

RESEARCH

Open Access



Open networks from within: input or output betweenness centrality of nodes in directed networks

Taichi Haruna 

Correspondence:

tharuna@lab.twcu.ac.jp

Department of Information and Sciences, Tokyo Woman's Christian University, 2-6-1 Zempukujji, Suginami-ku, 1678585 Tokyo, Japan

Abstract

New betweenness centralities of nodes in a directed network are proposed based on the idea that nodes in a network are processes rather than things. They are called input and output betweenness centralities. They measure importance of nodes as input and output for gluing arcs together as interface between processes, respectively. We demonstrate their use and discuss their meaning by calculating them in two toy directed networks and one real-world network. We also compare them with the existing centrality measures that reflect asymmetry of links in directed networks: out- and in-degrees and Hub and Authority scores. We found that input and output betweenness centralities behave differently from these measures in some nodes. It is suggested that they can effectively identify nodes that are less important in terms of existing measures but are noteworthy from the viewpoint that nodes are processes.

Keywords: Directed networks, Betweenness centrality, Category theory

Introduction

Categorical network theory is a general framework to study open networks, namely, networks with explicit input and output nodes such as electrical circuits (Baez and Fong 2015), signal flow diagrams (Baez and Erbele 2015; Bonchi et al. 2014) and chemical reaction networks (Baez and Pollard 2017). It thinks of networks as processes in contrast to network science where networks are thought of as things (Baez 2014). Indeed, the main challenges in network science are analysis and modeling of network structure found in nature and society (Newman 2010; Estrada 2012; Barabási 2016). On the other hand, the primary interest of categorical network theory is the behavior of open networks determined by the relation between inputs and outputs that is revealed by black-boxing the internal structure of networks (Baez and Fong 2015). The divide into networks as things and networks as processes is a natural consequence of the category theoretic perspective. As we explain in next section, a category has two-level structure: objects representing things and morphisms representing processes between things. In categorical network theory, networks are regarded as morphisms. The aim of this study is to bridge these two different approaches to networks and get a new insight into structure of networks. Our approach is based on a reformulation of our previous work (Haruna 2013b) that has been developed independently of categorical network theory. However, we reinterpret

our previous work as an internalization of the idea of categorical network theory into networks themselves and as a result we find new betweenness centralities that are obscure so far.

Identifying important nodes or links from a specific perspective is a basic task when analyzing networks found in social and natural sciences. A variety of centrality measures has been proposed to quantify importance of nodes or links (Borgatti and Everett 2006). In this paper, we focus on betweenness centrality among them since it seems to be the most natural one in our context. There are many variants of betweenness centrality such as the original one based on shortest paths (Anthonisse 1971; Freeman 1977) and those based on network flows (Freeman et al. 1991), random walks (Newman 2005) and percolation (Piraveenan et al. 2013). We will be faced with a path notion called lateral path different from the usual directed path in directed networks in the following attempt to internalize the idea of categorical network theory into networks themselves. Here, we only focus on the simplest betweenness centrality based on shortest lateral paths because extensions to other variants seem to be non-trivial and are out of the scope of this paper. We propose input (resp. output) betweenness centrality of nodes in a directed network quantifying importance of nodes as input (resp. output) from the viewpoint that nodes are processes.

This paper is organized as follows. In the second section, we explain how the idea of categorical network theory can be internalized into networks themselves and derive input or output betweenness centrality from the viewpoint that nodes are processes. In the third section, we calculate them in two toy directed networks and a real-world network. In the final section, conclusions are given.

Input or output betweenness centrality

In this section, we first explain the main idea of categorical network theory and then proceed to the way we internalize it into directed networks themselves. There are several approaches based on different techniques such as cospans (Fong 2015), props (Bonchi et al. 2014; Baez et al. 2017) and operads (Spivak 2013). Here, we follow the one based on cospans (Fong 2015). Finally, input or output betweenness centrality is introduced.

Idea of categorical network theory

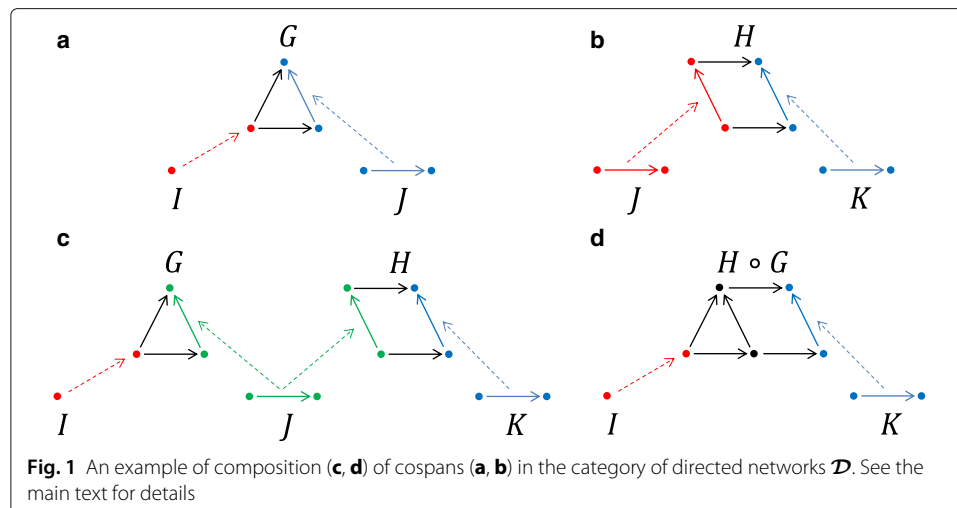
In categorical network theory, networks are regarded as processes, namely, a certain kind of action with input and output. This can be formalized in category theory. Here, we do not intend to go into the technical details. However, a few terminologies cannot be avoided to explain it without ambiguity. So, first we explain them.

