

Received March 17, 2020, accepted March 26, 2020, date of publication April 6, 2020, date of current version April 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2986017

# Distinguishing Arc Types to Understand Complex Network Strength Structures and Hierarchical Connectivity Patterns

CHUNG-YUAN HUANG<sup>ID1</sup> AND WEI CHIEN BENNY CHIN<sup>ID2</sup>

<sup>1</sup>Department of Computer Science and Information Engineering, School of Electrical and Computer Engineering, College of Engineering, Chang Gung University, Taoyuan 33302, Taiwan

<sup>2</sup>Applied Complexity Group, Singapore University of Technology and Design, Singapore 487372

Corresponding author: Wei Chien Benny Chin (wcchin.88@gmail.com)

This work was supported in part by the Republic of China Ministry of Science and Technology under Grant MOST-107-2221-E-182-069, and in part by the High Speed Intelligent Communication (HSIC) Research Center, Chang Gung University, Taiwan.

**ABSTRACT** Many real-world networks consisting of nodes representing (in)tangible asymmetric information or energy flows must be modeled as directed graphs (digraphs). Several methods for classifying non-directional edges in terms of strong or weak ties have been developed for well-known non-directional networks, but none specifically for directed networks. In almost all cases, definitions and identification methods are simple, incomplete, reliant on intuition, and based on the assumption that anything that is not weak must be strong. Researchers have generally failed to consider overlapping and hierarchical community properties that accurately reflect organizational structures or the functional components commonly found in real-world complex networks, resulting in multiple challenges to analyzing many types of directed networks. In this paper we describe a method that considers asymmetric definitions of arc strength, especially when arcs hold important directional significance. To more fully capture overlapping and hierarchical network community structures, we used hierarchy-based definitions to identify bond arcs,  $k$ th-layer local bridges, global bridges, and silk arcs and to create a hierarchical arc type analysis (HATA) algorithm. The algorithm employs a mix of common middle node measures and statistical parameters generated by randomized directed networks corresponding to the network being investigated. To test the HATA algorithm, we conducted four experiments involving a mix of arc rewiring and additions, multiple datasets associated with the Travian game, 56 empirical networks from previous studies, and 3 bird song transition networks. Our results indicate that HATA offers a novel perspective to understanding arc strengths and structures in directed complex networks.

**INDEX TERMS** Bond arcs, common middle node, digraphs, directed arcs, global bridges,  $k$ th-layer local bridges, silk arcs.

## I. INTRODUCTION

Researchers have constructed a large number of social and technology networks to study concepts and activities such as online trading, logistics, daily commutes, and the internet, and to monitor cash flows, electric power grids, opinion dynamics, and social networks [1-6]. These networks and related topics undergo continuous analysis and adjustment in step with changes in political and economic systems [7]. Over the past two decades, graph theory [8-9] and related analytical approaches have been used to establish a foundation for studying natural and artificial networks [7, 10-21]. Neuroscientists are using these approaches

to create neuronal maps and to study the plasticity of healthy and damaged brains, engineers are using them to study and design computer network systems associated with the Artificial Intelligence of Things (AIoT), programmers are using them to visualize relationships among Facebook participants and to predict future friend dynamics, and epidemiologists are using them to simulate disease epidemics to test various prevention and treatment policies. Clarification of complex network topological properties can help researchers develop cancer treatment methods or design more efficient devices and protocols for products such as packet classification routers for communication networks [22].

All networks contain two elements: nodes (also known as vertices or points) and edges (also called links). Nodes are

The associate editor coordinating the review of this manuscript and approving it for publication was Haipeng Yao<sup>ID</sup>.

used to represent basic network units (sometimes referred to as elements) such as proteins in protein-protein interaction networks, neurons in brain neural networks, individuals in acquaintance networks, companies in production networks, species in biological webs, base stations and mobile phones in communication networks, and user accounts in social media networks [23-24]. Edges are used to represent tangible or intangible asymmetric information or energy flow between source and target nodes [7]. Tangible examples include compound molecules, commodities, money, electricity, and bacteria; intangibles include trust, tracking, influence, and HTML hyperlinks.

Social network researchers and theorists have studied edge strengths/weaknesses and their roles ever since Granovetter [25] first described how a job candidate's weakest ties have utility in a job search [12, 26-34]. Definitions have been offered for four hierarchical edge strength types in an identification algorithm context: bond edges,  $k$ th-layer local bridges, global bridges, and silk edges [35]. Prior to their work, identification and functional analyses of strong- and weak-tie edge types were limited to well-known non-directional networks—that is, small numbers of nodes and directionless edges in networks identifiable by the naked eye or hand-drawn. In almost all cases, definitions and identification methods have been described as simple, incomplete, reliant on intuition, and based on an assumption that anything that is not weak must be strong. Researchers have generally failed to take into account overlapping and hierarchical community properties that accurately reflect the organizational structures and functional components of complex networks commonly found in real-world domains [2, 36-38]. Overlaps indicate that some nodes belong to more than one community, and hierarchies indicate a possibility of creating numerous small-scale communities with the potential to aggregate into larger ones.

These definitions and identification methods make it difficult to explore many types of directed networks. To make them less intuitive or overly simplistic, Huang *et al.* [35] analyzed non-directional graph scales as large as online friend networks and as small as social networks commonly found in instructional textbooks. In addition to providing detailed and quantifiable definitions for four hierarchical edge types, they developed a hierarchical edge type analysis (HETA) algorithm for determining the strengths and weaknesses of non-directional edges at multiple network hierarchy levels. They also described two practical applications with significant potential for future development: a network fingerprint analysis framework and a top-down hierarchical network community partition method. Their algorithms are capable of identifying edge types from bottom (micro-level) to top (macro-level) network layers. While HETA is useful for differentiating among non-directional edge types, its utility for analyzing directed networks is limited. This paper will consider expansions of Huang *et al.*'s definitions and HETA utility.

Many real-world complex networks consisting of nodes and directed arcs (or directed links) representing (in)tangible asymmetric information or energy flows between source and target nodes must be modeled as directed graphs (digraphs) [39-47]. Directed graph studies involving complex networks are found in engineering, the natural and social sciences, and the humanities. A partial list of specific topics might include brain neural networks, gene regulatory networks, protein-protein interaction networks, biological food webs, power networks, logistics, cash flow, trust networks, disease dissemination, online trading, rail transportation, and the web. Each involves large numbers of nodes and directed arcs, making it difficult to visually determine the strengths of arbitrarily directed arcs.

Several methods for classifying ties as strong or weak have been developed for well-known non-directional networks, but none have been specifically designed for directed networks. One possible but not recommended method entails removing all arc directions in order to change directed networks into non-directional ones, and then using non-directional methods such as the HETA algorithm to determine types. We have two reasons for not recommending this approach: first, for non-directional networks, the loss of arc direction can trigger the appearance of connected nodes that were not connected by directed paths of any length in their original graphs. Second, even though such information can be used for re-identification purposes when networks are anonymized without considering relation direction, removing arc direction can result in a significant loss of information regarding network topological structure. If a researcher then uses previously missing topological structure information to determine arc types, the resulting identifications will be highly inaccurate.

Edge-betweenness is another measure often used (with minor modifications) to describe the effects of directed arc strength and weakness levels in maintaining network connectivity properties [48-50]. This well-researched measure, which can be applied to directed networks using centrality algorithms available in most network analysis platforms [51], is defined as the number (or probability) of shortest paths between all node pairs running through a specific arc. The underlying assumption is that information or traffic travels along the shortest paths. Unlike arcs inside communities, arcs between communities tend to have higher numbers of shortest paths running through them, resulting in higher betweenness values. Deleting arcs with high betweenness values can break up a network while preserving intact community structures. All edge centrality measures are subject to weaknesses associated with a simplicity requirement; edge-betweenness is no exception. Information, for example, clearly does not prefer the shortest path in all cases, and vehicle drivers are often willing to travel longer distances if doing so circumvents traffic problems. Further, edge-betweenness centrality can be unstable, with the addition of only one shortcut arc having the potential to dramatically alter scores of other

network arcs [52]. As is the case with many network statistical methods used to address the shortest-distance problem, edge-betweenness ignores two essential points: real-world complex networks generally have overlapping and hierarchical community properties, and arcs at different network levels generally exhibit ranges of strong/weak properties and influences—an idea that conflicts with previous assertions of arcs having static strength properties. Accordingly, using edge-betweenness and converting networks from directed to non-directional are not suitable strategies for determining directed arc strength.

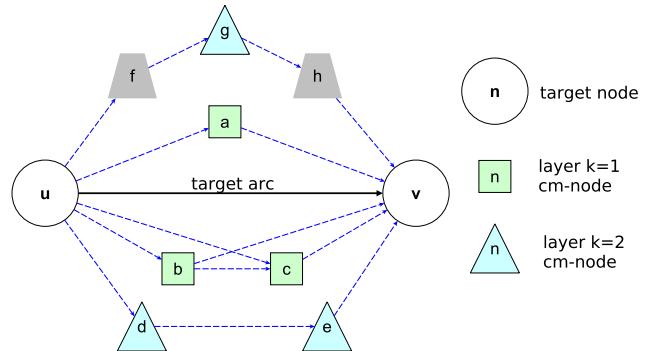
For a more stable and reliable approach, we suggest considering asymmetric definitions of arc strength, especially when arcs hold important directional significance. To determine a definitive solution for identifying four arc types in directed networks, and to more fully capture overlapping and hierarchical network community structures, we used hierarchy-based quantified definitions for determining bond arcs,  $k$ th-layer local bridges, global bridges, and silk arcs to develop a hierarchical arc type analysis (HATA) algorithm. HATA uses three factors to identify directed arc types in all hierarchy layers in a given directed network: arc direction, a common middle node measure involving source and target nodes [1, 47, 53], and statistical parameters generated from an ensemble of randomized directed networks corresponding to the network under investigation [35]. To test the HATA algorithm, we conducted four experiments involving a mix of arc rewiring and additions, multiple datasets associated with the Travian game, 56 empirical networks from previous studies, and 3 bird song transition networks. Our results indicate that HATA offers a novel perspective to understanding the arc strengths and structures of directed complex networks. In this paper we describe two applications with potential to offer insights to structural organizing principles and network functions: network fingerprint analysis, and hierarchical cluster analysis using network fingerprint results.

## II. METHOD

The proposed HATA algorithm is an extension of the HETA algorithm, which focused on non-directed edge type identification. HATA is designed to categorize directed connections as bond arcs,  $k$ th-layer local bridges, global bridges, or silk arcs.

### A. COMMON MIDDLE NODES

HATA is based on a common middle node concept—an extension of the common neighbor concept used in HETA. Common neighbors of node pairs in non-directional networks are targets that are accessible from both source nodes within a specified number of steps. Common middle nodes (hereafter referred to as cm-nodes) are based on the same concept but applied to directed networks. Given any pair of nodes connected by a directed arc, a common middle  $k$ th-layer node is found along a path linking the origin and destination; such nodes require  $k$  or  $k - 1$  steps to cover the distance from the origin to the cm-node and from the cm-node to



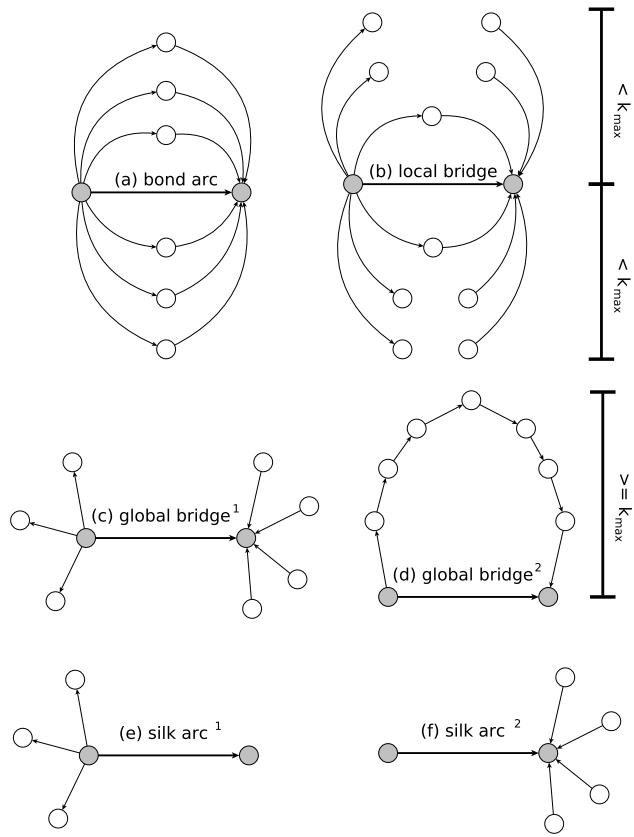
**FIGURE 1.** First-layer (squares) and second-layer (triangles) common middle nodes with arcs connecting node-u to node-v.

the destination. Assume node-u as the origin, node-v as the destination, node-a as the cm-node, and a directed arc connecting node-u to node-v. When shortest paths connecting node-u to node-a and node-a to node-v both require  $k$  steps, then node-a represents a cm-node for the u-v arc.

An example of  $k$ th-layer cm-nodes is shown as Fig. 1. Since nodes-a, -b and -c are one step farther from node-u and one step closer to node-v, they are considered first-layer cm-nodes. Node-d is one step farther from the origin and two steps closer to the destination, and node-e is two steps from the origin and one step from the destination, making them both second-layer cm-nodes. Although nodes-b and -c form a path similar to the u-d-e-v path, they are already considered first layer nodes, and therefore cannot concurrently be considered as second layer nodes. Node-g in the u-f-g-h-v path is two steps after the origin node and two steps before the destination node, and is therefore also considered second-layer. Node-f and node-h are not considered common middle nodes because they are either one step after the origin and three steps before the destination, or three steps after the origin and one step before the destination. Accordingly, they do not meet the  $k$  or  $k - 1$  step definition, and should be treated as part of the second-layer cm-node path.

