vec{x}_0 - \vec{v}t|^2}{4\chi t}}}{\sqrt{4\pi \chi t^d}} + T_0 \quad (12)$$

in the homogeneous case with the velocity field $\vec{v}(\vec{x}) = \vec{v}^{20}$. This solution takes its maximum value at

$$t_{max}(\vec{x}, \vec{x}_0) = \frac{\sqrt{\chi^2 + |\vec{v}|^2 |\vec{x} - \vec{x}_0|^2}}{|\vec{v}|^2}. \quad (13)$$

Network measures. To analyze the networks, we used the basic network measures degree and betweenness[2], as normalized measures to account for grid size effects:

The **degree** k_i of node i of a network with N nodes is given by the number of links attached to it,

$$k_i = \frac{\sum_{j=1}^{N} A_{ij}}{N - 1}, \quad (14)$$

and the shortest path **betweenness** b_i of a node i is defined as the number of all shortest paths that go through it,

$$b_i = \frac{\sum_{j, k \in 1, \dots, N, j \neq k} \frac{n_{jk}(i)}{n_{jk}}}{(N - 1)(N - 2)}. \quad (15)$$

Where n_{jk} is the number of shortest paths connecting k and j and $n_{jk}(i)$ is the number of those paths, that go through i.

Data. The daily anomaly SST data is based on the optimum interpolation data (OI.v2) as provided by NOAA/NCDC[21,22] and the averaged monthly current's velocity data was provided by the OSCAR Project Office (Earth and Space Research, Seattle). We used data from the region $120°$–$160°W$, $15°S$–$15°N$ and for the time period August 1996 to August 1997. The chosen year is neither an El Niño nor a La Niña-year, and the results we present in the following are largely robust against the choice of the particular year. The network is calculated by standard cross-correlation and allowing for a lag of up to one day.

1. Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–76 March (2001).
2. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. Complex networks: Structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).
3. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
4. Tsonis, A. A., Swanson, K. L. & Roebber, P. J. What Do Networks Have to Do with Climate? *Bull. Am. Meteorol. Soc.* **87**, 585–595 (2006).
5. Gozolchiani, A., Havlin, S. & Yamasaki, K. Emergence of El Niño as an Autonomous Component in the Climate Network. *Phys. Rev. Lett.* **107**, 1–5 (2011).
6. Berezin, Y., Gozolchiani, A., Guez, O. & Havlin, S. Stability of climate networks with time. *Sci. Rep.* **2**, 666 (2012).
7. Malik, N., Bookhagen, B., Marwan, N. & Kurths, J. Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. *Clim. Dyn.* **39**, 971–987 (2011).
8. Rehfeld, K., Marwan, N., Breitenbach, S. F. M. & Kurths, J. Late Holocene Asian summer monsoon dynamics from small but complex networks of paleoclimate data. *Clim. Dyn.* **41**, 3–19 (2012).
9. Donges, J. F., Zou, Y., Marwan, N. & Kurths, J. The backbone of the climate network. *Europhys. Lett.* **87**, 48007 (2009).
10. Guez, O., Gozolchiani, A., Berezin, Y., Brenner, S. & Havlin, S. Climate network structure evolves with North Atlantic Oscillation phases. *Europhys. Lett.* **98**, 38006 (2012).
11. Boers, N., Bookhagen, B., Marwan, N., Kurths, J. & Marengo, J. Complex networks identify spatial patterns of extreme rainfall events of the South American Monsoon System. *Geophys. Res. Lett.* **40**, 4386–4392 (2013).
12. Ebert-Uphoff, I. & Deng, Y. Causal Discovery for Climate Research Using Graphical Models. *J. Clim.* **25**, 5648–5665 (2012).
13. Yamasaki, K., Gozolchiani, A. & Havlin, S. Climate Networks around the Globe are Significantly Affected by El Niño. *Phys. Rev. Lett.* **100**, 228501 (2008).
14. Tsonis, A. A., Wang, G., Swanson, K. L., Rodrigues, F. A. & Costa, L. D. F. Community structure and dynamics in climate networks. *Clim. Dyn.* **37**, 933–940 (2010).
15. Steinhaeuser, K., Chawla, N. V. & Ganguly, A. R. An exploration of climate data using complex networks. *ACM SIGKDD Expl. Newsl.* **12**, 25 (2010).
16. Chatfield, C. *The analysis of time series: an introduction.* (CRC Press, Boca Raton, 2003).
17. Castrejón-Pita, A. A. & Read, P. L. Synchronization in a Pair of Thermally Coupled Rotating Baroclinic Annuli: Understanding Atmospheric Teleconnections in the Laboratory. *Phys. Rev. Lett.* **104**, 204501 (2010).
18. Timme, M. Revealing Network Connectivity from Response Dynamics. *Phys. Rev. Lett.* **98**, 224101 (2007).
19. Shandilya, S. G. & Timme, M. Inferring network topology from complex dynamics. *New J. Phys.* **13**, 013004 (2011).
20. Lebon, G., Jou, D. & Casas-Vàzquez, J. *Understanding non-equilibrium thermodynamics: foundations, applications, frontiers.* (Springer, Berlin, 2008).
21. Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S. & Schlax, M. G. Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *J. Clim.* **20**, 5473–5496 (2007).
22. NOAA/NCDC. NOAA Optimum Interpolation 1/4 Degree Daily Sea Surface Temperature Analysis. http://www.ncdc.noaa.gov/oa/climate/research/sst/oi-daily-information.php, (accessed 08.08.2013).

Acknowledgments

Author contributions

N.M., K.R., N.M. and J.K. designed and performed the research, analyzed the results and wrote the paper.

Additional information