In general, a category consists of objects and morphisms between objects. It is subject to a few axioms but we leave the details such as the rigorous definition of categories and their basic properties to introductory textbooks (Awodey 2010; Spivak 2014). Objects represent things of interest and a morphism between two objects represents an allowed process from one to the other. For example, in the category of sets, objects are sets and morphisms are maps. A map f from a set A to another set B (we write $f : A \rightarrow B$) specifies an element of B to which each element of A is transformed. As the map f has its domain A and codomain B , each morphism in a category has its domain and codomain. The domain and codomain of a morphism can be seen as input and output to the process

represented by the morphism, respectively. We denote a morphism m with domain D and codomain E in a category by $m : D \rightarrow E$ as in the case of maps. As two maps $f : A \rightarrow B$ and $g : B \rightarrow C$ can be composed and yield a new map $g \circ f : A \rightarrow C$, two morphisms $m : D \rightarrow E$ and $n : E \rightarrow F$ can be composed. The obtained morphism is denoted by $n \circ m : D \rightarrow F$.

Now, let us consider directed networks. We denote a directed network by $G = (N, A, s, t)$ where N is the set of nodes, A is the set of arcs and s and t are maps from A to N sending each arc to its source and target nodes, respectively. Directed networks form a category \mathcal{D} . Its objects are directed networks and morphisms are homomorphisms of directed networks, namely, “maps” preserving the structure of directed networks: A morphism $m : G_1 = (N_1, A_1, s_1, t_1) \rightarrow G_2 = (N_2, A_2, s_2, t_2)$ is a pair of maps $(m_N : N_1 \rightarrow N_2, m_A : A_1 \rightarrow A_2)$ such that $m_N \circ s_1 = s_2 \circ m_A$ and $m_N \circ t_1 = t_2 \circ m_A$.

In the category of directed networks \mathcal{D} , directed networks are objects and thus they are regarded as things, not processes. However, we can build a category whose morphisms are directed networks with input and output as follows. Let us call a pair of morphisms $(i : I \rightarrow G, o : J \rightarrow G)$ in \mathcal{D} a cospan from I to J . In Fig. 1a, an example is shown. If we have another cospan $(i' : J \rightarrow H, o' : K \rightarrow H)$ from J to K (Fig. 1b), we can glue them via $o : J \rightarrow G$ and $i' : J \rightarrow H$ (Fig. 1c). As a result, we obtain a new cospan $(\tilde{i} : I \rightarrow H \circ G, \tilde{o}' : K \rightarrow H \circ G)$ from I to K (Fig. 1d), where \tilde{i} and \tilde{o}' are homomorphisms to $H \circ G$ that are induced automatically from i and o' , respectively. Thus, we can compose cospans. It is known that we can form a category $\text{Cospan}(\mathcal{D})$ whose objects are objects of \mathcal{D} , namely, directed networks and morphisms are cospans (precisely, we should take isomorphism classes of cospans (Borceux 1994)). This fact itself is not used in the following, we have remarked it for completeness. Thus, a directed network G together with its input $i : I \rightarrow G$ and output $o : J \rightarrow G$ can be regarded as a process within the category $\text{Cospan}(\mathcal{D})$. This is one of the main ideas of categorical network theory. Note that here we only consider network topology. To include weights or dynamics on networks requires extra machineries (Fong 2015).