### B. FOUR ARC CATEGORIES

HATA's purpose is to identify the four arc categories shown in Fig. 2: bond arcs,  $k$ th-layer local bridges, global bridges, and silk arcs. The directed connection between gray nodes in Fig. 2a is a bond arc because it has many first-layer common middle nodes that form an alternative path in the same left-to-right direction. The directed connection in Fig. 2b is a local bridge because it does not have as high a ratio of cm-nodes as the bond arc, even though the cm-node ratio is higher than those for the other arcs. The directed connection in Fig. 2c is a global bridge because none of the outgoing nodes from the gray source node form a path to the gray destination node. The directed connection in Fig. 2d is also a global bridge because the first common middle node is found at the  $k_{max}$  layer, which is one-half the characteristic path length (1). Both connections in Figs. 2e and 2f are silk arcs because both the destination and source connections have



**FIGURE 2.** Examples of six arc types: (a) bond arcs with multiple cm-nodes; (b) local bridges with fewer cm-nodes than bond arcs; (c-d) global bridges with few low- or upper-layer cm-nodes unless there is a large number of cm-nodes on  $k_{max}$  layer; (e-f) one incoming or one outgoing silk arc.

only one incoming or outgoing arc.

$$k_{max} = \max(1, \text{floor}(\frac{1}{2} \times \frac{\sum_u^N \sum_{v \neq u}^N \text{Shortest Path Length}(u, v)}{N \times (N - 1)})) \quad (1)$$

where  $u$  and  $v$  are nodes and  $N$  the total number of nodes. If the calculation produces a floating value smaller than 1,  $k_{max}$  is set to 1. Since average shortest path length typically varies according to the logarithm of the number of nodes [10, 54],  $k_{max}$  is considered proportional to that logarithm rather than incrementally linear.

### C. HATA ALGORITHM

HATA consists of three steps: common middle node ratio calculation (part 1, yellow box), external threshold estimation (part 2, blue boxes), and arc category identification (part 3, orange boxes) (Fig. 3). The static input network has directed, non-weighted, and non-parallel arcs between nodes.

HATA implementation is provided under a Github open source license ([www.github.com/wcchin/HATA](https://www.github.com/wcchin/HATA)). The application was written in Python 3 to take advantage of that language's code readability and packaging support. Ultra-high performance computing was not a

concern because the current implementation was written for proof-of-concept purposes. Computational efficiency can be increased by using GPU-like parallel computing or multi-threading methods to improve cm-node identification and external threshold calculations. Since the latter require a large number of random networks to run cm-node ratio calculations, significant reductions in calculation time can be achieved by running random network generation and cm-node ratio calculations in parallel. Since cm-node identification and comparison are key parts of ratio calculations, the use of GPU-like computing and (sparse) matrix implementation can also significantly reduce run time.

#### 1) PART 1: COMMON MIDDLE NODE RATIO CALCULATION

Given a directed network  $G(V, E)$ , where  $V$  denotes nodes and  $E$  arcs between nodes, cm-nodes ratio ( $R_E^k$ ) can be calculated for each arc at all layers, where  $k \in \{1, 2, \dots, k_{max}\}$ . Specifically, common middle nodes at the  $k$ th-layer of any arc  $e_{u,v} \in E$  represent the intersection of an outgoing node set at  $k$  or  $k - 1$  steps past the origin node- $u$ , excluding the opposite node- $v$  ( $O_u^{k,v}$ , (2) and (3)). The incoming set is  $k$  or  $k - 1$  steps before the destination node- $v$ , excluding node- $u$  ( $I_v^{k,u}$ , (4) and (5)).

$$O_u^{k,v} = \begin{cases} OUT_u^{k,v} - \bigcup_{p=1}^{k-1} O_u^{p,v} - \{v_u, v_v\}, & \text{if } k > 1 \\ \{v_y | (v_u, v_y) \in E\} - \{v_u\}, & \text{otherwise} \end{cases} \quad (2)$$

$$OUT_u^{k,v} = \{v_y | \forall v_x \in O_u^{k-1,v} : (v_x, v_y) \in E\} \quad (3)$$

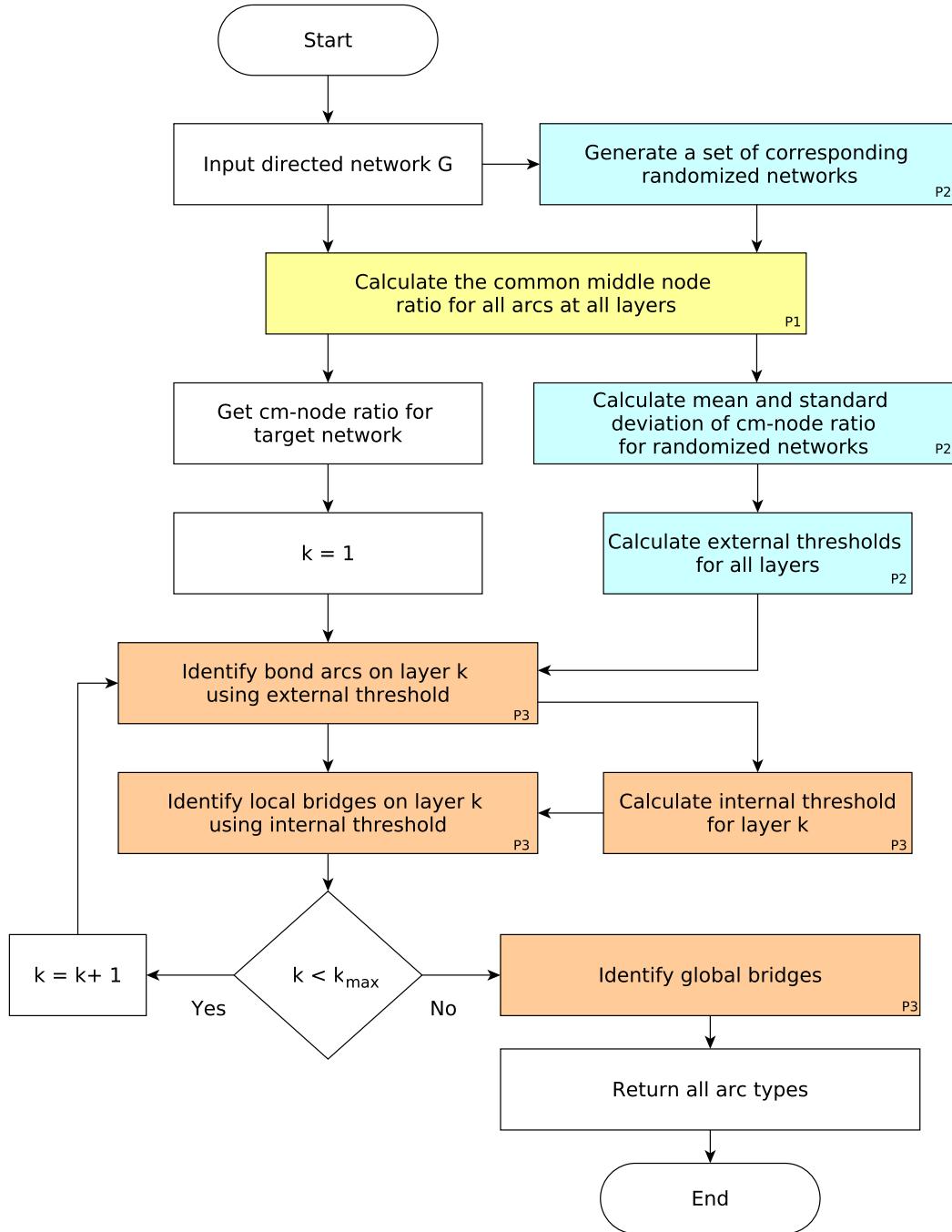
$$I_v^{k,u} = \begin{cases} IN_u^{k,v} - \bigcup_{p=1}^{k-1} I_v^{p,u} - \{v_u, v_v\}, & \text{if } k > 1 \\ \{v_x | (v_x, v_v) \in E\} - \{v_v\}, & \text{otherwise} \end{cases} \quad (4)$$

$$IN_u^{k,v} = \{v_x | \forall v_y \in I_v^{k-1,u} : (v_x, v_y) \in E\} \quad (5)$$

For arc  $e_{u,v}$ , let the outgoing node set from node- $u$  at the  $k$ th-layer (excluding the opposite node- $v$ ) be written as  $O_u^{k,v}$ , and let the incoming node set to node- $v$  at the  $k$ th-layer (excluding node- $u$ ) be written as  $I_v^{k,u}$ . The cm-node ratio  $R_{u,v}^k$  can be calculated using (6) to (8), which also includes the intersections of the outgoing node set at the  $(k - 1)$ th-layer ( $O_u^{k-1,v}$ ), the incoming node set at the  $k$ th layer ( $I_v^{k,u}$ ), the outgoing node set at the  $k$ th-layer ( $O_u^{k,v}$ ), and the incoming node set at the  $(k - 1)$ th-layer ( $I_v^{k-1,u}$ ). For each arc  $e_{u,v} \in E$ , a cm-node ratio set  $R_{u,v}^k$  is calculated as  $\{R_{u,v}^1, R_{u,v}^2, \dots, R_{u,v}^{k_{max}}\}$ , which is then used for arc category identification. The cm-node ratio calculation is used in two HATA procedures: one to calculate external thresholds in generated random networks (part 2), and one to determine target network arc types (part 3).

$$R_{u,v}^k = \begin{cases} CM_{u,v}^k / Neigh_{u,v}^k, & \text{if } CM_{u,v}^k > 0, \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where  $CM_{u,v}^k$  is the number of cm-nodes for an arc  $e_{u,v}$  at the  $k$ th-layer, which can be calculated using (7).  $Neigh_{u,v}^k$  is the number of potential cm-nodes for an arc  $e_{u,v}$  at the  $k$ th-layer,

**FIGURE 3.** HATA algorithm flowchart.

which can be calculated using (8).

$$CM_{u,v}^k = |O_u^{k,v} \cap I_v^{k,u}| + |O_u^{k,v} \cap I_v^{k-1,u}| + |O_u^{k-1,v} \cap I_v^{k,u}| \quad (7)$$

$$Neigh_{u,v}^k = \min(|O_u^{k,v}|, |I_v^{k,u}|) + \min(|O_u^{k,v}|, |I_v^{k-1,u}|) + \min(|O_u^{k-1,v}|, |I_v^{k,u}|) \quad (8)$$

## 2) PART 2: RANDOMIZATION FOR ESTIMATING EXTERNAL THRESHOLDS

HATA uses an external threshold to identify arcs with relatively high cm-node ratios. Thresholds are estimated using large numbers of randomized networks (e.g.,  $M = 1000$ ), each with the same structure as the target network in terms of number of nodes, number of arcs, and in- and out-degree distributions. Random networks are generated by an algorithm

that uses in-degree and out-degree node lists to generate random networks with the same degree distribution [54]. In the Python 3 version of HATA, a built-in NetworkX network generator (`directed_configuration_model`) is used for the randomization process.

Each generated randomized network is used to run the previously described procedure (part 1) to calculate cm-node ratios for each  $k$ th-layer ( $R_E^k$ ) arc. For external threshold estimations, the means ( $Mean_e(R_E^k)$ ) and standard deviations ( $SD_e(R_E^k)$ ) of cm-node ratios for all arcs in the randomized network are calculated and stored. After all randomized networks have been generated and calculated, the external threshold is estimated using (9), with means and standard deviations for each randomized network. In other words, when the  $M$  number of random networks is sufficiently large, the average and standard deviation of a graph's cm-node ratio is expected to resemble a normal distribution. Accordingly, if the cm-node ratio is above a point that is one standard deviation larger than the average, it is considered significantly high.

