Structural Hole Spanners Detection in Directed Social Networks: A Feed Forward Loop Motif Approach

Arindam Khanda*, Satyaki Roy†, Prithwiraj Roy‡, and Sajal K. Das*
* Department of Computer Science, Missouri University of Science and Technology, USA
† Department of Mathematical Sciences, University of Alabama in Huntsville, USA
‡ Global Action Alliance, USA
Email: *{akkcm, sdas}@mst.edu, †sr0215@uah.edu, ‡proy@globalactionalliance.net

GLOBECOM 2024 - 2024 IEEE Global Communications Conference | 979-8-3503-5125-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/GLOBECOM52923.2024.10901271

***Abstract*—Structural hole spanners (SHSs) are nodes that connect different communities to facilitate efficient information dissemination in complex networks. Existing efforts to identify SHS nodes have predominantly focused on undirected networks, rendering them inadequate to capture directional data flow. This paper presents a novel lightweight approach to motif span scores, called *mSpan* that leverages network substructures called *feed forward loop* (FFL) motifs, to detect SHS in directed, weighted as well as unweighted social networks. The proposed approach measures the spanning score of a node in terms of its participation in FFL motifs that bridge network communities. Our theoretical analysis establishes a strong association between the variants of the scores for a given node and the likelihood of its removal disrupting connectivity. We also utilize *mSpan* to detect spanner motifs that bridge the structural holes in social networks. We validate the efficacy of *mSpan* in detecting SHS in practical scenarios through comparative evaluations of three real-world social networks against existing spanner detection metrics.**

***Index Terms*—Spanners, Feed Forward Loop, Social Network**

I. INTRODUCTION

Analysis of social networks helps understand relationships among individuals and groups in complex networks. These networks are characterized by the small-world property, power-law degree distribution, homophily, and community structure [1]–[4]. *Structural hole spanners* (SHSs) in social networks play a crucial role in forming connections and bridges among communities. According to the structural hole (SH) theory [5], holes emerge when groups of individuals are disconnected from one another in social networks. Thus, SHS nodes acting as bridges or intermediaries between unconnected communities fill the holes [6]. Due to their unique positioning within social networks, these spanners have access to a rich diversity of information. This implies the importance of detecting SHSs in social and organizational applications for information diffusion, opinions, and rumor spreading.

The weight or *strength* of the links connecting individuals in a social network influences their structural properties as well as information flow dynamics [7]. While strong ties capture close interaction among entities within a social network community, they may limit individuals to acquire local information from shared common acquaintances, leaving little scope for the exchange of unknown information. The role of weak ties in connecting distant clusters, thereby tapping into fresh ideas, is explained in [8]. According to [5], a social group stands to benefit the most by pinpointing the presence of holes and bridging them to facilitate non-redundant information flow among myriad tightly networked clusters. This necessitates a combination of strong and weak ties to strike the balance of stability and innovation in networked systems.

While existing literature on SHS detection deals with undirected networks, the majority of interactions in complex social systems are often directional, implying the derived social networks are directed. have the following limitations. Since SHS detection is an NP-hard problem [9], exact (deterministic) algorithms are computationally expensive and do not scale to large networks. Moreover, the link weights in social networks contain key information (e.g., trust, nature, extent of the relationship) shared between the entities. Our work addresses a research gap in current solutions by incorporating the combined effect of direction and weights in SHS detection.

Contributions. To the best of our knowledge, this paper is the first to propose a lightweight node (vertex) scoring technique to find the top SHSs in weighted, directed social networks. Our novel contributions are as follows.

- We leverage the recurrent substructures in directed social networks, termed *network motifs*, for SHS identification. The proposed approach is called motif span or *mSpan*. It comprises two vertex scoring metrics (Z_1 and Z_2) that quantify the likelihood of a node being an SHS based on its participation in a 3-node *feed forward loop* (FFL) motif configuration of strong and weak ties capturing the ability to span network communities.
- We extend the concept of a spanner vertex to a spanner motif, thus developing a scoring technique where a high-scoring FFL effectively functions as a spanner.
- We theoretically analyze the proposed motif-based scoring approach. Experimental evaluation of *mSpan* demonstrates its efficacy against state-of-the-art baseline metrics on three real-world social network datasets.

The paper is organized as follows. Section II gives an overview of existing works related to SHS detection and FFL motifs. Section III introduces the background necessary for *mSpan* approach to identifying SHS nodes and motifs, including a theoretical analysis. Section IV experimentally evaluates the approach. The final section concludes the paper.

