

RESEARCH

Open Access



# Local bow-tie structure of the web

Yuji Fujita<sup>1\*</sup> , Yuichi Kichikawa<sup>2</sup>, Yoshi Fujiwara<sup>3</sup>, Wataru Souma<sup>4</sup> and Hiroshi Iyetomi<sup>2</sup>

\*Correspondence:

yuji@turnstone.jp

<sup>1</sup>Turnstone Research Inst. Inc.,

Kamakura 248-0004, Japan

Full list of author information is  
available at the end of the article

## Abstract

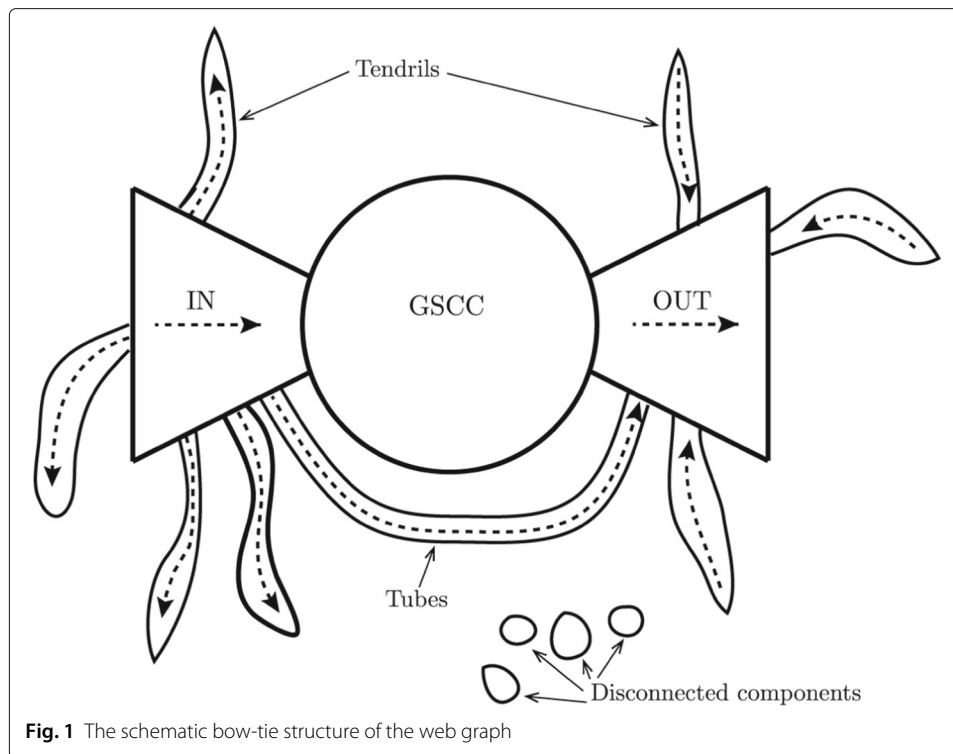
Social networks often has the graph structure of giant strongly connected component (GSCC) and its upstream and downstream portions (IN and OUT), known as a bow-tie structure since a pioneering study on the World Wide Web (WWW). GSCC, on the other hand, has community structure, namely tightly knitted clusters, reflecting how the networks developed in time. By using our visualization of enhanced multidimensional scaling (MDS) and force-directed graph drawing for large and directed graphs, we discovered that a bow-tie in the WWW usually has clusters, which are locally-located mini bow-ties that are loosely connected to each other, resulting in a formation of GSCC as a whole. To quantify the mutual connectivity among such local bow-tie, we define a quantity to measure how a local bow-tie connects to others in comparison with random graphs. We found that there are striking difference between the WWW and other social and artificial networks including a million firms' nationwide supply chain network in Japan and thousands of symbols' dependency in the programming language of Emacs LISP, in which a global bow-tie exists. Presumably the difference comes from a self-similar structure and development of the WWW speculated by others.

**Keywords:** Bow-tie, Locality, Visualization

## Introduction

Two decades studies of complex networks have revealed some aspects of complex systems. Up to now, many network indices (centralities) have been proposed to understand complex networks quantitatively. Many new methods for communities extraction have been also proposed and clarified the inner structure of complex networks. However, there is only one notion characterizing an overall structure of complex networks. That notion is a "bow-tie" structure discovered by Broder et al. (2000) when they investigated two AltaVista crawls in 1999 with over 200 million pages and 1.5 billion links (Broder et al. 2000). The bow-tie structure has been actively examined in the information science, especially in the study of the topology of world wide web. However, the bow-tie structure takes a vital role in a metabolic network in biology (for example, see Refs. (Csete and Doyle 2004; Kitano 2004; Zhao et al. 2006; Kawakami et al. 2016)).

A schematic bow-tie structure of the web is depicted in Fig. 1. This figure shows that the web consists of a giant connected component and many disconnected components. This giant connected component is called a giant weakly connected component (GWCC). Broder and colleagues found that the GWCC consists of the equal size of the giant strongly connected component (GSCC or SCC) (27.7%), the IN set (21.3%), the OUT set (21.2%), tendrils (21.5%), and tubes. The GSCC is made up of a single strongly connected component. The IN set contains nodes that can reach the GSCC but cannot be reached by



**Fig. 1** The schematic bow-tie structure of the web graph

nodes in GSCC. The OUT set contains nodes that can be reached by the nodes in GSCC but cannot reach nodes in the GSCC. Tendrils hang off the IN set and the OUT set, and contain nodes that are reachable from portions of the IN set or that can reach portions of the OUT set, without passing through the GSCC. Tubes contain nodes that passage from a portion of the IN set to a portion of the OUT set without touching the GSCC.

Some investigations have checked the bow-tie structure of the web as summarized in Table 1. Donato et al. (2005, 2008) investigated three national (Italy, Indochina, and UK) webs (collected by the “Language Observatory Project” and the “Institute Informatica e Telematica”) and suggested that these tree national webs almost consist of the GSCC and the OUT set, and also investigated the global web (collected by the WebBase project at Stanford in 2001) and suggested that the global web resembles the bow-tie structure (Donato et al. 2005; 2008). They also investigated the detailed structure of the IN set and the OUT set and found that IN and OUT sets are fragmented into a large number of small and shallow “petals” (weakly connected components; WCCs) hanging from the GSCC. They called such a structure of the web graph as “daisy” structure. They also found that different components (i.e., the GSCC, the IN set, and the OUT set) have a very distinct structure. This means that there is no self-similarity in the individual components of the web graph.

