Teleconnection Paths via Climate Network Direct Link Detection

Dong Zhou,[1,*] Avi Gozolchiani,[2] Yosef Ashkenazy,[2] and Shlomo Havlin[1]
[1]Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel
[2]Department of Solar Energy and Environmental Physics, BIDR, Ben-Gurion University, Midreshet Ben-Gurion 84990, Israel
(Received 7 September 2015; published 30 December 2015)

Teleconnections describe remote connections (typically thousands of kilometers) of the climate system. These are of great importance in climate dynamics as they reflect the transportation of energy and climate change on global scales (like the El Niño phenomenon). Yet, the path of influence propagation between such remote regions, and weighting associated with different paths, are only partially known. Here we propose a systematic climate network approach to find and quantify the optimal paths between remotely distant interacting locations. Specifically, we separate the correlations between two grid points into direct and indirect components, where the optimal path is found based on a minimal total cost function of the direct links. We demonstrate our method using near surface air temperature reanalysis data, on identifying cross-latitude teleconnections and their corresponding optimal paths. The proposed method may be used to quantify and improve our understanding regarding the emergence of climate patterns on global scales.

DOI: 10.1103/PhysRevLett.115.268501 PACS numbers: 92.60.hv, 05.40.-a, 89.75.-k

Teleconnections refer to persistent relations between climate anomalies of geographically separated regions [1–5]. Atmospheric or oceanic teleconnections dominate the climate variability on different time and spatial scales [1]. Teleconnections received much attention since they stand for the transport of energy and the climate dynamics on global scales (typically thousands of kilometers); as such they probably have an important role in global climate change and variability [1,6].

Various teleconnection patterns represent different preferred modes of large scale variability, in both the tropics and the extratropics [1,2,7]. These patterns can be captured by various methods, such as teleconnection maps based on Pearson correlation (linear) or mutual information (nonlinear), and principal component analysis [2,8]. The most prominent atmospheric patterns include the El Niño–Southern Oscillation (ENSO), the Arctic Oscillation (AO), the North Atlantic Oscillation (NAO), the North Pacific Oscillation–West Pacific pattern (NPO-WP), the Pacific–North America pattern (PNA), the Southern Annular Mode (SAM) and the Pacific–South American pattern (PSA) [1,2,7,9–12]. Some of these patterns influence each other such as the teleconnections that are related to ENSO [1,2,6].

Generally speaking, most teleconnections are caused by the transport of energy and propagation of waves [1]. However, the mechanisms of specific teleconnections are still not well understood, since these known patterns describe the mature phase of the variability instead of the developing phase [2]. Therefore, deeper investigations on the evolutionary process in the spatiotemporal domain are needed [2].

In the recent decade, climate network methods have been introduced to describe the spatiotemporal behavior of climate interactions. In climate networks, nodes represent geographical sites, and links are generated using correlation, mutual information, or other dependence measures, between the time series of two nodes [13–15]. Networks constructed in this way were already utilized to explore climate mechanisms or to predict extreme events [13–23].

Paths of influence that are associated with the teleconnections have not been analyzed systematically using climate network approaches. Naively applying search algorithms for these paths is, however, not possible. The reason being that most links in networks that are inferred from time series arise, in fact, mainly due to indirect interactions between two nodes through intermediate nodes or from "common-driver" nodes. This issue has arisen, and was shown to be important in various different disciplines (see, e.g., [24–30]). Different approaches to tackle this problem include partial correlation, conditional mutual information, and conditional Granger causality [26–31]. In the context of climate networks, this problem was thoroughly discussed in [32–34]. Runge *et al.* [32,34] also note that direct interactions need not be instantaneous but, instead, can bear time lags.

In this Letter, we develop a modified partial correlation method to calculate direct links. We identify the optimal direct link paths for the observed teleconnections by minimizing the total cost function based on direct link weights. This approach enables us to capture the dominant propagation paths associated with teleconnections.