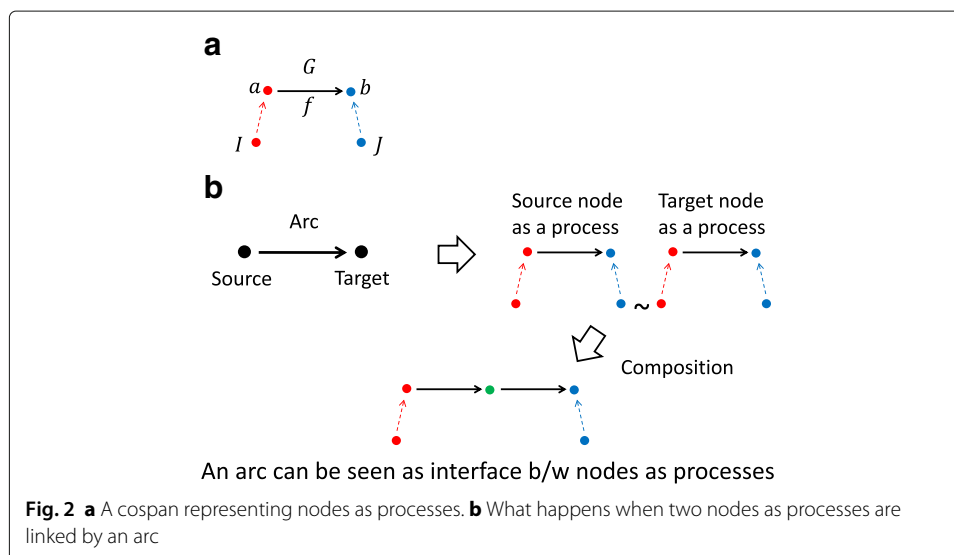


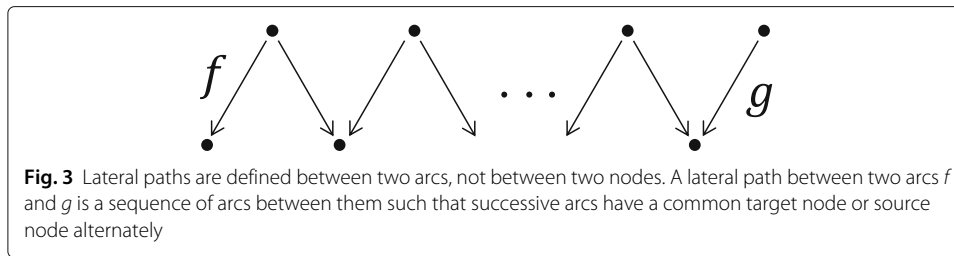
Internalization

In the previous subsection, we explain how a directed network together with its input and output can be seen as a process. In this subsection, we internalize this idea into networks themselves. The content of this subsection is a recapitulation of our previous work (Haruna 2013b) from the viewpoint of categorical network theory.

Our motivation is as follows. In real-world networks, nodes are not just points. They often have internal processes. In particular, this is obvious for some biological networks. For example, let us consider food webs. We will analyze one in next section. In a food web, nodes are biological taxa. They are living processes. Links represent prey-predator relations. From the physical viewpoint, they are passages of organic materials. However, if we think of nodes as processes, then links can be interpreted as interface between living processes. Thus, the idea that networks are processes should be internalized to each node in a given network. In the following, we explain how this idea can be formalized.

We would like to represent nodes as processes in the sense of categorical network theory. One of the simplest and natural way is to represent a node in a directed network by the cospan $(i : I \rightarrow G, o : J \rightarrow G)$ (Fig. 2a): G consists of two distinct nodes $\{a, b\}$ and an arc f from a to b . I and J are networks with a single node and no arcs. Maps i and o send the single nodes to a and b , respectively. Now let us consider two nodes in the network connected by an arc (Fig. 2b). We replace the source and target nodes of the arc by the two copies of the cospan $(i : I \rightarrow G, o : J \rightarrow G)$. The arc is interface between these two processes. This can be manifested by identifying the output of the cospan for the source node with the input of the one for the target node. Then, we can compose these two cospans and obtain a new cospan shown in the bottom of Fig. 2b. By forgetting the input and the output of the obtained cospan, we can see that the network at the left top in Fig. 2b is transformed to the one at the apex of the cospan at the bottom in Fig. 2b. Indeed, this procedure can be extended to the whole network and gives rise to a network transformation L described as follows (Haruna 2013b): Let $G = (N, A, s, t)$ be a directed network. $L(G) = (N', A', s', t')$ is a directed network such that $N' = N \times \{0, 1\} / \sim$, $A' = N$ and maps $s', t' : A' \rightarrow N'$ are defined by $s'(x) = [(x, 0)]$ and $t'(x) = [(x, 1)]$. Here, \sim

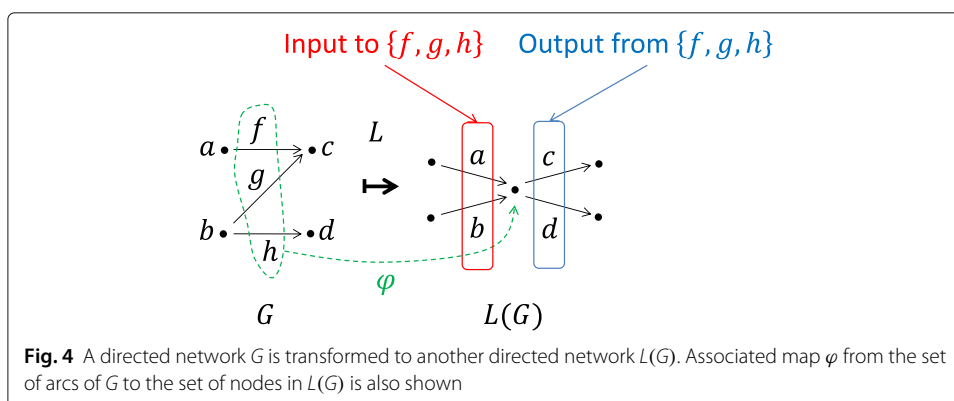




is an equivalence relation on the set $N \times \{0, 1\}$ generated by the relation R : We define $(x, 1)R(y, 0)$ when there is $f \in A$ such that $s(f) = x$ and $t(f) = y$. $[(x, i)]$ is the equivalence class containing (x, i) for $i = 0, 1$. In other words, nodes in G become arcs in $L(G)$ and they are glued up together along the relation induced by arcs in G . $(x, 0)$ and $(x, 1)$ correspond to the input and the output of the cospan in Fig. 2a, respectively. An example is shown in Fig. 4.

What is the precise relationship between the composition of cospans and $L(G)$? Both are examples of colimits (MacLane 1998). Colimits are a categorical construction to form an object by gluing parts together. The composition of cospans is a special type of colimits called pushouts (MacLane 1998). On the other hand, $L(G)$ is a more general colimit depending on the “shape” of G (Haruna 2013b).