Zhu et al. (2008) investigated the Chinese web in 2006 from the viewpoint of a hierarchy of three levels, i.e., page level (830 million pages), host level (17 million hosts), and domain level (0.8 million domains) (Zhu et al. 2008). They found that the page level web has “tea pod” structure (with a large size of the GSCC, a medium size of IN, and small size of OUT). They also found that the Chinese web becomes increasingly close to the daisy structure when the aggregation level is increased from the page level to the host levels and

**Table 1** Components of some web graphs

		# node	# link	SCC(%)	IN(%)	OUT(%)	Tendrils(%)	Disc.(%)
Broder	AltaVista	203.5M	1.45G	27.7	21.3	21.2	21.5	8.2
Donato	Italy	41.3M	1.15G	72.3	0.03	27.6	0.01	0
	Indochina	7.4M	194.1M	51.4	0.66	45.9	0.66	1.4
	UK	18.5M	298.1M	65.3	1.7	31.8	0.8	0.4
	WebBase	135.7M	1.18G	32.9	10.6	39.3	12.6	4.6
Zhu	China (page)	836.7M	43.28G	44.1	25.5	14.6		15.8
	China (host)	16.9M	43.28G	50.7	1.4	47.4		0.5
	China (domain)	16.9M	43.28G	63.3	0.7	34.9		1.1
Meusel	CC12 (page)	3.563G	128.786G	51	32	6	5	6
	CC12 (host)	101M	2043M	47	17	20	2	13
	CC12 (PLD)	43M	623M	52	8	31	1	8

the domain level. This fact means that the absence of self-similarity between page level and host/domain levels.

Meusel et al. (2014, 2015) investigated the publicly accessible crawl of the web gathered by the Common Crawl Foundation in 2012 (CC12) (Meusel et al. 2014; 2015). The CC12 is available to the public outside companies such as Google, Yahoo!, Yandex, and Microsoft, and contains over 3.5 billion web pages and 128.7 billion links. They analyzed the CC12 on three different levels of aggregation: page, host, and pay-level domain (PLD) (one "dot level" above public suffixes). They also obtained almost the same results founded by Zhu et al. (2008). Thus there is no self-similarity between page level and host/PDL levels.

When we study complex networks, it is rarely possible to get a complete set of nodes and edges of the networks that we are studying. Thus, we tend to abandon clarifying the gross structure of the network. However, if the network has self-similar property, it is anticipated that we can elucidate the overall structure by studying subgraphs. Although previous studies explained above suggest the absence of self-similarity, Dill et al. (2002) found that the web exhibits self-similarity, i.e., each *thematically unified region* (for instances, pages on a site or pages about a topic) plays the same characteristics as the web at large (Dill et al. 2002). This finding means that the thematically self-similarity of the web. Thus, we can elucidate the overall structure from thematically subgraphs, i.e., communities.

The purpose of this paper, therefore, is to extract communities of networks by using a modern algorithm of community extraction, such as the map equation, and show the self-similar property, i.e., the bow-tie structure of each community, by using visualization techniques and introducing the new measure quantifying the local bow-tie structure. In the next section, we explain the data set used in this paper. We summarize methods of network analysis used in this article in "Methods" section. "Analysis and results" section is the main part of this paper, and we show the self-similar property, i.e., the bow-tie structures of each community in the web. The last section is devoted to the conclusion and discussion.

## Data

Here we explain the network data analyzed in this paper. The web data is from the Web graphs datasets of Stanford Large Network Dataset Collection "SNAPNETS" (Leskovec

and Krevl 2014). Basic statistics of the data are listed in Table 2. These networks have contributed to revealing a notable feature of community size in Ref. (Leskovec et al. 2008).

Figure 2 depicts the complementary cumulative degree (number of links assigned to a vertex) distributions of four web data. This figure shows that the degree distributions of four webs follow almost same distributions except for a gap around the degree equals to 300 in the case of Notre-Dame.

As described in “Introduction” section, our purpose is to show the self-similar property, i.e., the bow-tie structures of each community in the web. Thus, to clarify whether this nature is typical in the web or not, we investigate other types of network data listed in Table 3.

Japanese production network consists of more than one million firms which are appropriately half of the total number of firms in Japan and five million links which are supplier-customer relations. Emacs24 (text editor “Emacs” of version 24) LISP means the relationship between functional definitions of symbols written in programming language LISP. Existence of GSCC in the symbol definition relation means there are direct or indirect recursive definitions. The size of GSCC is not particularly large and less than two percent of the whole network. The degree distribution of Japanese production network and Emacs24 lisp is shown in Fig. 3

## Methods

### Visualization

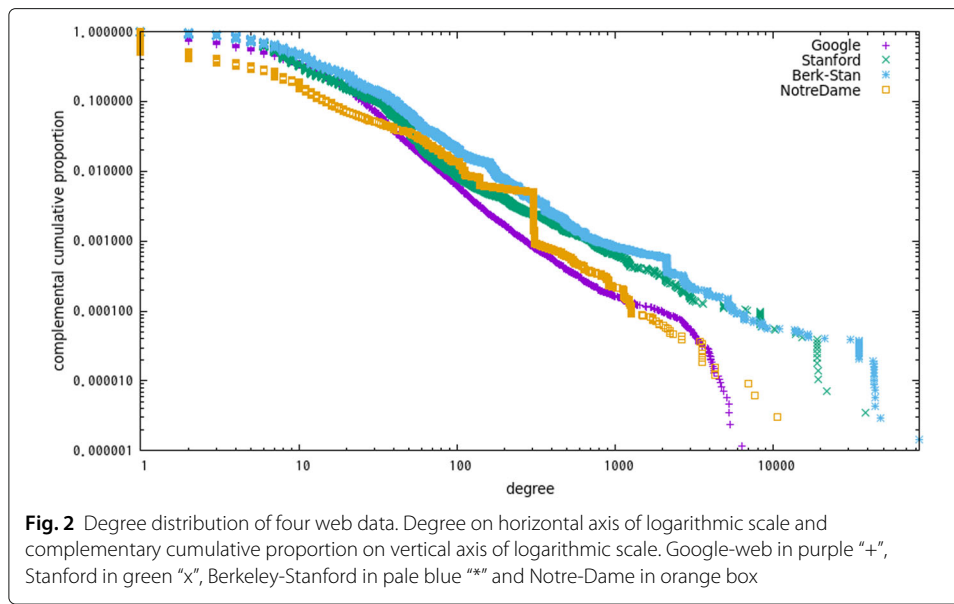
In order to visualize a large-scale structure of directed graph, we shall use our method of graph drawing, which is based on the so-called force-directed graph drawing (Fruchterman and Reingold 1991) with a nice property so as to calculate a layout for a large and directed graph in an aesthetically-pleasing way. Let us briefly explain our algorithm (see (Fujita et al. 2016) for more details).