II. RELATED WORK

This section reviews existing literature on SHS detection. It also covers studies on network motifs, as *mSpan* relies on it.

A. Structural Hole Spanners

The existing SHS detection approaches can be grouped into two classes. The first class centers around information dissemination, identifying key nodes crucial for maintaining information flow across various communities within the network [6], [10], [11]. In [11], the authors state that SHSs typically connect with core members of different communities and propose a kernel detection method for their identification. In [10], a link prediction method is introduced to identify SHSs by predicting the types of social relationships. Both these methods [10], [11] require community labels of nodes in advance. Lou *et al.* proposed two methods, HIS and MaxD, to identify top-K SHSs by considering the importance score of nodes and maximizing the drop of minimal cut, respectively. These approaches outperform community kernel detection and link prediction-based methods [10], [11].

The second class of approaches leverages centrality-based metrics. These algorithms identify SHSs by assessing the positions and significance of the nodes within social networks [8], [12], [13]. Song *et al.* developed a heuristic based on weak tie theory [12], while Rezvani *et al.* employed inverse closeness centrality to identify the top-K SHSs [13].

However, all existing approaches consider only undirected networks for SHS detection, neglecting edge weights that represent social aspects such as trust and relationship strength. Our work addresses these gaps in the current state-of-the-art.

B. Network Motifs

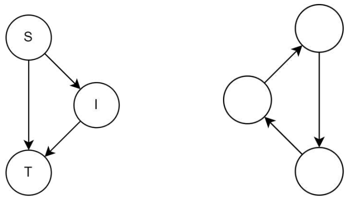
Network motifs are recurring patterns within complex networks that play a fundamental role in facilitating information flow across network types, including neuronal, protein-protein interaction, and social networks [14]. Motifs, namely, *dense overlapping regulons*, *bi fans*, *single input modules*, and *auto-regulation*, are crucial for combinatorial decision-making, signal integration, and sequential gene expression. They have been observed in various network types, including social networks derived from email communication and neuronal networks involved in memory formation and complex dynamics [15], [16]. Also, the role of motifs in optimizing network robustness and stability is an active research area [17]. Motif aggregation has been proposed as a means to typify complex networks, while motif participation serves as a criterion for selecting links to achieve immunization in social networks [18], [19]. Motif-based analysis has helped assess the robustness of power-grid networks against adversarial attacks and improve network resilience [20]. Additionally, studies have explored the role of motifs like the feed-forward loop (FFL) in improving performance objectives in communication networks, highlighting the practical implications of motif-based approaches in network optimization [21]–[23]. In this work, we propose a motif-based approach for detecting SHS.

III. PROPOSED APPROACH: *mSpan*

This section covers the preliminary concepts and proposed motif-based spanner detection approach, *mSpan*.

A. Preliminaries

1) Social Network: Let $G(V, E)$ be a directed, weighted social network, where each link $(u, v) \in E$ has a weight $W((u, v)) = \{+, -\}$ denoting strong and weak ties, respectively. If G is unweighted, then it is converted into a weighted graph by treating existing links as strong ties and nonexistent links as weak ties. From a communication standpoint in a social network, a strong tie serves as a primary channel for reliable, high-bandwidth information flow, while a weak tie represents a less reliable channel with lower bandwidth.



(a) Feed Forward Loop (b) Feedback Loop

Fig. 1: Closed 3-node motifs: Feed forward loop – S is the master regulator, I is the intermediate regulator, and T is the regulated node; and feedback loop.

2) Three-node triads: A feed forward loop (FFL) is a network motif with three nodes marked S, I, and T. As shown in Figure 1a, there is a direct link from S to T as well as an indirect path via an intermediate node I [24]. Unlike an FFL, which is an acyclic triangle, a feedback loop (FBL) is another 3-node motif that forms a cyclic triangle (see Figure 1b).

3) Community detection: A network community is a subset of nodes within a graph that are more densely connected to each other than to the rest of the network. We use Louvain community detection to find modules within G, attaching a community ID to each node $u \in V$. The best community configuration selected by maximizing *network modularity*, measuring the density of links inside communities compared to links between them, as:

$$Q = \frac{1}{2 \times |E|} \times \sum_{u,v} [A_{u,v} - \frac{d(u) \times d(v)}{2 \times |E|}] \delta(C(u), C(v)) \quad (1)$$

Here, A is the adjacency matrix of the undirected version of G, where $A_{u,v} = 1$ if u to v share a link, and 0 otherwise; $d(u)$ is the total degree (i.e., sum of in- and out-degree of node u; $C(u)$ is the community ID of node u; and $\delta(x, y)$ is 1 if $x = y$, and 0 otherwise.