Our analysis is based on global coverage of near surface (1000 hPa) daily mean air temperature data from the National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP-NCAR) reanalysis project [35]. The time series span 63 years (1948–2010). The original resolution of the data is 2.5° and

we consider here a subset of $n = 726$ nodes that approximately equally cover the globe; the distance between neighboring nodes is around 830 km in both zonal and meridional directions (see the node set in Fig. S1 in the Supplemental Material [36]).

From the temperature time series we obtain the temperature anomaly $S_i(t), t = 1, 2, \ldots, 63 \times 365$, by subtracting the calendar day's mean and the linear trend over 63 years. For each pair of nodes i and j, we calculate the cross-correlation between the two time series, $X_{i,j}(\tau)$, where $\tau = -\tau_{\max}, \ldots, \tau_{\max}$ is the time lag (in days), and $\tau_{\max} = 800$ days. To avoid edge effects, we use the anomaly time series from day $(\tau_{\max} + 1)$ to day $(63 \times 365 - 2\tau_{\max})$, i.e., total length $L = 20\,595$ days. Next, the observed positive and negative link weights are defined as $W_{i,j}^{\text{obs,pos}} = [(\max(X_{i,j}) - \text{mean}(X_{i,j}))/\text{std}(X_{i,j})]$, and $W_{i,j}^{\text{obs,neg}} = [(\min(X_{i,j}) - \text{mean}(X_{i,j}))/\text{std}(X_{i,j})]$ [17]. The observed time delays $\tau_{i,j}^{\text{obs,pos}}$ and $\tau_{i,j}^{\text{obs,neg}}$ are the values of τ where $X_{i,j}(\tau)$ reaches its maximum and minimum, respectively. Note that Runge *et al.* studied direct link strengths and time lags based on a time series graph method [32,34]. The analysis of Runge *et al.* differs from ours. First, we focus on finding paths of influence between distant sites interacting by teleconnections. Second, we engage in our analysis daily (compared to monthly in [34]) records, and a larger set of nodes covering the globe, since our aim is to study the teleconnections. Biases in the link identification due to strong autocorrelations are basically canceled due to the denominator, $\text{std}(X_{i,j})$, in the above definitions (see [37] for a deeper view on this issue, which is tackled somewhat differently also in [34]).

Here we base the removal of indirect effects and calculation of direct links on the classical partial correlation approach. For each pair of nodes i and j, we denote the time delay corresponding to the maximum of the absolute cross-correlation value $|X_{i,j}|$ by $\tau_{i,j}^{\text{obs,abs}}$. Then, we perform multivariate linear regressions for the time series $S_i(t)$ and $S_j(t)$:

$$S_i(t) = \sum_{k \neq i,j} a_{ik} S_k(t + \tau_{i,k}^{\text{obs,abs}}) + b_i + \varepsilon_i(t), \qquad (1)$$

$$S_j(t) = \sum_{k \neq i,j} c_{jk} S_k(t + \tau_{j,k}^{\text{obs,abs}}) + d_j + \varepsilon_j(t), \qquad (2)$$

where $t = \tau_{\max} + 1, \tau_{\max} + 2, \ldots, L + 2\tau_{\max}$. By the above multilinear regressions we remove the indirect effects of all the other nodes k from the time series of nodes i and j, and obtain two new time series $\varepsilon_i(t)$ and $\varepsilon_j(t)$. Based on $\varepsilon_i(t)$ and $\varepsilon_j(t)$, we define the positive and negative direct link weights $W_{i,j}^{\text{dir,pos}}$ and $W_{i,j}^{\text{dir,neg}}$, as well as the corresponding time delay values $\tau_{i,j}^{\text{dir,pos}}$ and $\tau_{i,j}^{\text{dir,neg}}$, similarly to the observed link definition. The main difference between this method and the classical partial correlation method is that we introduce the observed time delay $\tau_{i,k}^{\text{obs,abs}}$ and $\tau_{j,k}^{\text{obs,abs}}$ into the sum over all the other nodes k, instead of zero lag in classical methods [38]. The introduction of the delay is necessary to capture the strongest relation between nodes, a relation (correlation) that can be either positive or negative.

