Title
Correlated structural evolution within multiplex networks

Journal
Journal of Complex Networks, 8(2)

ISSN
2051-1310

Authors
Wu, Haochen
James, Ryan G
D'Souza, Raissa M

Publication Date
2020-04-01

DOI
10.1093/comnet/cnaa014

Peer reviewed

Journal of Complex Networks (2020) **2**, Advance Access Publication on 31 May 2020
doi: 10.1093/comnet/cnaa014

Correlated structural evolution within multiplex networks

HAOCHEN WU†
Department of Computer Science, University of California, Davis, CA 95616, USA
†Corresponding author. Email: hcwu@ucdavis.edu

RYAN G. JAMES AND JAMES P. CRUTCHFIELD
Department of Physics, Complexity Sciences Center, University of California, Davis, CA 95616, USA

AND

RAISSA M. D'SOUZA
Complexity Science Center, Department of Computer Science, Department of Mechanical and Aerospace Engineering, University of California, Davis, CA 95616, USA, and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Edited by: Manlio De Domenico

[Received on 17 October 2019; editorial decision on 27 March 2020; accepted on 30 March 2020]

Many natural, engineered and social systems can be represented using the framework of a layered network, where each layer captures a different type of interaction between the same set of nodes. The study of such *multiplex networks* is a vibrant area of research. Yet, understanding how to quantify the correlations present between pairs of layers, and more so present in their co-evolution, is lacking. Such methods would enable us to address fundamental questions involving issues such as function, redundancy, and potential disruptions. Here, we show first how the edge set of a multiplex network can be used to construct an estimator of a joint probability distribution describing edge existence over all layers. We then adapt an information-theoretic measure of general correlation called the conditional mutual information, which uses the estimated joint probability distribution, to quantify the pairwise correlations present between layers. The pairwise comparisons can also be temporal, allowing us to identify if knowledge of a certain layer can provide additional information about the evolution of another layer. We analyse datasets from three distinct domains—economic, political, and airline networks—to demonstrate how pairwise correlation in structure and dynamical evolution between layers can be identified and show that anomalies can serve as potential indicators of major events such as shocks.

Keywords: multiplex networks; dynamics of networks; information theory.

1. Introduction

Over the last two decades, network analysis has become a useful tool for understanding social, biological, physical and engineered complex systems [1]. At its most basic, a network is a set of nodes and edges, where edges denote pairwise interactions between nodes. Although this ignores many details and higher-order interactions, such as multivariate dependencies beyond dyadic, this network approximation has yielded important insights into the formation and dynamics of complex systems. Beyond a simple network, many real systems are composed of layers of individual networks, ranging from multimodal transportation networks to the Internet protocol stack and interpreting such a system as a single-layer network of homogeneous interactions is often an oversimplification [2]. In many instances each network layer contains the same set of nodes, but the edges in each distinct layer represent a distinct type of interaction between nodes. We use the terminology *multiplex network* to describe such a system. In the transportation setting, a multiplex network can be constructed where the nodes are geographic locations and each layer represents connectivity of a different transport type between locations, such as automobile, airplane, train, passenger ship, etc. In recent years, there has been a vibrant study of multiplex networks [3–5].

Generally, layers within a multiplex network are not independent. Consider the multiplex transportation network above. Due to geographic constraints, there will be little overlap between edges in the automobile and in the passenger ship layers, so the presence of an edge in one layer implies the likely absence in the other.

The straightforward approach of estimating the edge overlap rate between two layers (i.e. the frequency at which two layers contain an edge between the same pair of nodes) and then comparing that to a random network null model has demonstrated that correlation between layers are commonly found in multiplex networks [6]. Yet, basic edge overlap does not account for many important features such as anti-correlation in the location of edges which may depend on the characteristics of the system. For instance, the layers can variously either cooperate or compete with one another, or their structures can be complementary or redundant, which all influence edge overlap. Likewise, measures are still needed to quantify correlations present during evolution.

Such measures would allow a more nuanced understanding of the dynamics underlying multiplex networks. For instance, do some layers evolve independent of all others? Can we find temporal correlations indicating that one layer influences the evolution of another? In addition, we can use these measures to understand real-world multiplex networks across domains. For instance, there are principled arguments based on political and economic considerations that alliance treaties between nation states are related to their trade relationships [7]. Quantitative measures would allow us to establish this explicitly and also identify if specific types of goods are more dependent on the alliance than others. This, in turn, may reveal potential trade interventions that can impact the stability of alliances. Likewise, while diplomatic disputes between nations may lead to war, which among the various classes of disputes is most influential? In a different realm, airline companies compete and cooperate with each other, but to what extent does one company's decision influence another's?

Our focus here is on multiplex networks which evolve in time in a discrete manner. Our goal is to develop measures to quantify the correlation present between a pair of layers and in the co-evolution between a pair of layers, this includes anti-correlations. Our primary contributions are two-fold. First, we develop a method for constructing an estimator of the joint probability distribution describing the simultaneous existence of edges across layers of the multiplex network and also the discrete-time evolution of the edges. Second, using the joint probability distributions, we develop a mutual information measure to quantify correlations present between pairs of layers in a static multiplex network, and also a conditional mutual information measure to quantify the extent to which one layer influences another during their discrete co-evolution. We apply these measures to empirical datasets from airline, political and trade networks to explore interlayer relationships and also determine the temporal order of changes between layers. This reveals non-trivial relationships with some pairs of layers evolving in a more correlated manner than other pairs and also asymmetrical levels of influence between co-evolving layers. The details of the measures for particular datasets, including anomalous spikes, provide some insights useful for the questions posed above, such as trade and alliance relationships between nations.

The development is organized as follows. Section 2 discusses previous related work. Section 3 describes how we characterize a multiplex network and its correlated internal structural evolution with a joint probability distribution, and it also defines our conditional mutual information measure and tests for statistical significance. Section 4 then applies our method to multiple datasets, demonstrating its utility. Finally, Section 5 concludes by summarizing potential limitations and promising future directions.

2. Related work

Techniques from information theory offer quantitative methods to extract correlations in time series data but have not yet been extended to the multivariate setting required to analyse large networks. That said, techniques from network science for analysing multiplex networks provide insight into similarity of dynamics on layers or provide measures of ensemble properties, yet they do not extract correlations in structure and structural evolution. Here, we review these works organized along the two broad categories of approaches to provide some context and challenges.

2.1 *Information theory*

A multitude of existing techniques can be applied to time series data to quantify the correlations that are present, including delay-coordinate embedding, Granger analysis and time-delayed mutual information [8]. Unfortunately, these techniques have yet to be extended to the multivariate setting required for networks beyond a few nodes in size. Extending these informational measures to multivariate cases is an active area of current research [9, 10] which remains an open question. For example, transfer entropy [11, 12] measures the time-asymmetric information shared between two random processes. And, it subsumes Granger causality [12], which served for decades as the *de facto* detector of time-series causality [13]. As such, transfer entropy is now widely used in a variety of contexts including economic, biological and chemical processes [14, 15]. Yet, it was recently shown that such applications must take care to not interpret transfer entropy as detecting information flow or causal organization [9, 10]. As we scale up from two random processes to the size of networks, this will be an increasingly pertinent issue.

With respect to network systems, there is of course the classic discipline of network information theory which concerns itself with the *information transmission capacity* of a communication network [16]. There, given a network of rate-bound links, one measures the aggregate rate at which multiple sources can communicate without error to multiple receivers. One hallmark result is that for a single source and single receiver, the capacity is determined by the *max-flow min-cut theorem* which identifies bottlenecks due to network topology [16]. This approach, though, has a rather different focus from ours as we want to quantify information of correlated evolution *across* network layers.

One definition of information in a network, called *graph entropy*, uses the frequency of a node's occurrence in the orbits of the graph's automorphism group [17, 18]. Reference [19] gives a brief history and surveys its applications. Graph entropy is easily (and helpfully) interpreted for small graphs with substantial symmetry. It usually generates trivial or ambiguous results on large networks, however, due to their generic lack of perfect or near-perfect group symmetries.

As we will demonstrate, using information measures to quantifying relationships that arise in networks, and more generally in complex systems, is advantageous for several reasons. For one, informational measures are system-agnostic: so long as what is being studied is well-described by random variables, it matters little over what coordinates and with what units those variables are defined. For example, it is irrelevant if a random variable describes fluctuations in voltage, a child's gender or counts of chemical species. For another, as an extension to mathematical statistics, information measures quantify non-linear dependencies, expanding the common notions of correlation beyond their implied linear models. And so, information is model independent, operating directly on the data distributions with no assumptions as to the form of dependency. Finally, and key to our uses, information provides the ability to compare the relative strength of correlations across layer pairs.

2.2 *Network theory*

Approaching the analysis of multiplex networks from the network theory perspective, a natural consideration is the graph Laplacian of a network. For instance, in [20] they construct a probability distribution from a network's Laplacian. Once normalized, the Laplacian is mathematically similar to a density matrix, the object from which the von Neumann entropy is computed in quantum mechanics [21]. Using the Laplacian for each layer in a multiplex network as a probability distribution-like object for that layer, the similarity between layers can be quantified by using measures such as the Kullback–Leibler divergence between two probability distributions [22]. Since the Laplacian expresses the "local" curvature about nodes in a network, it is commonly used to analyse diffusion-like dynamics on multiplex networks [23]. As such, Laplacian-based analysis can capture the similarity of diffusion processes among layers. However, this does not reveal layer similarities and differences that are due to complex structures to which diffusion is insensitive.

Entropy methods have been developed to characterize multiplex network ensembles such as Ref. [24] which uses the ensemble entropy to analyse multiplex networks with correlated layer overlaps. Yet, ensemble considerations average over the detailed structures needed to describe pairwise interactions between layers in a multiplex network.

A stochastic block model also provides a probabilistic framework for a multiplex network ensemble, and very recently it was shown how edge-correlations could be incorporated into that framework [25]. This provides a rigorous model for generating correlated multiplex networks and a maximum-likelihood estimate of correlations present between layers. While such an approach is neat and well-defined, it is a form of linear Pearson correlation which can be limiting when non-linear correlation is present.

In making their approaches tractable several studies [26, 27] assumed bilayer networks in which, by definition, only one type of coupling between layers exists. Similarly, treating multiplex networks as tensors [28, 29] implicitly assumes different layers can be decomposed into linear, statistical-dependency structures. Our empirical analyses show that each layer may have distinct dynamical evolution and that there can be non-linear relationships between layers.

Finally, Ref. [30] introduced a multiplex Markov chain to model the correlated evolution between different layers in a multiplex network. The premise is that each multiplex edge in the network evolves according to an independent and identically distributed random process. One can then compare the difference between when that random process uses a multiplex-dependent null model to a null model that assumes each layer evolves independently. For two-layer networks this method can identify strong statistical correlations in structural evolution. Yet, this method does not scale well with the number of layers. Nor does it allow us to compare how strongly layers are coupled during structural evolution.

There is some progress in developing information-theoretic approaches to networks which rely on defining a variety of network-derived probability distributions. Degree distributions [31], deviation from mean degree distributions [32], motif distributions and even configuration distributions over a network ensemble [33] are all example probability distributions that represent distinct aspects of a network or network ensemble. Each captures some specific a priori selected features. Yet, these features do not capture the correlated structural evolution of layers in a multiplex network.

Our approach, complementary to the above, is to formulate a multiplex network as a joint probability distribution over the multiplex edge set including the discrete-time evolution of the network. We can then apply this approach to empirical datasets, which may contain non-linear correlations, to quantify the correlations present and track the enhanced predictive power that one layer can provide about the evolution of another layer.

3. Methods

3.1 *Notation*

We begin by introducing formal notation for defining multiplex networks. A multiplex network is a network with many layers which share the same node set. We use calligraphic letters such as $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$ to refer to the name of each individual layer. The multiplex network can be represented by the graph $G = \left(\mathcal{N}, E^{\mathcal{U}}, E^{\mathcal{V}}, E^{\mathcal{W}}, \ldots\right)$, where $\mathcal{N}$ is the node set and $E^{\mathcal{U}}, E^{\mathcal{V}}, E^{\mathcal{W}}$ are the edge sets for each of the different layers. $E^{\mathcal{U}}, E^{\mathcal{V}}, E^{\mathcal{W}} \subseteq [\mathcal{N}]^2$, where $[\mathcal{N}]^2$ denotes a Cartesian square of set $\mathcal{N}$ and, for example, $(i,j) \in E^{\mathcal{U}}$ if there is an edge between nodes i and j in layer $\mathcal{U}$. In the rest of the article, for simplicity, we consider only the case of undirected networks, so we have the additional constraint that if $(i,j) \in E^{\mathcal{U}}$ then $(j,i) \in E^{\mathcal{U}}$. However, our approach can be extended to directed networks in a straightforward manner.

