MA4M4: Bow-tie Decomposition of Vaccination Views in Social Media

Yueting Han

Mathematics for Real-World Systems CDT, University of Warwick

This manuscript was compiled on May 3, 2022

The bow-tie structure refers to a core-periphery structure of directed network, which has been used to describe the World Wide Web and biological networks. Recent literature focuses on its application in interpreting the social behavior of discursive communities, i.e. groups of people sharing some same identity, by decomposing their respective social networks into bow-tie structures. Based on the current literature, we develop a systematic framework for multi-scale bow-tie decomposition, including its definition, algorithms for detection and evaluation methods. Its application on the online social network is highlighted in particular. By implementing this approach on the polarized vaccination views in Facebook, we demonstrate that bow-tie decomposition can well explain the behavior of pro-, anti- and undecided vaccination view holders.

Online Social Network | Bow-tie Structure | Community Detection | Directed Network

1. Introduction

The concept of bow-tie structure was first proposed by A. Broder et al. in 2000 (1) to explore the structure of the directed networks, especially for the ones with large strongly connected component such as the World Wide Web. It consists of the strongly connected component(SCC), in-periphery component(IN) containing all nodes with a directed path to a node in SCC, out-periphery component(OUT) containing all nodes with a directed path from a node in the SCC, and other sets (OTHERS) for all the remaining nodes (2). This structure was later refined by Yang (3) through the introduction of TUBES, INTENDRILS and OUTTENDRILS as is displayed in Figure 1a. She also proved that any directed graph can be decomposed into bow-tie structure. Algorithms for bow-tie decomposition rely on either depth-first-search or eigenvector centrality (3, 4). Extensions allowing for recursive bow-tie decomposition are based on communities or different choices of SCC (3, 5, 6). Relevant applications include internet, biological and online social networks (OSN), such as Google web data (4), microbe metabolic networks (7, 8), Twitter social issues debate (9).

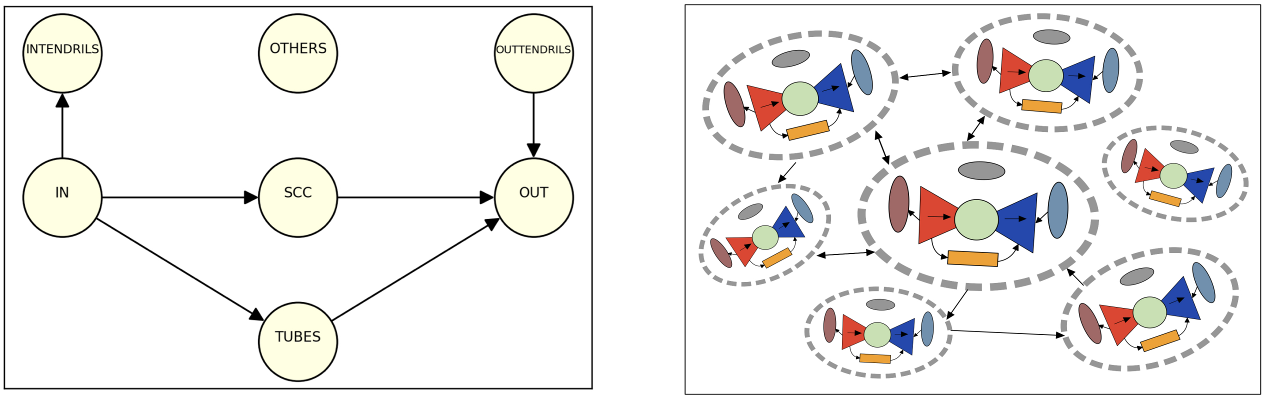
Even though the ubiquity of informative bow-tie structures in diverse types of networks were well demonstrated in these literature, the underlying meaning that can be interpreted from the bow-tie structure is not usually well explained. As a semi-hierarchical core-periphery structure evaluated by Csermely (10), understanding of bow-tie structure is undeniably crucial, especially for social networks. Mattei et al. (9) made a breakthrough by using it to characterize different discursive communities (groups of people who are inferred to share some same identity) in various static OSN via quantitative measures. Nevertheless, interactions between discursive communities were overlooked in this approach. In addition, relevant analysis results have not be statistically evaluated using the updated data set.