The algorithm has two steps of calculation. The first step is to determine the initial positions of nodes in a graph by using a multi-dimensional scaling (MDS) algorithm. We assume the initial positions are given in a two-dimensional Euclidean space in this paper. In the second step, we perform a physical simulation, in which nodes have electric charges (with a same sign, say plus) with repulsive forces between pairs of nodes by Coulomb’s law, while edges are regarded as “springs” exerting attractive forces between adjacent nodes by Hooke’s law. In addition, each edge has a magnetism and is aligned with a globally given magnetic field (see (Sugiyama and Misue 1995) for example). Frictional forces are additionally given to nodes so that the physical system will be relaxed into a quasi-stable configuration, which is the visualization result that we shall use.

We call the algorithm DMDS (direction-aware MDS), because it can capture the direction of edges appropriately in the following manner.

**Table 2** Basic statistics of the web graphs data

Name	# nodes	# edges	Description
Google	875,713	5,105,039	Web graph from Google
Stanford	281,903	2,312,496	Web graph from Stanford.edu
Berk-Stan	685,230	7,600,595	Web graph of Berkeley and Stanford
Notre-Dame	325,729	1,497,134	Web graph from Notre Dame



Let  $u, v \in V$  be arbitrary nodes in the set of vertices,  $V$ , of a directed graph  $G$ . Denote by  $d_u(u, v)$  the shortest distance between  $u$  and  $v$  when  $G$  is regarded as undirected graph. Let  $d_1(u, v)$  denote the shortest distance in the directed graph  $G$ . When  $v$  is not reachable from  $u$ ,  $d_1(u, v) = \infty$  by convention. Let us define

$$d_d(u, v) = \min\{d_1(u, v), d_1(v, u)\} \quad (1)$$

in order to define a general distance  $d(u, v)$ :

$$d(u, v) = \begin{cases} d_d(u, v) & \text{if } d_u(u, v) = d_d(u, v) \\ \frac{d_u(u, v)}{2} & \text{otherwise} \end{cases} \quad (2)$$

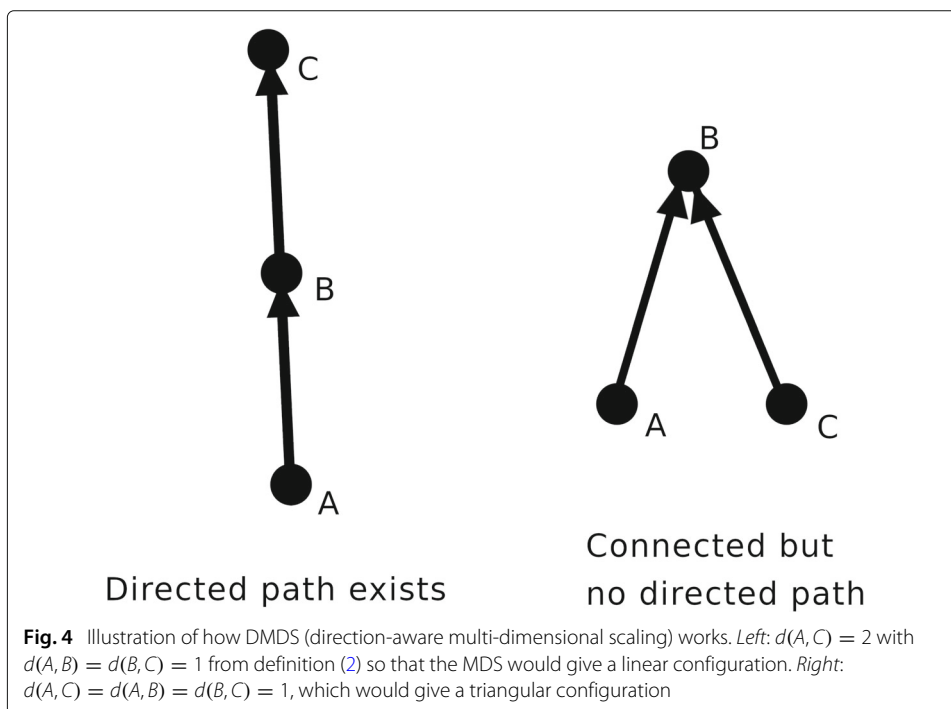
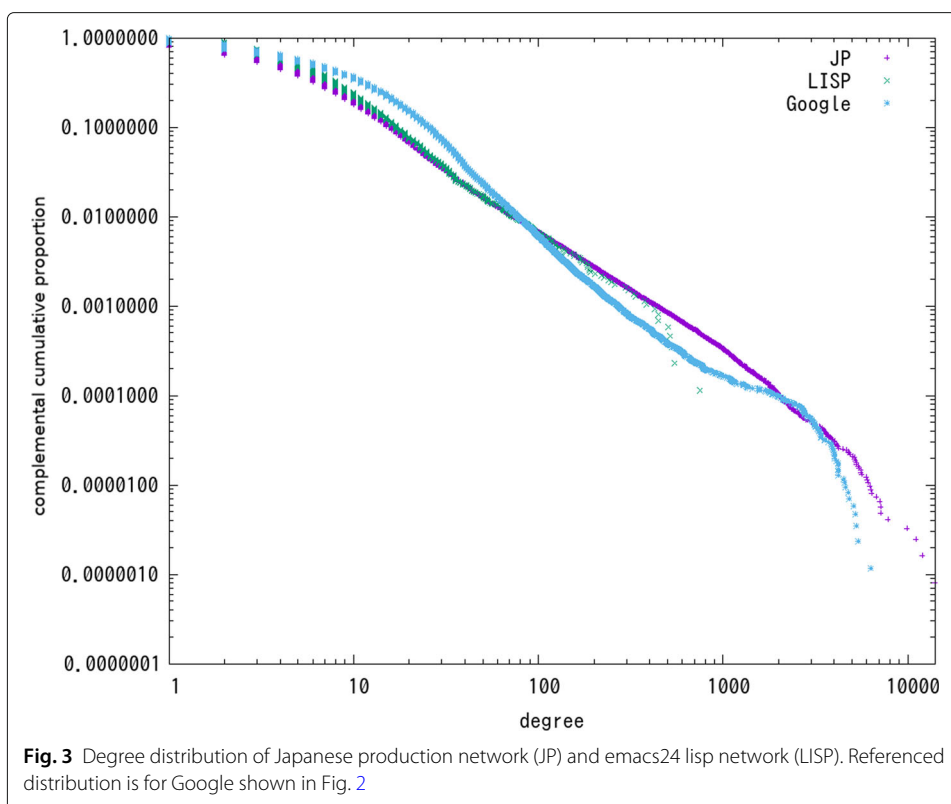
We use the definition (2) to determine similarity (or dissimilarity) between nodes in the standard calculation of MDS.