B. Feed Forward Loop-based SHS Detection

We leverage the feed forward loop (FFL) motifs for structural hole spanner (SHS) detection in directed and weighted social networks. The following are the reasons behind the choice of FFL motifs in SHS identification.

- FFLs are one of the most frequent 3-node motifs in directed complex networks [25], significantly more abundant than the feedback loop (FBL) as reported later in Tables I. They are the building blocks of directed networks, making them ideal for analyzing the connectivity and information flow characteristics of networks.
- FFLs have two directional paths for unidirectional data flow from source S to target T (see Figure 1a). A node participating in FFLs spanning two communities can act as a conduit for one-way communication between them.
- FFLs can be used to model the presence of a weak tie or the absence of a link (shown in dotted lines in Figure 2) between node pairs. This makes them generalizable to analyze weighted or unweighted directed complex networks and their inter-community information flow.

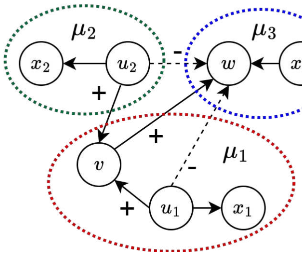


Fig. 2: Illustrative example of valid motif configuration.

The reasons behind the choice of FFLs are captured in an illustrative example. Consider a directed social network with 3 communities μ_1, μ_2, μ_3 (shown in red, green, and blue dotted lines), where the solid and dotted lines represent strong ties $+$ and weak ties $-$, respectively (see Figure 2). Node v participates in FFL motifs, (u_1, v, w), (u_2, v, w), each having $W((u_i, v)) = +$, $W((v, w)) = +$, $W((u_i, w)) = -$ (for $i = 1, 2$). It spans distinct community pairs (μ_1, μ_2) and (μ_1, μ_3), facilitating directional information flow node x_1 to x_3 belonging to those communities μ_1 and μ_2, respectively.

1) Valid Motif Configurations: It measures the number of instances of FFL motif triads (u, v, w) spanning different communities in a directed social network G that node v participates in.

$$Z_1(v) = \{(u, v, w) : \alpha(G, m) \times \beta(G, C, m) = 1 \\ \forall m \equiv (u, v, w) \in M \;\&\; u, w \in V\} \tag{2}$$

Here $\alpha(G, m) = 1$ if $W((u, v)) = +$ & $W((v, w)) = +$ & $W((u, w)) = -$, and 0 otherwise; $\beta(G, C, m) = 1$ if $C(u) \neq C(w)$, and 0 otherwise.

For instance, in Figure 2, node v is part of $Z_1(v) = 2$ FFLs (u_1, v, w) and (u_2, v, w) bridging distinct community pairs (u_1, u_3) and (u_2, u_3), respectively. Both FFLs are characterized by weak ties (u_1, w) and (u_2, w), shown in dotted lines, suggesting the absence of a strong connection (or no connection at all) between the communities.

$Z_1(v)$ computes the total connections between different communities $C(u)$ and $C(w)$ for all valid motifs $m \equiv (u, v, w)$ through vertex v. However, $Z_1(v)$ fails to measure the diversity of the communities connected through v. For example, it does not detect whether $C(u_1) = C(u_2)$ (or $C(w_1) = C(w_2)$) for two valid motifs (u_1, v, w) and (u_2, v, w) (or (u, v, w_1) and (u, v, w_2), respectively).

2) Vertex Score (Z_2): The quality of SHS vertices can be measured by its ability to connect diverse communities. Utilizing entropy as a measure of diversity in community participation, we propose a scoring mechanism (Z_2) for SHS detection. Additionally, in Z_2, we normalize Z_1 by the degree of node v, assigning a higher score to nodes that participate in many valid motif configurations, despite having low connectivity. The score is computed as follows.

$$Z_2(v) = \frac{|Z_1(v)|}{d_{in}(v) \times d_{out}(v)} \times H(\Gamma(v)) \tag{3}$$

Here, $H(\Gamma(v)) = \sum_{c \in \gamma(v)} p(c) log(p(c))$ is the entropy of a list (Γ) containing community IDs of the neighbors of node v forming valid FFL configurations. $H(\Gamma(v))$ is high when the probability distribution $p(c)$ of the unique IDs spanned by v is uniform, i.e., the IDs have a comparable presence in Γ.