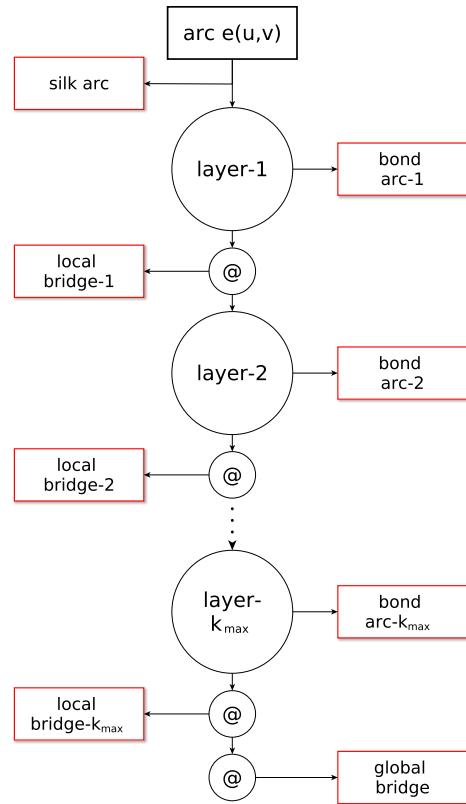
$$T_{ext}^k = Mean_r(Mean_e(R_E^k)) + Mean_r(SD_e(R_E^k)) \quad (9)$$

### 3) PART 3: ARC CATEGORY IDENTIFICATION

All arcs containing at least one endpoint are considered as having only one incoming or outgoing degree, and are therefore categorized as silk arcs (Fig. 4). The arc category identification process consists of iterations from  $k = 1$  to  $k = k_{max}$ . During each iteration, the cm-node ratios of unrecognized arcs (calculated in part 1) are compared to an external  $k$ th-layer threshold (calculated in part 2); arcs with equal or higher ratios (i.e.,  $R_E^k \geq T_{ext}^k$ ) are categorized as bond arcs. All other arcs are added to a candidate arc set and used to calculate an internal threshold ( $T_{int}^k$ , (10)) for comparison with the cm-node ratios of all candidate arcs; when the ratio exceeds the internal threshold ( $R_E^k > T_{int}^k$ ), the arc is categorized as a  $k$ th-layer local bridge. Remaining unrecognized arcs are again added to a candidate arc set to be used during the next iteration (i.e.,  $k = k + 1$  if the current  $k < k_{max}$ ). Iterations stop when the current  $k = k_{max}$  and all remaining candidate arcs are recognized as global bridges. At the end of the identification process, all arcs are recognized as silk arcs, bond arcs,  $k$ th-layer local bridges, or global bridges.

$$T_{int}^k = Mean_c(R_c) - SD_c(R_c) \quad (10)$$

Based on the arc categorization process just described, an arc may be identified as occupying a lower layer (e.g., first-layer bond arc or local bridge) if it exhibits a concrete and tight structure on a local scale—that is, from the perspective of the ego networks of the two endpoints. For arcs that are not identified as having smaller local scales, HATA increases  $k$  to determine whether arcs can be categorized according to their larger local scales. Iterations continue until  $k = k_{max}$ , which is approximately one-half the average path length required to connect all node pairs in a network. If an arc

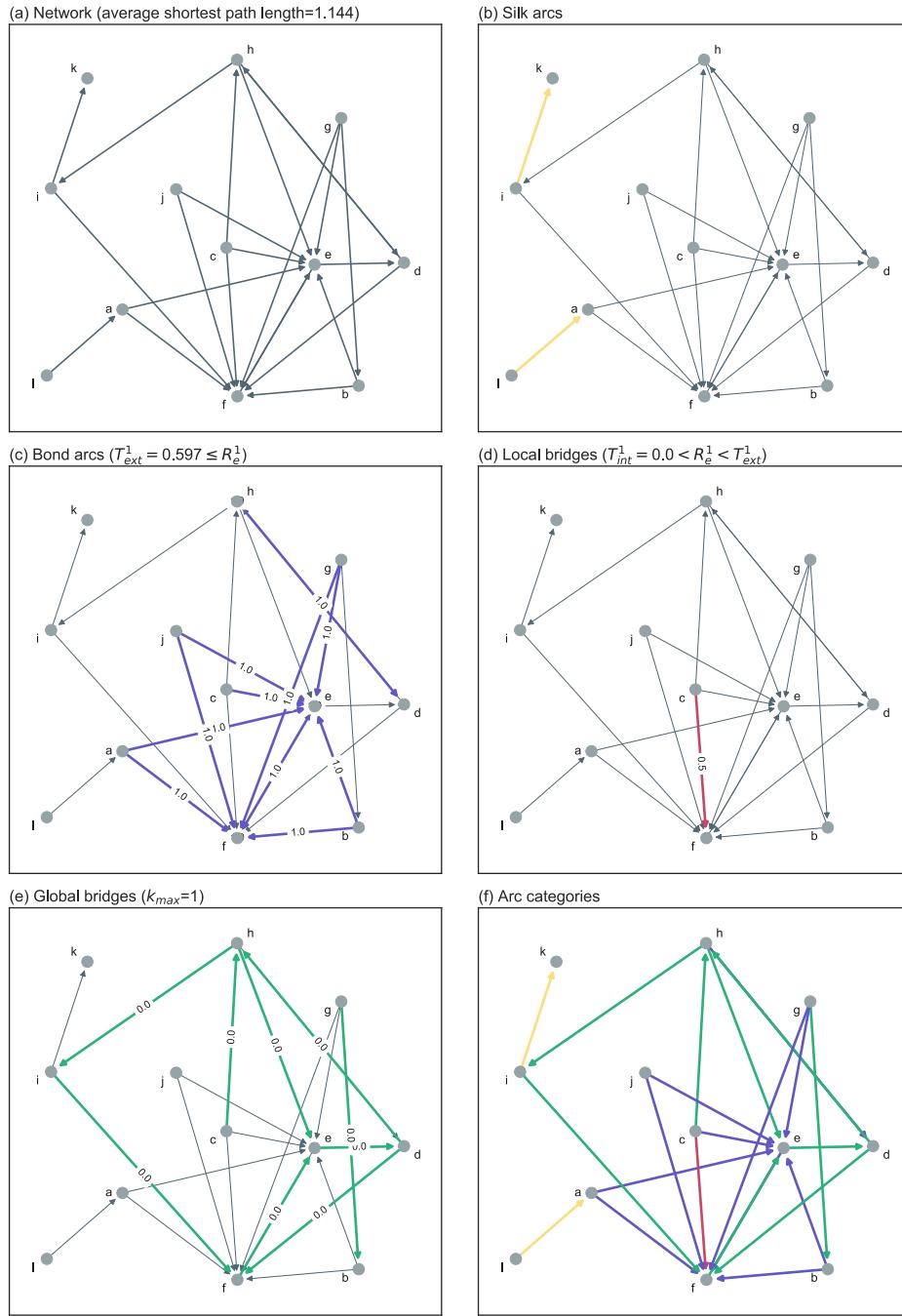


**FIGURE 4.** Arc category identification process. The process begins with an uncategorized arc and stops at any red box indicating a specific category. For any arc  $e_{u,v}$ , the purpose of the first step is to determine if it is a silk arc. Non-silk arcs are sent to the first layer, where they are compared to an external threshold to determine if they are first-layer bond arcs. Next, they are compared to an internal threshold to determine if they are first-layer local bridges. Arcs not belonging to either category are passed on to the second layer for comparison to a second-layer threshold. Any arc that is not identified as a silk arc,  $k$ th-layer bond arc, or  $k$ th-layer local bridge when  $k$ th-layer  $\leq k_{max}$  is categorized as a global bridge.

still cannot find a sufficient number of cm-nodes to become a bond arc or local bridge after determining post-origin and pre-destination half-lengths, the cross-node relationship is considered weak and therefore identified as a global bridge.

### D. DEMONSTRATION NETWORK

We will use a small network consisting of 12 nodes and 23 arcs for demonstration purposes. As shown in Fig. 5a, the average shortest path length value was 1.144, indicating a  $k_{max}$  of 1 according to (1). The first HATA step identified the silk arcs, with single arc endpoints having either one incoming or one outgoing arc (Fig. 5b). In the figure, arcs  $e_{i,k}$  and  $e_{l,a}$  are silk (yellow color) because node-k and node-l respectively have only one incoming and one outgoing arc. Post-randomization, an external threshold of 0.597 was calculated for the first layer. Arcs with cm-node ratios ( $R_E^1$ ) equal to one (blue color) were identified as bond arcs (Fig. 5c). Node-a has two outgoing nodes, node-e and node-f, resulting in arc  $e_{a,e}$  having one cm-node (node-f) because arc  $e_{f,e}$  forms an alternative path from node-a to node-e. Since node-f is the only outgoing node from node-a,  $R_{a,e}^1 = 1.0$ , indicating a

**FIGURE 5.** Examples of HATA algorithm arc categories.

single bond arc. The remaining candidate arcs were used to calculate an internal threshold  $T_{int}^1$ . Arcs with  $R_E^1 = 0.5$  (red color) were identified as first-layer local bridges (Fig. 5d)—for example, arc  $e_{c,f}$  was identified as a local bridge because node-c has two outgoing nodes (-e and -h) with only one cm-node (-e), hence  $R_{c,f}^1 = 0.5 > T_{int}^1$ . Since  $k_{max} = 1$ , remaining arcs (with  $R_E^1 = 0.0$ ) after silk arc, bond arc, and first-layer local bridge processing were identified as global bridges (Fig. 5e). Fig. 5f presents the final output of the HATA algorithm for this demonstration network.

Among the 23 arcs in the demonstration network, two were silk (8.7%) and 11 bond (47.8%), with one first-layer local bridge (4.3%) and nine global bridges (39.1%). These percentages served as a network fingerprint, with two sets of overlapping arcs pointed in opposite directions, the first set consisting of arcs between nodes-h and -d (one bond arc  $e_{h,d}$  and one global bridge  $e_{d,h}$ ) and the second consisting of arcs between nodes-e and -f (one bond arc  $e_{e,f}$  and one global bridge  $e_{f,e}$ ). In other words, there were different connectivity strengths for two arcs connected to the same nodes but in

opposite directions, one with a significantly high cm-node ratio and one with a single path connecting the origin and destination nodes.

### III. EXPERIMENT DESIGN

Four experiments were conducted to test HATA. Four types of directed networks were generated based on two small world network models: Watts and Strogatz [55] (designated WS) and Newman and Watts [56] (designated NW). The WS model randomly rewrites links, and the NW model randomly adds links to regular ring lattice networks according to a specified probability. A ring consisting of 50 nodes was generated for each network model, with each node connected to its eight nearest neighbors (four on the left, four on the right). Next, we converted edges into two types of directed arcs: bidirectional (eight outgoing and incoming arcs connecting the four neighbors on each side) and unidirectional (four outgoing to neighbors on the left and four incoming from neighbors on the right). We then used a probability series (0.001, 0.002, 0.004, 0.016, 0.032, 0.064, 0.128, 0.256, 0.384, 0.512, 0.640, 0.768, 0.896, 1.0) to randomly rewire and add arcs to the initial networks. We generated 100 networks for each probability for the two-way WS, one-way WS, two-way NW and one-way NW networks, and used them with HATA to identify arc types. Fingerprints were applied to determine differences between networks. The first experiment was conducted to (a) test HATA and investigate algorithm outcomes in terms of identifying arc structures and strengths in each network, and (b) understand changes in arc strengths resulting from connectivity changes using the WS and NW networks with the bi- and unidirectional ring networks.

The second experiment focused on arc types using three sets of networks from the Travian dataset [57-58]. Travian is a browser-based massively multi-player online game (MMOG) in which players build their own villages, defend them, and attack other villages. Players can create military and economic alliances with other players. Extracted Travian network datasets are used to establish attack, message, and trade networks, each containing 30 directed networks representing 30 continuous days within a 3.5-month game cycle. HATA was used with 90 networks to identify arc types and to determine network differences. The primary objective for the second experiment was to clarify arc strengths and structures in the three social relationship networks.

For the third experiment we used 56 real world networks to test HATA. They were divided into 4 groups: 25 animal-related, 10 citation, 10 hyperlink and p2p transfer, and a group consisting of 11 networks (5 social, 3 infrastructure, 2 metabolic and 1 neural). The largest weakly connected component (LWCC) was extracted from each dataset. Data were collected from three repositories: Pajek datasets (Vlado, <http://vlado.fmf.uni-lj.si/pub/networks/data/>), the Koblenz Network Collection (Konect, <http://konect.uni-koblenz.de/>), and the Stanford Large

**TABLE 1. Data sources for the 56 real world networks used in our experiments.**

Category	Data repository	Network case
animal	Konect	moreno_hens, moreno_sheep, maayan-foodweb
animal	Vlado	CrystalC, Everglades, gramwet, gramdry, Mondego, Narragan, CrystalD, ChesUpper, mangwet, mangdry, ChesMiddle, Rhode, Michigan, Maspalomas, ChesLower, baydry, Florida, baywet, cypdry, cypwet, StMarks, Chesapeake
citation	Konect	dblp-cite, subej-cora, HepPh, HepTh
citation	Vlado	SmallW, SmaGri, SciMet, Kohonen, Lederberg, Zewail
internet (hyper-link)	Konect	moreno_blogs
internet (p2p)	Snap	Gnutella08, Gnutella09, Gnutella06, Gnutella05, Gnutella04, Gnutella25, Gnutella24, Gnutella30, Gnutella31
others (social)	Konect	moreno_taro, moreno_highschool, moreno_innovation, ego-gplus, ego-twitter
others (infrastructure)	Konect	opsahl-openflights, openflights, maayan-faa
others (metabolic)	Konect	maayan-Stelzl, maayan-figeys
others (neural)	Snap	C-elegans-frontal

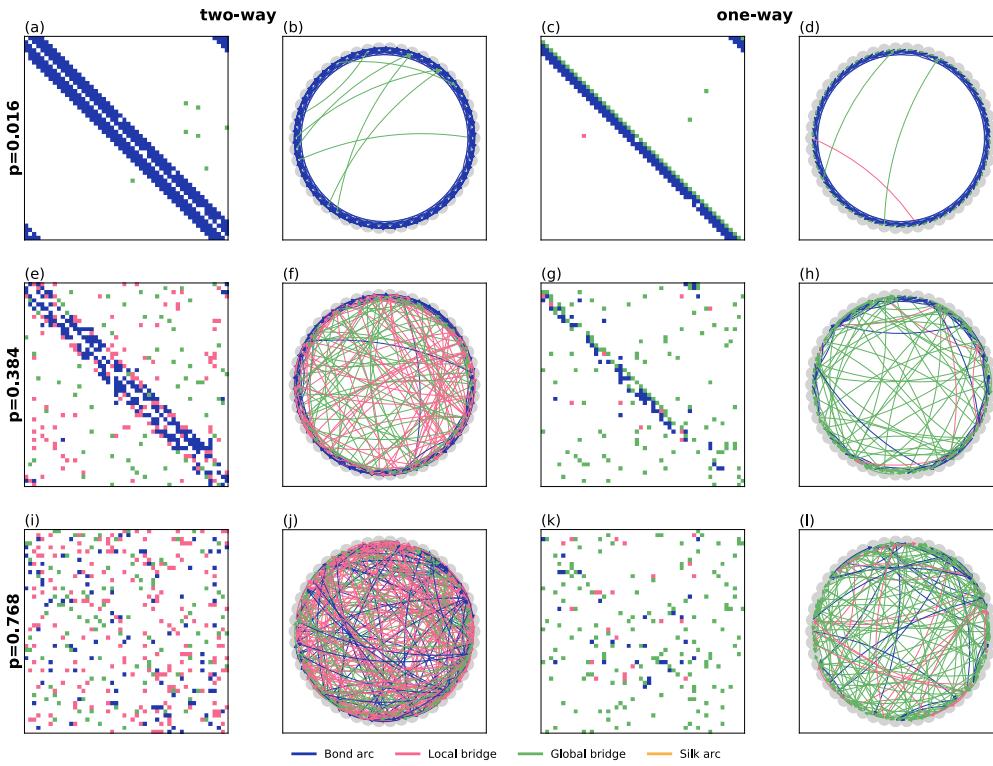
Network Dataset Collection (Snap, <https://sparse.tamu.edu/SNAP/>) (Table 1).