To illustrate how (2) works, consider two cases of three nodes depicted in Fig. 4. In the left-hand case, in which there exists a path from node  $A$  to node  $C$ , one has  $d(A, C) = 2$  with  $d(A, B) = d(B, C) = 1$  from (2). If the MDS yields a spatial configuration, one expects that the three nodes are aligned along a line. In the right-hand case, there exists no path from  $A$  to  $C$  (and also from  $C$  to  $A$ ) so that  $d(A, C) = d(A, B) = d(B, C) = 1$ . One can expect that the three nodes are placed with equal distance forming a triangle.

Thus we can determine the initial configuration by using the MDS in our first step of graph layout. Practically, the computational cost of MDS, both in space and time (memory

**Table 3** Basic statistics of Japanese production network and Emacs24 lisp network

Name	# nodes	# edges	Description
Japanese production	1,234,687	5,481,427	Production network of Japan
Emacs24 LISP	8,689	40,127	LISP symbol relations



usage and computational time), can be reduced by using a sophisticated method invented by (Brandes and Pich 2006).

Also, because the calculation of the Coulomb interaction can be naively proportional to the square of the number of nodes, one can employ a well-known algorithm developed in astronomy (Barnes and Hut 1986) to reduce the computational time significantly. Additionally, we used a recent technique of Phantom-GRAPe (Ataru Tanikawa 2012) to accelerate the computation.

### Community detection

For our purpose, we need a community detection algorithm that can be applied to a large and directed graph. We shall use the well-known Infomap algorithm, first proposed by (Rosvall and Bergstrom 2008), which optimizes the so-called map equation that is a flow-based method and operates on dynamics on the network. The algorithm works for directed links and can cluster tightly interconnected nodes into modules (two-level clustering) or the optimal number of nested modules (multi-level clustering) in a hierarchical way (Martin Rosvall 2011). See the original papers and references therein. We employ the code given by the original authors who invented the algorithm.

### Bow-tie analysis

To examine the graph structure of giant strongly connected component (GSCC) and its upstream and downstream portions (IN and OUT), we used a well-known algorithm of graph search in the following way.

Let  $Fw(A)$  be a set of vertices reachable from vertex  $A$  in a directed graph  $G$ , and  $Bk(A)$  be a set of vertices reachable from vertex  $A$  by following links in a *backward* manner. Then a strongly connected component (SCC) can be found simply by

$$SCC(A) = Fw(A) \cap Bk(A) \quad (3)$$

Any pair of vertices that belong to the set  $Fw(A) \cap Bk(A)$  are connected by some path, as one can easily prove.

We utilize the well-known algorithm of breadth-first search to calculate (3), and then identify the GSCC by finding the largest SCC found in the search.

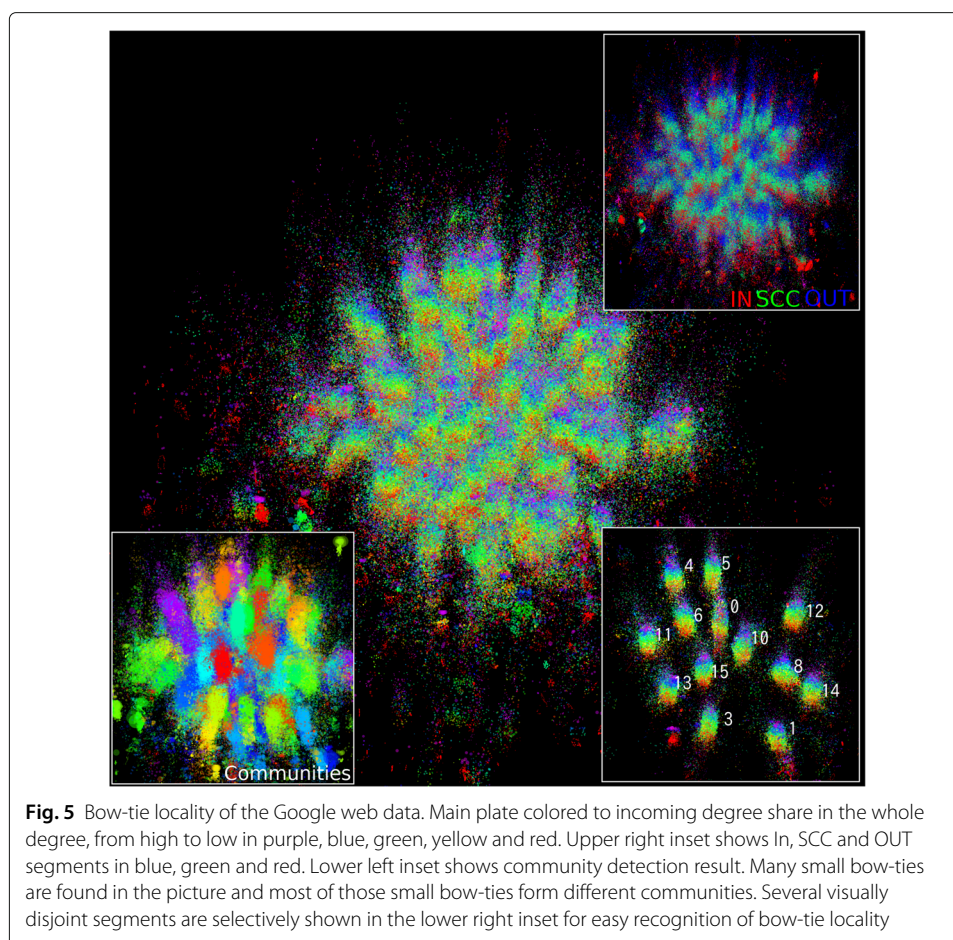
### Analysis and results

In this section we present our analysis and claim a remark on the feature of community structure which is particular to the network of the World Wide Web.

### Visualization and qualitative analysis

Figure 5 is a visualization of the Google Web network data from (Leskovec and Krevl 2014) by using the visualization method described in the “Visualization” section. The link direction is shown as the relative vertical position of the vertices so that the link direction faces upward in the picture. In the main plate and lower-right inset the color is derived from incoming link share, higher in purple, medium in green and lower in red.