The idea that arcs are interface between nodes as processes can be formalized as a map $\varphi : A \rightarrow N'$ defined by $\varphi(f) = [(s(f), 1)]$ or equivalently, $\varphi(f) = [(t(f), 0)]$. In Fig. 4, arcs f, g and h in G are mapped to a single node in $L(G)$. In general, $\varphi(f) = \varphi(g)$ holds for $f, g \in A$ if and only if f and g are connected by a lateral path (Fig. 3). A lateral path between two arcs f and g in a directed network is a sequence of arcs such that the first and last arcs are f and g , respectively, and successive arcs in the sequence have a common target node or source node alternately. Note that lateral paths are not defined between nodes but arcs in this paper. Lateral paths between nodes have been considered in the literature (Crofts et al. 2010) to reveal the bipartite community structure of directed networks. The map φ has a characterization in terms of a category theoretic universality (Haruna 2013b): It is the “minimum” map materializing the idea that arcs are interface between nodes as processes. We also note that the directed network transformation L is a kind of dual transformation to the operation of taking the line-graph: the line graph $R(G)$ of a directed network $G = (N, A, s, t)$ is a directed network such that the set of nodes is A and arcs are



directed paths of length 2 in G . In category theoretic terms, both L and R can be made into functors from \mathcal{D} to itself and L is left adjoint to R (Pultr 1979; Haruna and Gunji 2007).

Definition of input or output betweenness centrality

Let us consider a directed network G shown on the left-hand side in Fig. 4. In G , a is the input to f and c is the output from f in the sense of cospan (isolate f and its source a and target c from G and consider the cospan like Fig. 2a). The same holds for g and h . From the result of the previous subsection, we can regard a, b, c and d as processes and f, g and h as interface between them by applying L to G and considering the map φ . By the map φ, f, g and h are sent to the same node at the center of $L(G)$. From this viewpoint, we can say that a and b are input to the set $\{f, g, h\}$ and c and d are output from $\{f, g, h\}$ (Fig. 4). A natural question is, how important are they as input or output? Since f, g and h are related by lateral paths, one could use lateral paths to measure importance of nodes in G with respect to cohesiveness among f, g and h . One way is to introduce analogues of betweenness centrality (Anthonisse 1971; Freeman 1977). Namely, if a node is the source (resp. target) of arcs in many shortest lateral paths, then it would be important as input (resp. output) for retaining cohesiveness of arcs mapped to the same node in $L(G)$. To quantify importance of nodes as input (resp. output) in this sense, first we calculate the betweenness centrality of arcs with respect to lateral paths and then project them to their source (resp. target) nodes.

Let $G = (N, A, s, t)$ be a directed network. The lateral betweenness centrality (LBC) of an arc $f \in A$ is (Haruna 2013b)

$$LBC_f = C \sum_{g,h \in A, l_{gh} > 0} \frac{l_{gh}^f}{l_{gh}}, \tag{1}$$

where l_{gh} is the number of shortest lateral paths between g and h , l_{gh}^f is the number of shortest lateral paths between g and h that pass through f and $C = \sum_{g,h \in A, l_{gh} > 0} (d_{gh} + 1)$ is the normalization constant such that $\sum_{f \in A} LBC_f = 1$. d_{gh} denotes the length of shortest lateral paths between g and h . The length of a lateral path is the number of nodes that are passed through between g and h . In particular, $d_{gg} = 0$. $\sum_{f \in A} LBC_f = 1$ follows from the equality $\sum_{f \in A} l_{gh}^f = l_{gh}(d_{gh} + 1)$ for $g, h \in A$ such that $l_{gh} > 0$. Indeed, both sides of the equality are two different ways to count the number of arcs on shortest paths from g to h with repetition. Note that in the summation in Eq. (1), g, h are an ordered pair. Thus, the same shortest lateral path is counted twice if $g \neq h$: one is from g to h and the other is from h to g .

The input betweenness centrality (IBC) of a node $x \in N$ is defined by summing all LBC_f s such that the source of f is x :

$$IBC_x = \sum_{s(f)=x} LBC_f. \tag{2}$$

The output betweenness centrality (OBC) of x is defined similarly:

$$OBC_x = \sum_{t(f)=x} LBC_f. \tag{3}$$

Note that we do not directly focus on the structure of $L(G)$ to define LBC, IBC and OBC. Lateral paths induce an equivalence relation on the set of arcs in G to form the

nodes in $L(G)$. These measures evaluate contribution of each arc or node for gluing arcs along lateral paths and yielding nodes in $L(G)$. Since lateral paths are derived from the map φ representing arcs as interface between nodes as processes, we suggest that IBC (resp. OBC) can be used to identify natural input (resp. output) nodes of a given directed network from the viewpoint that nodes are processes.

The LBCs of all the arcs can be calculated by slightly modifying the Brandes-Newman algorithm (Brandes 2001; Newman 2001): For each arc f , we construct the shortest path tree with respect to lateral paths and run the algorithm to calculate the contribution to LBC of shortest lateral paths starting from all the arcs and ending at f . The time complexity to calculate LBC_f for all the arcs f in G is $O(|A|^2)$ or $O(|N|^2)$ for sparse networks. The same holds for the calculation of IBCs or OBCs of all the nodes.