In this report, we seek to develop a systematic framework of using bow-tie decomposition for network analysis. It involves the definition and recursive detection for bow-tie structure, as well as its evaluation through bow-tie component proportions and their statistical significance. Especially, its application in online social network will be detailedly demonstrated. Data for investigation describes the online competition of vaccination views on February and October 2019 in Facebook, which was initially analyzed by Johnson et al. (11).

Our report is organized as follows. In section 2, we will introduce definition of bow-tie structure and algorithm for (recursive) bow-tie decomposition. Section 3 mainly focus on the evaluation of bow-tie structure, including empirical criterion for informative and strong/weak bow-tie structure, as well as statistical method to obtain significance of bow-tie component via random network simulation. Section 4 explains the Facebook OSN dataset and relevant background, such as discursive

Significance Statement

Network modelling allows for big data analysis, which can provide a systematic understanding of behavioural ecology in areas such as world wide web, biology and sociology. This report seeks to organize a network-based methodology framework, by considering a network structure called bow-tie. It works well for the analysis of online social behavior, demonstrated by an example of its application about the competition between different vaccination views in Facebook.



(a) Yang's bow-tie structure (9)

(b) Schema of recursive bow-tie structure. Color represents bow-tie component.

Fig. 1. Bow-tie Structure Concept

community, echo chamber and filter bubble. Section 5 contains data analysis based on bow-tie decomposition for Facebook OSN on both Feb and Oct in 2019. Section 6 summarizes our main results and discusses limitations of our research.

2. Definition and Algorithm for Bow-tie Decomposition

We only consider directed and weighted network with neither multi-edge or self-loop here. For a directed network $G = (V, A)$, V is the set for all nodes and $A = (A_{u,v})_{u,v \in V}$ is the weighted adjacency matrix, where $A_{u,v} = w > 0$ if there is an edge from node u to v with weight w and $A_{u,v} = 0$ otherwise.

Ordinary bow-tie decomposition will disregard graph weight. However, we may be able to consider it for recursive bow-tie decomposition based on community detection.

A. Definition of Bow-tie Structure/Decomposition. Here we use the refined definition of bow-tie structure from Yang (3) (See Figure 1a). Assume S is a strongly connected component of G. The bow-tie structure of G with respect to S consists of the following sets of nodes:

$$SCC = S, \quad IN = \{v \in V - S \,|\, \text{S is reachable from } v\}, \quad OUT = \{v \in V - S \,|\, v \text{ is reachable from S}\}$$

$$TUBES = \{v \in V - S - IN - OUT \,|\, v \text{ is reachable from IN and OUT is reachable from } v\}$$

$$INTENDRILS = \{v \in V - S \,|\, v \text{ is reachable from IN and OUT is not reachable from } v\}$$

$$OUTTENDRILS = \{v \in V - S \,|\, v \text{ is not reachable from IN and OUT is reachable from } v\}$$

$$OTHERS = V - S - IN - OUT - TUBES - INTENDRILS - OUTTENDRILS$$

Yang also proved that the sets of bow-tie components above are mutually disjoint and thus form a partition of the nodes. Therefore, we can conclude that any directed graph can be decomposed into bow-tie structure.

B. Algorithm for Bow-tie Decomposition. Let G^T be the transpose graph of G, which has the same nodes as G but the directions of the edges are reversed compared with G. For any $v \in V$, $DFS_G(v)$ denotes the set of nodes that v can reach, which is found by a depth-first search.

Consider the bow-tie decomposition of graph G with respect to the strongly connected component S. It can be computed as follows (3):

1. Set $SCC = S$.
2. Choose any $v \in S$. Then $OUT = DFS_G(v) - S$ and $IN = DFS_{G^T}(v) - S$.
3. For each $v \in (V - S - IN - OUT)$, compute the following two Boolean values:

$$IBV = (IN \cap DFS_{G^T}(v) \neq \emptyset), \quad VRO = (OUT \cap DFS_G(v) \neq \emptyset)$$

   - IRV and VRO $\Rightarrow v \in TUBES$
   - IRV and not VRO $\Rightarrow v \in INTENDRILS$
   - not IRV and VRO $\Rightarrow v \in OUTTENDRILS$
   - not IRV and not VRO $\Rightarrow v \in OTHERS$

The runtime for the algorithm above is $O(|V|^2 + |V||A|)$, which is relatively efficient. Here $|V|$ is the number of nodes in graph G and $|A|$ is the number of edges in graph G. Code for its implementation is provided by this paper (2), available on their GitHub.