If you see the upper right inset, we can obviously see that the network has no bow-tie structure as its global feature. The network is a union of tightly connected communities, within which local bow-tie structure is found. These localized bow-ties are actually communities, which we can see in the lower left inset. To look into the communities we placed



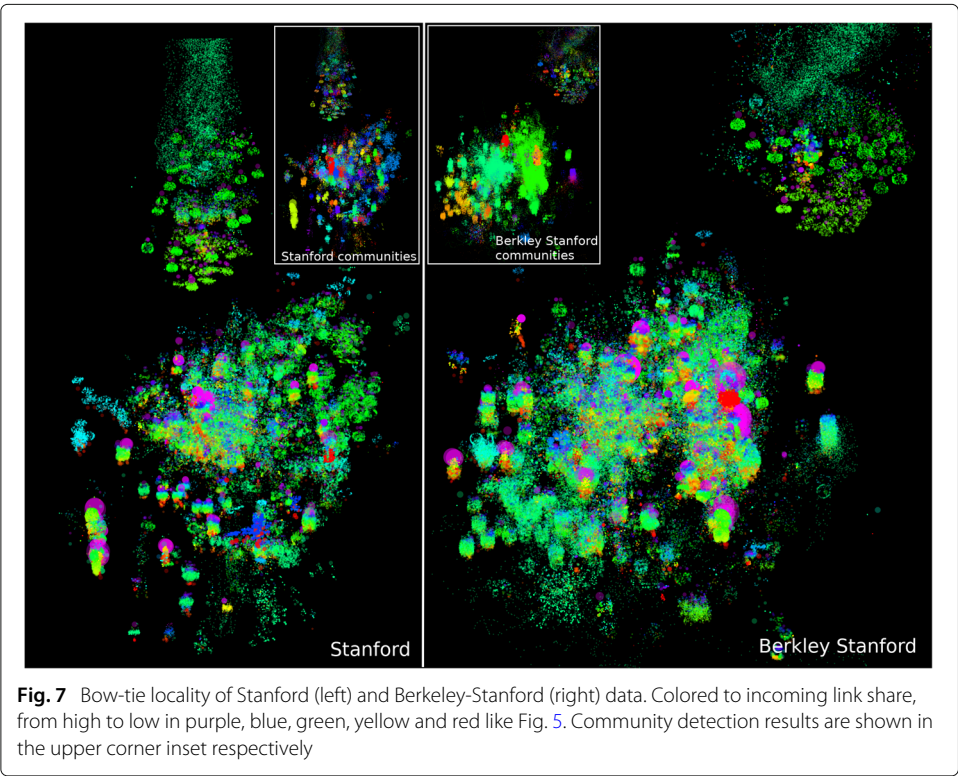
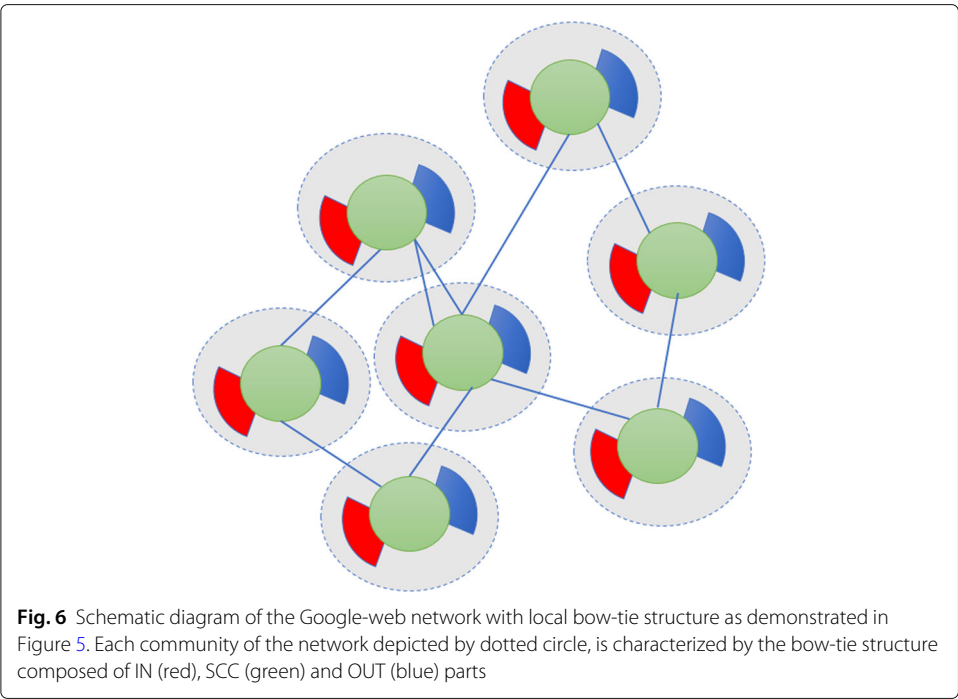
lower right inset to show several of the communities, where we can see each communities have their own bow-ties.

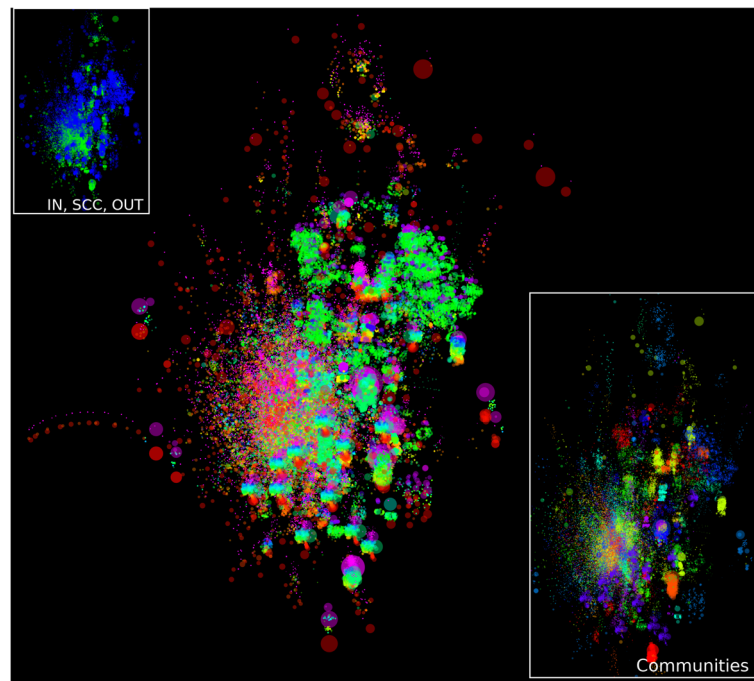
Figure 5 suggests that the Google-web network has a structure composed of a number of local bow-tie structures. We call the network with such a structure a “bow-tie locality” or “local bow-tie structure” which means the network has the IN-SCC-OUT structure not as global but as locally limited attribute within a community. Figure 6 illustrates the suggested local bow-tie network in a schematic way.

As there are several other web network data available at (Leskovec and Krevl 2014), we also analyzed Stanford, Berkeley-Stanford and Notre Dame web data. Stanford and Berkeley-Stanford data are visualized in a single Fig. 7 and Notre Dame data is visualized in Fig. 8. Each figure has insets to show community separation result. Additionally Notre Dame figure (Fig. 8) has an inset in the upper left corner to show that it lacks “IN” segment. Bow-tie locality can also be seen in other three web data.