Figures 1(a) and 1(b) show the two-dimensional distribution of direct and observed link weights for positive and negative links, respectively. Figures 1(c) and 1(d) further show the probability density function (PDF) of these observed or direct positive and negative links. It is clear that there is no strong relation between the observed and direct link weights as a wide range of observed link weights is associated with the same direct link weight. Figure 2 depicts the link weights (both positive and negative) as a function of the geographic distances $D_{i,j}$ of the global observed and direct links. The results are in accordance with the results of [17] that associate the distance between peaks in Figs. 2(a) and 2(b) with Rossby wave wavelength. However, as shown in Figs. 2(c) and 2(d), these peaks are not present in the direct link network. Strong direct positive links are related to short distances (neighboring nodes), suggesting that the long observed positive links are the results of information propagation through short direct positive links. Below, we investigate the direct link paths for the observed positive teleconnections.

To isolate the significant long-distance and cross-latitude links (teleconnections), we focus on positive links with (i) distance $D_{i,j} \geq 5000$ km, (ii) link weight $W_{i,j}^{\text{obs,pos}} \geq 9$ [as depicted by the dashed line in Fig. 2(a)], and (iii) latitude distance larger than 20°. These 226 observed links (out of $726 \times 725/2 = 263\ 175$ possible links) are depicted in Fig. 3. It is apparent that many of them are along the

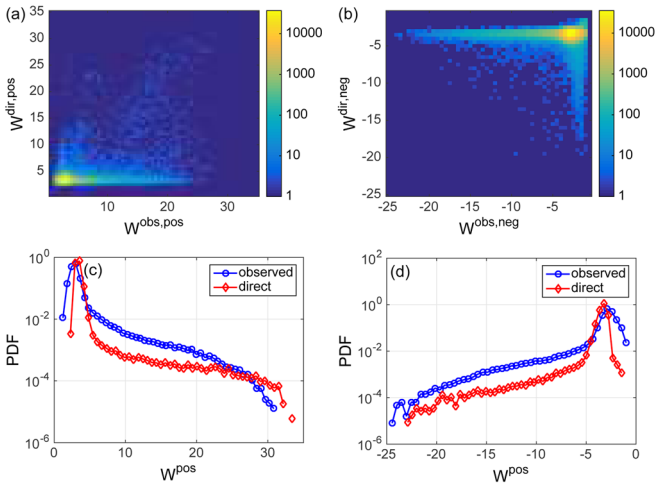

FIG. 1 (color online). (a) The direct positive link weight $W_{i,j}^{\text{dir,pos}}$ versus the observed one $W_{i,j}^{\text{obs,pos}}$. Two-dimensional frequency of links in the $W_{i,j}^{\text{obs,pos}}$ versus $W_{i,j}^{\text{dir,pos}}$ space. (b) Same as (a) for negative links. (c) PDF of $W_{i,j}^{\text{obs,pos}}$ (blue) and $W_{i,j}^{\text{dir,pos}}$ (red). (d) PDF of $W_{i,j}^{\text{obs,neg}}$ (blue) and $W_{i,j}^{\text{dir,neg}}$ (red).