B.1. Recursive Bow-tie Decomposition. Current approaches to recursive bow-tie decomposition either rely on different choices of SCC or community detection (3, 5, 6). Using different choices of SCC is easier, but highly dependent on the number and size of SCCs in the empirical graph data. Therefore, combination with community detection as a more general way will be implemented in our report. In this section, we will firstly describe its general framework, and then elucidate the relevant community detection method we choose.

Framework for recursive bow-tie decomposition is simple (see schema in Figure 1b):

1. Perform community detection to assign all the nodes in graph G into k disjoint sets $\{C_1, ..., C_k\}$, where k is the optimal number of communities determined by this algorithm.

2. Construct separate subgraph $G_i = (C_i, A_i)$ for each community C_i, where A_i denote all the edges between C_i nodes in G and $i = 1, ...k$.

3. Decompose each subgraph G_i into bow-tie structure using the same algorithm mentioned before.

The community detection method we choose here applies for directed and weighted networks, first proposed by Rosvall and Bergstorm in 2008 (12). It uses the flow-based random walks on a network as a proxy for information flows, and then optimizes the entropy-based map equation to partition the entire network into communities. More details can be seen in the reference paper above. Algorithm implementation involves Python package CDlib.

The reasons for us to perform this community detection algorithm rather than the traditional modularity maximization (13) are listed as follow:

- It works well for module partition of large-scale network in a hierarchical way (14), which modularity maximization cannot attain. This is the same reason for Fujita et al (5) to choose it in their research of local bow-tie structure detection.

- It may be more preferable compared with modularity maximization in the case where network links represent patterns of movement rather than relationship among nodes (12). The OSN dataset in this report represent dynamic recommendation for Facebook pages, which will be detailed explained in section 4.

3. Evaluation of Bow-tie Structure

As we mentioned before, every directed graph can be represented by bow-tie structure. Therefore, how informative such representation is becomes a key problem. Both empirical and statistical evaluation methods are introduced in this report. They were mostly developed by Mattei et al. (9). Exceptions will be specified.

A. Empirical Criterion. By rule of thumb, bow-tie structure can be evaluated by proportion of each component.

A.1. Informative Bow-tie Structure. The graph G is said to have informative bow-tie structure if it has less than half of nodes in OTHERS; otherwise, the bow-tie structure it has is uninformative.

A.2. Strong/Weak Bow-tie Structure. For an informative graph G, it is said to have strong bow-tie structure if its OTHERS block is smaller than SCC; otherwise, we say it has weak bow-tie structure.

B. Statistical Significance. We would like to distinguish whether the presence of bow-tie structure merely results from the different roles each node plays. For example, nodes with high out-degrees and low in-degrees may well be in the IN sector; by contrast, nodes with low out-degrees and high in-degrees are highly likely to be in the OUT sector. Thus, we will use a benchmark model of random graphs to generate the distribution for the number of nodes in each bow-tie component, and then obtain its statistical significance for real graph data. Notice that originally Mattei et al. measures it by two-tailed p-value without specifying too big or small the component is. That is why we replace it with rank here.

Method skeleton is described below:

1. Run a benchmark model for 1000 times and obtain corresponding 1000 random graphs.

2. Perform bow-tie decomposition on these random graphs and real graph data.

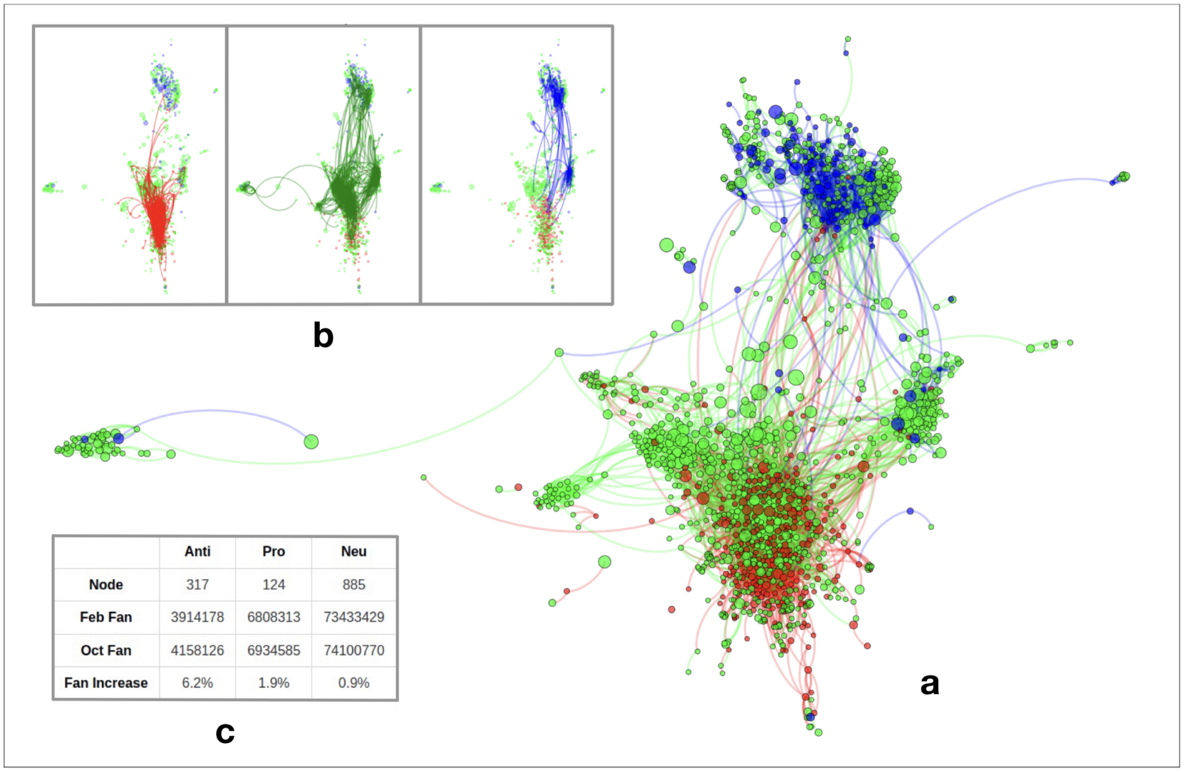
3. For each bow-tie component, sort all the results in an ascending order according to the number of nodes they contain.

4. For each bow-tie component, calculate the rank for real graph data. Either too low or high rank suggests its statistical significance.

B.1. Benchmark Model. An entropy-based direct configuration model was applied by Mattei et al in their research, initially proposed by Mastrandrea et al. (15). Due to its enormous computational complexity, we use Newman's directed configuration model instead (16).

The pipeline for implementation is stated as follows:

1. Make sure that the average in- and out-degree are the same in the observed graph, that is there is no dangling edge.



	Anti	Pro	Neu
Node	317	124	885
Feb Fan	3914178	6808313	73433429
Oct Fan	4158126	6934585	74100770
Fan Increase	6.2%	1.9%	0.9%

Fig. 2. Online ecology of vaccine views. **a**, Reproduced snapshot of the connected component on Feb 2019. Node color: red-anti, blue-pro, green-neu. Node layout: ForceAtlas2. Node size: fan size. Edge color: source node color (11). **b**, the edges that increase during Feb-Oct 2019 with source nodes from different vaccine groups. **c**, node data from Feb to Oct 2019.

2. Regenerate random in/out pairs of edges such that the resulting random graph has the same nodes and in/out degree distribution as the observed graph.
3. Parallel edges and self-loops are removed afterwards.

Algorithm is performed via Python package NetworkX.

4. Data and Relevant Background

Bow-tie decomposition will be applied to analyze the dataset provided by Johnson et al. (11). Before that, dataset and its background will be explained.

A. Data. The dataset describes the snapshots of online competition between different vaccination views on February and October 2019, involving nearly 100 million individuals in Facebook from across countries, continents and languages. They can be represented by two networks with respect to different times in the following ways (see Figure 2):

- Each node is a Facebook page with attribute fan size, that is the number of members who subscribe to the Facebook page, along with attribute polarity including anti-vaccination, pro-vaccination and neutral. Whereas its polarity remains the same for February and October snapshots, its fan size will change.
- An edge from node A to B means A recommends B to all its members at the page level, weighted by the fan size of B as a measure of influence. In addition, each snapshot record the accumulated recommendation data, which means edges existing in February snapshot must appear in October snapshot, but the other way around does not necessarily hold. Thus the other attribute of edge is a boolean value to denote whether it exist in February snapshot.

B. Background. Sociology concepts relevant with our analysis are introduced as below.

B.1. Discursive Community. The term ‘discursive community’ was generated from ‘discourse community’. While discourse community describes a group of people ‘sharing some set of communicative purposes’ (17), discursive community is simply an inference of discourse community by characterize the group members via observed data (9).