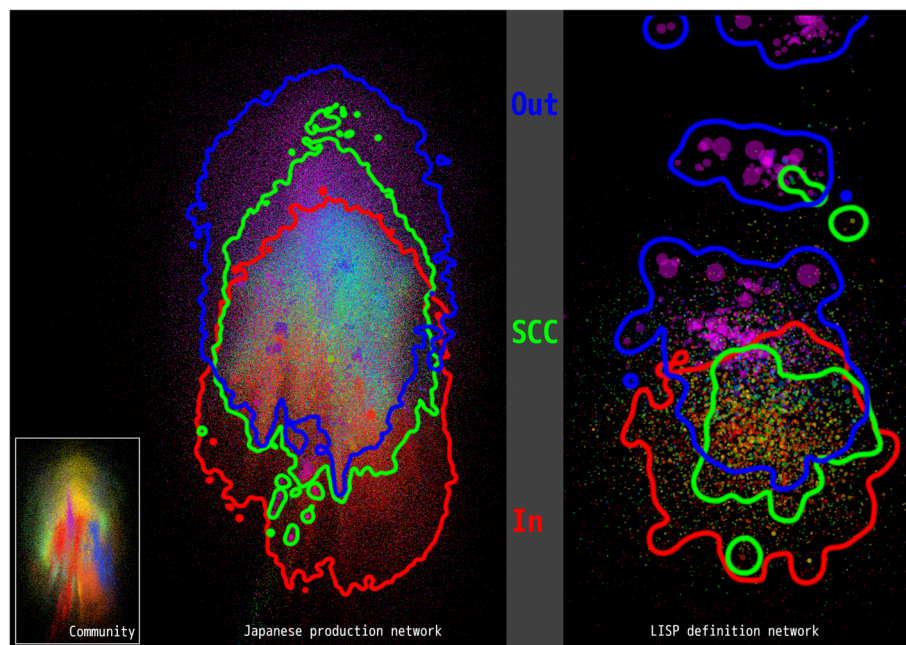
For comparison we show a visualization of Japanese production network and symbol definition relation network of a programming language in Fig. 9. We can recognize that both Japanese production network and symbol definition network have bow-tie structure as their overall feature. However, the locality of the bow-tie structure is not observable in these two networks.

To begin with, we first checked the community size distribution of the four web data, which is shown in Fig. 10. We can see that Google web network has a sharp community

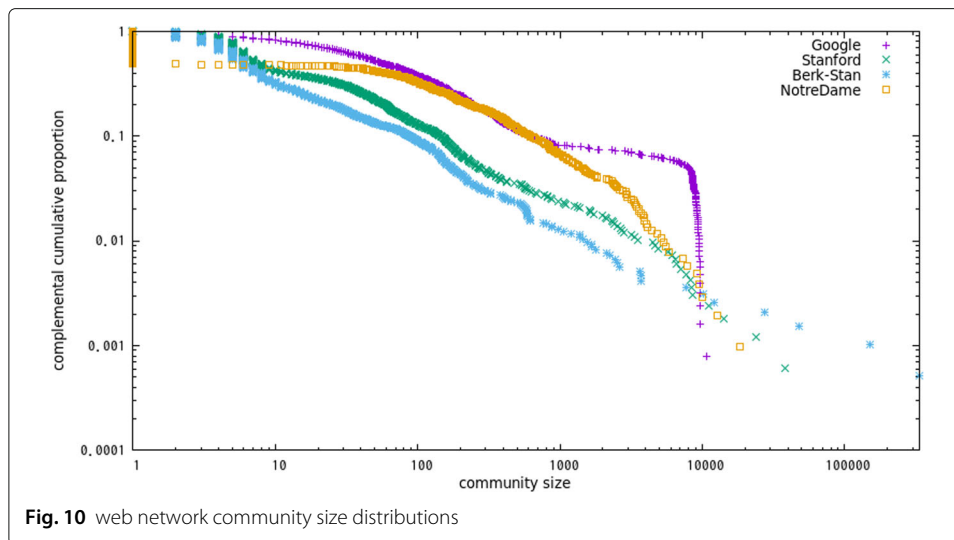




**Fig. 8** Bow-tie locality in Notre Dame web network. Lower right inset shows community detection result, and upper left shows only SCC (green) and OUT (blue) segments as the network lacks IN segment. Main plate is colored to the same scheme as Figs. 5 and 7



**Fig. 9** Bow-tie structure exists in Japanese production network and Emacs LISP network as their global features. IN, SCC and OUT segments' locations are shown by blue, green and red lines. Link direction upward, main plate colored to incoming link share like Fig. 5. Lower left inset shows community detection result

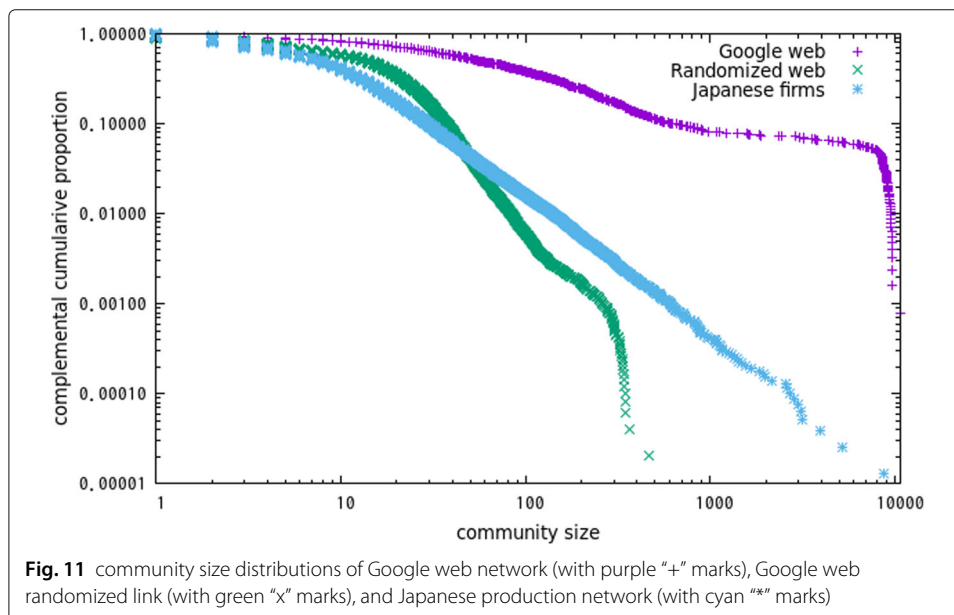


size limit around ten thousand nodes, while other networks show natural power-law distribution.

As Fig. 11 shows clear difference in community size distribution between Google web data and its randomized control data with identical degree distribution, community structure of the web network does not come simply by accident.