In the fourth experiment, three small networks used for two analyses and visual observations were generated from Cassin's Vireo song sequences [59]: sample g1236 (11 nodes, 27 arcs), sample g1421 (25 nodes, 51 arcs), and sample g1682 (34 nodes, 75 arcs). The two analyses were compared to other methods and a non-directional network test. Four similarity measures based on the common middle node concept were used: common neighbor counts and the Jaccard, cosine, and minimum similarity indexes [60]. Edge-betweenness, a key bridge concept measure for edge-based characteristics, was also used for comparison purposes. The goal for the second part of experiment 4 was to clarify the effects of HATA calculations for a non-directional network. The three small networks were used to convert directed to non-directional networks for HETA calculation purposes, and then convert them to bidirectional networks for HATA calculations. The two analyses helped clarify differences between HATA and HETA and between HATA and other network measures.

## IV. RESULTS

### A. EXPERIMENT 1

Six WS small world networks were generated from three random rewiring probabilities: 0.016 (low, first row), 0.384 (medium, second) and 0.768 (high, third). The WS small world networks in the two columns on the left side of Fig. 6 were generated using bidirectional arcs, and in the two



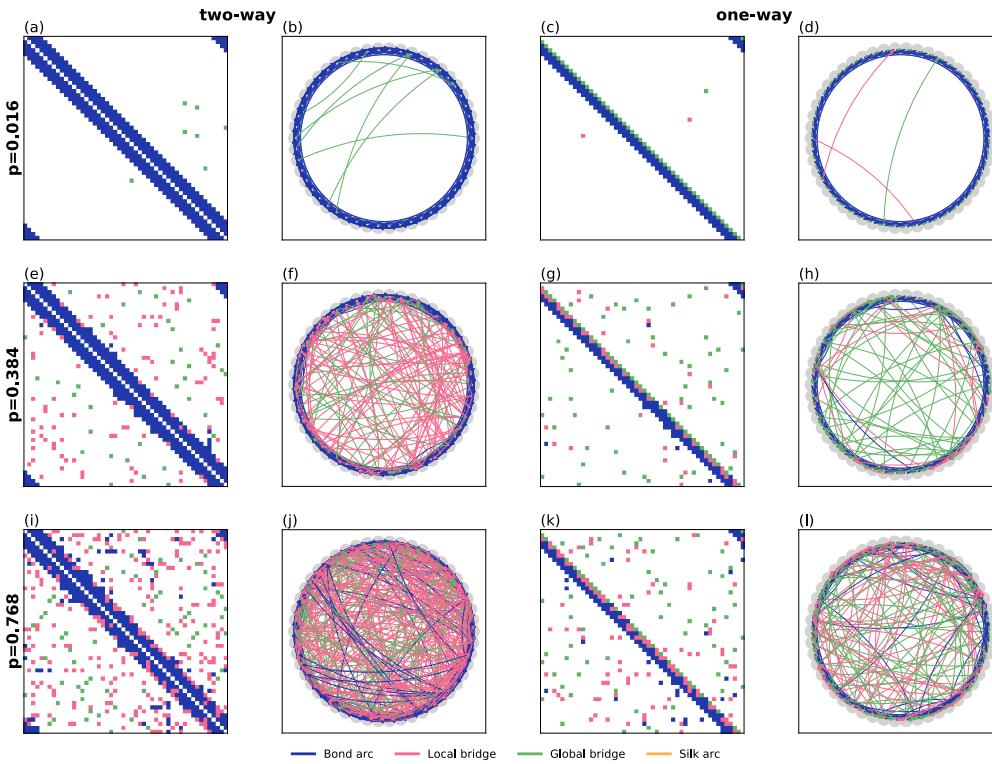
**FIGURE 6.** Examples of generated WS small world networks and corresponding adjacency matrices. First and third columns indicate network adjacency matrices for networks in the second and fourth columns. Rows 1-3 show WS small world networks generated using low (0.016), medium (0.384), and high (0.768) random rewiring probabilities. Blue, bond arcs; red, local bridges; green, global bridges; yellow, silk arcs.

columns on the right using unidirectional arcs. Blue indicates bond arcs, green global bridges, and red local bridges. Since the network at the lowest probability resembles a regular ring lattice, its structure is preserved. Global bridges emerge from rewired arcs consisting of green cells in the matrix that appear at random points away from the diagonal section, and green arcs that connect nodes across the interior space. The ring structure was still observable in the diagonal part of the adjacency matrix when the probability increased to 0.384, with many local and global bridges appearing in the two-way directed network and many global bridges with smaller numbers of local bridges appearing as rewired arcs. The WS small world ring structure disappeared as the probability increased to 0.768, and the network structure became disorganized in both bi- and unidirectional small world networks. According to our results from the low and medium random rewiring probabilities, most of the rewired arcs were global or local bridges.

One difference between the bi- and unidirectional WS small world networks concerns the types of arcs found on the ring structures, especially the diagonal sections. In the bidirectional network, the closest neighbors 1 through 4 were all identified as bond arcs (Figs. 6a and 6e). In contrast, the first neighbor in the unidirectional network was identified as a global bridge (with cells closer to a diagonal line of adjacency matrices), while neighbors 2-4 were identified as

bond arcs (Figs. 6c and 6g)—a counter-intuitive but certain observation. We found that for each directed arc between any nearest node pair, the source node had three outgoing arcs connected to the next three neighbors situated ahead of the target node, hence they did not form an alternative connective path from source to target. In other words, alternative paths between nearest node pairs did not exist in the unidirectional directed small world networks we examined. However, our data indicate that directed arcs between nearest node pairs can serve as part of the alternative paths connecting second-, third-, and fourth-nearest neighbors.

Fig. 7 shows six NW small world networks based on the same probabilities and layout structures as in Fig. 6. Since NW networks generate additional arcs without removing any existing arcs, the initial ring lattice structure is preserved at all probabilities, as seen along the diagonal lines in the adjacency matrices. The partial initial arc types along the ring become local bridges when new arcs are added to the network; as more arcs are added, the number of initial arcs changing to local bridges increases. Added arcs at all probabilities were primarily global or local bridges. For the bidirectional networks, initial arcs in ring structures were identified as bond arcs when the probability of additional arcs was low ( $p=0.016$ ). A large number of nearest neighbor relationships (starting with the fourth nearest neighbor) were identified as local bridges as the probability increased to 0.384, and an

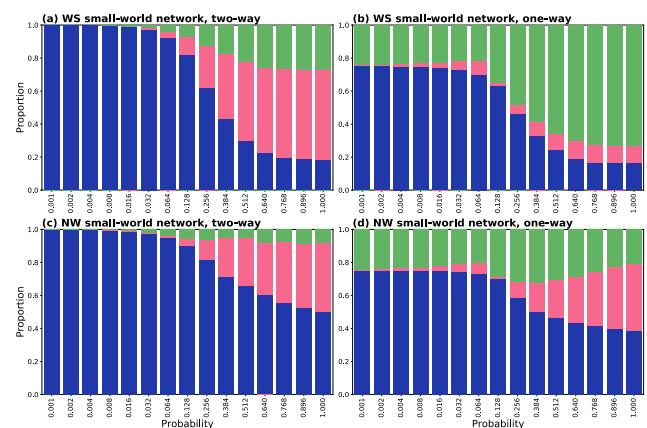


**FIGURE 7.** Examples of generated NW small world networks and corresponding adjacency matrices. First and third columns indicate adjacency matrices for the networks in the second and fourth columns. Rows 1-3 show NW small world networks generated using low (0.016), medium (0.384), and high (0.768) arc addition probabilities. Blue, bond arcs; red, local bridges; green, global bridges; yellow, silk arcs.

even larger number of initial arcs became local bridges as the probability increased to 0.768. For the unidirectional networks, our results were similar to those for the WS networks: immediate neighbor relationships involved global bridges, while arcs for neighbors 2-4 were bond types when the probability was low ( $p=0.016$ ). As probabilities increased to 0.384, some arcs connecting second-nearest neighbors became local bridges, while bond arcs connecting neighbors 3 and 4 did not change. As probabilities increased to 0.768, local bridges started to appear between nearest neighbors 2 and 3, while bond arcs connected to fourth nearest neighbors did not change.

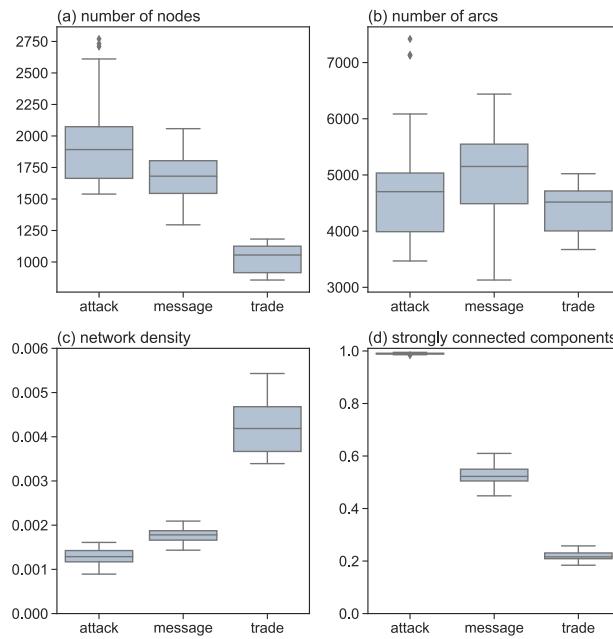
Since total arc numbers increased in step with the probability of additional arcs, external and internal threshold values for evaluating NW network arc types differed across various probabilities. Common middle node values changed as arcs were added to the network, therefore arc types in the initial ring structure changed as probabilities increased. Among unidirectional NW networks, relationships involving second-nearest neighbors were closer to first-nearest neighbor arcs than those involving third- and fourth-nearest neighbors. For reasons similar to those for unidirectional WS networks, it was easier for nearest-neighbor arcs to become local bridges.

Fig. 8 presents the mean fingerprint from the four experiments (15 probabilities and 100 repetitions per experiment).



**FIGURE 8.** Fingerprint analyses of four network sets: (a) two-way WS, (b) one-way WS, (c) two-way NW, (d) one-way NW.

In all cases, more arcs were identified as bridges (global or local) and fewer as bond arcs as probabilities increased. As shown, bond arc proportions decreased (a) from 1.0 to 0.18, (b) 0.75 to 0.16, (c) 1.0 to 0.5 and (d) 0.75 to 0.39 as probabilities increased from 0.001 to 1.000. Changes for the NW networks were fewer than those observed in the WS network experiment, perhaps because the ring structure was maintained in the former but not in the latter. For bidirectional experiments involving both network models, more local than



**FIGURE 9.** Metrics for three network sets. (a) Number of nodes, (b) number of arcs, (c) network density, (d) normalized number of strongly connected components.

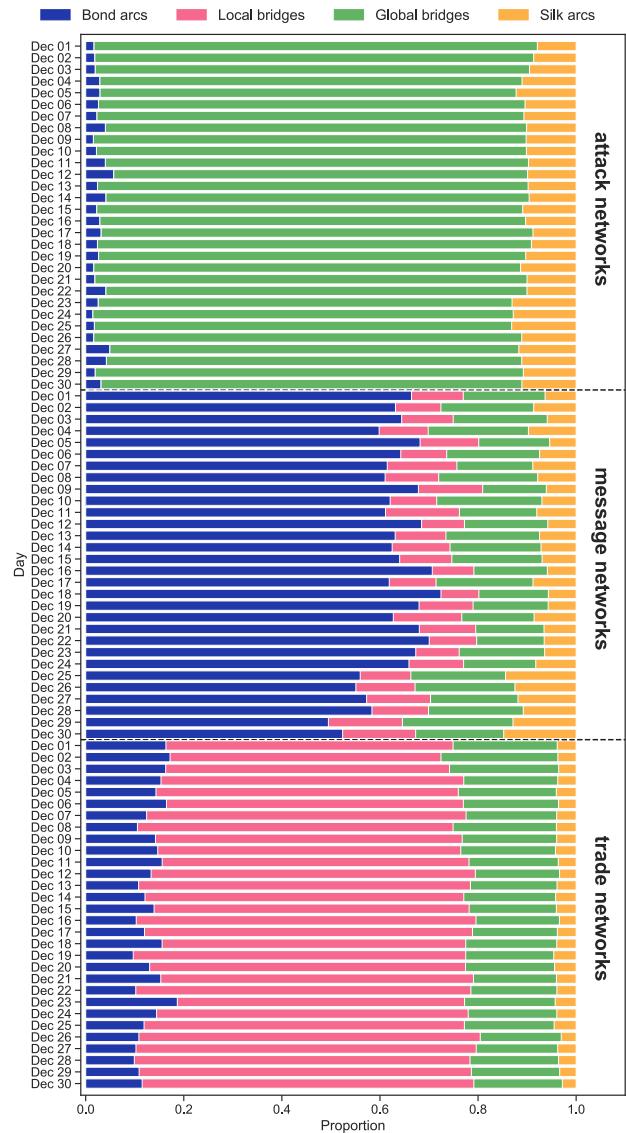
global bridges emerged as the probability of rewired or additional arcs increased. For the WS experiment, the local and global bridge proportions for bidirectional networks at a 1.0 probability were 0.55 and 0.28, respectively, and in the NW experiment they were 0.41 and 0.08, also respectively. For the unidirectional networks, more global bridges appeared in WS settings (0.73 at a 1.0 probability) and more local bridges in NW settings (0.40 at a 1.0 probability).