We define the *multiplex edge vector* for a pair of nodes i and j in an l-layered multiplex network as $\boldsymbol{e}_{ij} = e^1_{ij}e^2_{ij}\ldots e^l_{ij}$, where each vector element $e^{\mathcal{U}}_{ij} = 1$ if $(i,j) \in E^{\mathcal{U}}$ and $e^{\mathcal{U}}_{ij} = 0$ otherwise. The layers are ordered in an arbitrary but fixed manner.

A particular multiplex edge vector $\boldsymbol{e}_{ij} = e^1_{ij}e^2_{ij}\ldots e^l_{ij}$ between nodes i and j can be represented as an l-gram, where l is the number of layers. When the number of layers is small, rather than using $e^{\mathcal{U}}_{ij}$, $e^{\mathcal{V}}_{ij}$ and $e^{\mathcal{W}}_{ij}$, we will use u_{ij}, v_{ij} and w_{ij}, or even simply u, v and w to refer the element in the vector corresponding to layer $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$, respectively.

An illustrative example is shown in Fig. 1, where the multiplex edge vectors $\boldsymbol{e}_{ij}$ for all possible pairs of nodes i and j are enumerated for a particular example with $l = 3$ layers. Since there are only three layers in this specific example, we could simplify notation and, rather than write $\boldsymbol{e}_{ij} = e^{\mathcal{U}}_{ij}e^{\mathcal{V}}_{ij}e^{\mathcal{W}}_{ij}$, instead write $\boldsymbol{e}_{ij} = uvw$.

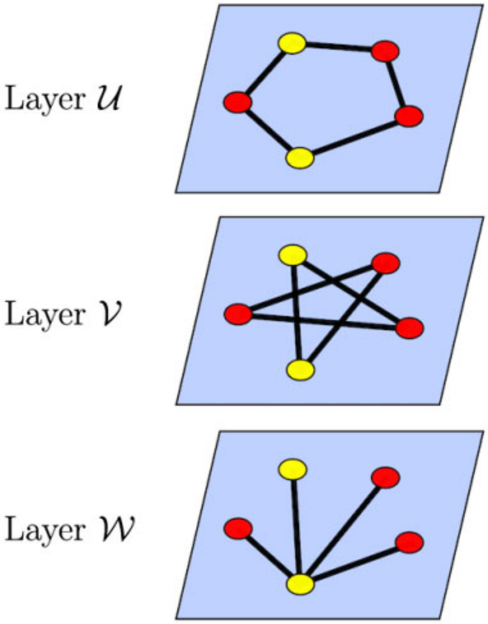
In what follows, we will need to consider a random instance of $\boldsymbol{e}_{ij}$, and we will use capital letters U, V, W to refer the random variables corresponding to u, v, w. Thus, we use different forms of the same letter depending on context. For instance, we use $\mathcal{U}$, u and U, to refer to different concepts associated with that layer: namely $\mathcal{U}$ for the name of the layer; $u \in \{0, 1\}$ for the value of the element in the edge-existential vector corresponding to that layer; and U for the random variable corresponding to that element in the joint probability distribution describing the multiplex network.

We also use some basic notation from information theory. Following the tradition in the literature, we use H to denote Shannon entropy and I to denote mutual information.

Briefly, we are interested in the interactions between random variables. Let X denote a random variable which takes on values x drawn from a discrete set, that is, an alphabet $\mathcal{X}$, with probability $p(x)$. The entropy of a random variable, $H[X]$, is defined as:

$$H[X] = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x), \tag{3.1}$$

it measures the average uncertainty of random variable X. H is also used for defining joint entropy. Given a set of discrete random variables $X_1, ..., X_n$ and their joint distribution $p(x_1, ..., x_n)$, the joint entropy



uvw	Count	$\hat{P}(uvw)$ = Freq.
000	0	0
001	0	0
010	3	0.3
100	3	0.3
011	2	0.2
101	2	0.2
110	0	0
111	0	0

FIG. 1. An example of a 3-layer multiplex network and how to build the associated joint probability distribution for the multiplex edge vectors. For each pair of nodes i, j, we consider the l-gram, $\boldsymbol{e}_{ij}$, describing the presence or absence of an edge between them respectively in each layer of the network. If an edge exists in a particular layer, we denote the corresponding element in the l-gram as 1, otherwise 0. We then take this edge as an instantiation of a joint probability distribution. For instance, between the two yellow nodes, there is no edge in layer $\mathcal{U}$, but there are edges in layers $\mathcal{V}$ and $\mathcal{W}$, so we have an instance of the 3-gram $\boldsymbol{e} = 011$ which contributes to the tally of counts in the table row for l-gram 011. We then repeat this process for all possible pairs of nodes and use the final counts to estimate the probability of having different l-grams. The right-hand column gives the values for these estimators of the joint probability distribution of the l-grams describing the multiplex edge vectors, as shown formally in Eq. 3.5.

$H[X_1, ..., X_n]$ is defined as

$$H[X_1, ..., X_n] = -\sum_{x_1 \in \mathcal{X}_1} ... \sum_{x_n \in \mathcal{X}_n} p(x_1, ..., x_n) \log_2 p(x_1, ..., x_n). \tag{3.2}$$

The mutual information $I[X; Y]$ between two random variables X and Y is defined as:

$$I[X; Y] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}. \tag{3.3}$$

It measures how much information one random variable contains about the other.

For convenience, we also often use $H(p_1, ..., p_n)$ to denote the entropy of a random variable with probability $p_1, ..., p_n$ for each of its possible values. For example, $H(0.3, 0.7)$ represents the entropy of a random variable that has two possible outcomes where the first outcome has a probability of 0.3 and the second has a probability of 0.7.

We will define more complex information measures as needed and the interested reader can refer to the classic text by Cover and Thomas [34] for more information on this topic.

3.2 *The joint probability distribution of a multiplex network*

Here, we show how it is possible to characterize a multiplex network by a joint probability distribution. The overall idea is straightforward, given that all layers have the same node set. In the classical Erdös–Rényi random graph model with N nodes, each edge is independently included in the network with probability p [35]. If we only care about the existence of an arbitrary edge, the probability of existence for that particular edge is drawn from a Bernoulli distribution with probability p. Similarly in a multiplex network with l layers, there is an analogous construction where all of the l elements of a particular $\boldsymbol{e}_{ij}$ can be drawn from an arbitrary joint probability distribution over l discrete events.

First we must introduce a basic formulation of a multiplex network. Consider the following simple model to generate a N-node multiplex network, starting from N isolated nodes. Assume that for each pair of nodes i and j, the multiplex edge vector $\boldsymbol{e}_{ij}$, is formed following the same independent stochastic process. Recall that each element of the multiplex edge vector $\boldsymbol{e}_{ij}$ indicates whether there is an edge or not in the corresponding layer. Let $P(uvw\ldots)$ denote the probability that a randomly chosen $\boldsymbol{e}_{ij}$ is equal to the particular l-gram uvw..., where $P(u=1)$ and $P(u=0)$ are respectively the marginalized probability that i and j are connected or not connected in layer $\mathcal{U}$. Then, we can generate random $\boldsymbol{e}_{ij}$'s drawn from this distribution for all i,j pairs and from that construct a corresponding instance of an N-node multiplex network.

Under these same assumptions, given a real system, we can get an estimate of the distribution $P(uvw...)$ from the given data. We assume that each $\boldsymbol{e}_{ij}$ follows the same distribution independently between all i,j pairs, thus each multiplex edge vector observed in a real network can be treated as a sample for inference. In a N-node multiplex network, there are $\frac{N(N-1)}{2}$ pairs of nodes, therefore there are $\frac{N(N-1)}{2}$ number of $\boldsymbol{e}_{ij}$'s. Note, that there are only 2^l distinct values that the $\boldsymbol{e}_{ij}$'s can take on (since each vector element must be either 0 or 1), and each distinct value can be written as a distinct l-gram. Note that we may not see the occurrence of all possible l-grams in a specific real-world instance of a multiplex network.

We next will use the frequency of occurrence for each distinct l-gram to construct an estimator of $P(uvw...)$. First we count the number of times a particular l-gram occurs, and introduce the following function to do so:

$$\begin{aligned}\text{count}(uvw\ldots) = \sum_{i,j\in\mathcal{N}} & \left[\left[(i,j)\in E^{\mathcal{U}}\right]=u\right] \\ & \left[\left[(i,j)\in E^{\mathcal{V}}\right]=v\right] \\ & \left[\left[(i,j)\in E^{\mathcal{W}}\right]=w\right] \\ & \ldots,\end{aligned} \tag{3.4}$$

where we have used Iverson brackets, $[P]$, a generalization of the Kronecker delta; it evaluates to 1 if the proposition P inside it is True and 0 otherwise [36].

With the counts in place, we can construct an estimator of the probability distribution $P(uvw...)$, explicitly:

$$\begin{aligned}&\hat{P}\left(U=u \text{ and } V=v \text{ and } W=w\ldots\right) \\ &\qquad = \frac{\text{count}(uvw\ldots)}{N(N-1)/2},\end{aligned} \tag{3.5}$$

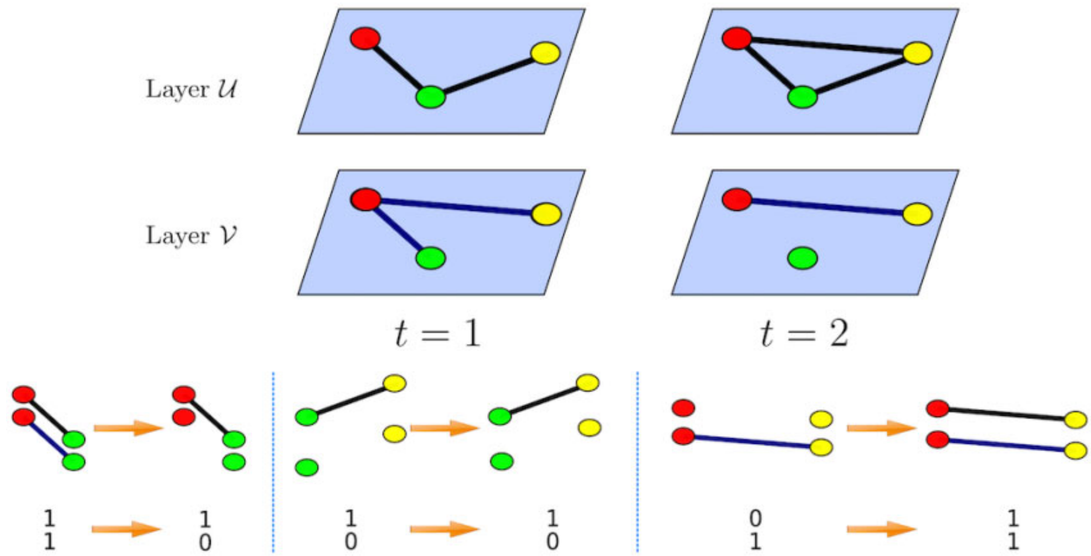


FIG. 2. An example of how to construct random variables describing the discrete-time evolution of a two-layer multiplex network. In this simple example there are three pairs of nodes and the evolution of each pair can be encoded by a 4-gram describing the presence or absence of edges. Each 4-gram encodes a specific pattern of evolution. For instance, the 4-gram 1110 describes the evolution of the edges between the red and green nodes, while 1010 describes the evolution of the edges between the green and yellow nodes. A count of the 4-grams observed across all pairs of nodes provides an estimator of the joint probability distribution for each possible 4-gram.

where $N = |\mathcal{N}|$. Once we have the estimated probability distribution, we can construct random variables and directly calculate information-theoretic measures.

Figure 1 shows an example of how the estimator of a joint probability distribution can be constructed from an instance of a multiplex network with three layers $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$. It is established by counting the frequency of occurrence for the different values of the multiplex edge vectors (which are 3-grams for this example). Each layer of the multiplex network has a corresponding random variable in such a joint probability distribution which we denote by U, V and W, respectively. The random variables take on the values of either one or zero indicating, respectively, the presence or absence of an edge in the corresponding layer.

A similar method can also be applied to analyse the discrete-time dynamics of a multiplex network. For each time step t, $t+1$, ..., all of the layers can be included in a composite multiplex network $G = \left(\mathcal{N}, E^{\mathcal{U}^t}, E^{\mathcal{U}^{t+1}}, E^{\mathcal{V}^t}, E^{\mathcal{V}^{t+1}}, \dots\right)$. We can then construct a time-labelled probability distribution for the network at each of these time steps. This process is demonstrated in Fig. 2 for a two-layered multiplex network with layers $\mathcal{U}$ and $\mathcal{V}$. For each particular pair of nodes, we can denote their specific evolution over two consecutive time steps by a 4-gram defined as $u^t v^t u^{t+1} v^{t+1}$, where $u^t = 0$ if there is no edge in layer $\mathcal{U}$ between these two nodes in time step t and $u^t = 1$ if there is an edge (and respectively for v^t and v^{t+1}). For example, the 4-gram 1011 represents the case that in time step t there is an edge between these two nodes in layer $\mathcal{U}$ but not in layer $\mathcal{V}$ and there are edges in both layers $\mathcal{U}$ and $\mathcal{V}$ in time step $t+1$. By counting the frequency of these 4-grams among all pairs of nodes in the multiplex networks, we can have an estimator of the joint probability distribution $P(u^t v^t u^{t+1} v^{t+1})$.