In our dataset, we regard users who subscribe to anti-vaccination Facebook pages as a discursive community. Similarly for pro-vaccination and neutral cases.

B.2. Echo Chamber and Filter Bubble. Echo chamber and filter bubble are prominent factors that can lead to polarization in online social media.

Echo chamber can be interpreted as an enclosed media space, where people inside tend to share consistent views and are almost insulated from rebuttal. Influence of messages that agree with their opinions can even by magnified by this phenomenon. Such circumstance can result from confirmation bias, that is people prefer information that confirms their preexisting beliefs (9).

Filter bubble traps people inside an information bubble, which is similar with echo chamber. But filter bubble is result from the personalized content-based recommendations, imposed by social networking platforms such as Ticktock, Instagram, Twitter and Facebook (9).

Although we cannot distinguish the effects of these two factors in this report, we seek to use bow-tie decomposition to assess their overall impact on different vaccine view groups.

5. Data Analysis based on Bow-tie Decomposition

Analysis from Johnson et al. (11) shows that pro-vaccination view holders exhibit more trend of echo chamber/filter bubble compared with anti-vaccination group. We would like to consolidate this outcome in a different way using bow-tie decomposition. Meanwhile other results mentioned in their research may also be spot here.

The general role key bow-tie components play in our context will be illustrated first, followed by explanation for the necessity of statistical evaluation. Then analysis based on multi-scale bow-tie decomposition will be explained, where we focus on the proportion of each bow-tie component with its statistical significance as an auxiliary tool. For each way of bow-tie decomposition, analysis will start with the snapshot in Feb 2019 and then use the one in Oct 2019 to confirm our findings.

A. Roles of Key Bow-tie Component. OUT, SCC and OTHERS are often the sectors with high proportions in our context (see Figure 3), which also play an important role in echo chamber/filter bubble impact. Thus we will detailedly explain their effect using this reference (9). Notice that these interpretations may hold better for bow-tie structure of discursive communities, since this is the study focus in that reference.

A.1. OUT. Facebook pages (nodes) in OUT sector are recommended by pages in SCC, OUTTENDRILS and TUBES, whose influence will especially magnified by SCC, dependent on how large the SCC is as well. In this sector, new and popular content tend to be produced if the network displays an informative bow-tie structure.

We could say an OUT-dominate bow-tie structure is more democratic, open to information, but meanwhile prone to misinformation. By contrast, a bow-tie structure with small OUT sector is likely to indicate echo chamber/filter bubble trend.

A.2. SCC. Pages in SCC sectors are frequently recommended by each other and pages in IN sector. Trustworthy though these pages appear, the quality of their content is relatively low, which means they may simply repeat what other pages mentioned before.

Notice that for a bow-tie structure with small OUT sector, its echo/filter bubble impact tends to be magnified by large SCC.

A.3. OTHERS. For networks with uninformative bow-tie structure, users may be less involved in the vaccine topics such that OTHERS dominates the overall graph.

B. Necessity of Statistical Evaluation. As we mentioned before, the implementation of the statistical evaluation method is to distinguish whether the presence of bow-tie structure is mostly attributed to degree distribution. The meaning of applying it in our case is stated as follows. For example, the Facebook pages of bloggers as nodes can naturally have low in-degrees and high out-degrees, thus they are likely to be divided into OUT. But if they only repeats ideas from other people without his/her own novel points, there is a fairly high chance for them to be included in SCC in reality. If such events happen a lot, then our SCC will be uncommonly large, leading to echo chamber/filter bubble phenomenon. That is the case we would like to investigate in particular.

C. Analysis. We initially apply the analysis approach by Mattei (9) for bow-tie decomposition based on discursive community. Then we extend it by considering interactions between discursive communities (about 40% of all recommendations), as well as recursive bow-tie detection for finer-scale study. Results are depicted in Figure 3.

C.1. Bow-tie Decomposition based on Discursive Community. Each discursive community is extracted from the entire graph as an independent subgraph and then decomposed by bow-tie structure. In this process, edges between different discursive communities are disregarded.