3) Motif Score ($\mathcal{M}$): We envision that an FFL motif can serve as a single spanner unit. For a given FFL $m \equiv (u, v, w)$, we calculate the spanning score using the simple sum of the Z_1 scores of u, v, and w. This sum inherently captures the connectivity among different communities. However, the method could be improved by incorporating entropy and using a weighted sum, which we plan to explore in future research.

$$\mathcal{M}(m \equiv (u, v, w)) = Z(w) + Z(v) + Z(u) \tag{4}$$

Algorithm 1: Compute Z_1-score

```
Initialize an array $Z_1$ of size $|V|$ with each element 0
for each edge $(u, v) \in E^+$ do
    $itr_u \leftarrow 0$, $itr_v \leftarrow 0$ // Initialize iterators
    $count \leftarrow 0$ // Initialize local count
    while $itr_u < |N_u^-|$ and $itr_v < |N_v^+|$ do
        if $N_u^-(itr_u) < N_v^+(itr_v)$ then
            $itr_u \leftarrow itr_u + 1$
        else if $N_u^-(itr_u) > N_v^+(itr_v)$ then
            $itr_v \leftarrow itr_v + 1$
        else
            $w \leftarrow N_v^+(itr_v)$
            if $C(u) \neq C(w)$ then
                $count \leftarrow count + 1$
            $itr_u \leftarrow itr_u + 1$
            $itr_v \leftarrow itr_v + 1$
    $Z_1(v) \leftarrow Z(v) + count$
```

C. Proposed mSpan Algorithm

Let the set of strong and weak ties of graph G be E^+ and E^- respectively. Let N_u^- and N_u^+ denote the sorted lists of out-neighbor vertices connected to vertex u via weak and strong ties, respectively. Algorithm 1 utilizes these neighbor lists to efficiently compute Z_1-score for all the vertices.

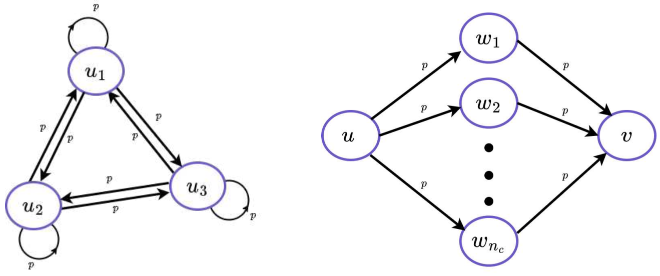
According to Eq. (2), the contribution of a strong tie (u, v) towards the Z_1-score of vertex v depends on the number of common neighbors $w : w \in N_u^- \cap N_v^+$. Therefore, the algorithm processes each edge $(u, v) \in E^+$ and computes the intersection of the two sorted sets N_u^- and N_v^+. This computation can be performed using a merge-like method, requiring at most $|N_u^-|+|N_v^+|$ comparisons. We use two iterators, itr_u and itr_v, that operate on N_u^- and N_v^+, respectively, and find the common neighbor $w = N_u^-(itr_u) = N_v^+(itr_v)$. Algorithm 1, at Line 12, checks the condition for β as specified in Eq. (2).

Complexity Analysis. Let d_{out}^+ and d_{out}^- represent the maximum out-degree of a graph G associated with strong ties and weak ties, respectively. Algorithm 1 visits each element of the lists N_u^- and N_v^+ at most one time to find valid triads and count common vertices w for each strong tie (u, v). In the worst case, this operation takes $O(d_{out}^+ + d_{out}^-)$ time per strong tie. Consequently, computing Z_1 score for all vertices or finding valid motifs across the entire graph requires $O(|E^+| \times (d_{out}^+ + d_{out}^-))$ time.

D. *Expected Spanning Potential of Nodes with High Z_1-Score*

Theorem 1. *The number of node pairs disconnected by removing a given node in a random, directed clustered network is expected to be proportional to its Z_1-score.*

Proof. Let us consider a N-node random, directed, and clustered network $G(V, E)$ with C communities, where each community $c = 1, 2, \cdots, C$ has n_c nodes (s.t. $n_1+n_2+\cdots+n_C = N$). The probability of an edge between any pair of (a) intra-community nodes is p and (b) inter-community nodes ≈ 0.