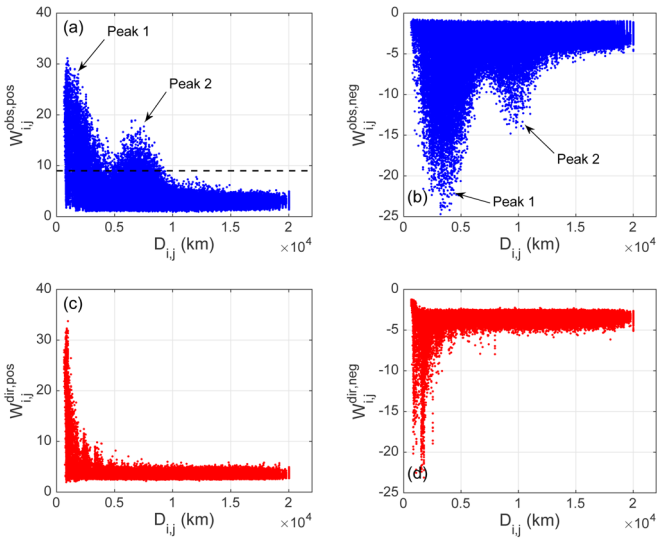

FIG. 2 (color online). Link weights as a function of geographical distances. (a) Observed positive links. The horizontal dashed line at $W_{i,j}^{\mathrm{obs,pos}} = 9$ is the threshold for the observed positive link weights we considered later in Fig. 3. The two arrows indicate the two positive link peaks related to Rossby waves. (b) Observed negative links. The two arrows indicate the two negative link peaks related to Rossby waves. (c) Direct positive links. (d) Direct negative links.

pathway of atmospheric Rossby waves, especially in the southern hemisphere. We focus on two examples of teleconnections which are marked by red in Fig. 3. Example 1 is the connection between the Norwegian Sea and southern Japan. Example 2 connects the Drake Passage with western Australia. We consider these two teleconnections since they are typical examples that are related to different climate processes, and they enable detailed comparison with the previous literature. To identify the dominant direct link paths between two nodes, we consider a subset of direct positive links which satisfy (i) distance $D_{i,j} \leq 1000$ km, and (ii) link time delay $\tau_{i,j}^{\mathrm{dir,pos}} \geq 0$. The second condition

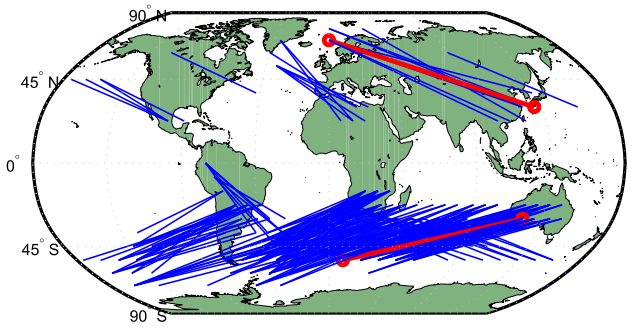

FIG. 3 (color online). 226 strong and long observed positive links which have (i) distance larger than 5000 km, (ii) $W_{i,j}^{\mathrm{obs,pos}} \geq 9$, and (iii) latitude difference above 20°. The red circles and lines indicate the two examples considered in the text and in Fig. 4.

ensures that all steps have the same direction (with nonnegative time delay). In this subset of weighted and directed links, we use $1/W_{i,j}^{\mathrm{dir,pos}}$ as the cost value for each link. For each pair of nodes i and j, we can determine the directed optimal path (sum of costs minimal) using the Dijkstra algorithm [39]. In this way, we identify the preferred paths for information spreading in the two-dimensional space of near surface air temperature. In addition, the related measure of "betweenness" (basically the number of shortest paths through a node) is presented in the Supplemental Material [36] (Fig. S1). The shortest paths represent the optimal paths of transfer between specific pairs of nodes, while nodes with the largest betweenness values play the most important role in the global information flows. Here, we mainly focus on positive teleconnections. Two more positive teleconnections seen in Fig. 3 are discussed in the Supplemental Material [36] (Figs. S2 and S3). In fact, our approach is also applicable for negative teleconnections that are related to wave propagations. After propagating along the wave for a period of time (corresponding to half wavelength), although each step is a positive link, one should observe a negative long link when this dominates the cross-correlation. The optimal paths in these cases can be captured similarly to positive teleconnections. We show two examples of negative teleconnections in the Supplemental Material [36] (Figs. S4 and S5).