3.3 *Correlations between layers: mutual Information*

Given the estimator of the joint probability distribution we can construct information measures using that inferred distribution. Most important in our context is the mutual information between layers which provides us a way to quantify the extent of their correlation. This has a clean null model that all layers are statistically independent, in which case the mutual information between them is 0. Another advantage of utilizing the mutual information is that it also captures anti-correlation, the scenario where existence of edge in one layer signals the decreased likelihood of having an edge between the same pair of nodes in another layer, which is not captured by basic edge overlap considerations [6].

For the example shown in Fig. 1, the mutual information between the random variables U, V and W can be constructed in a pairwise manner. With the standard Shannon entropy and mutual information notation [34] the mutual information between layer $\mathcal{U}$ and $\mathcal{V}$ is:

$$\begin{aligned} I[U;V] &= H[U] + H[V] - H[U,V] \\ &= H(0.5, 0.5) + H(0.5, 0.5) \\ &\quad - H(0.0, 0.5, 0.5, 0.0) \\ &= 1 \text{ bit} \end{aligned} \tag{3.6}$$

$H[U]$ is obtained by counting the 1-grams present in layer $\mathcal{U}$. There we have 5 edges ('1's) and 5 non-edges ('0's) and therefore $H[U] = H(5/10, 5/10)$. $H[V]$ is obtained the analogous manner. For $H[U,V]$, we count the 2-grams formed by layers $\mathcal{U}$ and $\mathcal{V}$. We then have 0 '00' and '11's values, 5 '01's values and 5 '10's values. Therefore, $H[U,V] = H(0/10, 5/10, 5/10, 0/10)$. The resulting mutual information of 1 bit is consistent with intuition: layers $\mathcal{U}$ and $\mathcal{V}$ are complementary, and therefore there is maximal mutual information between them.

We can also calculate the mutual information between layer $\mathcal{U}$ and layer $\mathcal{W}$ and between layer $\mathcal{V}$ or layer $\mathcal{W}$:

$$\begin{aligned} I[U;W] &= H[U] + H[W] - H[U,W] \\ &= H(0.5, 0.5) + H(0.4, 0.6) \\ &\quad - H(0.3, 0, 2, 0.3, 0.2) \\ &= 0 \text{ bit} \end{aligned} \tag{3.7}$$

$$\begin{aligned} I[V;W] &= H[V] + H[W] - H[V,W] \\ &= H(0.5, 0.5) + H(0.4, 0.6) \\ &\quad - H(0.3, 0.2, 0.3, 0.2) \\ &= 0 \text{ bit} \end{aligned} \tag{3.8}$$

Thus, for this example, knowing whether an edge is in either layer $\mathcal{U}$ or layer $\mathcal{V}$ is not helpful for predicting the existence of the edge in layer $\mathcal{W}$ and vice versa; the pairwise mutual information between $\mathcal{U}$ and $\mathcal{W}$ and between $\mathcal{V}$ and $\mathcal{W}$ is 0.

Note that this is a case where even though layers $\mathcal{U}$ and $\mathcal{W}$ have overlaps, it is due to simple randomness and therefore the existence of an edge in one layer is uninformative as to the existence of the same edge in the other layer.

We next turn to the main focus of the manuscript, which is conditional mutual information. We do further consider mutual information and apply it to real data, with the details found in Appendix B. There we show that some pairs of layers are much more correlated than others.

3.4 *Correlated structural evolution: conditional mutual information*

In this section, we introduce conditional mutual information and establish how to use this to develop information-theoretic measures to quantify the correlations present in the structural evolution of multiplex networks.

Given three random variables X, X' and Y, the conditional mutual information $I\left[X'; Y|X\right]$ is defined as the relative entropy between the joint probability distribution of X' and Y and the product of distributions of X' and Y each conditioned on X. Formally:

$$I\left[X'; Y|X\right] = \sum_{\substack{x' \in \mathcal{X}' \\ y \in \mathcal{Y}}} p\left(x', y|X\right) \log \frac{p\left(x', y|X\right)}{p\left(x'|X\right) p\left(y|X\right)}. \tag{3.9}$$

This quantifies the amount of additional information available to predict X' knowing both Y and X, beyond simply knowing X alone. This is related to the notion of transfer entropy discussed briefly in Section 2 which is widely used in time series analysis to quantify if one time series can be used to predict another. One nicety of this measure is that if Y and X' are correlated to some other confounding variable X, conditioning on X can filter out such effects.

We introduce the notion of *information-theoretic influence* (denoted I-INF or simply IINF) which is calculated by applying conditional mutual information to the correlated structural evolution of a multiplex network. Using the notation introduced in Section 3.1, a pair of layers at time t is represented by the random variables U^t, V^t. Information-theoretic influence (IINF) from layer $\mathcal{U}$ to $\mathcal{V}$ then can be defined as the mutual information between layer $\mathcal{U}$ at time step t and layer $\mathcal{V}$ at time step $t+1$ conditioned on layer $\mathcal{V}$ at time step t, that is the conditional mutual information $I[U^t; V^{t+1}|V^t]$. Formally,

$$\text{IINF}_{\mathcal{U}\to\mathcal{V}}^{t\to t+1} = I[U^t; V^{t+1}|V^t]. \tag{3.10}$$

Unlike mutual information discussed in Section 3.3, IINF is asymmetric due to the existence of chronological order among the random variables. In general $\text{IINF}_{\mathcal{U}\to\mathcal{V}}^{t\to t+1}$ is not equal to $\text{IINF}_{\mathcal{V}\to\mathcal{U}}^{t\to t+1}$. Intuitively, what $\text{IINF}_{\mathcal{U}\to\mathcal{V}}^{t\to t+1}$ quantifies is the amount of extra information available to predict layer $\mathcal{V}$ in time $t+1$ if we also have information of layer $\mathcal{U}$ in time t in addition to information of layer $\mathcal{V}$ at time t. Here, conditioning on layer $\mathcal{V}$ at time t allows for the filtering out of some effects generated by node level factors or exogenous events that happen in layer $\mathcal{V}$.

We then use the IINF defined in equation 3.10, which is essentially a one-step *transfer entropy*, to quantify the degree to which the evolution of network layer $\mathcal{U}$ affects layer $\mathcal{V}$. The intuition here is that we wish to quantify the influence that U^t has on V^{t+1} above and beyond the influence of V^t. This is qualitatively similar to Granger causality [13], where vector auto-regression is used to determine if observations of a time-series X improves predictions of a time-series Y beyond utilizing only observations of Y. By definition, the IINF from a layer to itself $IINF_{\mathcal{U}\to\mathcal{U}} = I[U^t; U^{t+1}|U^t]$ is always zero – no additional information is gained through redundant knowledge of the network itself. Of course, the history of a particular layer may best inform the evolution of that layer, but in order to measure only the influence between layers we condition on knowledge of that particular layer.

The Python implementations of above measures are made available [37], which makes use of the `dit` information theory package [38].

3.5 *Illustrative examples*

IINF measures information-theoretic influence from one layer to another. The larger the value of IINF from layer $\mathcal{U}$ to layer $\mathcal{V}$, the better that we can predict layer $\mathcal{V}$ by also knowing layer $\mathcal{U}$'s history, beyond knowing layer $\mathcal{V}$'s history alone. From our construction, IINF is an information-theoretic measure applied to a binary random variable and therefore the value is always between 0 and 1. However, dependent upon the edge density of the layers, the theoretical maximum is sometimes smaller than 1. IINF complements other common information-theoretic measures for binary random variables, such as entropy or mutual information. As it is difficult to normalize IINF across different systems, it is often more useful to compare IINF between different pairs of layers or time steps in the same multiplex network. More discussion about possible normalization of IINF are provided in Section 5.

Consider the following simple cases which should serve as illustrative scenarios:

(1) Layer $\mathcal{U}$ and layer $\mathcal{V}$ are both independent Erdös–Rényi networks and they evolve to other independent Erdös–Rényi networks in the next time step. Let us denote their probability for having an edge between two nodes in each respective layer and time step as $p(\mathcal{U}^t)$, $p(\mathcal{V}^t)$ and $p(\mathcal{V}^{t+1})$. This represents the extreme case where two layers are totally independent and there is no information-theoretic influence at all from layer $\mathcal{U}$ to layer $\mathcal{V}$.

In such case, the information-theoretic influence is:

$$\begin{aligned}\mathrm{IINF}_{\mathcal{U}\to\mathcal{V}}^{t\to t+1} &= I\left[U^t; V^{t+1}|V^t\right] \\ &= I\left[U^t; V^{t+1}\right] \\ &= H\left[U^t\right] + H\left[V^{t+1}\right] - H\left[U^t, V^{t+1}\right] \\ &= H\left(0, p\left(\mathcal{U}^t\right)\right) + H\left(0, p\left(\mathcal{V}^{t+1}\right)\right) \\ &= 0 \text{ bit.}\end{aligned} \tag{3.11}$$

(2) Layer $\mathcal{V}$ is an Erdös–Rényi network with a static structure that does not evolve in time. Then no matter what layer $\mathcal{U}$ is, the information-theoretic influence is:

$$\begin{aligned}\mathrm{IINF}_{\mathcal{U}\to\mathcal{V}}^{t\to t+1} &= I\left[U^t; V^{t+1}|V^t\right] \\ &= H\left[U^t|V^t\right] + H\left[V^{t+1}|V^t\right] \\ &\quad - H\left[U^t, V^{t+1}|V^t\right] \\ &= H\left[U^t|V^t\right] + 0 - H\left[U^t|V^t\right] \\ &= 0 \text{ bit.}\end{aligned} \tag{3.12}$$

This represents the extreme case where layer $\mathcal{V}$ can be perfectly predicted from itself in the previous time step and there is nothing more that we can learn from another layer no matter what.

(3) Layer $\mathcal{U}$ and layer $\mathcal{V}$ are both independent Erdös–Rényi networks and layer $\mathcal{V}$ mimics layer $\mathcal{U}$ in next time step, or formally $\mathcal{V}^{t+1} = \mathcal{U}^t$. In this case, layer $\mathcal{V}$ fully depends on layer $\mathcal{U}$ in the previous time step, therefore, it has maximum information-theoretic influence from layer $\mathcal{U}$ to layer $\mathcal{V}$.

The information-theoretic influence then will be:

$$\begin{aligned}\mathrm{IINF}_{\mathcal{U}\to\mathcal{V}}^{t\to t+1} &= I\left[U^t; V^{t+1}|V^t\right] \\ &= H\left[U^t|V^t\right] + H\left[V^{t+1}|V^t\right] \\ &\quad - H\left[U^t, V^{t+1}|V^t\right] \\ &= H\left[U^t\right] + H\left[U^t\right] - H\left[U^t\right] \\ &= H\left[U^t\right] \\ &= H\left(p\left(\mathcal{U}^t\right)\right).\end{aligned} \tag{3.13}$$

(4) Layer $\mathcal{U}$ and layer $\mathcal{V}$ are both independent Erdös–Rényi networks and layer $\mathcal{V}$ is a combination of layer $\mathcal{U}$ and layer $\mathcal{V}$ in next time step. For the extreme case, say in the next time step the existence of an edge between two nodes in layer $\mathcal{V}$ is the xor of the existence of the edge between the same pair of nodes in layer $\mathcal{U}$ and layer $\mathcal{V}$ in the previous time step, where layer $\mathcal{V}$ is determined by a synergy effort of both layer $\mathcal{U}$ and layer $\mathcal{V}$.

The information-theoretic influence then will be:

$$\begin{aligned}IINF_{\mathcal{U}\to\mathcal{V}}^{t\to t+1} &= I\left[U^t; V^{t+1}|V^t\right] \\ &= H\left[U^t|V^t\right] + H\left[V^{t+1}|V^t\right] \\ &\quad - H\left[U^t, V^{t+1}|V^t\right] \\ &= H\left[U^t\right] + H\left[V^{t+1}\right] - H\left[U^t\right] \\ &= H\left[V^{t+1}\right] \\ &= H\left(\rho\left(\mathcal{V}^{t+1}\right)\right),\end{aligned} \tag{3.14}$$

where $\rho\left(\mathcal{V}^{t+1}\right)$ is the density of layer $\mathcal{V}$ in time step $t+1$.

In all four extreme scenarios described above, our measure agrees with intuition, and for all other scenarios the IINF will fall between these extreme cases. A few more practical examples are provided in Appendices F, G and H.