By observation of Figure 3a, all three discursive communities have informative bow-tie structures. Nevertheless, anti-vaccination group demonstrates stronger bow-tie structure, with both SCC and OUT dominate its bow-tie structure. It implies that this group may be active in popular content introduction, mentioned by Johnson et al. as 'good at telling contagious stories'. Therefore, this group is vulnerable when facing misinformation. The node number in its OTHER sector ranks fairly high among simulated random graphs, that is probably because anti-vaccine view holder is distributed more geographically global (11).

For comparison, pro-vaccination group has larger SCC and smaller OUT. It suggests relatively low efficiency and quality of content production, thus may exhibit echo chamber/filter bubble trend. Such feature makes them 'have the wrong impression of winning' according to Johnson et al. But they can also suffer less from misinformation in this case. Neutral group has larger OTHERS, appearing less concerned about vaccination.

All observations above seem to hold in general when it comes to the snapshot in October 2019. Moreover, the change of rank in OTHERS displays that both pro-vaccination and neutral views appear to be gradually marginalized. It confirms the finding from Johnson et al. that anti-vaccination views grow prosperous.

C.2. (Recursive) Bow-tie Decomposition for the Entire Network. To consider recommendation of pages among different discursive communities, we apply bow-tie decomposition on the entire graph. Since we would like to avoid the difference between node numbers of each group to impede our analysis (see Figure 2c), this bow-tie decomposition is regarded as a partition for each group, calculating the percentage each section holds.

By Figure 3b, all three discursive communities exhibit this trend: SCC > OUT > OTHERS, which can be classified as strong bow-tie structures. The emergence of large SCCs can be explained as a result of considering more edges. Although less difference may be observed compared with Figure 3a, we can still see that pages with anti-vaccination and neutral views are more involved in OUT section. While such trend for neutral opinions can be more attributed to the higher degrees their nodes hold, anti-vaccination group demonstrates more initiative by checking the rank colors.

For finer-scale study, community detection is introduced to partition the entire graph and then the main community is extracted for bow-tie decomposition. See Figure 3c for results. In this way, the distribution of bow-tie component in anti-vaccination and neutral appears more different. While SCC is slightly larger than OUT in anti-vaccination, neutral holds the opposite trend. It seems that neutral pages are more engaged in producing content, which was verified by Johnson et al. as well. Pro-vaccination only contains 5 nodes, yet none of them gets involved in SCC. They mostly concentrate in IN and OUTTENDRILS, thus acting more like listeners in this community.

It is worth noticing that the bow-tie component proportion of October snapshot varies less compared with the discursive community case. Thus most findings above can still hold. In addition, we may say the network possesses more stable bow-tie structure in this case.

6. Summary and Limitation Discussion

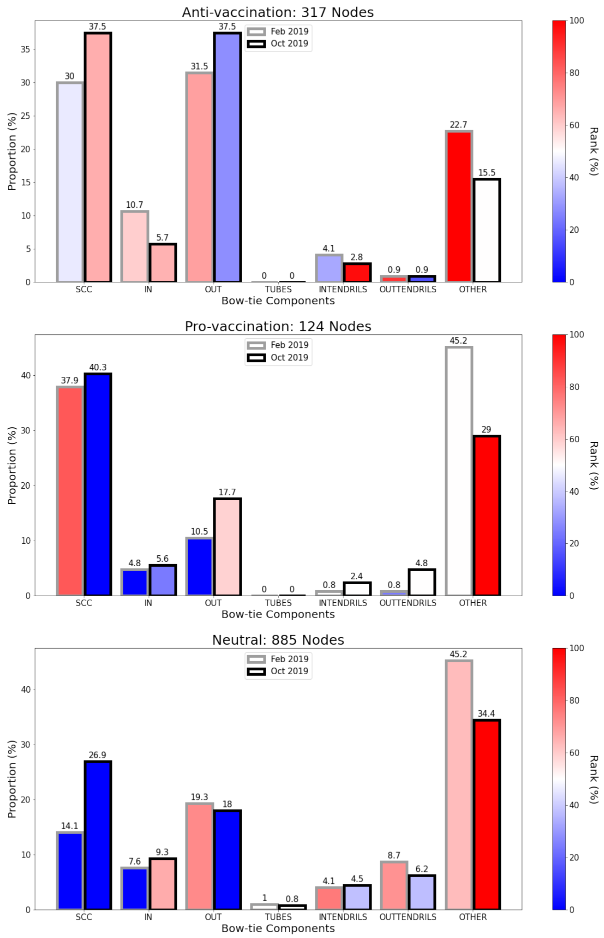
To summarize, a systematic framework is organized for bow-tie structure analysis. Its application in online social network is illustrated in particular. As to the framework, we introduce the bow-tie structure definition refined by Yang (3). The algorithms of recursive bow-tie detection based on communities by reference (3, 5) are also elaborated. The evaluation of bow-tie structure is highlighted by considering bow-tie component proportions and their statistical significance, first proposed by Mattei et al (9). As to its application in online social network, roles of key bow-tie components are explained according to Mattei. Then we apply bow-tie decomposition for Facebook vaccine view data, where multi-scale bow-tie decomposition is performed for comprehensive analysis. This includes discursive community based, entire network based, and local community based. In this way, we demonstrate that bow-tie structure can be applied to interpret online behavior of different vaccine views.