Figures 4(a) and 4(c) show the optimal direct link paths for the two examples of observed positive teleconnections that are indicated by the red color in Fig. 3. Figures 4(b) and 4(d) depict the corresponding cross-correlation functions of the observed links (to obtain a clearer view it is limited to $\tau = -50, \ldots, 50$). For each example, we show both the observed link (indicated by the black line) and the directed shortest path (indicated by the colored circles and lines); the arrow indicates the direction of both the observed link and the directed optimal path. The color of each step along the optimal path indicates the "performance ratio," which is the ratio between the weight of the chosen direct link at this step and the mean direct link weight over all neighboring nodes. Therefore, if all neighboring nodes have the same weight the ratio will be 1, and thus a ratio larger than 1 indicates that the chosen link is better than the average at this step. (The extreme maximal ratio is equal to the number of neighboring nodes and it occurs when the weights of the neighboring links except the chosen one are all 0.)

As shown in Figs. 4(a) and 4(c), the observed teleconnection in Example 1 has a six-day time lag from the Norwegian Sea to southern Japan, which agrees with the wave train pattern generated by NAO from the North Atlantic to East Asia [40–42]. The obtained direct link shortest path also follows the pathway of this wave propagation (resembling a much broader path shown in Fig. 2 of [40] and Fig. 3 of [41]). In Example 2, the observed link has a five-day time lag from the Drake

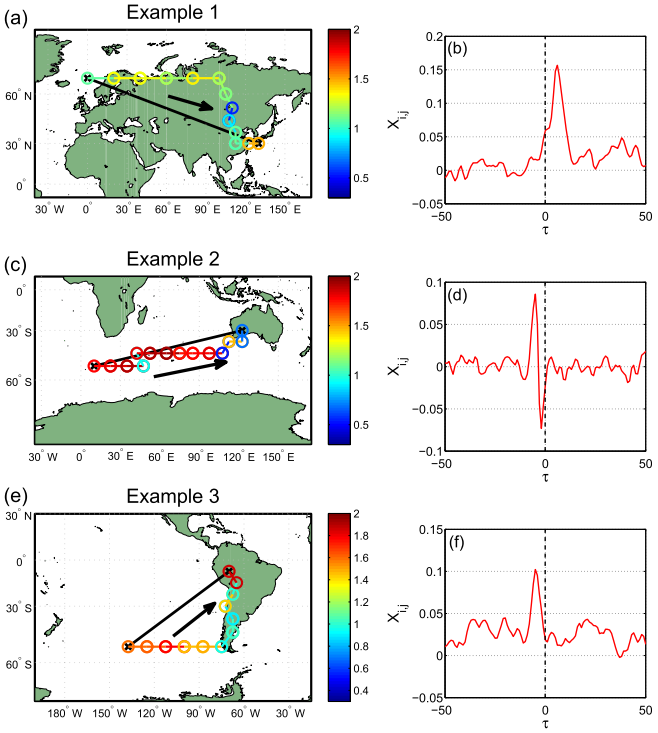

FIG. 4 (color online). Three examples of observed links and their direct link optimal paths. (a) Example 1 (denoted on the upper red circle and line in Fig. 3). (b) Cross-correlation $X_{i,j}$ of the observed link in Example 1. Here i and j are the northern and southern nodes, respectively. (c),(d) Same as (a),(b) but for the second example (denoted on the lower red circle and line in Fig. 3). (e),(f) Same as (a),(b) but for a link that is not included in the links shown in Fig. 3. The time delay values $\tau_{i,j}^{\mathrm{obs,pos}}$ of these three observed teleconnections are 6, −5, and −5 days, respectively. The color bar indicates the performance ratio discussed in the text.