3.6 *Testing for statistical significance*

When applied to empirical data, it is also important to be able to distinguish true signal from random fluctuations. Fortunately, methods for statistical testing of information-theoretic measures have been established and can be adopted easily.

According to Goebal *et al.* [39], in a joint probability distribution ω, consider three random variables X, Y, Z. If X and Y are independent when conditioned on Z, then we can simply use the frequency as an estimator for probability. The inferred conditional mutual information $\hat{I}[X;Y|Z]$ by M independent samples generated from ω is approximately gamma distributed:

$$\hat{I}[X;Y|Z] \sim \Gamma\left(\frac{|\mathcal{Z}|}{2}(|\mathcal{X}|-1)(|\mathcal{Y}|-1), \frac{1}{M\ln 2}\right), \tag{3.15}$$

where the alphabet for these random variables are $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$ respectively. Applying this to network data, where the presence of an edge is binary, therefore $|\mathcal{X}| = |\mathcal{Y}| = |\mathcal{Z}| = 2$, the gamma distribution then reduces to an exponential distribution, we find:

$$\hat{\mathrm{IINF}}^{t\to t+1}_{\mathcal{U}\to\mathcal{V}} \sim \mathrm{Exp}\left(\frac{1}{M\ln 2}\right), \tag{3.16}$$

where $M = N(N-1)/2$ is the number of possible edges in a network with N nodes. Then, for a given pair of layers with $\hat{\mathrm{IINF}}^{t\to t+1}_{\mathcal{U}\to\mathcal{V}} = a$, the p-value would be the probability that $\hat{I}\left[U^t; V^{t+1}|V^t\right] \geq a$ if U^t and V^{t+1} are conditionally independent given V^t. This provides a method to calculate p-values present in real datasets. Section 4 uses it to establish when a correlation observed in the structural evolution is statistically significant.

4. Applications: correlated structural evolution

Now that we have developed a measure to quantify the enhanced predictive power that one layer provides about another during structural evolution of a multiplex network, we can show the applicability of the method.

In this article, we restrict our analysis to only networks evolving at consecutive time steps in order to give a simple and clear picture. This is analogous to a Markov assumption, or the even weaker assumption that the most recent time step provides the most predictive single measurement. However, it is straight forward to extend our framework to incorporate more time steps with more data as well. Some details about such extensions are discussed later.

We present results for three kinds of multiplex networks that have intrinsically different correlation patterns between their layers: political interactions between nation states, the commercial US airline network made of multiple carriers and trade and alliance networks between nations. The differences between these networks are highlighted here with details later. In the political events network, layers are different types of actions and the correlations among the different actions are relatively stable over time. In the airline network, layers represent the flight route maps of individual airline companies within the USA and the IINF between them can change abruptly around certain events such as mergers. In the ally-trade network of nation states, we have two categories of layers: one layer representing the alliances, and then many other layers representing trade of distinct types of goods respectively, with correlations among two trading layers distinct from correlations between the alliances layer and a specific trading layer.

In all three domains, we demonstrate that IINF can quantify the correlated structural evolution and be used to detect anomalies. We find that, in general, there is a strong statistically significant information theoretic influence present in the correlated structural evolution between the layers of real-world multiplex networks. Of course there are limitations and we will discuss them in Section 5.

Although we do not currently have prior knowledge about how to quantify this measure across different multiplex networks, it is very useful when applied to a specific multiplex network. Indeed, the

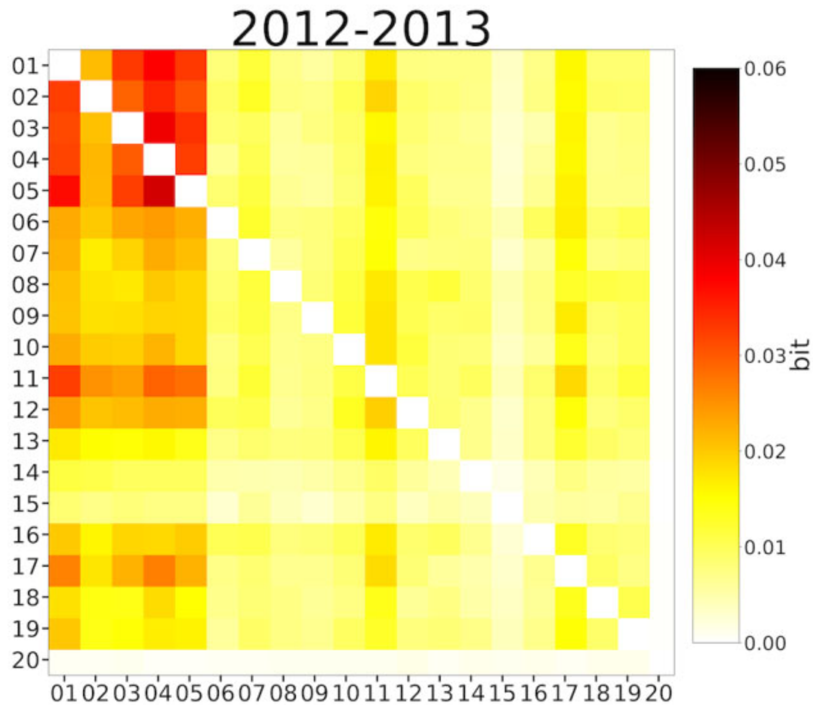


FIG. 3. IINF between 20 different classes of events during the period from 2012 to 2013, with event-types labelled by their corresponding CAMEO code [40] (given in Appendix A). Each pixel represents the IINF from the event type in X-axis to the event type in Y-axis. The values that are not statistically significant from 0 in a $p = 0.001$ level are omitted. The IINF from 03 (express intent to cooperate) to 05 (engage in diplomatic cooperation) is much higher from 03 to 15 (exhibit military posture), which indicates that knowing whether countries expressed intent to cooperate in the previous time step allows us to better predict whether they actually engage in diplomatic cooperation, but knowing so will not help with better predicting actions such as military postures. We can also see that actions coded by 01 to 05, in general, provide more power for predicting other layers but that the relationship is not symmetrical.

IINF varies greatly for different pairs of layers within the same network. The IINF of the most strongly coupled pairs of layers can be several orders of magnitude larger than the IINF of the majority of the rest, which indicates a significant potential connection between how those layers choose to create or destroy edges. Moreover, we also find a correspondence between IINF spikes and major events occurring in some of the networks, which makes IINF a tool for probing potential shocks to network structures.

4.1 *ICEWS events network*

The Integrated Crisis Early Warning System (ICEWS) is an automatically generated dataset of international events [41]. The data contain multiple different types of political interactions between nation states ranging from making statements about one another to conducting military operations against one another. Using these data, we build a series of snapshots of this multiplex network of nation states over distinct years, where each layer corresponds to a distinct type of interaction. During the 17-year time period of our data, spanning from 1997 to 2013, the IINF pattern is stable with just small fluctuations (see Appendix C). We show in Fig. 3 the typical behaviour of the pattern for the transition between two recent years. Thus we can use such stable patterns to help promote future predictions. This is in contrast to what we will show next in the airline networks, where IINF patterns can rapidly spike. The yearly transitions of a full 17-year period for the ICEWS data, in addition to the weekly transitions of a recent period, can be found in Appendices C and D.

We find that the relative correlation strength observed between layers is consistent with intuition. As mentioned in a previous section, the diagonal elements measure the IINF from a layer to itself $\text{IINF}_{\mathcal{U}\to\mathcal{U}}$ and they are always zeros. IINF is a directional measure and is not symmetric, which means extra information can be more easily gained from one direction of evolution over the other. For example, we find that the IINF from both action 03 (express intent to cooperate) or action 04 (consult) to action 05 (engage in diplomatic cooperation) is quite high when compared to other actions, indicating that knowing whether countries expressed intent to cooperate or consult in the previous time step allows us to better predict whether they actually engage in diplomatic cooperation in the current time step. In contrast, the IINF from action 03 or action 04 to action 14 (protest) or action 15 (exhibit military posture) is relatively low, indicating that knowing whether countries expressed intent to cooperate or consult in the previous time step does not allow us to better predict the onset of actions such as protests and military posturing. Note that our measure only quantifies the strength of the influence, it does not establish whether the influence is in the positive or negative direction.

Consistent with the use of information theory, IINF can be related to how much information a source layer has and how much extra information can be possibly gained with this knowledge in a target layer. This explains some of other features in Fig. 3. For instance, events with code 20 (use of unconventional mass violence) happen rarely and contain little information. As such, they are very hard to predict from the occurrence of other actions (i.e. the row for code 20 has entries that are mostly close to 0). Code 20 actions also provide little information for predicting other actions (i.e. the column for code 20 has entries that are mostly close to 0). Actions with codes 01 to 05 contain more information useful for predicting events such as 11, but in general those actions are difficult to predict given other actions (i.e. the columns for codes 01 to 05 have higher values than other columns, but the rows do not).

4.2 *US airline network*

The USA Department of Transportation, Bureau of Transportation Statistics maintains a public database of the monthly report it receives from all certified USA air carriers [42]. Every domestic flight segment is recorded therein. For our purposes, we focus on the ‘scheduled passenger service’ flights as this is representative of the air carriers’ regular flight network structure. We do not include flights such as ‘non-scheduled passenger service’ flights and flights with no passengers as they occasional and *ad hoc* and do not seem to reflect a carrier’s network-building strategy.

We find that the spikes observed in IINF often are important signals, revealing the interactions between different layers. To demonstrate such a result, we first show that when a relatively high spike in IINF is observed between two layers, we often find co-occurring real-world events associated with this spike. Conversely, we also provide evidence that when an expected event with high impact happens, such as a merger between carriers, we see a spike in the IINF measure.

Figure 4 shows how IINF behaves between the 15 major airline companies during a 17-year period. In general, there are statistically significant information flows between all carrier pairs. However, when comparing this to the IINF values present in the ICEWS events network, the magnitude of the values in the airline network are generally much lower, which suggests that the amount of influence between layers is much smaller in airline networks than in the ICEWS network. Another notable observation is that a transition happened during 2001 which could be related to the September 11 attacks (also referred to as

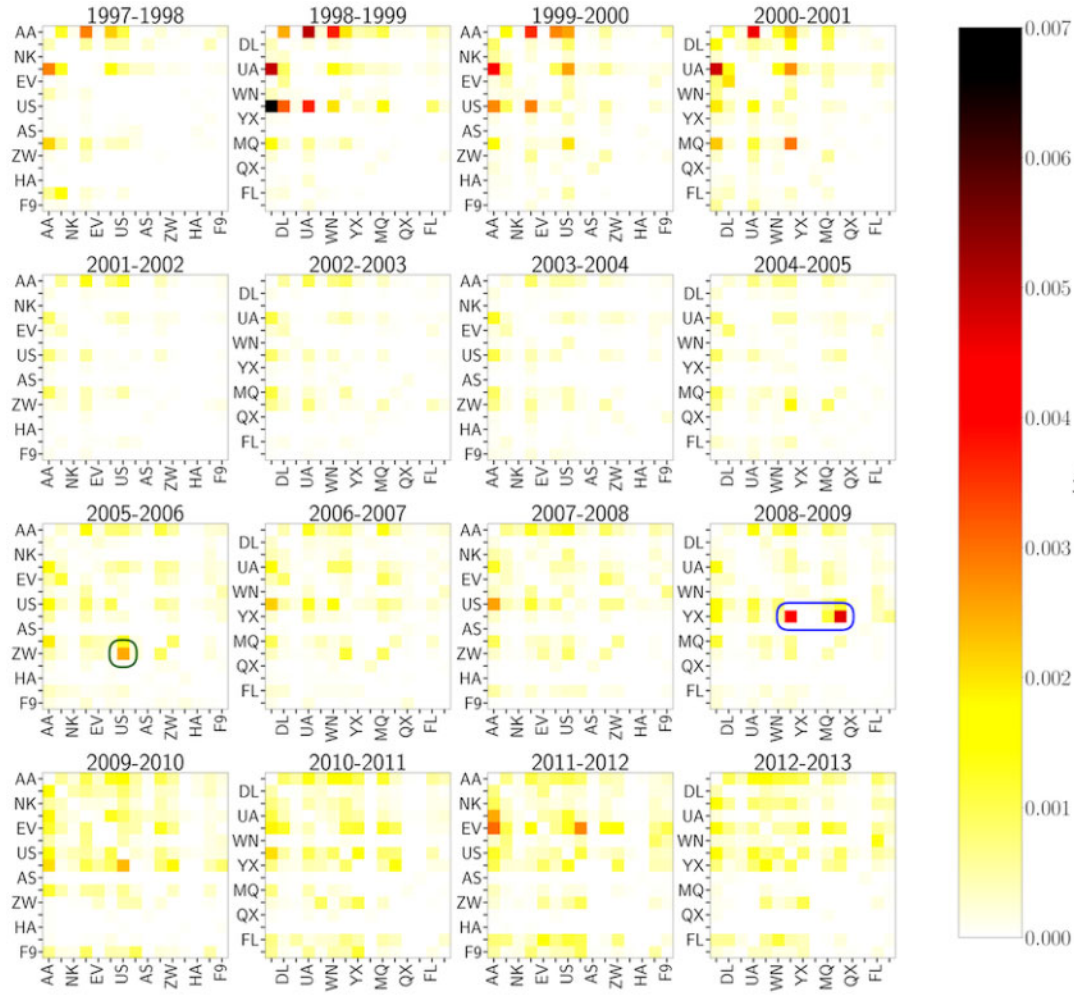


FIG. 4. IINF between 15 major carriers from 1997 to 2013. (The IATA codes are indicated for every other carrier in each figure, the full list from top to bottom and also left to right are: AA, DL, NK, UA, EV, WN, US, YX, AS, MQ, ZW, QX, HA, FL, F9.) Each distinct panel is the IINF between two distinct consecutive years, with each pixel representing the strength of IINF from the carrier in X-axis to the carrier in Y-axis. Values that are not statistically significant from 0 at the $p = 0.001$ level are omitted. In general, we can see that the interactions among carriers decreased significantly after the 9/11 attacks. After that, there are a few cases with unusual spikes in IINF. These are generally explainable by large events. For example, in 2009 (blue circle), Midwest Airlines (YX) is acquired by Republic Airways. The latter inherited the same IATA code from the former. They then adjusted the flight routes to compete with US Airways (US) and Air Wisconsin (ZW). Also in 2005 (green circle), Air Wisconsin (ZW) invested heavily into US Airways (US) and signed a long-term contract operating as US Airways Express.