(a) (b)

Fig. 3: Directed weighted networks: a) 3-node directed complete network G; b) Transition of a random walker from node u to v in 2-hops along links with transition probability p.

Let us generate an infinite set of such N-node networks, each having C communities. Let the subset of nodes participating in each $c = 1, 2, \cdots, C$ be identical across all the networks. Combining all these networks, we get a single network $\mathbf{G}$ where the fraction of times any link (u_1, u_2) (s.t., $u_1, u_2 \in V$ and belongs to the same community) exists is given by p. We term this directed, complete network $\mathbf{G}$ to be the *expected network* (see Figure 3a), where the probability of transitioning from any node to any other is p.

Lemma 2. *The probability of reaching from any node u to v belonging to the same community c in $k > 0$ hops is given by $n_c^{k-1} \times p^k$.*

Proof. We know that the probability of traversing from any node u to any other v within the same community is given by the likelihood of link existence p.

Since $\mathbf{G}$ may have self-loops, it is possible for any random walker to stay at the current node. Thus, the probability of moving from u to v in $k = 2$ hops must be via a node w, where w can equal u and v due to self-loops. As depicted in Figure 3b, $u \to w_i$ and $w_i \to v$ have a transition probability p. Since there are n_c nodes, the probability of possible 2-hop walks equal $\sum_{i=1}^{n_c} p \times p = n_c \times p^2$. Generalizing to k-hop walks, the probability is $n_c^{k-1} \times p^k$.

The above probability can also be derived from the adjacency matrix of $\mathbf{G}$, say $\mathbf{A}$, where each element $\mathbf{A}_{u,v}$ (the probability of transitioning from u to v) is equal to p. Here the probability of a walk from u to v in 2 hops can be inferred from a matrix product: $\mathbf{A}_{u,v}^2 = n_c \times p^2$ (as shown below). Generalizing to k hops we get a probability of an intra-community walk equal to $\mathbf{A}_{u,v}^k = n_c^{k-1} \times p^k$.

$$\begin{bmatrix} p & \cdots & p \\ \vdots & & \\ p & & \end{bmatrix} \times \begin{bmatrix} p & \cdots & p \\ \vdots & & \\ p & & \end{bmatrix} = \begin{bmatrix} n_c \times p^2 & \cdots & n_c \times p^2 \\ \vdots & & \\ n_c \times p^2 & & \end{bmatrix}$$

□

Lemma 3. *The probability of a walk from any node u to v belonging to the same community c is given by $\frac{p}{1-n_c \times p}$ (s.t., $n_c \times p < 1$).*

Proof. From Lemma 2, let $P^k(u, v) = n_c^{k-1} \times p^k$ be the probability of a path from node u to v in k-hops. This makes the total probability of moving from u to v:

$$\mathbf{P}(u, v) = P^1(u, v) + P^2(u, v) + \cdots, +P^\infty(u, v) \tag{5}$$

$$\mathbf{P}(u, v) = p + n_c p^2 + n_c^2 p^3 + \cdots \tag{6}$$

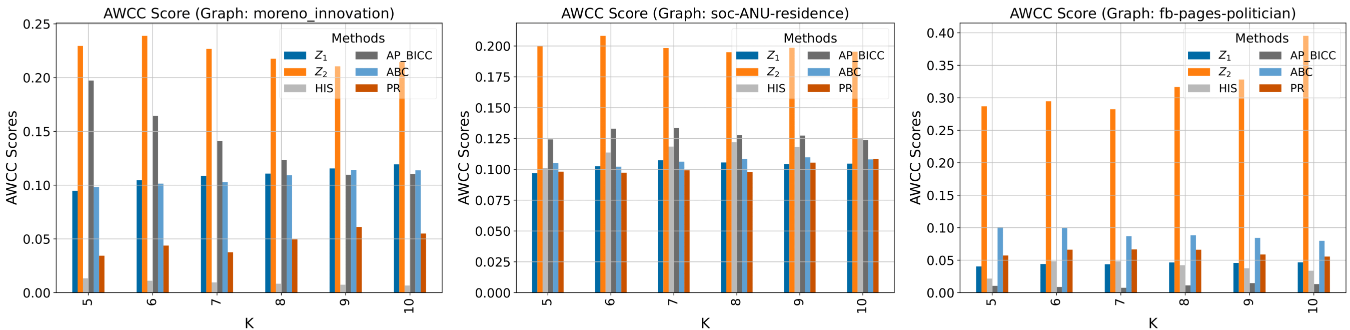
$$\mathbf{P}(u, v) = \frac{p}{1 - n_c \times p} \tag{7}$$