Examples and an application

In this section, we calculate IBC and OBC of two toy networks and one real-world network. We compare them with existing centrality measures reflecting asymmetry of links. The aim of this section is not a thorough analysis of a specific network but demonstration of their use.

Toy examples

First, let us calculate IBC and OBC of nodes in a small network for an illustration (Fig. 5a). Since $l_{ij} = 1$ for all $i, j \in \{f, g, h\}$, $d_{ff} = d_{gg} = d_{hh} = 0$, $d_{fg} = d_{gf} = d_{gh} = d_{hg} = 1$ and $d_{fh} = d_{hf} = 2$, we have $C = 3(0+1)+4(1+1)+2(2+1) = 17$. Since $l'_{ff} = l'_{fg} = l'_{gf} = l'_{fh} = l'_{hf} = 1$, $LBC_f = 5/17$. Similarly, we find $LBC_g = 7/17$ and $LBC_h = 5/17$. Thus, we have $IBC_1 = LBC_f = 5/17$, $IBC_2 = LBC_g + LBC_h = 12/17$ and $IBC_3 = IBC_4 = 0$. Similarly, $OBC_1 = OBC_2 = 0$, $OBC_3 = LBC_f + LBC_g = 12/17$ and $OBC_4 = LBC_h = 5/17$.

Next, let us consider a larger but still a small network consisting of 10 nodes and 12 arcs (Fig. 5b). Here, thickness of arcs is proportional to LBC, size of red nodes IBC and size of blue node OBC. OBC of node i for $1 \leq i \leq 5$ and IBC of node j for $6 \leq j \leq 10$ are 0. Let us call a set of arcs forming dense lateral connections lateral community of arcs (LCA). In Fig. 5b, we could identify two such communities by visual inspection: the set of arcs from nodes 1,2,3 to 6,7 and the set of arcs from nodes 4,5 to 8,9,10. From this example, it is suggested that nodes bridging LCAs from input and output sides of them have high IBC and OBC, respectively. This can be expected from the definitions of IBC and OBC since shortest lateral paths between two LCAs must pass through an arc bridging them as in the case of the classical betweenness centrality. The values of IBC and OBC of all the nodes are shown in Table 1 together with out-degree, in-degree, Hub score and Authority score for comparison. Recall that the out-degree of a node in a directed network is the number

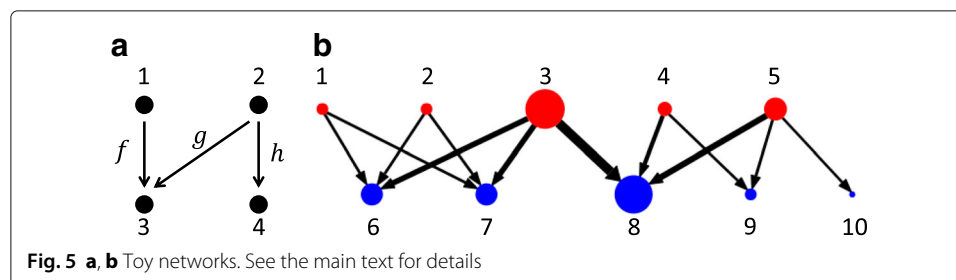


Fig. 5 a, b Toy networks. See the main text for details

Table 1 Centrality measures for the toy network in Fig. 5b

Node	Out-degree	In-degree	Hub	Authority	IBC	OBC
1	2	0	0.461162	0.000000	0.112108	0.000000
2	2	0	0.461162	0.000000	0.112108	0.000000
3	3	0	0.640115	0.000000	0.405830	0.000000
4	3	0	0.309048	0.000000	0.230942	0.000000
5	2	0	0.263439	0.000000	0.139013	0.000000
6	0	3	0.000000	0.600224	0.000000	0.221973
7	0	3	0.000000	0.600224	0.000000	0.221973
8	0	3	0.000000	0.465831	0.000000	0.392377
9	0	1	0.000000	0.118723	0.000000	0.051570
10	0	2	0.000000	0.219926	0.000000	0.112108

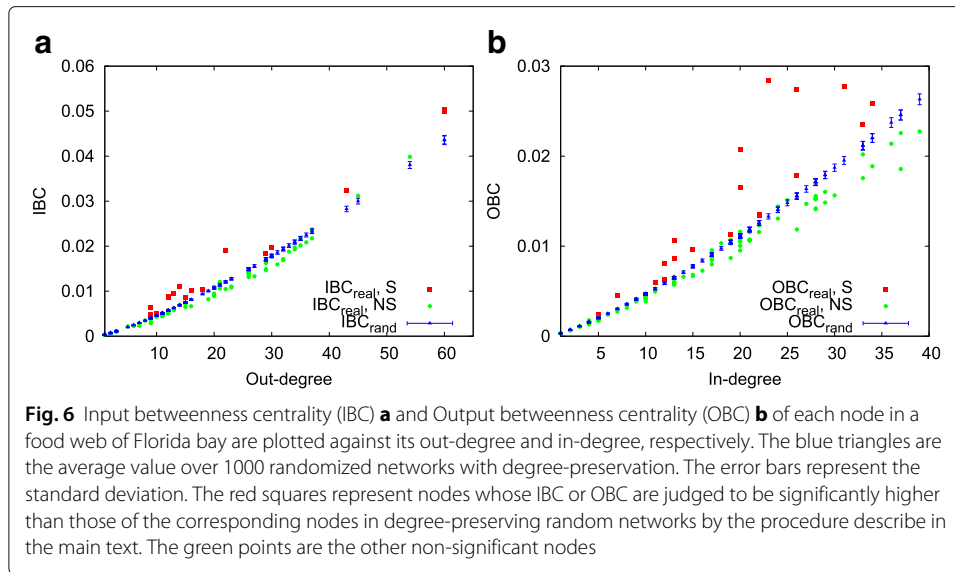
of outgoing arcs from the node and the in-degree of a node is the number of incoming arcs to the node. Authority and Hub scores were originally proposed as a method to find authoritative pages about a specific topic on the WWW together with hubs collecting such authoritative pages (Kleinberg 1999). The idea is that a page with a high Hub score has many links toward pages with a high authority score on one hand, a page with a high Authority score receives many links from pages with a high Hub score on the other hand. In this paper, we apply the HITS algorithm (Kleinberg 1999; Newman 2010) to calculate Authority and Hub scores of all the nodes in a given network.