#### Quantitative identification of locality

In order to identify the local bow-tie structure quantitatively, we first pay attention to the GSCC parts of the Google-web network and Japanese production network. If the GSCC is well divisible into pieces, those would be cores of local bow-tie structures. We carry out the bow-tie decomposition of subnetworks isolated by the communities as shown in Figs. 5 and 9. The Table 4 shows the number of communities and the modularity of the



**Table 4** Summary of the community detection using the Infomap algorithm for the Google-web network and the Japanese production network, showing the number of communities, the modularity of the GSCC corresponding to the Infomap partition, and the ratio of the number of the GSCC nodes in the subnetworks to that in the original network

	Google-web	Japanese production network
#Communities	172	20
Modularity	0.887	0.471
GSCC ratio	86.2%	54.9%

GSCC partitioned by the communities. Also, we give the result for ratio of the number of the GSCC nodes in the subnetworks to that in the original network. The modularity of the GSCC of the Google-web network is close to the unity. This means the GSCC is extremely modular. On the other hand, the GSCC of the Japanese production network is not so modular as compared with that of Google-web. The highly large ratio of the GSCC in Google-web indicates that the GSCC is dominated by local loops and also confirms that each of the subnetworks has a well-defined GSCC. As shown below, most of IN and OUT nodes in a given community are exclusively connected to the GSCC nodes within the same community. We can thus establish the locality of the bow-tie structure of the Google-web network.

Next, let us further dig into the communities and see how those sub-networks are linked to other part of the network.

For the purpose of finding the connection characteristic, we developed bow-tie modularity index as follows to see the “IN” and “OUT” segments’ acting.

The bow-tie modularity index is defined to each subgraph.  $G$  be a graph and  $G_p$  be a subgraph of  $G$ . Let  $Out(G_p)$ ,  $SCC(G_p)$ ,  $In(G_p)$  be the bow-tie segment (if any) of the subgraph  $G_p$ .

Let  $Bf(G_p)$  be a set of links that bridges  $Out(G_p)$  and outside of  $G_p$  in *forward* (from  $Out(G_p)$  view point) direction. Likewise  $Bb(G_p)$  be a set of links that bridges  $In(G_p)$  and  $G - G_p$  *backward*. We put letter  $f$  ( $b$ ) on  $Bf$  ( $Bb$ ) to signify that it is a “forward” (or “backward”) bridge.

To be precise

$$(v_i, v_j) \in Bf(G_p) \leftrightarrow v_i \in Out(G_p) \wedge v_j \in (G - G_p) \quad (4)$$

and

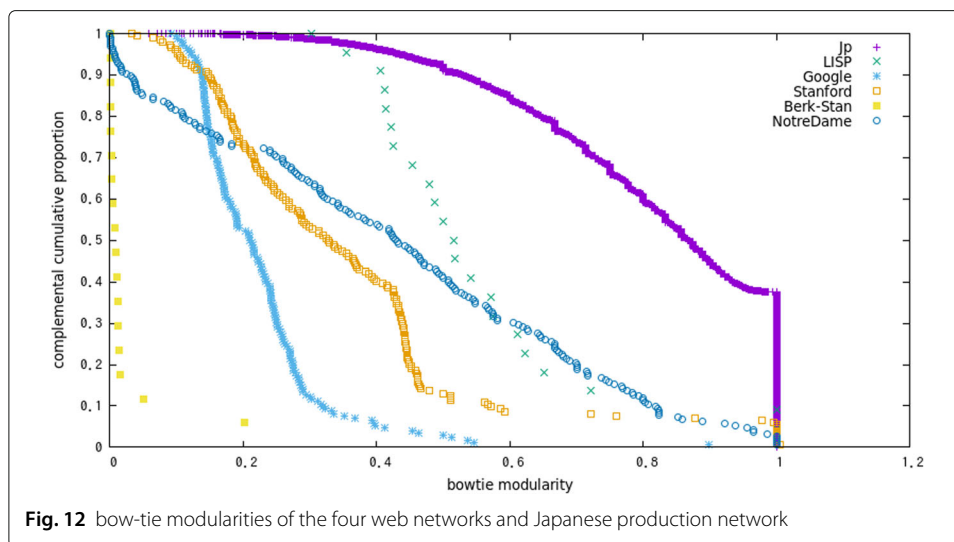
$$(v_i, v_j) \in Bb(G_p) \leftrightarrow v_j \in In(G_p) \wedge v_i \in (G - G_p) \quad (5)$$

Let  $E_p$  be the link set of the bow-tie of subgraph  $G_p$ , then bow-tie modularity  $BTM(G_p)$  of the subgraph  $G_p$  is

$$BTM(G_p) = \frac{|Bf(G_p)| + |Bb(G_p)|}{|E_p|} \quad (6)$$

Intuitively, Eq. 6 measures the ratio of outside connectivity through IN or OUT segment relative to the inside connectivity of the bow-tie. If this index is small, the bow-tie of concerning subgraph (or community) is relatively independent to the external segment.

Here we present the comparison plot of the index values of communities of Japanese production network, LISP definition network and four web networks as Fig. 12. To obtain these results we started checking from largest community to proceed smaller one, until



the checked community population share reaches 90 percent of the total node number. So, not all communities are inspected here.

We can see that the index value is consistently lower in web network than Japanese production network. Emacs LISP network comes in between four web networks and Japanese production network. Average value of the locality index values are listed in Table 5 for convenience.

web are respectively.

It is known the community of Japanese production network is closely related to the industrial sectors (see (Chakraborty et al.)). Industrial sector has definite economic role within the production network. That particular role gives appropriate niche to the collection of nodes within the whole economic network. For example, wholesale sector is a customer of the manufacturing sector and a supplier of the retail sector. As we have seen in the Introduction section, the IN and OUT segment of bow-tie structure works as a interface to the external of the network. Consequently IN or OUT segment of wholesale sector should be tightly connected to the outside, which is manufacturing and retail sector's community. The meaning of "external" depends on the network. For a production sub-network, the external existence is either 1) other part of the production network of that segment or 2) economic agent located outside of production network (such as consumers or governmental organizations).

In case of the web network, IN or OUT part have less need to be connected to other part of the network. Maybe viewer comes directly to the IN segment rather than by following link from other web contents, although we have no way to check this supposition so far.

## Conclusion

Bow-tie structure was first proposed almost twenty years ago as the overall feature of the World Wide Web. Since then this structure served several different purposes including

**Table 5** Average values of the locality index

	JP	Emacs	Google	Stanford	Berke-Stan	Notre Dame
Average	0.82	0.54	0.22	0.35	0.02	0.42

the connectivity evaluation of directional network, dividing network into disjoint parts using connectivity information only, estimate or analyze robustness of the network and overall network feature analysis.

In this study we discovered that the bow-tie structure of web network, upon which the structure was originally proposed as a overall feature, is actually a locally limited structure within relatively densely connected sub-networks (communities). It was found by using originally developed two-staged directed graph visualization method. We named this property as “bow-tie locality” and developed an index that can evaluate the degree of locality quantitatively. Locality is quantitatively confirmed by the fact that all the web network data housed in SNAP Dataset ((Leskovec and Krevl 2014)) shows more locality than economic network (Japanese production network) or programming symbols’ network (Emacs LISP function definition relation).