Finally, there are two limitations of our report that should be pointed out. When we discuss the statistical significance of bow-tie component, we choose Newman's directed configuration model as our benchmark without checking its robustness. According to Mattei, this method may also be biased concerning information entropy. The other problem is that we extend the interpretations of bow-tie component based on discursive communities to the entire graph, considering the interactions between discursive communities. But currently there is no empirical evidence to support such hypothesis. Regarding these limitations, further research is needed for investigation.

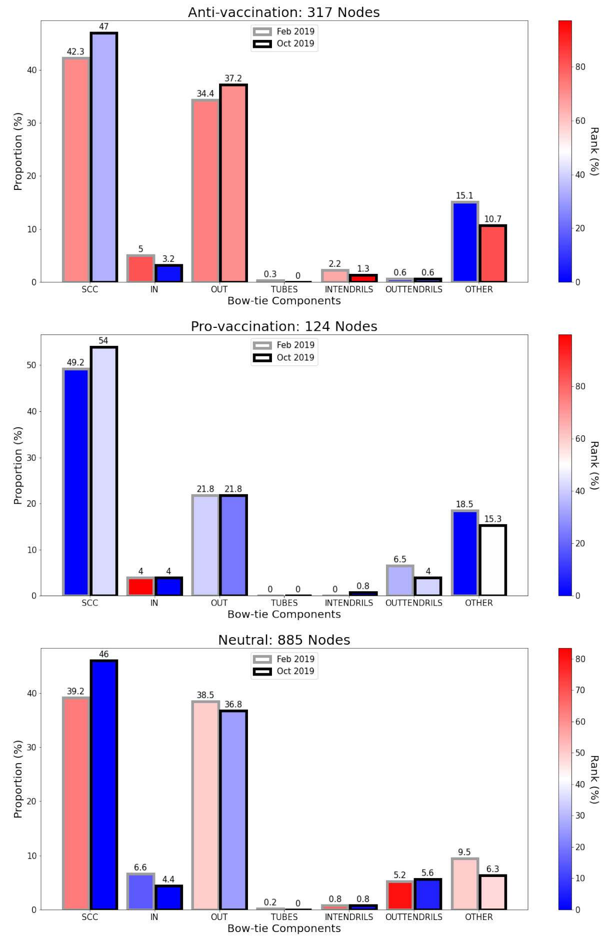
Supplementary Information

This report uses Python for programming. Code for method implementation and data analysis is available on my GitHub.