In summary, HATA captured the influences of rewired and additional arcs, and identified which arc types changed during rewiring or addition processes. The first experiment clarified the different impacts of rewiring changes (WS) and additions (NW) to a network's arc structure, which can also affect arc strength. The HATA bidirectional and unidirectional network results reveal differences in the arc strength structures of the two directed small world networks we examined. The rewired WS arcs and additional NW arcs represent shortcuts, and the varying WS and NW probabilities indicate that HATA is capable of identifying these shortcuts and distinguishing among their various strengths.

## B. EXPERIMENT 2

For the second experiment we used a social network consisting of data for three types of interactions in the Travian online multiplayer game: attacks involving individual players, messages exchanged between players, and resource transaction behaviors (Fig. 9).

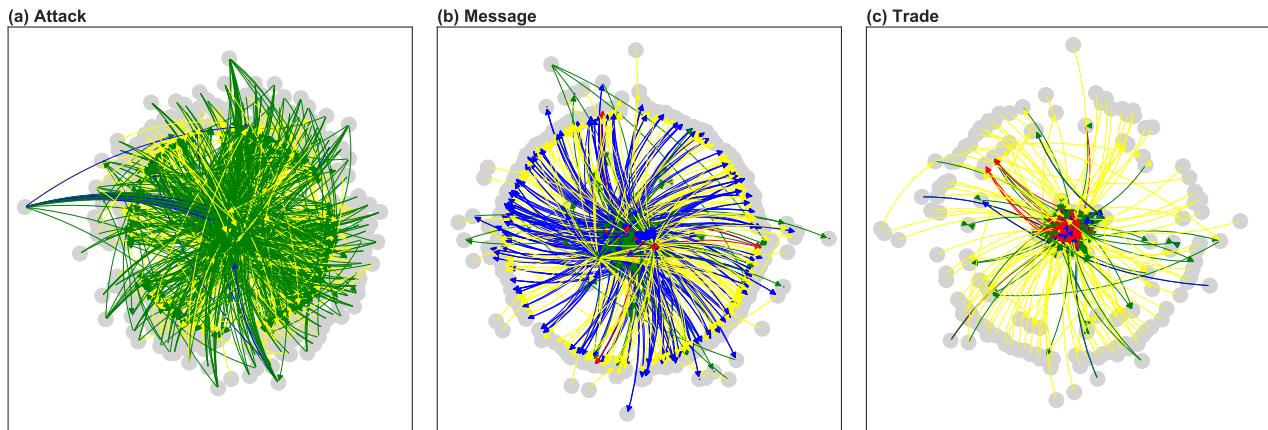
Each dataset represented 30 networks, one each for 30 days in a month. Data included player numbers in each network, arc numbers within each network set, network density (number of arcs divided by number of possible arcs, i.e.  $n \times n - 1$ , with  $n$  denoting node number), and normalized number of



**FIGURE 10.** Fingerprint derived from 90 networks divided into three groups.

strongly connected components within a network (i.e., number of strongly connected components divided by number of nodes, with the normalized value = 1 when all nodes forming separate individual components). Attack network node numbers were slightly higher than those in message networks, and both were higher than trade network node numbers. Arc numbers indicate slightly more message than attack interactions, and roughly equal trade and attack interactions. The large number of strongly connected attack network components indicates that attacks were mostly aimed at fixed numbers of players, possibly forming hierarchical and acyclic relationships consisting of stronger players attacking weaker ones. The low number of strongly connected components involved in trade interactions indicates the presence of a large and active trade chain.

Our arc type proportion data from fingerprint analyses involving the three datasets show network fingerprint consistency within datasets and differences across them (Fig. 10).



**FIGURE 11.** Examples of three Travian networks using data for games played on December 10, 2009: (a) attack, (b) message, (c) trade. Blue, bond arcs; red, local bridges; green, global bridges; yellow, silk arcs.

Attack networks contained much larger percentages of global bridges ( $\approx 85\%$ ) than bond ( $< 5\%$ ) and silk arcs ( $\approx 10\%$ ), and zero local bridges. Message networks had many bond arcs ( $\approx 60\%$ ), and much fewer bridges (20% global, 10% local) and silk arcs ( $< 10\%$ ). Different patterns were noted in the trade networks: mostly local bridges ( $\approx 60\%$ ) followed by global bridges ( $\approx 20\%$ ), bond arcs ( $\approx 15\%$ ), and silk arcs ( $\approx 5\%$ ).

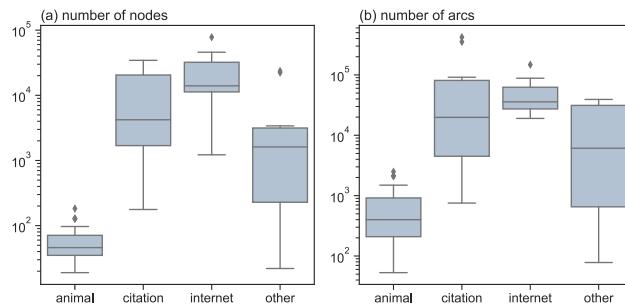
Fig. 11 shows one sample network from the 30 networks constituting each dataset, using the same color scheme as in Figs. 5-7. Node positions were calculated using the ForceAtlas2 algorithm [61], a Python module found at <https://github.com/bhargavchippada/forceatlas2>. Curved and directed arcs in Fig. 11a indicate source-to-target attack behaviors. The finding that most arcs were global bridges (green arrows) and silk arcs (yellow arrows) might be explained by two attack behavior characteristics. First, players tend to attack weaker rather than stronger opponents, resulting in unidirectional internode relationships. This hierarchical structure may increase characteristic path lengths, thus making the task of finding alternative paths more difficult. Second, situations in which two players are assailed by the same attackers indicate that they have identical enemies with similar or identical strengths who are unlikely to attack each other. The resulting unique arcs make it difficult to find alternative paths between linked nodes, meaning that most arcs are global bridges or silk arcs. This agrees with a previous finding that in MMOG attack networks, player separation tends to be greater and clustering coefficients generally lower [57].

Most messaging relationships were found within rather than across alliances (Fig. 11b). While communication mostly involved players with close relationships, communication network nodes contained some communities (e.g., friendship zones or player groups within alliances) with strong connections, resulting in most observed arcs being the bond type. In some large alliances, players might communicate with one or more small groups with whom

they have close relationships (possibly via group leaders or communication intermediaries), but not with all alliance members, resulting in a number of local bridges. Further, global bridges might have players who are connected across different alliances, or players who are spatially close but not in the same alliance; such intra-network connections are potential global or local bridges.

Most of the arcs shown in Fig. 10 are local bridges, but only silk (yellow) arcs are visible in the trade network shown in Fig. 11c. Results from the ForceAtlas2 node-positioning algorithm show a relatively dense collection of non-silk arcs at the center, possibly because interactions between node pairs involving local bridges, global bridges, and bond arcs were stronger than silk arc interactions. Global bridges represented the second highest proportion after local bridges, indicating that trade relationships were generally non-mutual, or at least not immediately bilateral. An example might consist of player-a buying resources from player-b, who has another customer in player-c. In this scenario, players-a and -c are unlikely to have a buying-selling relationship because players consuming the same kinds of resources might be training similar types of armies, and are therefore unwilling to sell the same resources they consume in large quantities. The absence of immediate bilateral trade interactions can result in a lack of short alternative paths, and therefore in fewer bond arcs. While our trade interaction structure did not exhibit a strong hierarchical pattern similar to that for attack relationships, the trade network did have longer alternative paths compared to the attack network, resulting in more local than global bridges.

The Travian networks represent three common MMOG relationships with distinctive differences. Attack networks consist of aggressive action structures involving individual players or player groups, message networks reveal information-sharing structures, and trade networks reveal relationships entailing resource and production chains involving players with different roles within or across alliances. Our datasets provide ways for understanding real-world network



**FIGURE 12.** Box plots for two network metrics for the four network groups in experiment 3.

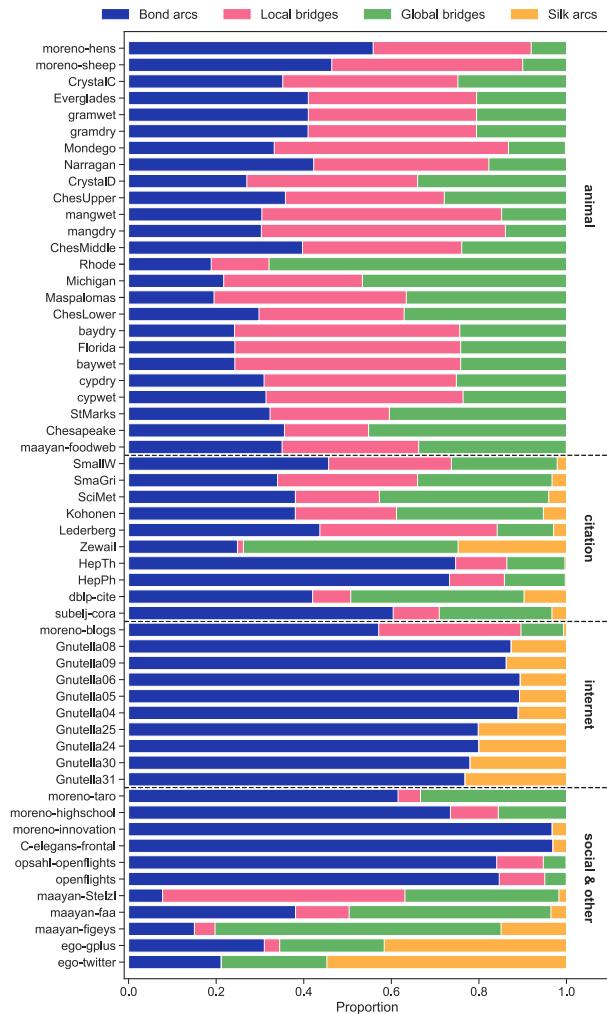
structures: the strong-weak hierarchical structure found in attack relationships produces higher percentages of global bridges, the social proximity pattern found in message interactions produces larger numbers of communities and bond arcs, and the non-bilateral structure of trade relationships produces higher percentages of bridges, with more local than global.

### C. EXPERIMENT 3

For the third experiment we tested HATA with 56 directed networks used in previous studies. As shown in Fig. 12, animal networks had small node numbers (<100 for 100–1,000 arcs) but dense connectivity structures (>0.1), while citation networks had large ranges of sizes (2,000 to 20,000) and arc numbers (5,000 to 90,000). Compared to citation networks, internet networks were slightly larger in terms of node (10,000 to 20,000) and arc numbers (30,000 to 60,000), but with smaller within-group variation. In terms of network density, citation networks were denser ( $4 \times 10^{-4}$  to  $3 \times 10^{-3}$ ) than internet networks ( $6 \times 10^{-5}$  to  $3 \times 10^{-4}$ ).

Fig. 13 shows the overall fingerprint and arc type proportions for the 56 networks. Most of the 25 animal networks consisted of 30% to 40% bond arcs,  $\approx$  40% local bridges, and 20% to 30% global bridges. The animal network arc composition indicates strong interactions, perhaps due to high network structure densities. Citation network fingerprint variation is indicative of different connectivity and arc strength levels. Among the 10 citation networks, the Zewail network had the lowest numbers of bond arcs and local bridges and the highest numbers of global bridges and silk arcs, while the HepTh and HepPh networks had > 70% bond arcs,  $\approx$  10% local bridges, a smaller percentage of global bridges, and a very small percentage of silk arcs.