From Table 1, one can see that there is no monotone relation between IBC and Hub score or between OBC and Authority score. For example, OBC of node 8 is the highest OBC but its Authority score is not. Nodes 1 and 2 have the lowest IBC but their Hub scores are the second highest. This example quantitatively suggests that IBC and OBC measure importance of nodes that cannot be captured by Hub and Authority scores.

Florida Bay food web

In this subsection, we consider a food web of Florida Bay (Ulanowicz et al. 1998). It consists of 121 nodes and 1767 arcs. To focus on the prey-predator relation, we excluded the two detrital nodes and the node representing roots from the original data (Ulanowicz 2002) from the analysis. In Fig. 6a and b, IBC and OBC are plotted against out-degree and in-degree, respectively. One can find overall positive correlation. However, we can identify several nodes with significantly high IBC or OBC values by the following procedure: First, we prepared 1000 randomized networks with degree-preservation as a null model. Second, we calculated the p -value of b where b denotes IBC or OBC of a given node. It was calculated from its z -score if the distribution of b in the null model can be approximated by a normal distribution. We tested this by the Kolmogorov-Smirnov test. If the p -value of the KS test is greater than 0.10, then we adopted the normal approximation. Otherwise, the p -value of b is simply the proportion of the degree-preserving randomization trials in which b exceeds the value in the real-world network. Third, we applied the Benjamini-Hochberg-Yekutieli procedure for arbitrary dependency at level 0.05 for the multiple comparisons correction (Benjamini and Yekutieli 2001). This means that we keep the false discovery rate less than 0.05.

In Fig. 7, IBC and OBC are plotted against Hub and Authority scores, respectively. We again found positive correlations. However, there are several nodes that are deviated from



the line of the best linear fittings. They mostly overlap with those judged to be significant in Fig. 6.

The nodes that are judged to have significantly high IBC and OBC are listed in Tables 2 and 3, respectively. One noticing point is that 6 out of 7 phytoplankton taxa whose out-degree is relatively low appear in Table 2, but the other taxa of primary producers such as seagrasses that are important in the carbon cycle (Ulanowicz et al. 1998) do not have significantly high IBC. Another point is that Table 3 consists of mostly invertebrates and fishes that occupy intermediate level of the food chain. The taxa at the top of the food chain such as mammals, and some of reptiles and birds, namely, nodes with zero out-degree, do not have significantly high OBC. The table of centrality measures of all the nodes in the food web is available as an Additional file 1.

How can we interpret these results? In our previous work, we found that robustness of the largest connected components with respect to lateral paths for ten food webs is higher than that for randomized ones (Haruna 2013a). We suggested that the non-random

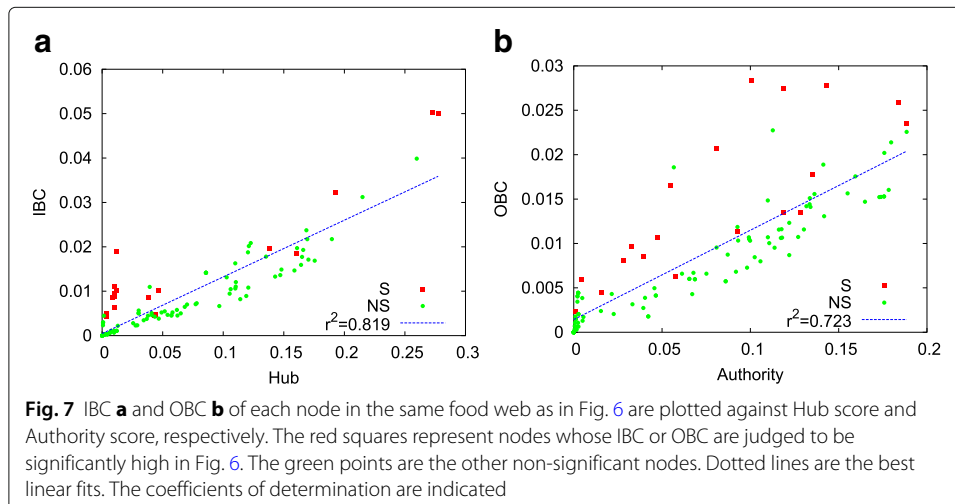


Table 2 Nodes with significantly high IBC in the Florida Bay food web. The values of out-degree and Hub score are also shown