9/11) [43]. The information-theoretic influence was higher before the attacks and fell-off dramatically after it. This suggests that heavy regulation after 9/11 may have had a significant impact, preventing carriers from adjusting their route map relative to other carriers.

We now demonstrate a correspondence between spikes in IINF and significant real-world events. We manually identified the top three, post-9/11, IINF hot-spots and corroborated that each one corresponds

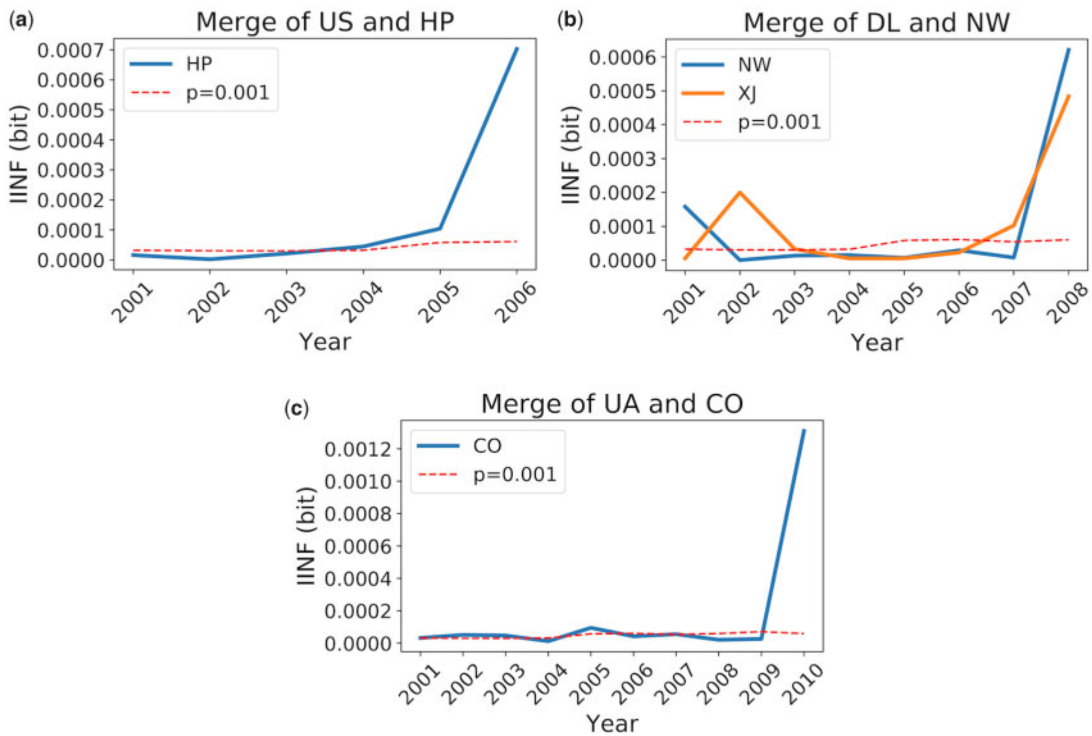


FIG. 5. Changes in IINF during three different large merger events between air carriers showing a spike in IINF as the carriers merge. (Note that carriers are required to continue reporting separately for one more year beyond the official merger date.) American West Airlines (HP) merged with US Airways (US) in 2005. Northwest Airlines (NW) merged with Delta Air Lines (DL) in 2008 (The additional green solid line is Mesaba Airlines (XJ) who was operating routes for NW). Continental Airlines (CO) merged with United Airlines (UA) in 2010. The statistical significance level for IINF is also included.

to some associated event, including an acquisition and the signing of a long-term cooperation contract, as explained in the caption of Fig. 4. This is consistent with our expectation that carriers adjust their flight routes to take into account the routes of the carriers that they acquire or sign major contracts with. When it is know in advance that mergers are underway, it is expected that the merging carriers will adjust their flight networks accordingly, and we hypothesize that this will result in an increase in IINF. Note that after a merger, carriers are still required to separately report their flight information for one additional year which allows us to corroborate this hypothesis with our dataset. Figure 5 shows details for three different airline carrier mergers.

We also provide a heatmap of IINF among all 60 air carriers in recent years in Appendix E. It is interesting to note that IINF between major carriers is generally larger than among smaller carriers and with more frequent spikes.

4.3 *Alliance and trade network*

To study the alliance and trade network between nations, we combined two different datasets. The trade network is compiled from the publicly accessible COMTRADE data maintained by the United

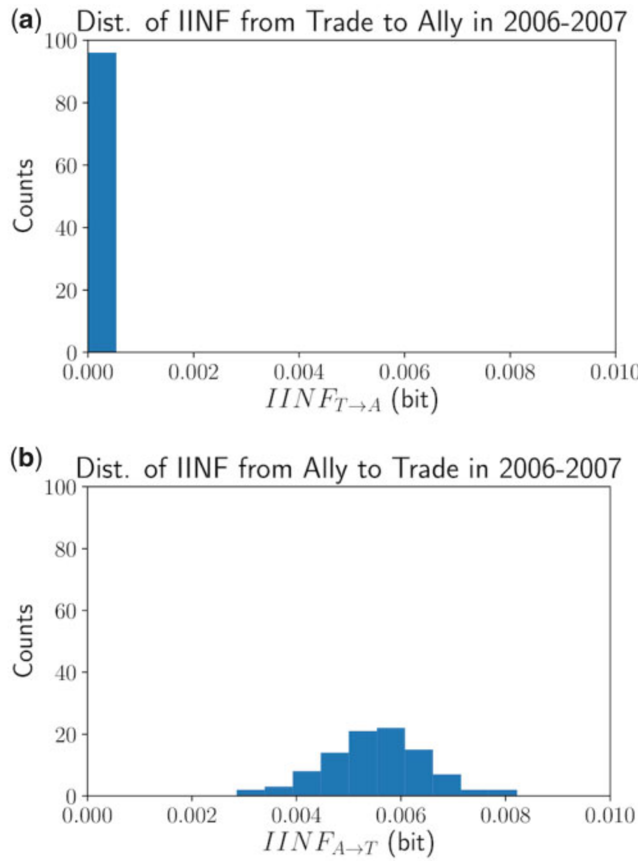


FIG. 6. (a) The IINF from all the trade layers to the alliance layer is not statistically significant at a $p = 0.001$ level. (b) The values of IINF from the alliance layer to the trade layers are statistically significant at a $p = 0.001$ level. and normally distributed around non-zero values, which indicates the information flows from alliance to trade relationships are unidirectional and somewhat ubiquitous. (The $p = 0.001$ level corresponds to an IINF of 0.0034. It falls within the first bin and we omit it for aesthetic reasons.) This could be due to the relative stability of the alliance network in the short term and quantifies prior arguments based on political and economic reasoning.

Nations [44]. This dataset includes yearly trade information for many different categories of goods which are hierarchically classified into a 6-digit system. For example, code 260111 represents 'Iron ore, concentrate, not iron pyrites, unagglomerate', 2601XX represents 'Iron ores' and 26XXXX represents 'general ores and concentrates'. In this research, we limited ourselves to an aggregation to the first two digits to get a denser and more reliable network, which results in a 96-layer network where each layer is a distinct trade category of commodities.

The alliance network is generated from the Alliance Treaty Obligations and Provisions Project [45], containing the alliance treaties signed by nation states. We manually matched these two datasets to construct a multiplex network with one alliance layer and 96 trade layers and then studied the IINF between all the distinct alliance and trade layer pairs.

We find that for this discrete-time formulation, the IINF is unidirectional, from the alliance network to the trade network, corroborating prior research establishing this fact from political and economic considerations [7]. Figure 6 shows that at this yearly time scale there is no statistically significant non-zero IINF from any commodity trade network to the alliance network, but for different commodities there is typically some information that can be gained from knowledge of the alliance network in the previous year. This means that in the short term we can use the alliance network to help predict the change in the trade network but not vice versa. We also notice that the information flows from the alliance network to trade networks are small, and there is no statistically significant difference when comparing IINF values to different commodity layers. Thus we can say, at least at this level of aggregation of the commodity categories, each category receives roughly the same IINF from the alliance network.

5. Conclusions

We have shown that it is possible to use the edge set of a multiplex network to construct a joint probability distribution characterizing the network. Information-theoretic measures over this probability distribution enable us to quantify correlations between pairs of layers, including temporal considerations. To specifically capture the extent of correlation present in the structural evolution between pairs of layers in a multiplex network, we introduce a measure called the information-theoretic influence (IINF) which is based on conditional mutual information. Applying this to several empirical datasets, we find that the extent of information sharing between different pairs of layers can vary dramatically in real-world multiplex networks with some sets of layers evolving in a highly correlated manner while other layers evolve independently, especially when the number of layers is large.

In addition, we show that IINF also detects asymmetric relationships between layers. For instance, political scientists theorize that for short-term considerations the influence between trade and alliance networks is unidirectional: that the alliance network drives the trade network, but that there is significantly less influence the other direction [7]. Our IINF measure quantifies this phenomena showing that, conditioned on the previous time step, a trade network provides no information for predicting the alliance network in the next step, but the alliance network does provide information for the evolution of trade networks.

Furthermore, our approach of mapping a multiplex network onto a joint probability distribution allows for many other information measures to be calculated. One potential direction is to use the newly developed autonomy of three-way mutual information, related to synergy and redundancy of information, to divide three-way mutual information into two types of factors [46]. One might be able to identify different signatures for different types of correlations present, such as cooperative or competitive. Likewise, one may be able to use these three-way measures to identify the higher-order organization in a system, beyond the dyadic organization.

We used IINF to understand the structure and structural evolution internal to several real-world networks, but we do not currently use IINF to compare different networks. One of the major challenges is that the edge density in a particular multiplex network can affect the magnitude of all of our measures. This does not pose a problem for the results here, since we compare features within the same multiplex network. However, extending our method to comparing measures among different multiplex networks requires a proper way to normalize IINF across networks of different sizes and types of probability distributions. There are many ways to normalize the results so that different perspectives can be brought into consideration. For instance, the IINF $I\left[U^t; V^{t+1}|V^t\right]$ can be normalized by either the entropy $H\left[V^{t+1}\right]$ or the conditional entropy$H\left[V^{t+1}|V^t\right]$. These, respectively, would consider the ratio of IINF to the maximum information that can actually be obtained from the data with the layer $\mathcal{V}$ in time step $t+1$ itself, or the layer $\mathcal{V}$ in time step $t+1$ conditioned on layer $\mathcal{V}$ in time step t. As our intent here is to use IINF within an individual multiplex network, we leave normalization considerations for future work.

Of course there are many ways to refine the considerations introduced herein. For example, the assumption that all the multiplex edge vectors are drawn from a same joint probability distribution is not always valid. Such an assumption is analogous to the foundational Erdös–Rényi model but ignores important characteristics found in empirical networks such as degree distribution or clustering coefficient. In future works, it may be possible to extend our method to include such features by building a probability distribution over the multiplex edge set of higher-order than the simple edge-existence considered here. Similarly, the weaknesses [9, 10] of the conditional mutual information should be addressed explicitly in future efforts. However, we believe this provides a useful framework for quantifying correlations present between layers in a multiplex network including in their co-evolution.