Passage to western Australia, and based on Fig. 4 of [17], it is probably related to Rossby waves activity in the southern hemisphere. The optimal path here also suggests an eastward propagation. Besides the observed teleconnections highlighted in Fig. 3, we consider another example in the South Pacific; the weight value of this link is relatively small, while its maximal cross-correlation is still high. As shown in Figs. 4(e) and 4(f), the observed teleconnection in Example 3 has a five-day time lag, and it is directed from the South Pacific to the northwestern South America. The obtained optimal path follows the mean winds (westerlies) in the region and also the Humboldt Current (the northward current adjusted to the western side of South America).

For these three paths, we calculate both the total cost value and the fraction of steps with performance ratios larger than 1 ("good steps"), which are two different ways to evaluate the optimal paths. In the Supplemental Material [36], we estimate the significance of each path by comparing the total cost with the total costs of the surrogate paths, and we find that all these three optimal paths are significant. As a supplementary measurement of the goodness of the paths, we also find that the fractions of good steps for the three examples are large enough, as they are larger than or equal to the 95% quantile. Moreover, in these three examples, the second shortest paths are very close to the first shortest paths. These three findings (significantly small total costs, large fractions of good steps and similarity between the first and second shortest paths) strengthen the climatic interpretation of our results.

In summary, teleconnections are very important processes in the climate system as they connect very remote regions. It is thus clear that improvement of their characteristics will improve our description and understanding of the climate system. Here, we developed a climate network approach that detects significant teleconnections and their associated optimal paths in a systematic and quantitative way. Our method is based on the partial correlation approach and it is applied on the global near surface air temperature network. Unlike the observed network, the direct link network exhibits strong correlation only between neighboring nodes; we thus relate the observed long links (teleconnections) to the information propagation along direct link paths along neighboring nodes. The shortest path is identified by minimizing a total cost function that is the sum over the reciprocals of the direct link weights. We presented three examples for which the observed teleconnections are related to waves, winds, or currents. We find highly significant direct link optimal paths that capture the preferred propagation paths for these teleconnections. We believe that our approach is unique and that it will facilitate the study of directed paths which might lead to better understanding of the global climate system. Our approach could be also extended to study and improve our knowledge in other fields such as neuroscience, systems biology, and financial systems.

We thank the LINC project (No. 289447) funded by the EC's Marie-Curie ITN program (FP7-PEOPLE-2011-ITN) and the Israel Science Foundation for financial support. We thank Y. Wang and J. Runge for helpful discussions.

*Corresponding author.
zhoudongbnu@gmail.com

[1] Z. Liu and M. Alexander, Rev. Geophys. **72**, Y1 (2007).
[2] S. Nigam and S. Baxter, in *Encyclopedia of Atmospheric Sciences*, 2nd ed. (Elsevier Ltd, London, 2015), Chap. Teleconnections, pp. 90–109.
[3] J. Bjerknes, Mon. Weather Rev. **97**, 163 (1969).
[4] J. M. Wallace and D. S. Gutzler, Mon. Weather Rev. **109**, 784 (1981).
[5] H. von Storch and F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, 1999).
[6] A. A. Tsonis, K. Swanson, and S. Kravtsov, Geophys. Res. Lett. **34**, L13705 (2007).