Node	Taxon	Classification	Out-degree	Hub	IBC
1	2um Spherical Phytoplankton	Primary Producers	14	0.010182	0.011130
2	Synedococcus	Primary Producers	22	0.012086	0.019034
3	Oscillatoria	Primary Producers	9	0.009904	0.006438
5	Big Diatoms (> 20um)	Primary Producers	13	0.009851	0.009418
6	Dinoflagellates	Primary Producers	12	0.008580	0.008619
7	Other Phytoplankton	Primary Producers	12	0.010108	0.008840
8	Benthic Phytoplankton	Primary Producers	16	0.011732	0.010121
22	Other Zooplankton	Invertebrates	18	0.046550	0.010272
23	Benthic Flagellates	Invertebrates	10	0.003484	0.004930
24	Benthic Ciliates	Invertebrates	9	0.003462	0.004358
25	Meiofauna	Invertebrates	15	0.038636	0.008634
30	Bivalves	Invertebrates	43	0.192335	0.032299
37	Macrobenthos	Invertebrates	30	0.138043	0.019683
42	Herbivorous Shrimp	Invertebrates	60	0.272954	0.050273
43	Predatory Shrimp	Invertebrates	60	0.277423	0.049966
63	Brotalus	Fishes	9	0.043694	0.004823
80	Mojarra	Fishes	29	0.160359	0.018436
90	Mullet	Fishes	9	0.043322	0.004516

structures of real-world food webs contribute to their robustness. Since nodes that have significantly high IBC or OBC are expected to lie at boundaries between LCAs that are destroyed by degree-preserving randomization, they could play a key role to enhance robustness of food webs as a collection of living processes joined via prey-predator interactions. In particular, we here identified phytoplankton as such nodes as input in the

Table 3 Nodes with significantly high OBC in the Florida Bay food web. The values of in-degree and Authority score are also shown

Node	Taxon	Classification	In-degree	Authority	OBC
20	Other Copepoda	Invertebrates	5	0.000975	0.002374
27	Coral	Invertebrates	7	0.015479	0.004491
29	Echinoderma	Invertebrates	20	0.054952	0.016491
35	Predatory Polychaetes	Invertebrates	13	0.047606	0.010619
36	Suspension Feeding Polychaetes	Invertebrates	11	0.004143	0.005912
44	Pink Shrimp	Invertebrates	15	0.032820	0.009682
65	Needlefish	Fishes	12	0.057367	0.006321
66	Other Killifish	Fishes	20	0.080793	0.020724
68	Rainwater killifish	Fishes	23	0.100857	0.028408
77	Pompano	Fishes	33	0.188513	0.023484
83	Pinfish	Fishes	26	0.118824	0.027448
84	Scianids	Fishes	34	0.183795	0.025882
85	Spotted Seatrout	Fishes	22	0.118785	0.013489
88	Parrotfish	Fishes	13	0.039724	0.008584
90	Mullet	Fishes	12	0.028262	0.008111
95	Flatfish	Fishes	26	0.135255	0.017806
98	Other Pelagic Fishes	Fishes	19	0.092812	0.011312
99	Other Demersal Fishes	Fishes	31	0.143135	0.027782
109	Omnivorous Ducks	Birds	22	0.128449	0.013457

Florida Bay food web. This is unexpected in the literature (Ulanowicz et al. 1998) and cannot be derived by using out-degree or Hub score. Thus, IBC (resp. OBC) can be used as an exploratory tool to find important nodes as input (resp. output) that are overlooked by the conventional measures.

Conclusions

In this paper, we bridged two existing approaches to networks, categorical network theory and network science. The former regards networks as processes, on the other hand, the latter does networks as things. We internalized the idea that networks are processes into networks themselves: Nodes are processes rather than things. Based on the category theoretic representation of this viewpoint, we proposed betweenness centralities of nodes as input and output called IBC and OBC, respectively. We discussed their meaning through toy directed networks and demonstrated their use in real-world directed networks by calculating them for a food web. We also compared them with existing centrality measures that reflect asymmetry of links in directed networks. We found that IBC and OBC take quite different values from out- and in-degrees or Hub and Authority scores for some nodes.

In this paper, IBC and OBC are defined based on shortest lateral paths between arcs. Thus, contributions to betweenness from the other lateral paths are ignored. Their contributions can be included by considering random walks along lateral paths (Newman 2005), which is left as future work. Another future direction is an analytic study of the behavior of IBC and OBC in random networks generated by the configuration model as has been investigated in the case of the classical betweenness centrality (He et al. 2009; Guo et al. 2010). Finally, we only consider unweighted directed networks in this paper. Proposing a natural way to define analogues of IBC and OBC in weighted directed networks seems to be a non-trivial task. We hope that further development of the idea presented in this paper will open a new perspective in network science.