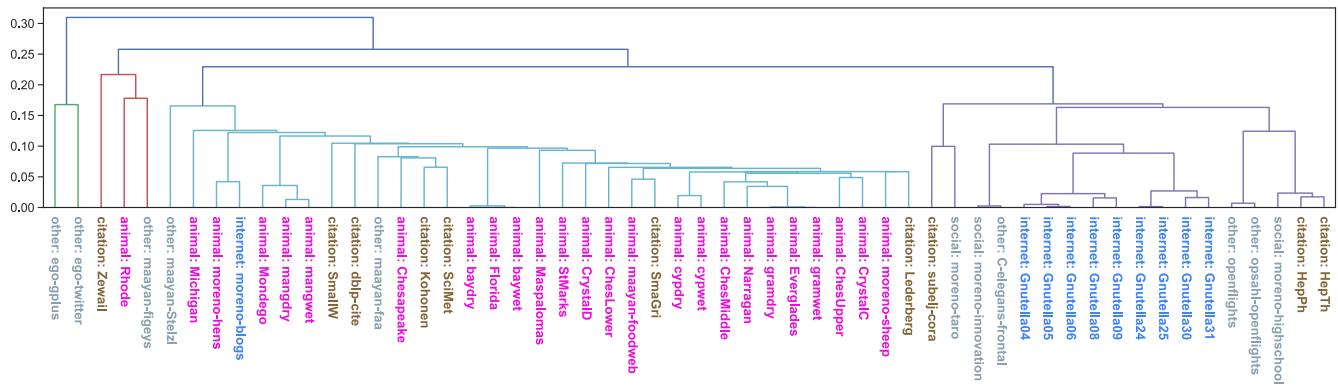
Among the internet networks, the moreno-blogs network fingerprint differed from those for the nine single-day Gnutella network snapshots. The moreno-blogs network consists of connective hyperlinks based on urls found on the first pages of individual blogs; the network structure is divisible into two main communities, with internal community connectivity denser than intra-community connectivity [62]. This network category was found to consist of 57% bond arcs, 32% local bridges, 10% global bridges, and < 1% silk arcs,



**FIGURE 13.** Fingerprint analysis results for 56 real world networks. Network orders in each group are sorted as most-to-least densely connected.

a composition likely reflecting the tendency of individuals who hold similar views to interact online. In contrast, Gnutella networks emphasize peer-to-peer (p2p) file-sharing interactions [63], which may explain the exceptionally high percentage ( $\approx$  80%) of bond arcs, with the rest identified as silk arcs. The absence of local and global bridges in these networks indicates tight connections and sufficient short alternative paths between nodes, with some nodes only exchanging specific file types with the same user. Since the Gnutella protocol was designed to let nodes send requests to actively connected nodes, the connectivity structure of these p2p networks is strong and stable, resulting in a large percentage of bond arcs.

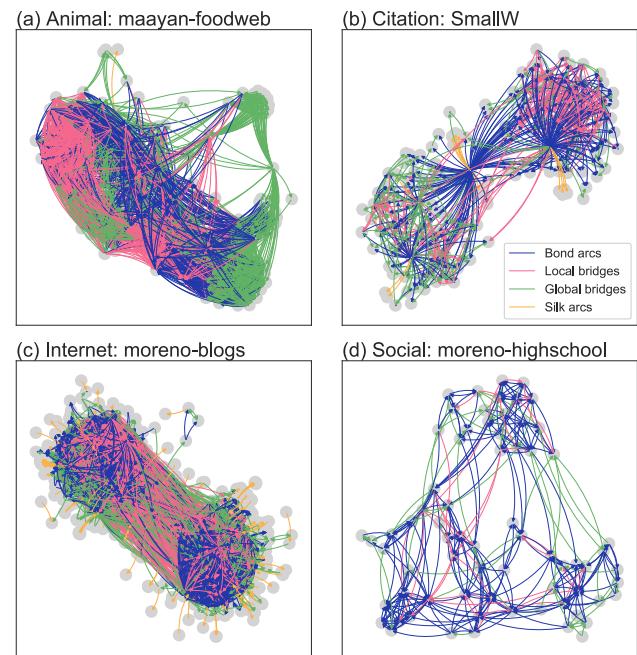
The multiple categories of other networks resulted in numerous fingerprint varieties for analysis. The moreno-taro social network, which consists of gift exchanges between households, had 60% bond arcs and 40% global or local bridges [64]. The moreno-highschool network captures male friendships within the same high school, based on a survey in which students were asked to identify their friends [65].



**FIGURE 14.** Cluster analysis results for 56 real world networks.

The data were surprisingly similar to those for a Travian message network: a large percentage of bond arcs plus a few local and global bridges. The moreno-innovation network represents academic interactions and discussion relationships involving physicians [66]. Due to the tendency of members to only elicit comments and opinions from individuals they believe are capable of providing professional-level information, the tight network connection again results in a large number of bond arcs and a small number of silk arcs. Note that silk arcs frequently indicate the presence of individuals who only ask one question or talk to one specific person. The complete absence of local and global bridges makes this network unique, and perhaps representative of all academic social circles and interactions. The other two social networks, ego-gplus and ego-twitter, are based on follower relationship samples from the Google Plus and Twitter social network platforms [67]. Note that follower relationships differ from mutual friendship relationships. The idea that individuals can choose to follow any other participant results in large numbers of silk arcs with few, if any, connecting nodes. The smaller number of bond arcs ( $< 30\%$ ) might be explained by the small number of arcs aimed at friends; such arcs are more likely to be found in social communities consisting of mutual friendship relationships.

Fig. 14 presents data from a cluster analysis using finger-print results. Colors indicate animal, citation, internet, and “other” categories. Whereas the animal networks showed high degrees of similarity (resulting in tighter clusters), the citation networks showed high degrees of cluster separation. The distinct Gnutella p2p network patterns clustered at lower dendrogram hierarchies, while the political blogs network shared similarities with some animal networks. The two ego social networks were similar to each other but significantly different from the other 54 networks. Our cluster analysis results also show that, with the exception of the two ego networks, all networks in the “other” category (taro giving, high school friendship, and innovation social networks, three infrastructure networks, two metabolic networks, and a single neural network) were dispersed across different positions within the dendrogram and connected to each other only at higher levels.



**FIGURE 15.** Examples of networks used in experiment 3 with arc types noted for each category.

Four sample networks from each category are illustrated in Fig. 15, with arc colors indicating HATA results. The maayan-foodweb network shown in Fig. 15a illustrates food-web relationships at Little Rock Lake in Wisconsin [68]. Each node represents an autotroph, herbivore, carnivore or decomposer, and each link represents the movement of a food source from producer to primary consumer and from prey to predator. In this hierarchical network there is a large number of arcs connecting a large number of nodes between top left and bottom right; HATA identified these arcs as bonds, local bridges, and global bridges. One node in the middle of the right side appears to be an origin node with no incoming but numerous outgoing arcs. All outgoing arcs from this node, including those at the top and bottom right, were identified as global bridges.

The presence of these global bridges in the maayan-foodweb is due to the clear hierarchical structure of the

relationships between nodes—that is, the first species serves as a food source for the second, which serves as a food source for the third. Note that the third species does not eat the first. Although the nodes at the top right are not interconnected, they all have some global bridge or local bridge arcs pointed toward the largest group of nodes, indicating shared food sources, similar groups of predators, and a shared position on the same hierarchy level (probably an herbivore). This network shares a similarity with the Travian attack network—that is, the foodweb lacks cross-level consumption relationships, as indicated by the sole presence of weak arcs (global bridges). Instead, the foodweb consists of more complex relationships consisting of all bond arcs, all local bridges, or all global bridges. The HATA results for this network indicate a mix of inter-species relationships labeled as strong (bond arcs), weak (global bridges), or somewhere in-between (local bridges), again influenced by the hierarchical network structure.

The SmallW network shown in Fig. 15b is representative of the citation networks used in Milgram's Small World Problem [69]. SmallW has a common source—an initial journal article influencing all other studies. Subsequent studies can be divided into two main topics, as indicated at the top right and bottom left in the figure. This citation network implies a time of publication for each article, resulting in a strong hierarchical structure from older-to-more recent studies. This structure differs slightly from that of the foodweb network in that most (if not all) of the articles cite the same common origin study. Accordingly, SmallW contains multiple cross-level links and alternative paths, resulting in higher numbers of bond arcs and lower numbers of global bridges compared to the maayan-foodweb. One interesting SmallW network outcome is that most of the weak arcs were found within each group rather than between groups. Most global bridges are in the lower left, and most local bridges in the upper right.

The moreno-blogs network shown in Fig. 15c consists of hyperlinks between political blogs published during the 2004 American presidential election [62]. Unlike the foodweb and citation networks, the nodes in this network (each representing one blog contributor) did not share a strong hierarchical structure because each writer was considered as having equal status, although social interactions sometimes resulted in certain nodes gaining greater influence. The network pattern consisted of two major groups of nodes, one per political party preference. Interaction strength is indicated by the prevalence of bond arcs; cross-group interactions are indicated by the large numbers of local and global bridges. Our data show a higher number of bond arcs, reflecting the tendency among writers to reference other writers in the same group, thus emphasizing their common core values. However, they also reference members of the opposing party when replying to (mostly attacking) their opponents' posts. In these cases, cross-group interactions were weaker compared to within-group interactions.

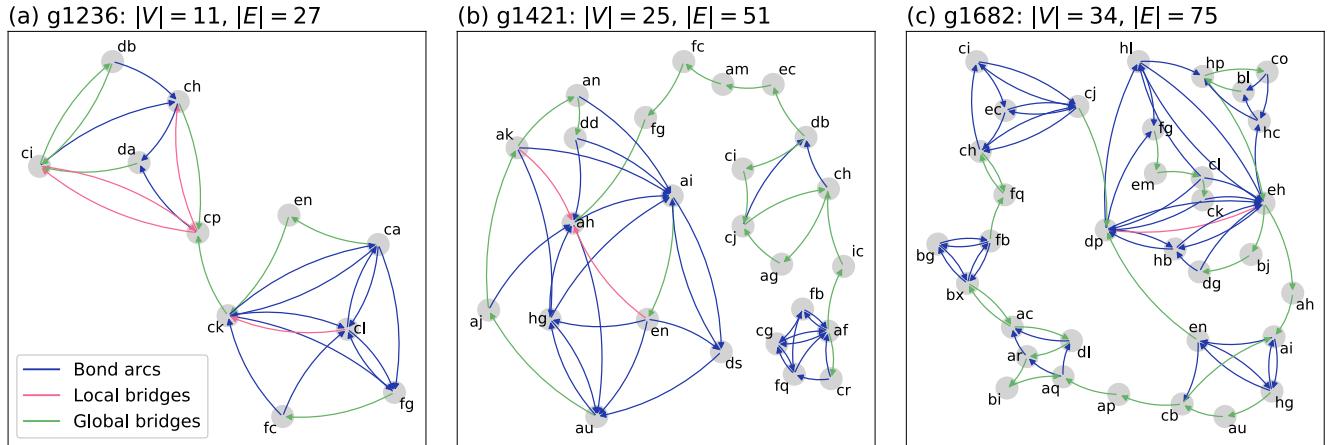
The moreno-highschool network shown in Fig. 15d consists of a friendship network of students attending the same high school [65]. Similar to the blogs network, all nodes in this network are considered as having equal status, with no hierarchical restrictions. As mentioned in an earlier section, the fingerprint of this network is similar to that of the Travian message network; the interaction structure is readily observable in Fig. 15d. Three groups of nodes are visible: one group at the top and one each at bottom left and bottom right. The bottom two groups have a strong connectivity relationship consisting of a large number of bond arcs and a smaller number of local bridges. All of the bottom-to-top arcs were global bridges, but some of the top-to-bottom arcs were the bond type.

In summary, all of the networks shown in Fig. 15 contain arc type information for describing cross-node interactions and for analyzing how their connectivity structure (i.e., hierarchical or equal status) affects the connectivity strength of directional arcs. Our data indicate that the HATA algorithm is capable of providing information in support of detailed analyses of network structure.

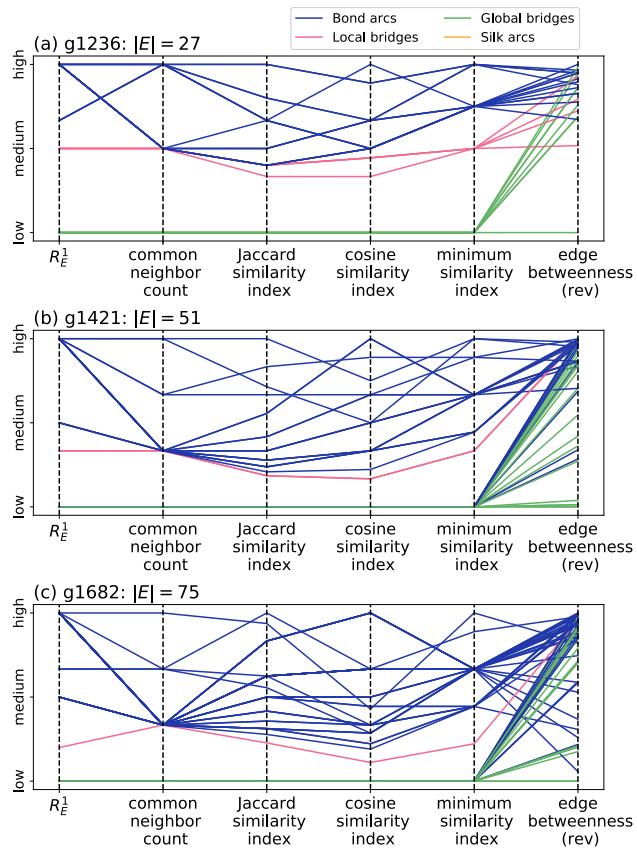
#### D. EXPERIMENT 4

We tested HATA with three small directed bird song networks chosen for their visual representations of calculation results. The  $k_{max}$  values for the three networks were 1 for g1236, 2 for g1421, and 3 for g1682. According to our HATA arc type calculation results, the g1236 network (Fig. 16a) had two major node groups (upper left and lower right) connected by a global bridge-type arc  $e_{ck, cp}$ . Other arcs in the two groups were identified as global bridges. Note that arcs  $e_{ca,en}$  and  $e_{en,ck}$  were identified as global bridges because no alternative paths existed between node-ca and node-en or from node-en to node-ck. This "no alternative path" (NAP) situation also occurred at arcs  $e_{fg,fc}$  and  $e_{ci,db}$ . The situation was different for arcs  $e_{db,ci}$  and  $e_{ch,cp}$ : even though alternative paths existed at the second layer, they were categorized as global arcs because the network  $k_{max} = 1$ .