[7] C. Deser, M. A. Alexander, S.-P. Xie, and A. S. Phillips, Annu. Rev. Mar. Sci. **2**, 115 (2010).
[8] J. Hlinka, D. Hartman, M. Vejmelka, D. Novotná, and M. Paluš, Clim. Dyn. **42**, 1873 (2014).
[9] H. A. Dijkstra, *Nonlinear Climate Dynamics* (Cambridge University Press, Cambridge, 2013).
[10] K. C. Mo and R. W. Higgins, Mon. Weather Rev. **126**, 1581 (1998).
[11] A. Gershunov and T. Barnett, Bull. Am. Meteorol. Soc. **79**, 2715 (1998).
[12] M. A. Alexander, I. Bladé, M. Newman, J. R. Lanzante, N. C. Lau, and J. D. Scott, J. Clim. **15**, 2205 (2002).
[13] A. A. Tsonis, K. L. Swanson, and P. J. Roebber, Bull. Am. Meteorol. Soc. **87**, 585 (2006).
[14] K. Yamasaki, A. Gozolchiani, and S. Havlin, Phys. Rev. Lett. **100**, 228501 (2008).
[15] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, Europhys. Lett. **87**, 48007 (2009).
[16] A. Gozolchiani, S. Havlin, and K. Yamasaki, Phys. Rev. Lett. **107**, 148501 (2011).
[17] Y. Wang, A. Gozolchiani, Y. Ashkenazy, Y. Berezin, O. Guez, and S. Havlin, Phys. Rev. Lett. **111**, 138501 (2013).
[18] N. Boers, B. Bookhagen, N. Marwan, J. Kurths, and J. Marengo, Geophys. Res. Lett. **40**, 4386 (2013).
[19] V. Stolbova, P. Martin, B. Bookhagen, N. Marwan, and J. Kurths, Nonlinear Processes Geophys. **21**, 901 (2014).
[20] M. van der Mheen, H. A. Dijkstra, A. Gozolchiani, M. den Toom, Q. Feng, J. Kurths, and E. Hernandez-Garcia, Geophys. Res. Lett. **40**, 2714 (2013).
[21] J. Ludescher, A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber, Proc. Natl. Acad. Sci. U.S.A. **111**, 2064 (2014).
[22] Q. Feng and H. A. Dijkstra, arXiv:1503.05449.
[23] N. Boers, B. Bookhagen, H. M. Barbosa, N. Marwan, J. Kurths, and J. A. Marengo, Nat. Commun. **5**, 5199 (2014).
[24] R. De Smet and K. Marchal, Nat. Rev. Microbiol. **8**, 717 (2010).
[25] S. Feizi, D. Marbach, M. Médard, and M. Kellis, Nat. Biotechnol. **31**, 726 (2013).
[26] A. de la Fuente, N. Bing, I. Hoeschele, and P. Mendes, Bioinformatics **20**, 3565 (2004).
[27] S. C. Curtin and K. Wilkes, Int. J. Tourism Res. **9**, 131 (2007).
[28] K. Ashok, S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, J. Geophys. Res. C **112**, C11007 (2007).
[29] L. M. de Campos, J. Mach. Learn. Res. **7**, 2149 (2006).
[30] D. Y. Kenett, M. Raddant, L. Zatlavi, T. Lux, and E. Ben-Jacob, Int. J. Mod. Phys. Conf. Ser. **16**, 13 (2012).
[31] R. Salvador, A. Martínez, E. Pomarol-Clotet, S. Sarró, J. Suckling, and E. Bullmore, NeuroImage **35**, 83 (2007).
[32] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Phys. Rev. Lett. **108**, 258701 (2012).
[33] J. Hlinka, D. Hartman, M. Vejmelka, J. Runge, N. Marwan, J. Kurths, and M. Paluš, Entropy **15**, 2023 (2013).
[34] J. Runge, V. Petoukhov, and J. Kurths, J. Clim. **27**, 720 (2014).
[35] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph, Bull. Am. Meteorol. Soc. **77**, 437 (1996).
[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.115.268501 for additional information regarding the betweenness pattern, more examples of positive teleconnections, negative teleconnections, and the significance of optimal paths.
[37] O. C. Guez, A. Gozolchiani, and S. Havlin, Phys. Rev. E **90**, 062814 (2014).
[38] C. E. Weatherburn, *A First Course In Mathematical Statistics* (Cambridge University Press, Cambridge, 1949).
[39] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms* (MIT Press, Cambridge, 1990).
[40] M.-K. Sung, W.-T. Kwon, H.-J. Baek, K.-O. Boo, G.-H. Lim, and J.-S. Kug, Geophys. Res. Lett. **33**, L21713 (2006).
[41] X. Zhu, O. Bothe, and K. Fraedrich, Quat. Int. **236**, 151 (2011).
[42] J. Sun and H. Wang, J. Geophys. Res. **117**, D08110 (2012).