Eq. (7) gives the sum of an infinite geometric progression with the first term p and common ratio $n_c \times p$. The common ratio $n_c \times p < 1$ ensures the convergence of the sum. □

From Eq. (2), the definition of Z_1-score of a node $w \in V$ ($Z_1(w)$) is the number of inter-community node pairs it connects. Since the probability of inter-community link in the expected random clustered network $\mathbf{G} \approx 0$, the number of node pairs disconnected by the removal of w is $Z_1(w) \times \frac{p}{1-n_{\mu(c)} \times p} \times \frac{p}{1-n_{\mu(c)} \times p}$, where $\mu(c)$ is the mean community size in $\mathbf{G}$. Finally, if $\kappa = \frac{p}{1-n_{\mu(c)} \times p} \times \frac{p}{1-n_{\mu(c)} \times p}$, the expected number of node pairs disconnected by the removal of w is $\kappa \times Z_1(w)$, i.e., proportional to its Z_1-score in a random, directed and clustered network. □

TABLE I: Networks

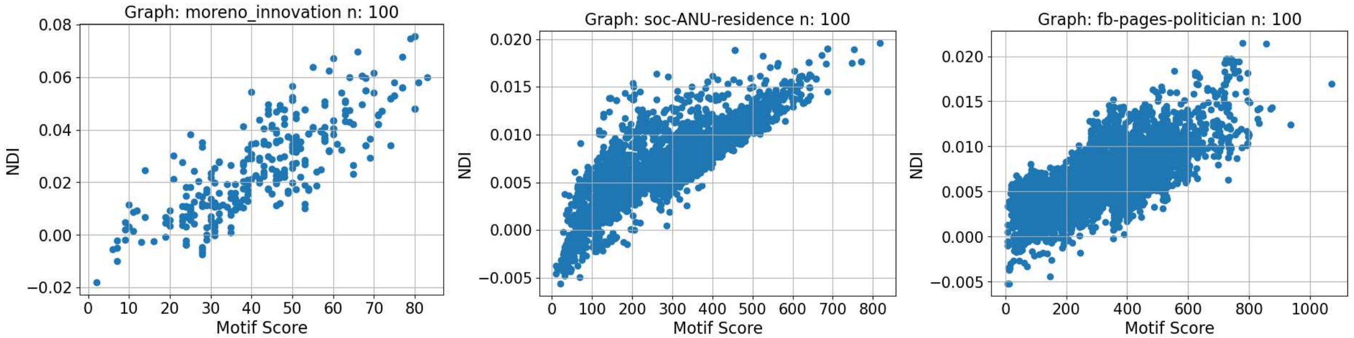
Network	$\lvert V \rvert$	$\lvert E \rvert$	**FFL**	**FBL**
moreno_innovation [26]	241	1098	1144	636
soc-ANU-residence [27]	217	2672	10727	9003
fb-pages-politician [27]	5908	41729	174957	23


AWCC Score (Graph: moreno_innovation)

AWCC Score (Graph: soc-ANU-residence)

AWCC Score (Graph: fb-pages-politician)

Fig. 4: Analysis of different spanning scoring using Average Weighted number of Connected Community (AWCC).


Graph: moreno_innovation n: 100

Graph: soc-ANU-residence n: 100

Graph: fb-pages-politician n: 100

Fig. 5: Motif score ($\mathcal{M}$) vs node disruption index (NDI) that evaluates the extent to which the removal of network motifs with varying robust spanning scores disrupts the shortest communication paths between the social network communities.

IV. EXPERIMENTAL EVALUATION

The experiments were designed using Python on an Intel Core i7 CPU equipped with 12 threads and 16GB DDR4 RAM. Since a valid motif configuration (refer to Sec. III-B1) requires FFLs to span communities, we employ the *Louvain approach* (see Sec. III-A3) to identify network communities as a preprocessing step. The spanner scoring and structural hole spanners (SHS) motif detection have been performed on three real social network datasets. Refer to Table I for the order, size, and a higher relative abundance of FFLs in comparison with another 3-node motif type, FBL (see Sec. III-A2).

We use the following metrics to quantify the likelihood of a set of nodes to be SHS [13], [28].