The g1421 network (Fig. 16b) consisted of three groups of nodes: two five-node groups at the lower right and middle right, and a ten-node group to the left. The three groups were linked via several arcs, including an af-ic-ch path connecting the first two groups and a db-ec-am-fc-fg-ah path connecting the latter two. The arcs on both paths were all global bridges. The NAP situation occurred on all other global bridge arcs in the network. In comparison, the g1682 network shown in Fig. 16c consisted of five groups: a four-node group at the top left, a three-node group at the middle left, a five-node group at the bottom left, a five-node group at the bottom right, and a large fourteen-node group at the top right. The five groups were connected by several paths, with all arcs identified as global bridges. Similar to the g1421 network, the other global bridges were the result of a NAP situation.



**FIGURE 16.** HATA calculation results for the three networks in experiment 4.  $|V|$  and  $|E|$  indicate numbers of nodes and arcs.

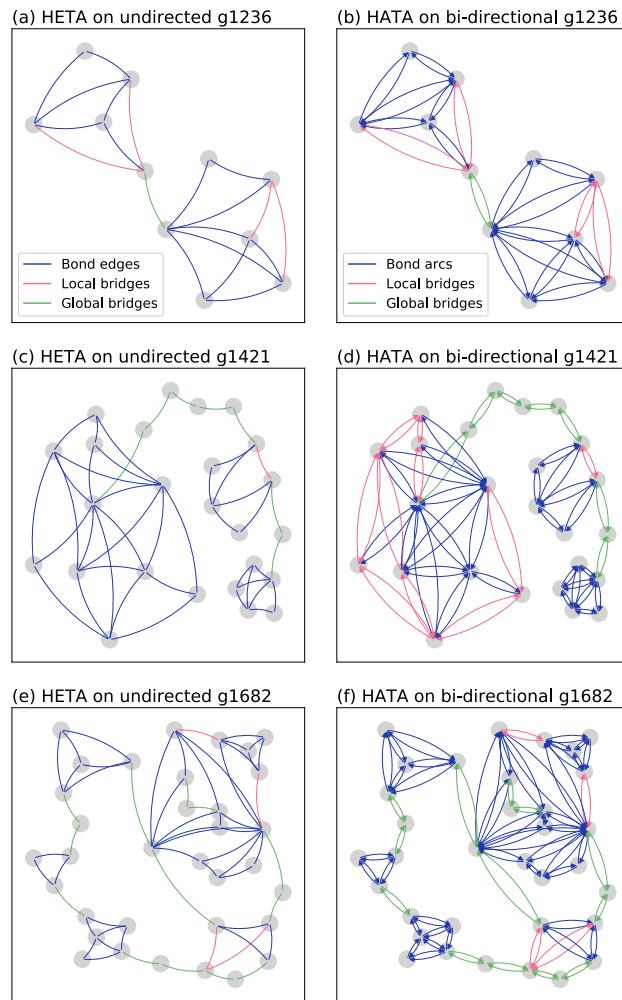


**FIGURE 17.** Results from a comparison of first-layer cm-node ratio ( $R_E^1$ ) and other network measurements in experiment 4. Each line represents a single network arc. Betweenness variables were reversed for comparisons with  $R_E^1$  and similarity indexes.  $|E|$  indicates the number of arcs in a network.

Fig. 17 presents the results of our comparison of first layer cm-node ratios ( $R_E^1$ ) for all arcs with other node-pair similarity and edge-based network measures. Each variable was normalized to a range between 0 and 1. Edge betweenness values were reversed because the concept moves in

an opposite direction compared to other cm-node ratio and similarity measures—that is, the higher the edge-betweenness value of an arc, the more likely it is a bridge type. A lower cm-node ratio indicates a weaker arc that may be a bridge. Each line in Fig. 17 represents an arc in the network, with various colors used to indicate arc type category as determined by HATA. According to the parallel plot in Fig. 17, the  $R_E^1$  values were low for the global bridges and the four similarity indexes. Some of the bond arcs in Figs. 17b and 17c were found to have low  $R_E^1$  and similarity index values; these arcs were identified as bond arcs at higher layers ( $k > 1$ ). All of the other bond arcs had medium-to-high  $R_E^1$  and similarity index values that fluctuated between the indexes (x-axis). Only a few arcs were identified as local bridges with  $R_E^1$  values that were uniformly lower than those for the bond arcs. The parallel plots reflect a strength of HATA: the ability to differentiate between local bridges and bond arcs and between first-layer weak arcs and higher-layer strong arcs.

To clarify the effects of using HATA with non-directional networks, we performed a test involving the modification of networks to create non-directional and bi-directional networks prior to HETA and HATA execution. Test results for the two sets of networks are shown in Fig. 18. According to the bi-directional network HATA results, two arcs connecting the same pair of nodes were identified as the same arc type, with both having identical cm-node ratios ( $R_E^k$ ) due to sharing the same endpoints. A HETA-HATA comparison indicates that the patterns of the non-directional g1236 and g1682 networks respectively shown in Figs. 18a and 18e were similar to those for their bidirectional opponents (Fig. 18b and Fig. 18f, also respectively). HETA and HATA also identified the same edge/arc types for links connecting the same node pairs. Results for the g1421 network revealed some differences between the HETA (Fig. 18c) and HATA test results (Fig. 18d), with global bridges remaining the same for both, but with some of the bond links in the HETA results being identified by HATA as local bridges. Results



**FIGURE 18.** First column, HETA results for non-directional networks. Second column, HATA results for bi-directional networks.

from an external check indicate that the  $R_E^k$  values for the edges in the non-directional network were all equal to that for the corresponding arcs in the bidirectional network. This was expected because the actual and potential numbers of cm-nodes for the bi-directional network (the numerator and denominator in (6)) were double the corresponding numbers for the non-directional network, resulting in the same cm-nodes ratio.

Apparently the main reason for the HETA-HATA difference is the external threshold, which was calculated from the randomization procedure (part 2). When converting a network from non-directional to bidirectional, the doubling of the number of links indicates that for each pair of previously connected nodes there are two connecting links, resulting in a symmetrical network with adjacency matrixes that are identical to those for the non-directional network. The randomization procedure is executed with an assurance that the in/out degree distribution is maintained without retaining a symmetrical structure; this structure is not mandatory for any directed network. As a result, a randomized directed

**TABLE 2.** First-layer external ( $T_{ext}^1$ ) and internal ( $T_{int}^1$ ) HETA and HATA thresholds for the three networks in experiment 4. HATA external thresholds were all higher than HETA external thresholds, but HATA internal thresholds were all lower than HETA external thresholds. Accordingly, when the  $R_E^1$  value was between the HETA and HATA external thresholds, HETA identified them as bond arcs but HATA identified them as local bridges.

Network	HETA $T_{ext}$	HATA $T_{ext}$	HETA $T_{int}$	HATA $T_{int}$
g1236	0.6674	0.7707	0.2667	0.2667
g1421	0.4588	0.5576	0.0	0.0434
g1682	0.3813	0.4629	0.0	0.0

network that might be asymmetrical has greater variability and greater potential for different connection structures. The standard deviation in part 2 is therefore higher than that for the corresponding non-directional network, as is the external threshold (Table 2). The  $R_E^k$  values for all arcs identified as local bridges fell between the two thresholds—that is, higher than the HETA  $T_{ext}^1$ —therefore they were identified as bond links by HETA. Since they were lower than the HATA  $T_{ext}^1$ , they were not identified as bond arcs by HATA, but were still higher than the internal threshold, leading to a local bridge result (Table 2). In summary, this experiment showed that (a) results from running HETA with converted non-directional networks (Fig. 18, left) differed from those for HATA with original directed networks (Fig. 16); and (b) results from running HATA with converted bi-directional networks (Fig. 18, right) differed from those for HATA with non-directional networks (Fig. 18, left). We therefore suggest using HATA with directed networks and HETA with non-directional networks.

## V. DISCUSSION AND CONCLUSION

Despite their importance and number, very little is known about overlapping and hierarchical network community structures or their topological properties. Three key research goals in network science during the past two decades have been to (a) understand network topological properties in order to apply them to structures consisting of large numbers of nodes, (b) understand complex network development and construction, and (c) learn how to predict their evolution over time.

Directed complex networks constitute the majority of real world networks. To understand their connection characteristics and the functions of directed arcs at different complex network layers, we modified an existing algorithm, HETA, which was originally created to study non-directional networks. HETA is based on the common neighbor concept and HATA on a similar common middle node concept. According to results from our four experiments, HATA is capable of accurately identifying the strong/weak properties of arcs in a directed network, while also specifying weak link type (silk arcs,  $k$ th-layer local bridges, global bridges) and the network layers to which they belong. Although it is fast in calculating the common middle node of any two nodes connected by a directed arc, HATA suffers from bottlenecks when determining directed arc characteristics. The reason is that a

large number of randomized directed networks corresponding to the target network must be generated in order to calculate external thresholds and identify bond arcs at each layer.

Even slight increases in temporal and spatial algorithm complexities are considered drawbacks when working with large real world directed networks. There are several solutions, the first being the use of a compiled programming language such as C or C++ instead of Python; in addition to improving overall HATA performance, this can save computing memory via a user-designed data structure. However, this solution requires the redesign of complex network-specific data structures and related access operators, thereby increasing the chances of program bugs and logical errors, and reducing the flexibility required to further expand the algorithm. Accordingly, we do not recommend this solution.

A second solution, one focused on local improvements, starts with two modifications: constructing a randomized directed network corresponding to the target network, and calculating common middle nodes and arcs at each layer in both the target network and corresponding randomized directed network. As independent operations, these construction and calculation steps do not interact during execution. Methods such as multithreading or GPU-like parallel computing can be used to simultaneously generate the large number of randomized directed networks required by the algorithm. During network construction, HATA calculates all common middle nodes and their external thresholds at all layers and in all networks, including the target and corresponding randomized directed networks. Regardless of whether multithreading or GPU-like parallel computing is used, each local improvement must increase implementation level complexity without altering core algorithm concepts. Readers are welcome to assist with the task of creating multiple HATA implementation versions to enhance performance (see [www.github.com/wcchin/HATA](http://www.github.com/wcchin/HATA) for the original version).

A third solution involves working with bottlenecks "from the ground up." One possible approach consists of a series of steps aimed at determining distribution rules and variations for the external thresholds of all layers in complex networks that have different topological properties. The goal is to find the best external threshold calculation method for each target network layer without having to construct large numbers of randomized directed networks. This solution offers significant reductions in randomized directed network construction time and memory storage requirements. When combined with the second solution, this solution can simultaneously identify all strong/weak properties of directed arcs in the same layers of the target network, thereby accelerating HATA performance and identifying new algorithm applications—for example, quickly adjusting the hierarchical strong/weak properties of individual arcs for evolving networks.

We examined several possible HATA applications, some of which were identified for HETA (HATA's predecessor [35]). The first involves a fingerprint database for directed networks, using proportions of the four arc types as determined by HATA plus other topological properties as a network

fingerprint. The fingerprint can be used to compare degrees of similarity in topological structures across complex networks, organize complex network groups, and create hierarchy trees from fingerprint databases. This process could eventually be used to identify topological similarities between known complex networks and target networks, or to quickly detect the characteristics of unknown target networks.

The second potential application is hierarchical network partitioning, sometimes called hierarchical community detection. In directed networks, close small groups are connected internally and externally by strong bond arcs and low-level local bridges, while large groups consisting of several smaller groups are connected to each other by higher-level local bridges or global bridges so as to form single network elements. Once the strengths and weaknesses of all directed arcs at all layers have been determined and categorized, HATA can be used to partition target networks from top to bottom to determine directed arcs via a recursive procedure that is the opposite of the one described above. Each network element can be disassembled into several larger groups that are in turn disassembled into smaller and tighter groups, with steps repeated until further disassembly is no longer feasible. It may be possible to achieve two goals concurrently: clarifying the roles and responsibilities of individual arcs at each layer in a complex network, and partitioning networks from top to bottom to establish a hierarchical group structure that can be used for other purposes.

A third potential application is aimed at identifying network nodes with key characteristics—for example, identifying super-spreader nodes in directed networks. HATA can assist in super-spreader agent detection and determination or the development of new computational epidemiology approaches. Super infectious agents are capable of transmitting ideas or diseases among network nodes within the same group or to other groups in a manner that infects the largest possible number of nodes. In addition to having more neighbor than other-network nodes, super infectious agents must have strong bond arc, local bridge, and global bridge connections with neighbor nodes in order to transmit as many ideas or diseases as possible to nodes within the same close group, to nodes in small groups belonging to the same large group as the super infector, and to nodes in other large groups within the network.

There are at least three possible HATA algorithm extensions, the first involving the use of arc weights when analyzing arc strengths. In addition to directed arcs and asymmetrical structures, many real world complex networks have different weighted arcs with various connection strengths. Modifying HATA to consider arc weights would require redefining the strength concept in weighted networks, with different weight types (flow, capacity, cost) exerting different effects on strength measurements. A second possible HATA extension involves hierarchical network structures. Since HATA is capable of classifying arc types, researchers may be interested in using that feature to conduct expanded analyses of node positions and roles, and to study recurring patterns in

complex networks. A third possible extension would involve link prediction. Our HATA analysis is based on the hierarchical common middle node concept with pairs of connected nodes. This concept can be applied to non-connected node pairs to determine their relationships in support of the link prediction process.

## ACKNOWLEDGMENT

The authors wish to thank Dr. Yu-Hsiang Fu and Prof. Yu-Shiuan Tsai for participating in detailed discussions during the early stages of this project.