Large-scale complex networks consists of multiple networks connected together and not a monolithic structure in general, which has been studied by various community detection methods. Bow-tie locality concept is useful to analyze the way how these components (communities) are connected with each other, which has so much diversity that it is often difficult to have any research direction. Japanese production network, which is one of the reference data in this study, is a complex of heterogeneous communities. We will apply bow-tie locality to understand the internal structure of the network and analyze the relation between industrial sectors in the near future.

#### Acknowledgements

We would like to thank Hideaki Aoyama for helpful discussions and encouragement. We also thank Atsushi Kawai for his kind support in using Phantom GRAPE and providing GRAPE interface.

#### Funding

This study was supported in part by the Project “Large-scale Simulation and Analysis of Economic Network for Macro Prudential Policy” undertaken at Research Institute of Economy, Trade and Industry (RIETI), MEXT as Exploratory Challenges on Post-K computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena), Grant-in-Aid for Scientific Research (KAKENHI) by JSPS Grant Number 17H02041.

#### Availability of data and materials

The datasets used and/or analysed during the current study are publicly available from (Leskovec and Krevl 2014) for Web network. The programming language symbol definition relation data is available from the corresponding author on reasonable request. Japanese production network data of this study are available from Tokyo Shoko Research, LTD. (TOKYO SHOKO RESEARCH 2016) but restrictions apply to the availability of this data, which is used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Tokyo Shoko Research, LTD. The analysis tools (computer program) of this research is publicly available from author’s repository <https://github.com/fjt/ngraph/>.

#### Authors’ contributions

YF: Visualization algorithm development, algorithm implementation, data acquisition, analysis and interpretation of data. YK: Initial analysis and interpretation of data, modularity analysis. YF: Acquisition of funding, visualization algorithm, data acquisition and important advice for analysis. WS: Acquisition of funding, data acquisition and analysis design. HI: Overall analysis design and important advice on the overall logical structure of research. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>Turnstone Research Inst. Inc., Kamakura 248-0004, Japan. <sup>2</sup>Department of Mathematics, Niigata University, Niigata 950-2181, Japan. <sup>3</sup>Graduate School of Simulation Studies, University of Hyogo, Kobe 650-0047, Japan. <sup>4</sup>Nihon University, Funabashi 274-0063, Japan.

Received: 12 November 2018 Accepted: 2 April 2019

Published online: 18 April 2019

## References

- Ataru Tanikawa e. a (2012) Phantom-grape: numerical software library to accelerate collisionless n-body simulation with simd instruction set on x86 architecture. *arXiv.org/astro-ph/*arXiv:1203.4037
- Barnes J, Hut P (1986) A hierarchical  $O(N \log N)$  force-calculation algorithm. *Nature* 324(6096):446–449
- Brandes U, Pich C (2006) Eigensolver Methods for Progressive Multidimensional Scaling of Large Data. *LNCS* 4372. [https://doi.org/10.1007/978-3-540-70904-6\\_6](https://doi.org/10.1007/978-3-540-70904-6_6)
- Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, Tomkins A, Wiener J (2000) Graph structure in the web. *Comput Netw* 33:309–320
- Chakraborty A, Kichikawa Y, Iyetomi H, Iino T, Inoue H, Fujiwara Y, Aoyama H Hierarchical communities in the walnut structure of japanese production networks. *PLoS ONE* 13. <https://doi.org/10.1371/journal.pone.0202739>
- Csete M, Doyle J (2004) Bow ties, metabolism and disease. *Trends Biotechnol* 22:446–50. <https://doi.org/10.1016/j.tibtech.2004.07.007>
- Dill S, Kumar R, McCurley KS, Rajagopalan S, Sivakumar D, Tomkins A (2002) Self-similarity in the web. *ACM Trans Internet Technol (TOIT)* 2(3):205–223
- Donato D, Leonardi S, Millozzi S, Tsaparas P (2005) Mining the inner structure of the web graph. *8th Int Work Web Database*:145–150
- Donato D, Leonardi S, Millozzi S, Tsaparas P (2008) Mining the inner structure of the web graph. *J Phys A Math Theor* 41(22):224017
- Fruchterman TMJ, Reingold EM (1991) Graph Drawing by Force-directed Placement. *Softw Pract Experience* 21(11):1129–1164
- Fujita Y, Fujiwara Y, Souma W (2016) Large directed-graph layout and its application to a million-firms economic network. *Evol Inst Econ Rev* 13(2):397–408. <https://doi.org/10.1007/s40844-016-0059-9>
- Kawakami E, Singh VK, Matsubara K, Ishii T, Matsuoka Y, Hase T (2016) Network analyses based on comprehensive molecular interaction maps reveal robust control structures in yeast stress response pathways. *npj Syst Biol Appl* 2. <https://doi.org/10.1038/npjbsa.2015.18>
- Kitano H (2004) Biological robustness. *Nat Rev Genet* 5(11):826
- Leskovec J, Krevl A (2014) SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>. Accessed 12 Oct 2018
- Leskovec J, Lang KJ, Dasgupta A, Mahoney MW (2008) Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *arXiv.org*:0810.1335
- Martin Rosvall CTB (2011) Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS ONE* 6(4). <https://doi.org/10.1371/journal.pone.0018209>
- Meusel R, Vigna S, Lehmborg O, Bizer C (2014) Graph structure in the web—revisited: a trick of the heavy tail. In: *Proceedings of the 23rd international conference on World Wide Web*. ACM. pp 427–432. <https://doi.org/10.1145/2567948.2576928>
- Meusel R, Vigna S, Lehmborg O, Bizer C (2015) The graph structure in the web: Analyzed on different aggregation levels. *J Web Sci* 1(1):33–47
- Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *PNAS* 105(4):1118–1123. <https://doi.org/10.1073/pnas.0706851105>
- Sugiyama K, Misue K (1995) Graph drawing by the magnetic spring model. *J Vis Lang Comput* 6:217–231
- TOKYO SHOKO RESEARCH LTD. (2016). <http://www.tsr-net.co.jp>. Accessed Dec 2016
- Zhao J, Yu H, Luo J-H, Cao Z-W, Li Y-X (2006) Hierarchical modularity of nested bow-ties in metabolic networks. *BMC Bioinformatics* 7(386). <https://doi.org/10.1186/1471-2105-7-386>
- Zhu JJH, Meng T, Xie Z, Li G, Li X (2008) A teapot graph and its hierarchical structure of the Chinese web. In: *Proceedings of the 17th international conference on World Wide Web*. ACM. pp 1133–1134. <https://doi.org/10.1145/1367497.1367692>

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)