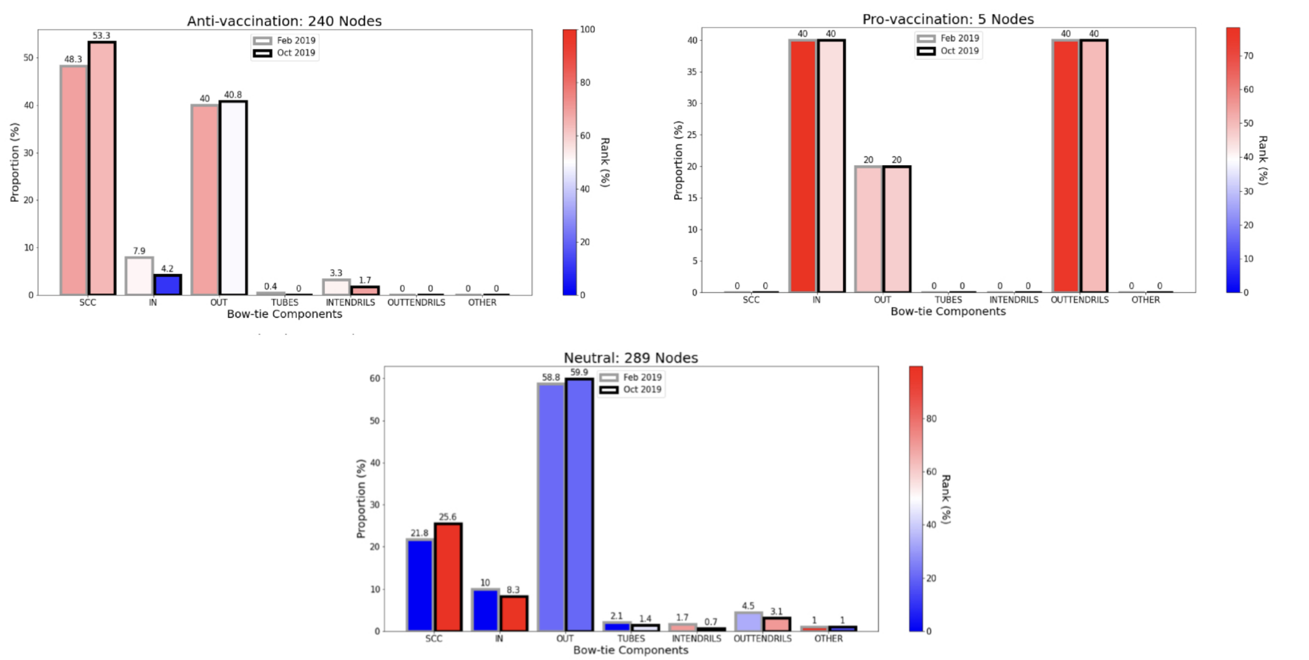
1. A Broder, et al., Graph structure in the web. *Comput. Networks* **33**, 309–320 (2000).
2. A Elliott, A Chiu, M Bazzi, G Reinert, M Cucuringu, Core–periphery structure in directed networks. *Comput. Networks* **476**, 20190783 (2020).
3. R Yang, L Zhuhadar, O Nasraoui, Bow-tie decomposition in directed graphs in *14th International Conference on Information Fusion.* pp. 1–5 (2011).
4. NP Lu, Using eigenvectors of perturbed and collapsed adjacency matrices to explore bowtie structures in directed networks. *J. Chin. Inst. Eng.* **39**, 936–945 (2016).
5. Y Fujita, Y Kichikawa, Y Fujiwara, W Souma, H Iyetomi, Local bow-tie structure of the web. *Appl. Netw. Sci.* **4** (2019).
6. D Chakrabarti, C Faloutsos, Graph mining: Laws, generators, and algorithms. *ACM Comput. Surv.* **38**, 2–es (2006).
7. J Zhao, H Yu, J Luo, ZW Cao, YX Li, Hierarchical modularity of nested bow-ties in metabolic networks. *BMC Bioinform.* **7**, 386 (2006).
8. HW Ma, AP Zeng, The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* **19**, 1423–1430 (2003).
9. M Mattei, M Pratelli, G Caldarelli, M Petrocchi, F Saracco, Bow-tie structures of twitter discursive communities (2022).
10. P Csermely, A London, LY Wu, B Uzzi, Structure and dynamics of core/periphery networks. *J. Complex Networks* **1**, 93–123 (2013).
11. VNRNea Johnson, N.F., The online competition between pro- and anti-vaccination views. *Nature* pp. 230–233 (2020).
12. M Rosvall, CT Bergstrom, Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105**, 1118–1123 (2008).
13. EA Leicht, MEJ Newman, Community structure in directed networks. *Phys. Rev. Lett.* **100** (2008).
14. M Rosvall, CT Bergstrom, Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLOS ONE* **6**, 1–10 (2011).
15. R Mastrandrea, T Squartini, G Fagiolo, D Garlaschelli, Enhanced reconstruction of weighted networks from strengths and degrees. *New J. Phys.* **16**, 043022 (2014).
16. MEJ Newman, SH Strogatz, DJ Watts, Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64** (2001).
17. M Kehus, K Walters, M Shaw, Definition and genesis of an online discourse community. *Int. J. Learn.* **17**, 67–85 (2010).



(a) Discursive community



(b) Entire graph



(c) Community detection

Fig. 3. Bow-tie decomposition based on discursive community, entire graph and detected community. Each bar denotes the proportion of each bow-tie component, whose face color represents its rank among simulation results and edge color represents the date(Feb/Oct) of the network snapshot.