Acknowledgements

We thank Brandon Kinne and Martin Hilbert for many useful discussions and for providing us with the ICEWS and COMTRADE datasets. We are grateful to a number of others from the SPINS, complexnets and Complexity Sciences Center at UC Davis who provided insightful discussions and feedback.

Funding

U.S. Army Research Office under Multidisciplinary University Research Initiative (Award No. W911NF-13-1-0340) and Cooperative Agreement No. (W911NF-09-2-0053); the U.S. Department of Defense Minerva (W911NF-15-1-00502); DARPA (Award No. W911NF-17-1-0077); and Intel Corporation through the Intel Parallel Computing Center at the UC Davis Complexity Sciences Center.

References

1. Newman, M. (2010) *Networks: An Introduction*. New York, NY, USA: Oxford University Press.
2. Zanin, M. (2015) Can we neglect the multi-layer structure of functional networks? *Phys. A*, **430**, 184–192.
3. Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. & Porter, M. A. (2014) Multilayer networks. *J. Compl. Netw.*, **2**, 203–271.
4. Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z. & Zanin, M. (2014) The structure and dynamics of multilayer networks. *Phys. Rep.*, **544**, 1–122.
5. Bianconi, G. (2018) *Multilayer Networks: Structure and Function*. New York, NY, USA: Oxford University Press.
6. Battiston, F., Nicosia, V. & Latora, V. (2014) Structural measures for multiplex networks. *Phys. Rev. E*, 89:032804.
7. Mansfield, E. D. & Bronson, R. (1997) Alliances, preferential trading arrangements, and international trade. *Am. Polit. Sci. Rev.*, **91**, 94–107.
8. Kantz, H. & Schreiber, T. (2004) *Nonlinear Time Series Analysis*, vol. 7. New York, NY, USA: Cambridge University Press.
9. James, R. G., Barnett, N. & Crutchfield, J. P. (2016) Information flows? A critique of transfer entropies. *Phys. Rev. Lett.*, **116**, 238701.
10. James, R. G. & Crutchfield, J. P. (2017) Multivariate dependence beyond Shannon information. *Entropy*, **19**, 531.
11. Schreiber, T. Measuring information transfer. (2000) *Phys. Rev. Lett.*, **85**, 461–464.
12. Barnett, L., Barrett, A. B. & Seth, A. K. (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.*, **103**, 238701.

13. Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438.
14. Bauer, M., Cox, J. W., Caveness, M. H., Downs, J. J. & Thornhill, N. F. (2007) Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Trans. Control Syst. Technol.*, **15**, 12–21.
15. Wibral, M., Rahm, B., Rieder, M., Lindner, M., Vicente, R. & Kaiser, J. (2011) Transfer entropy in magnetoencephalographic data: quantifying information flow in cortical and cerebellar networks. *Progr. Biophys. Mol. Biol.*, **105**, 80–97.
16. El Gamal, A. & Kim, Y.-H. (2011) *Network Information Theory*. New York, NY, USA: Cambridge University Press.
17. Rashevsky, N. (1955) Life, information theory, and topology. *Bull. Math. Biophys.*, **17**, 229–235.
18. Mowshowitz, A. (1968) Entropy and the complexity of graphs: I. An index of the relative complexity of a graph. *Bull. Math. Biophys.*, **30**, 175–204.
19. Dehmer, M. & Mowshowitz, A. (2011) A history of graph entropy measures. *Inform. Sci.*, **181**, 57–78.
20. Braunstein, S. L., Ghosh, S. & Severini, S. (2006) The Laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Ann. Combin.*, **10**, 291–317.
21. Von Neumann, J. (1955) *Mathematical Foundations of Quantum Mechanics*. Princeton, NJ: Princeton University Press.
22. De Domenico, M., Nicosia, V., Arenas, A. & Latora, V. (2015) Structural reducibility of multilayer networks. *Nat. Commun.*, 6.
23. Gomez, S., Diaz-Guilera, A., Gomez-Gardenes, J., Perez-Vicente, C. J., Moreno, Y. & Arenas, A. (2013) Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.*, **110**, 028701.
24. Bianconi, G. (2013) Statistical mechanics of multiplex networks: entropy and overlap. *Phys. Rev. E*, **87**, 062806.
25. Pamfil, A., Howison, S. D. & Porter, M. A. (2019) Edge correlations in multilayer networks. https://arxiv.org/abs/1908.03875.
26. Granell, C., Gómez, S. & Arenas, A. (2013) Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Phys. Rev. Lett.*, **111**, 128701.
27. Pósfai, M., Gao, J., Cornelius, S. P., Barabási, A. & D'Souza, R. M. (2016) Controllability of multiplex, multi-time-scale networks. *Phys. Rev. E*, **94**, 032316.
28. Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. & Onnela, J. (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science*, **328**, 876–878.
29. De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gómez, S. & Arenas, A. (2013) Mathematical formulation of multilayer networks. *Phys. Rev. X*, **3**, 041022.
30. Vijayaraghavan, V. S., Noël, P.-A., Maoz, Z. & D'Souza, R. M. (2015) Quantifying dynamical spillover in co-evolving multiplex networks. *Sci. Rep.*, 5.
31. Lin, Y., Desouza, K. C. & Roy, S. (2010) Measuring agility of networked organizational structures via network entropy and mutual information. *Appl. Math. Comput.*, **216**, 2824–2836.
32. Solé, R. V. & Valverde, S. (2004) Information theory of complex networks: on evolution and architectural constraints. *Complex Networks*. *Lecture Notes in Physics* (E. Ben-Naim, H. Frauenfelder & Z. Toroczkai eds). vol 650. Berlin: Springer, pp. 189–207.
33. Ji, L., Bing-Hong, W., Wen-Xu, W. & Tao, Z. (2008) Network entropy based on topology configuration and its computation to random networks. *Chin. Phys. Lett.*, **25**, 4177.
34. Cover, T. M. & Thomas, J. A. (2012) *Elements of Information Theory*. Hoboken, NJ: John Wiley & Sons, Inc.
35. Erdös, P. & Rényi, A. On random graphs, I. (1959) *Publ. Math. (Debrecen)*, **6**, 290–297.
36. Knuth, D. E. (1992) Two notes on notation. *Am. Math. Monthly*, **99**, 403–422.
37. Wu, H. (2017) Multinet. http://github.com/wuhaochen/multinet (accessed 20 April 2020).
38. James, R. G., Ellison, C. J. & Crutchfield, J. P. (2018) DIT: discrete information theory in Python. https://github.com/dit/dit (accessed 20 April 2020).

39. Goebel, B., Dawy, Z., Hagenauer, J. & Mueller, J. C. (2005) An approximation to the distribution of finite sample size mutual information estimates. *2005 IEEE International Conference on Communications, 2005. ICC 2005, Seoul.* vol. 2. IEEE, pp. 1102–1106.

40. Gerner, D. J., Schrodt, Abu-Jabr, R. & P. A., Yilmaz, Ö. (2002) Conflict and mediation event observations (CAMEO): a new event data framework for the analysis of foreign policy interactions. *International Studies Association, New Orleans*. Paper presented at the 43rd Annual Convention of the International Studies Association, New Orleans, LA, 24–27 March.

41. Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J. & Ward, M. (2016) ICEWS coded event data.

42. Transtat. -100 domestic segment. (2014) http://www.transtats.bts.gov/Fields.asp?Table_ID=259 (accessed 22 April 2020).

43. Ito, H. & Lee, D. (2005) Assessing the impact of the september 11 terrorist attacks on us airline demand. *J. Econ. Bus.*, **57**, 75–95.

44. Gaulier, G. & Zignago, S. (2010) BACI: international trade database at the product-level. the 1994-2007 version. Working Papers 2010-23, CEPII research center. https://ideas.repec.org/p/cii/cepidt/2010-23.html.

45. Leeds, B., Ritter, J., Mitchell, S. & Long, A. (2002) Alliance treaty obligations and provisions, 1815-1944. *Int. Interact.*, **28**, 237–260.

46. Williams, P. L. & Beer, R. D. (2010) Nonnegative decomposition of multivariate information. https://arxiv.org/abs/1004.2515.

47. Stark, C., Breitkreutz, B., Reguly, T., Boucher, L., Breitkreutz, A. & Tyers, M. (2006) Biogrid: a general repository for interaction datasets. *Nucleic Acids Res.*, 34(suppl_1):D535–D539.

48. Duval, R. D. & Thompson, W. R. (1980) Reconsidering the aggregate relationship between size, economic development, and some types of foreign policy behavior. *Am. J. Polit. Sci.*, **24**, 511–525.

Appendix

A. CAMEO code for ICEWS data

Table A1 gives the codebook for different layers classified with CAMEO code [40] in the ICEWS data. Each layer corresponds to a different type of interaction between nation states. Refer to the codebook cited for more details.

Table A2 is a sample of the explanation taken from the codebook for a subcategory of 01 (MAKE PUBLIC STATEMENT).

B. Mutual information in real multiplex networks

Mutual information explained in Section 3.3 can be used to quantify the correlation between two layers in a multiplex network when defined as described in the main text. It provides a principled way of quantifying the relationships between layers without assumptions such as linear correlation and it frees us from consulting a null model to verify that the correlation is not from sheer randomness. With the maturity of information theory, we can also easily obtain many statistical tools to test the significance of the results.

We apply this measure to different data sets and report here in Figs B1, B2 and B3. Figure B1 shows the mutual information for four biological interaction network built from the BioGrid dataset [47]. The multiplex networks represent different types of interactions between proteins/genes in four different species. Figures B2 and B3 give the mutual information between layers in the Transtat and ICEWS datasets described in Sections 4.2 and 4.1, respectively. Note that mutual information between layers is

TABLE A1 *CAMEO codes used in the ICEWS dataset.*

Code	Meaning
01	MAKE PUBLIC STATEMENT
02	APPEAL
03	EXPRESS INTENT TO COOPERATE
04	CONSULT
05	ENGAGE IN DIPLOMATIC COOPERATION
06	ENGAGE IN MATERIAL COOPERATION
07	PROVIDE AID
08	YIELD
09	INVESTIGATE
10	DEMAND
11	DISAPPROVE
12	REJECT
13	THREATEN
14	PROTEST
15	EXHIBIT FORCE POSTURE
16	REDUCE RELATIONS
17	COERCE
18	ASSAULT
19	FIGHT
20	USE UNCONVENTIONAL MASS VIOLENCE

TABLE A2 *Sample CAMEO code from codebook.*

CAMEO	**011**
Name	**Decline comment**
Description	Explicitly decline or refuse to comment on a situation.
Usage Notes	This event form is a verbal act. The target could be who the source actor declines to make a comment to or about.
Example	NATO on Monday declined to comment on an estimate that Yugoslav army and special police troops in Kosovo were losing 90 to 100 dead per day in NATO air strikes.

symmetric:

$$I[U;V] = I[V;U]. \tag{B.1}$$

The diagonal elements are mutual information between one layer and itself,

$$I[U;U] = H[U] \tag{B.2}$$

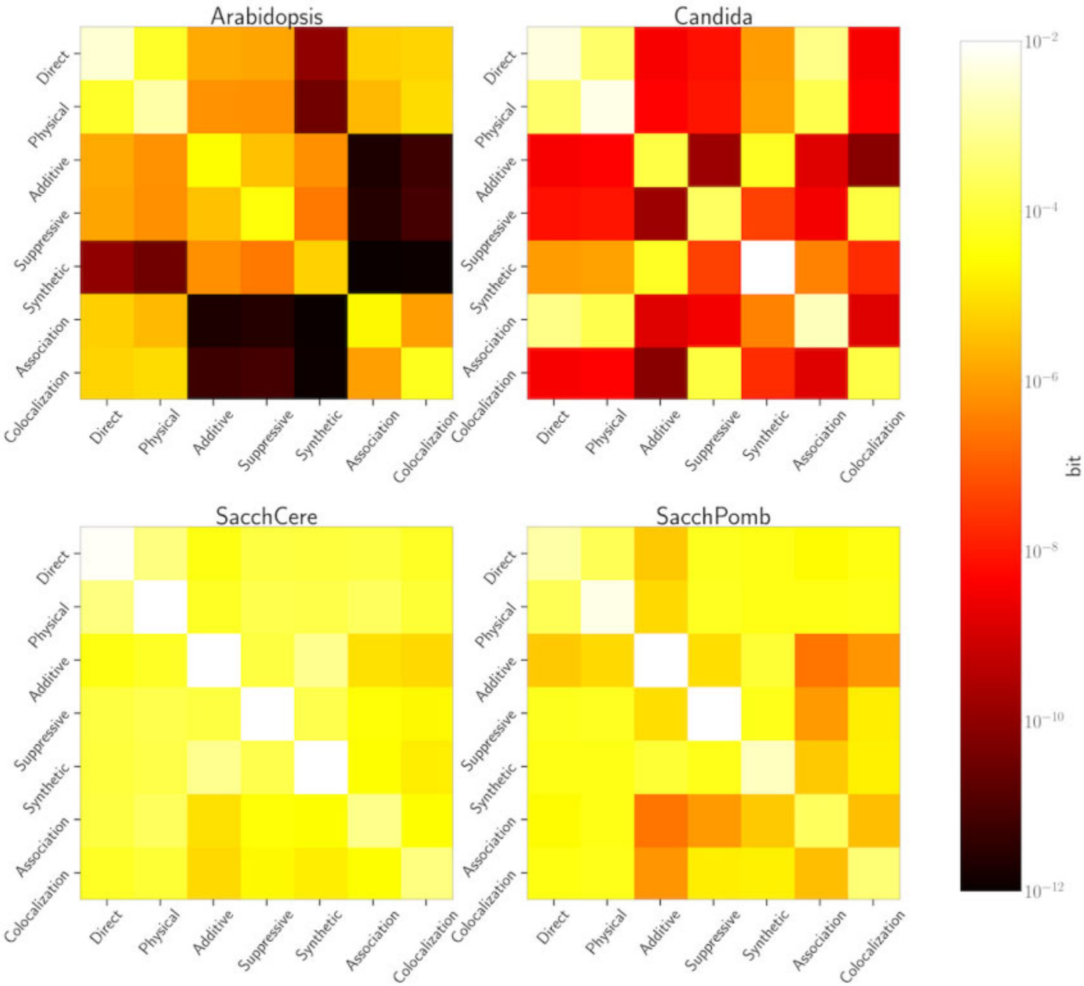


FIG. B1. Mutual Information between layers in four different biology networks. The elements on the diagonal can be seen as layers' entropy.