Additional file

Additional file 1: Table of the centrality measures for the Florida Bay food web. The values of Out-degree, In-degree, Hub score, Authority score, IBC and OBC for all the nodes in the Florida Bay food web are listed in `table_floridabay.csv`. (CSV 14 kb)

Abbreviations

LBC: Lateral Betweenness Centrality. It measures importance of arcs when nodes are regarded as processes. The value of it for a give arc is proportional to the number of shortest lateral paths that pass through it; IBC: Input Betweenness Centrality. It measures importance of nodes as input when nodes are regarded as processes; OBC: Output Betweenness Centrality. It measures importance of nodes as output when nodes are regarded as processes; LCA: Lateral Community of Arcs. It is a set of arcs forming dense lateral connections

Acknowledgements

The author is grateful to the anonymous reviewers for their helpful comments to improve the manuscript.

Funding

The writing of this paper was partially supported by JSPS KAKENHI Grant Number 18K03423.

Availability of data and materials

The data set used in this article is available from the cited reference. The result of analysis is available as an Additional file 1.

Author's contributions

TH designed and carried out the research. The author 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.

Received: 15 February 2018 Accepted: 20 June 2018

Published online: 09 July 2018

References

- Anthonisse JM (1971) The rush in a directed graph. Technical Report BN 9/71. Stichting Mathematisch Centrum, Amsterdam
- Awodey S (2010) *Category Theory*, Second Edition. Oxford University Press Inc., New York
- Baez JC (2014) Network theory: Overview. Talk at Centre for Quantum Mathematics and Computation, University of Oxford. http://math.ucr.edu/home/baez/networks_oxford/. Accessed 14 Feb 2018
- Baez JC, Erbele J (2015) Categories in control. *Theory Appl Categ* 30:836–881
- Baez JC, Fong B (2015) A compositional framework for passive linear networks. arXiv:1504.05625
- Baez JC, Pollard BS (2017) A compositional framework for reaction networks. *Rev Math Phys* 29:1750028
- Baez JC, Coya B, Rebro F (2017) Props in network theory. arXiv:1707.08321
- Barabási AL (2016) *Network Science*. Cambridge University Press
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:1165–1188
- Bonchi F, Sobociński P, Zanasi F (2014) A categorical semantics of signal flow graphs. In: CONCUR, 2014, Springer Lecture Notes in Computer Science 8704. Springer, Berlin. pp 435–450
- Borceux F (1994) *Handbook of Categorical Algebra 1*. Cambridge University Press
- Borgatti S, Everett M (2006) A graph-theoretic perspective on centrality. *Social Networks* 28:466–484
- Brandes U (2001) A faster algorithm for betweenness centrality. *J Math Sociol* 25:163–177
- Crofts JJ, Estrada E, Higham DH, Taylor A (2010) Mapping directed networks. *Elec Trans Num Anal* 37:337–350
- Estrada E (2012) *The Structure of Complex Networks: Theory and Applications*. Oxford University Press
- Fong B (2015) Decorated cospans. *Theory Appl Categories* 30:1096–1120
- Freeman LC (1977) A set of measures of centrality based upon betweenness. *Sociometry* 40:35–41
- Freeman LC, Borgatti SP, White DR (1991) Centrality in valued graphs a measure of betweenness based on network flow. *Soc Networks* 13:141–154
- Guo D, Liang M, Wang L (2010) Betweenness centrality of an edge in tree-like components with finite size. *J Phys A: Math Theor* 43:485003
- Haruna T (2013a) Robustness and directed structures in ecological flow networks. In: Liò P, Miglino O, Nicosia G, Nolfi S, Pavone M (eds). *Advances in Artificial Life, ECAL 2013, Proceedings of the Twelfth European Conference on the Synthesis and Simulation of Living Systems*. MIT Press. pp 175–181
- Haruna, T (2013b) Theory of interface: Category theory, directed networks and evolution of biological networks. *BioSystems* 114:125–148
- Haruna T, Gunji YP (2007) Duality between decomposition and gluing: A theoretical biology via adjoint functors. *BioSystems* 90:716–727
- He S, Li S, Ma H (2009) Betweenness centrality in finite components of complex networks. *Physica A* 388:4277–4285
- Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. *J ACM* 46:604–632
- MacLane S (1998) *Categories for the Working Mathematician*, 2nd edition. Springer-Verlag, New York
- Newman MEJ (2001) Scientific collaboration networks. II. shortest paths, weighted networks, and centrality. *Phys Rev E* 64:016132
- Newman, MEJ (2005) A measure of betweenness centrality based on random walks. *Social Networks* 27:39–54
- Newman MEJ (2010) *Networks: An Introduction*. Oxford University Press Inc., New York
- Piraveenan M, Prokopenko M, Hossain L (2013) Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks. *PLoS ONE* 8:e53095
- Pultr A (1979) On linear representations of graphs. In: *Fundamentals of computation theory (Proc. Conf. Algebraic, Arith. And Categorical Methods in Comput. Theory, Berlin/Wendisch-Riets, 1979)*, Math. Res. 2. pp 362–369
- Spivak DI (2013) The operad of wiring diagrams: Formalizing a graphical language for databases, recursion, and plug-and-play circuits. arXiv:1305.0297
- Spivak, DI (2014) *Category Theory for the Sciences*. The MIT Press
- Ulanowicz R (2002) A sample data collection. <https://www.cbl.umces.edu/ulan/networks.html>. Accessed 14 Feb 2018
- Ulanowicz R, Bondavalli C, Egnotovitch M (1998) Network analysis of trophic dynamics in south Florida ecosystem, FY 97: The Florida bay ecosystem. Ref No [UMCES]CBL-98-123. Chesapeake Biological Laboratory, Solomons