## REFERENCES

- [1] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. London, U.K.: Cambridge Univ. Press, 2010.
- [2] M. E. J. Newman, “The structure and function of complex networks,” *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.
- [3] M. Nielsen, *Reinventing Discovery: The New Era of Networked Science*. Princeton, NJ, USA: Princeton Univ. Press, 2011.
- [4] M. T. Thai and P. M. Pardalos, Eds., *Handbook of Optimization in Complex Networks: Theory and Applications*. New York, NY, USA: Springer-Verlag, 2012.
- [5] L. Trajkovic, “Analysis of Internet topologies,” *IEEE Circuits Syst. Mag.*, vol. 10, no. 3, pp. 48–54, 3rd Quart., 2010.
- [6] X. Yu, C. Cecati, T. Dillon, and M. G. Simoes, “The new frontier of smart grids,” *IEEE Ind. Electron. Mag.*, vol. 5, no. 3, pp. 49–63, Sep. 2011.
- [7] A. Barabási and M. Pósfai, *Network Science*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
- [8] J. A. Bondy and U. S. R. Murty, *Graph Theory*. Berlin, Germany: Springer-Verlag, 2008.
- [9] N. Deo, *Graph Theory With Applications to Engineering and computer Science*, North Chelmsford, MA, USA: Dover, 2017.
- [10] R. Albert and A. Barabási, “Statistical mechanics of complex networks,” *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.
- [11] G. Chen, X. F. Wang, and X. Li, *Fundamentals of Complex Networks: Models, Structures and Dynamics*. Singapore: Wiley, 2015.
- [12] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, “On Facebook, most ties are weak,” *Commun. ACM*, vol. 57, no. 11, pp. 78–84, Oct. 2014.
- [13] A. Gurzawska, “Towards responsible and sustainable supply chains—innovation, multi-stakeholder approach and governance,” *Philosophy Manage.*, pp. 1–29, Jun. 2019, doi: 10.1007/s40926-019-00114-z.
- [14] C.-Y. Huang, “An agent-based epidemic simulation of social behaviors affecting HIV transmission among Taiwanese homosexuals,” *Comput. Math. Methods Med.*, vol. 2015, Mar. 2015, Art. no. 867264.
- [15] C. Y. Huang, C. T. Sun, J. L. Hsieh, and H. Lin, “Simulating SARS: Small-world epidemiological modeling and public health policy assessments,” *J. Artif. Soc. Social Simul.*, vol. 7, no. 4, pp. 1–2, 2004.
- [16] C.-Y. Huang, C.-T. Sun, J.-L. Hsieh, Y.-M.-A. Chen, and H. Lin, “A novel small-world model: Using social mirror identities for epidemic simulations,” *Simulation*, vol. 81, no. 10, pp. 671–699, Oct. 2005.
- [17] C. Y. Huang, C. T. Sun, and H. C. Lin, “Influence of local information on social simulations in small-world network models,” *J. Artif. Soc. Social Simul.*, vol. 8, no. 4, pp. 1–27, 2005.
- [18] C.-Y. Huang, C.-L. Lee, T.-H. Wen, and C.-T. Sun, “A computer virus spreading model based on resource limitations and interaction costs,” *J. Syst. Softw.*, vol. 86, no. 3, pp. 801–808, Mar. 2013.
- [19] C.-Y. Huang, T.-H. Wen, and Y.-S. Tsai, “FLUed: A novel four-layer model for simulating epidemic dynamics and assessing intervention policies,” *J. Appl. Math.*, vol. 2013, Aug. 2013, Art. no. 325816.
- [20] C.-Y. Huang and Y.-S. Tsai, “Effects of friend-making resources/costs and remembering on acquaintance networks,” *Phys. A, Stat. Mech. Appl.*, vol. 389, no. 3, pp. 604–622, Feb. 2010.
- [21] M. E. J. Newman, *Networks: An Introduction*. London, U.K.: Oxford Univ. Press, 2010.
- [22] C. L. Lee, C. Y. Huang, T. C. Hsiao, C. Y. Wu, Y. C. Chen, and I. C. Wang, “Impact of vehicular networks on emergency medical services in urban areas,” *Int. J. Environ. Res. Public Health*, vol. 11, no. 11, pp. 11348–11370, 2014.
- [23] R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler, “A model of large-scale proteome evolution,” *Adv. Complex Syst.*, vol. 5, no. 1, pp. 43–54, Mar. 2002.
- [24] A. Wagner, “The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes,” *Mol. Biol. Evol.*, vol. 18, no. 7, pp. 1283–1292, Jul. 2001.
- [25] M. S. Granovetter, *Getting a Job: A Study of Contacts and Careers*. Chicago, IL, USA: Univ. of Chicago Press, 1995.
- [26] E. Gilbert, “Predicting tie strength in a new medium,” in *Proc. ACM Conf. Comput. Supported Cooperat. Work (CSCW)*, Seattle, WA, USA, 2012, pp. 1047–1056.
- [27] E. Gilbert and K. Karahalios, “Predicting tie strength with social media,” in *Proc. 27th Int. Conf. Hum. Factors Comput. Syst. (CHI)*, Boston, MA, USA, 2009, pp. 211–220.
- [28] M. E. Helander and S. McAllister, “The gravity of an edge,” *Appl. Netw. Sci.*, vol. 3, May 2018, Art. no. 7, doi: 10.1007/s41109-018-0063-6.
- [29] I. Kahanda and J. Neville, “Using transactional information to predict Link strength in online social networks,” in *Proc. 3rd Int. Conf. Weblogs Social Media*, San Jose, CA, USA, vol. 2009, pp. 74–81.
- [30] G. Kossinets and D. J. Watts, “Empirical analysis of an evolving social network,” *Science*, vol. 311, no. 5757, pp. 88–90, Jan. 2006.
- [31] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, “Structure and tie strengths in mobile communication networks,” *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 18, pp. 7332–7336, May 2007.
- [32] S. Sintos and P. Tsaparas, “Using strong triadic closure to characterize ties in social networks,” in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2014, pp. 1466–1475.
- [33] N. Takahashi and N. Inamizu, “Logical weakness of ‘the strength of weak ties,’” *Ann. Bus. Administ. Sci.*, vol. 13, no. 2, pp. 67–76, 2014.
- [34] R. Xiang, J. Neville, and M. Rogati, “Modeling relationship strength in online social networks,” in *Proc. 19th Int. Conf. World Wide Web (WWW)*, Raleigh, NC, USA, 2010, pp. 981–990.
- [35] C.-Y. Huang, W.-C.-B. Chin, Y.-H. Fu, and Y.-S. Tsai, “Beyond bond links in complex networks: Local bridges, global bridges and silk links,” *Phys. A, Stat. Mech. Appl.*, vol. 536, Dec. 2019, Art. no. 121027, doi: 10.1016/j.physa.2019.04.263.
- [36] L. Du, Z. Lu, Y. Wang, G. Song, Y. Wang, and W. Chen, “Galaxy network embedding: A hierarchical community structure preserving approach,” in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Stockholm, Sweden, Jul. 2018, pp. 2079–2085.
- [37] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, “Detect overlapping and hierarchical community structure in networks,” *Phys. A, Stat. Mech. Appl.*, vol. 388, no. 8, pp. 1706–1712, Apr. 2009.
- [38] X. Wang, P. Chi, J. Wang, J. Pei, W. Zhu, and S. Yang, “Community preserving network embedding,” in *Proc. 21st AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, 2017, pp. 203–209.
- [39] J. E. Cohen, F. Briand, and C. M. Newman, *Community Food Webs: Data and Theory*. Berlin, Germany: Springer-Verlag, 1990.
- [40] B. Dutta and M. O. Jackson, “The stability and efficiency of directed communication networks,” *Rev. Econ. Des.*, vol. 5, no. 3, pp. 251–272, 2000.
- [41] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, “The large-scale organization of metabolic networks,” *Nature*, vol. 407, no. 6804, pp. 651–654, Oct. 2000.
- [42] H. Ma and A.-P. Zeng, “Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms,” *Bioinformatics*, vol. 19, no. 2, pp. 270–277, Jan. 2003.
- [43] M. Pascual and J. A. Dunne, Eds., *Ecological Networks: Linking Structure to Dynamics in Food Webs*. Oxford, U.K.: Oxford Univ. Press, 2006.
- [44] R. Pastor-Satorras and A. Vespignani, *Evolution Structure of the Internet: A Statistical Physics Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [45] D. B. Stouffer, J. Camacho, W. Jiang, and L. A. Nunes Amaral, “Evidence for the existence of a robust pattern of prey selection in food webs,” *Proc. Roy. Soc. B, Biol. Sci.*, vol. 274, no. 1621, pp. 1931–1940, Aug. 2007.
- [46] R. Tanaka, “Scale-rich metabolic networks,” *Phys. Rev. Lett.*, vol. 94, no. 16, Apr. 2005, Art. no. 168101.
- [47] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1994.
- [48] U. Brandes, “A faster algorithm for betweenness centrality,” *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, 2001.

- [49] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.
- [50] C. Mavroforakis, R. Garcia-Lebron, I. Koutis, and E. Terzi, "Spanning edge centrality: Large-scale computation and applications," in *Proc. 24th Int. Conf. World Wide Web (WWW)*, Florence, Italy, 2015, pp. 732–742.
- [51] W. De Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [52] M. E. J. Newman, "A measure of betweenness centrality based on random walks," *Social Netw.*, vol. 27, no. 1, pp. 39–54, Jan. 2005.
- [53] N. Meghanathan, "A greedy algorithm for neighborhood overlap-based community detection," *Algorithms*, vol. 9, no. 1, p. 8, 2016, doi: [10.3390/a9010008](https://doi.org/10.3390/a9010008).
- [54] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 2, 2001, Art. no. 026118, doi: [10.1103/PhysRevE.64.026118](https://doi.org/10.1103/PhysRevE.64.026118).
- [55] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998, doi: [10.1038/30918](https://doi.org/10.1038/30918).
- [56] M. E. J. Newman and D. J. Watts, "Renormalization group analysis of the small-world network model," *Phys. Lett. A*, vol. 263, nos. 4–6, pp. 341–346, Dec. 1999, doi: [10.1016/S0375-9601\(99\)00757-4](https://doi.org/10.1016/S0375-9601(99)00757-4).
- [57] A. Hajibagheri, K. Lakkaraju, G. Sukthankar, R. T. Wigand, and N. Agrawal, "Conflict and communication in massively-multiplayer online games," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling, Predict.*, vol. 9021, N. Agarwal, K. Xu, and N. Osgood, Eds. Cham, Switzerland: Springer, 2015, pp. 65–74.
- [58] A. Hajibagheri, G. Sukthankar, K. Lakkaraju, H. Alvari, R. T. Wigand, and N. Agarwal, "Using massively multiplayer online game data to analyze the dynamics of social interactions," in *Social Interactions in Virtual Worlds: An Interdisciplinary Perspective*, K. Lakkaraju, G. Sukthankar, and R. T. Wigand, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2017, pp. 375–418.
- [59] R. W. Hedley, "Composition and sequential organization of song repertoires in Cassin's vireo (Vireo cassinii)," *J. Ornithol.*, vol. 157, no. 1, pp. 13–22, Jan. 2016, doi: [10.1007/s10336-015-1238-x](https://doi.org/10.1007/s10336-015-1238-x).
- [60] E. A. Leicht, P. Holme, and M. E. J. Newman, "Vertex similarity in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 73, no. 2, Feb. 2006, Art. no. 026120, doi: [10.1103/PhysRevE.73.026120](https://doi.org/10.1103/PhysRevE.73.026120).
- [61] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software," *PLoS ONE*, vol. 9, no. 6, 2014, Art. no. e98679, doi: [10.1371/journal.pone.0098679](https://doi.org/10.1371/journal.pone.0098679).
- [62] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 U.S. election: Divided they blog," in *Proc. 3rd Int. Workshop Link Discovery (LinkKDD)*, 2005, pp. 36–43.
- [63] M. Ripeanu, I. Foster, and A. Lamnitchi, "Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design," *IEEE Internet Comput. J.*, vol. 6, no. 1, pp. 50–57, Jan./Feb. 2002.
- [64] P. Hage and F. Harary, *Structural Models in Anthropology*. Cambridge, U.K.: Cambridge Univ. Press, 1983.
- [65] J. S. Coleman, *Introduction to Mathematical Sociology*. Glencoe, IL, USA: London Free Press, 1964.
- [66] J. Coleman, E. Katz, and H. Menzel, "The diffusion of an innovation among physicians," *Sociometry*, vol. 20, no. 4, pp. 253–270, 1957.
- [67] J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 556–584.
- [68] N. D. Martinez, "Artifacts or attributes? Effects of resolution on the little rock lake food Web," *Ecol. Monographs*, vol. 61, no. 4, pp. 367–392, Feb. 1991, doi: [10.2307/2937047](https://doi.org/10.2307/2937047).
- [69] E. Garfield, I. H. Sher, and R. J. Torpie, *The Use of Citation Data in Writing the History of Science*. Philadelphia, PA, USA: The Institute for Scientific Information, 1964, doi: [10.21236/ad0466578](https://doi.org/10.21236/ad0466578).



**CHUNG-YUAN HUANG** was born in Taiwan, in 1970. He received the M.S. degree in computer information and science and the Ph.D. degree in computer science from National Chiao Tung University, Taiwan, in 2000 and 2005, respectively. He is currently a Professor with the Department of Computer Science and Information Engineering, Chang Gung University, where he is also a member of the Research Center for Emerging Viral Infections. His research interests include complex network analysis, agent-based modeling and simulations for social science research, and computational epidemiology.



**WEI CHIEN BENNY CHIN** was born in Malaysia, in 1988. He received the bachelor's, master's, and Ph.D. degrees in geography from National Taiwan University, in 2011, 2014, and 2018, respectively. He is currently a Postdoctoral Researcher with the Applied Complexity Group, Singapore University of Technology and Design. His previous studies have focused on the development of novel space-time analysis algorithms and complex network measures. He is also participating in projects associated with complex human movement and the horizontal and vertical dimensions of spatial structures.

• • •