therefore can been seen as the entropy of the layer and all other mutual information are strictly less than the entropy. As mentioned in Section 3.3, these results show a great variety of correlation strength between different layers. Note that these figures use log scales and there could often be several orders of magnitudes difference between different layers.

C. Stable patterns in ICEWS data

In Section 4.1, we show a typical pattern of the IINF in ICEWS data, here we provide a 17-year period of IINF in ICEWS data demonstrating the patterns of IINF among different years remain to be similar. This is indicating that the underlying mechanism of how different types of interactions affect each other holds constant over the studied time period. In contrast, we see that in airline networks the influence of

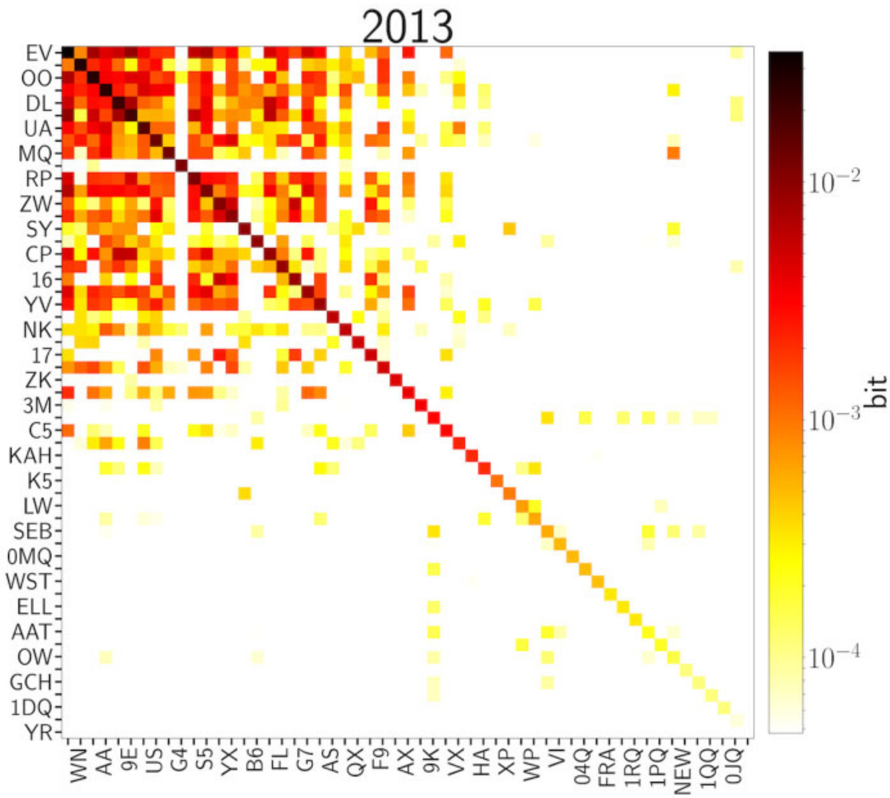


FIG. B2. Mutual Information between airline carriers in 2013. Carriers are ordered from left to right and top to bottom and every other carriers are labelled. The values that are not statistically significant from 0 in a $p = 0.001$ level are truncated. The elements on the diagonal can be seen as layers' entropy.

one carrier upon another changes over time which reflects the changing interaction among carriers over different years. This can be seen in Fig. C1.

D. Time scale and aggregation

When applying the information-theoretic influence to quantifying the correlated structural evolution of multiplex networks, one must also be aware that this measure is sensitive to the choice of time step like transfer entropy. Here, we provide a weekly IINF of ICEWS data for comparison in Figs D1 and D2.

The result from weekly snapshots is qualitatively similar to what we have from yearly snapshots. The patterns remain stable with slight variations. Quantitatively, the magnitude of weekly IINF is smaller which indicate a weaker influence between layers. This is suggesting that for a shorter time period, how one types of interaction helps predicting another is qualitatively similar to a longer time period, but the predicting power is weaker and more volatile.

Another notable factor is how the layer is constructed from the data. Often times there are many ways to interpret data into a multiplex network, the correlated structural evolution according to those different constructions are likely to behave differently as well. As an example, ICEWS data can also

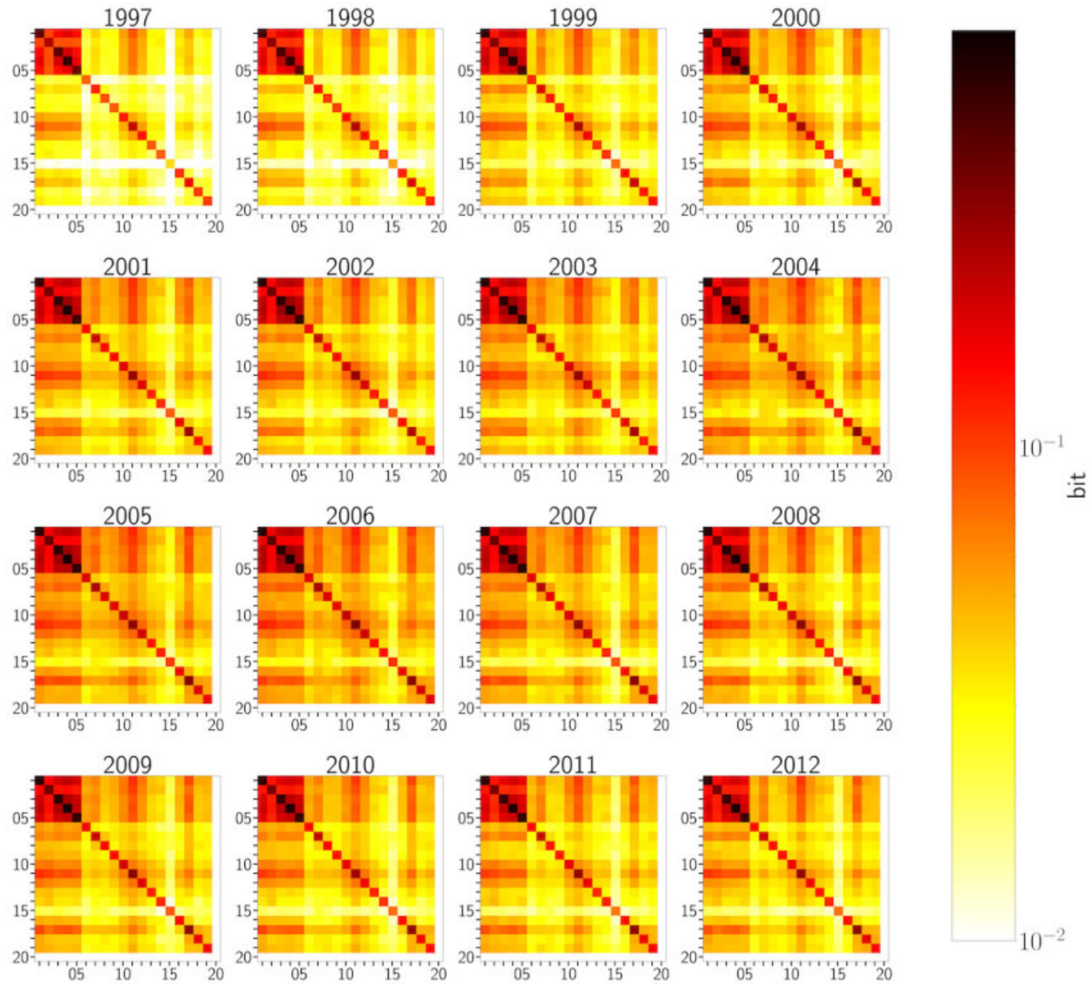


FIG. B3. Mutual Information between layers in ICEWS events network from 1997 to 2012. Types of interactions are labelled by their CAMEO codes from 01 to 20, ordered from left to right and top to bottom. The elements on the diagonal can be seen as layers' entropy.

be classified into a multiplex network using events' penta class, which is a higher level aggregation of CAMEO code [48]. The results with this method is also presented here. This specific observation also shows a potential direction we could pursue. By minimizing the information-theoretic influence between layers, we might be able to divide a multiplex network into a few relatively independent components and study them independently.

E. IINF of airline network

See Fig. E1 for the information-theoretic influence among airlines between the years 2012 and 2013. There are two notable points as mentioned in Section 4.2: the IINF varies greatly among layers and most

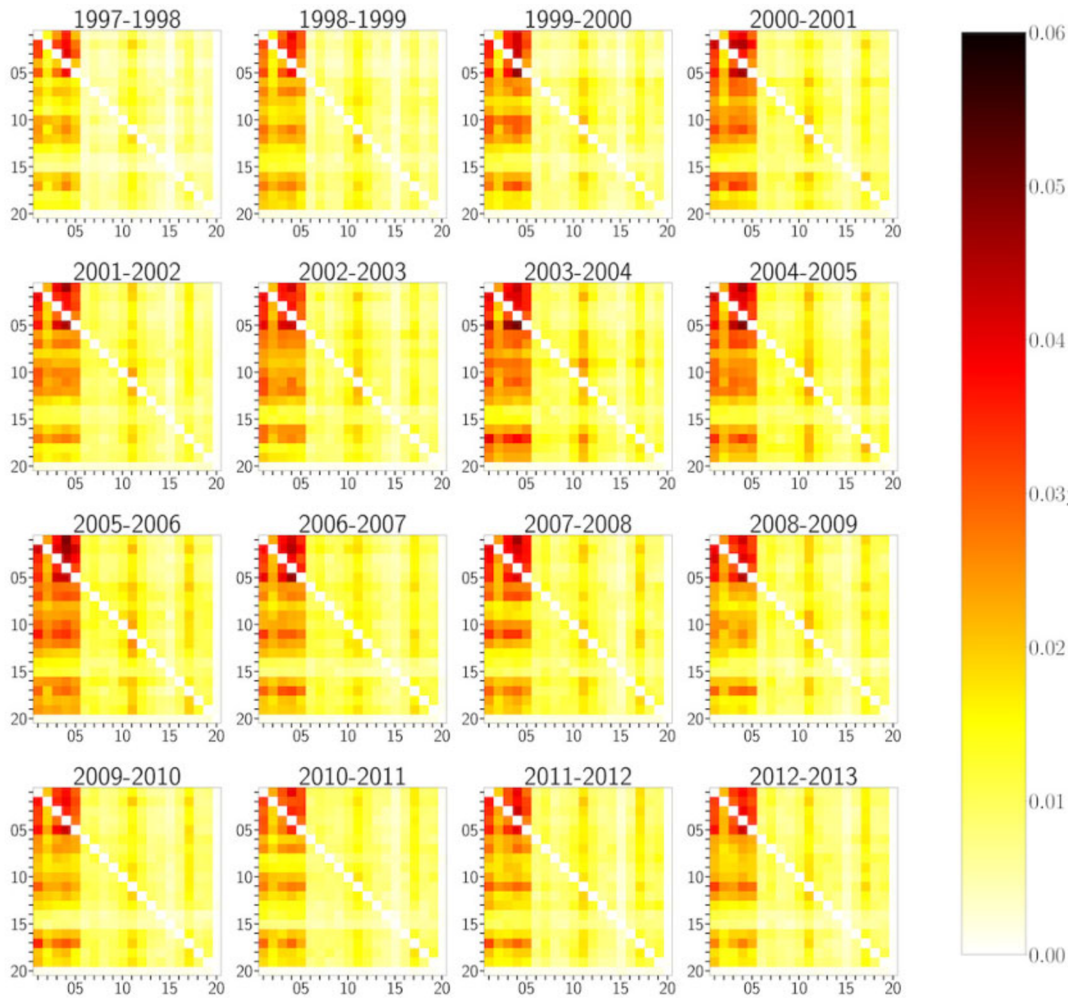


FIG. C1. IINF between layers in ICEWS events network from 1997 to 2013. Types of interactions are labelled by their CAMEO codes from 01 to 20, ordered from left to right and top to bottom. The elements on the diagonal are all 0 by definition.

of influence happens between major carriers, suggesting that some carriers influence each other much more and those influence happen more frequently among major carriers that may be caused by their cooperation or competition.

F. MI and IINF in network formation

At the first glance, it may not be clear whether this framework also works for network formation models beyond Erdös–Rényi, but a carefully thought could show that it is not unreasonable to apply it to others. Let's consider a configuration model where two layers are independently generated from two arbitrary degree sequences. For every pairs of nodes i, j, the probability that there is an edge between i and j in

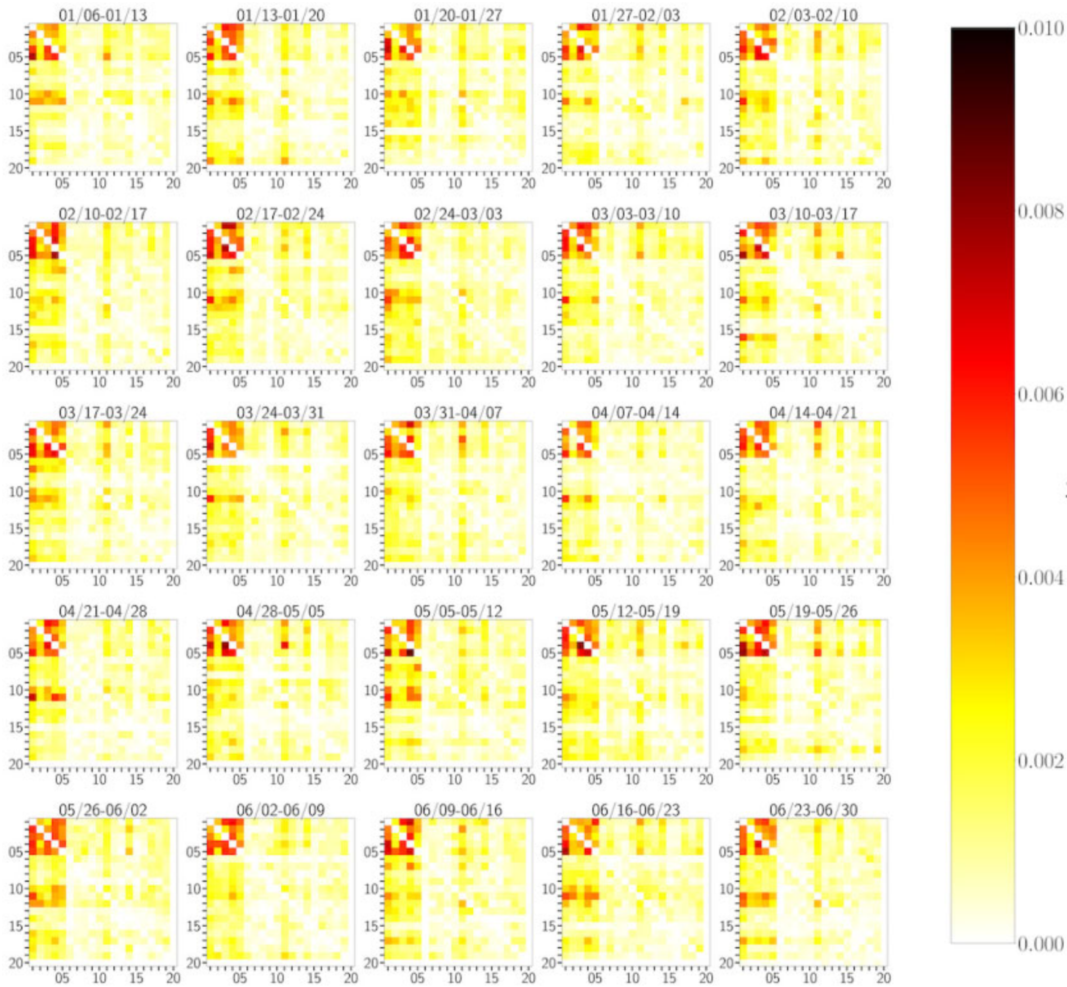


FIG. D1. IINF between layers in ICEWS events network during first 26 weeks of 2014. Types of interactions are labelled by their CAMEO codes from 01 to 20, ordered from left to right and top to bottom. The elements on the diagonal are all 0 by definition.

layer U is $p\left(U_{ij}\right)$. The estimated mutual information and information-theoretic influence through our method then will still be 0 since $p\left(U_{ij}|V_{ij}\right) = p\left(U_{ij}\right)$ for all the edges.

G. MI and IINF for random rewiring

Consider a network with N nodes and M edges, let $\Sigma = N(N-1)/2$ and $\rho = M/\Sigma$. If we randomly rewire k edges in such a network, the mutual information between the network before rewiring and after rewiring is then:

$$I\left(G;G'\right) = H\left(\rho\right) - \left[\rho H\left(\frac{k}{M}\right) + (1-\rho)\, H\left(\frac{k}{\Sigma - M}\right)\right]. \tag{G.1}$$

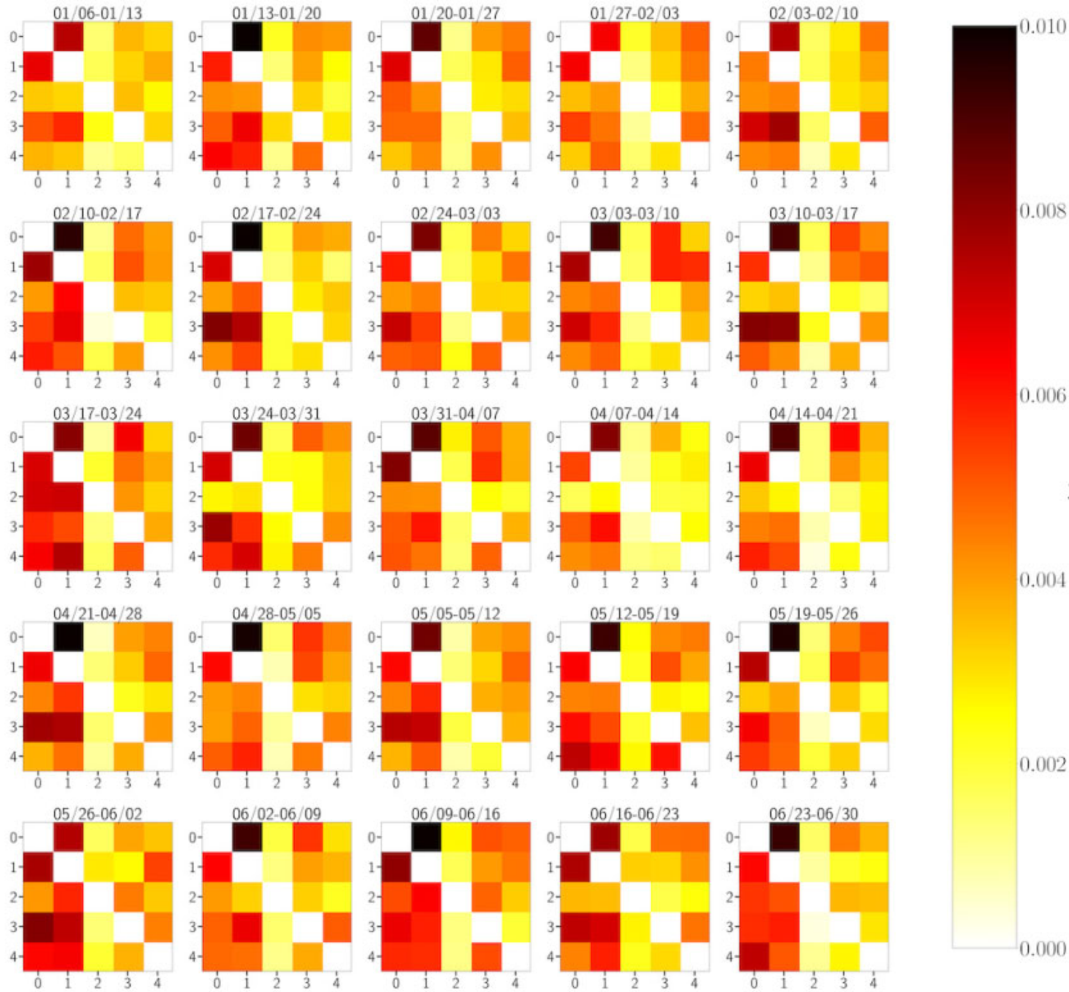


FIG. D2. IINF between layers in ICEWS events network during first 26 weeks of 2014. Aggregated to penta class [48]. The elements on the diagonal are all 0 by definition.

Suppose instead we have two networks G and H both with N nodes and M edges and rewire k edges of G results G' that is the same as H, then the information-theoretic influence is:

$$\begin{aligned} \mathrm{IINF}_{H\to G} &= I\left(H : G'|G\right) \\ &= \rho H\left(\frac{k}{M}\right) + (1-\rho)\, H\left(\frac{k}{\Sigma - M}\right). \end{aligned} \tag{G.2}$$

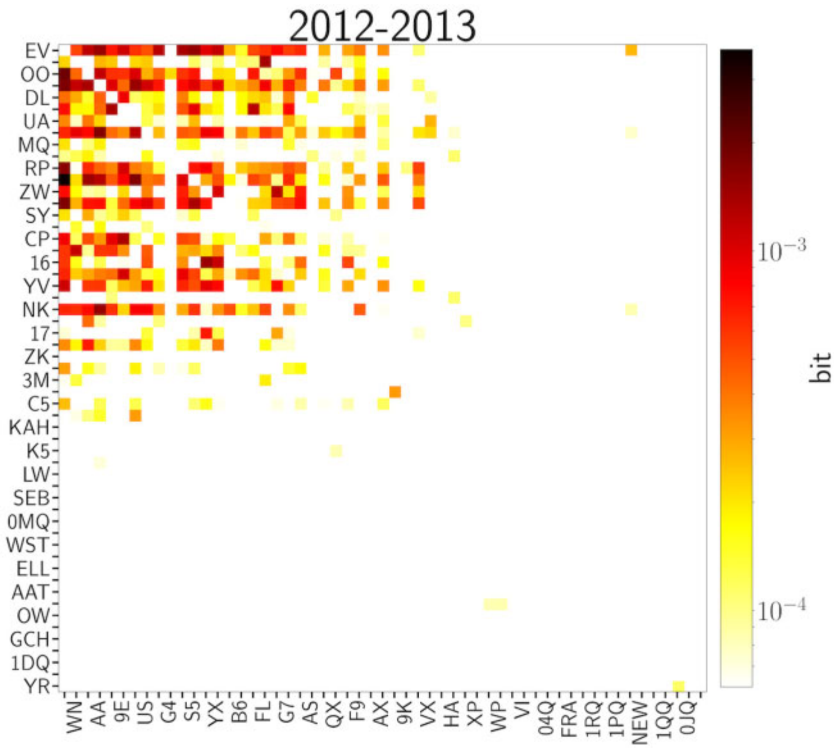


FIG. E1. IINF between all pairs of carriers in 2012–2013 in log scale. Carriers are ordered from left to right and top to bottom and every other carriers are labelled. The values that are not statistically significant from 0 in a $p = 0.001$ level are truncated. Here, we can see that the information-theoretic influence are mainly between those major carriers. Smaller carriers are generally neither influential nor influenced by others. The elements on the diagonal are all 0 by definition.

H. IINF during merge

Consider two networks G and H with N nodes and M_1 and M_2 edges, respectively. Suppose the number of overlapped edges are k and G' is the simple aggregation between G and H, then we have:

$$\begin{aligned}\mathrm{IINF}_{H\to G} &= I\left(H : G'|G\right) \\ &= H\left(\frac{M_2 - k}{\Sigma - M_1}\right) \\ &- \frac{M_1}{\Sigma}\left[H\left(\frac{M_2 - k}{\Sigma - M_1}\right) + H\left(\frac{k}{M_1}\right)\right]\end{aligned} \tag{H.1}$$