1) Average Weighted number of Connected Community (AWCC): AWCC of a given set of potential structural holes $\mathcal{S}$ is measured in terms of the set of communities connected to it as $\frac{1}{|\mathcal{S}|}\sum_{v\in\mathcal{S}}\frac{|\zeta(v)|}{d(v)}$, where $\zeta(v)$ is the set of community IDs of neighbors of v.

2) Network Disruption Index (NDI): NDI measures how the shortest communication paths among the communities can be disrupted due to substructure removal. NDI of a substructure, such as an FFL motif $m \equiv (u, v, w)$, is $\mathcal{D}'(G) - \mathcal{D}'(G - m)$, where $\mathcal{D}'(G) = avg.\{\frac{1}{\mathcal{D}(u,v)}\forall(x, y) : x, y \in V, C(x) \neq C(y)\}$, $\mathcal{D}'(G - m) = avg.\{\frac{1}{\mathcal{D}(x,y)}\forall(x, y) : x, y \in \{V - u, v, w\}, C(x) \neq C(y)\}$, and $\mathcal{D}(x, y)$ is the shortest distance between nodes x and y. Specifically, we leverage NDI to capture the intuition that the shortest path length between a pair of communities may increase, or they may potentially be disconnected upon the removal of a network substructure acting as a spanner in a network.

A. Spanning Score for a Set of Nodes

We apply the proposed FFL-based SHS detection technique, *mSpan*, on three social networks. We compute Z_1, and Z_2 scores for each node and select top-K (ranging from 2 to 10) scoring nodes in each category as SHS. We also find top-K SHSs using HIS [6] and AP_BICC [13] methods and top-K nodes with high approximate betweenness centrality (ABC) and PageRank (PR) [29]. Next, we evaluate each such set of top spanners/important vertices using the AWCC metric. Figure 4 depicts a comparison of the AWCC scores for Z_1, Z_2, HIS, AP_BICC, ABC, and PR for varying values of K. We observe that Z_2 consistently achieves the highest AWCC scores showing the effectiveness of our proposed method.

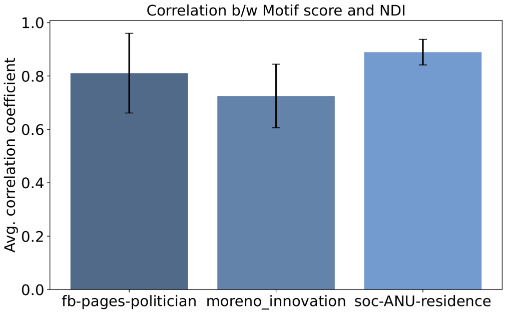

Correlation b/w Motif score and NDI

Fig. 6: Pearson Correlation between motif score and NDI.

B. Spanning Potential of Feed Forward Loop Motifs

In Sec. III-B3, we intuit that a set of nodes constituting a network motif may collectively serve as effective spanners. We examine the spanning potential of FFL motifs using their node disruption index (NDI). It is worth noting that calculating

NDI is a computationally expensive operation since it entails the identification of all pair shortest paths where the vertices in each pair belong to different communities. Therefore, we apply NDI to subnetworks with 50, 100, and 150 vertices, sampled from the networks in Table I. These subnetworks are obtained by randomly selecting seed vertices and their ego networks. Before comparing NDI results, we compute the motif scores using the proposed motif scoring function (Eq. 4).

These experiments aim to demonstrate that the impact of removing an FFL on inter-community connectivity correlates with its motif score. In other words, removing a high-scoring motif is more likely to disrupt the connections between communities in a directed social network. We present the results in Figure 5, where FFLs with varying motif scores are removed one by one, and NDI is computed. We observe a consistent linear relationship between motif scores and NDI.

Finally, we summarize this positive association by estimating the Pearson correlation coefficient (PCC) between $\mathcal{M}$ and NDI in 100-node subnetworks of the same datasets across 15 runs. The high mean PCC (along with error bars) depicted in Figure 6 provide further evidence that *mSpan* offers a useful measure of motif spanning in complex social networks.

V. CONCLUSION

We introduced a network motif-based spanner detection mechanism, *mSpan*, for directed social networks. It comprises metrics measuring the spanning potential of a node in terms of its participation in a type of network motif, termed *Feed Forward Loop* (FFL), that spans different network communities. Our theoretical analysis shows that the spanning score of a node quantifies the number of inter-community node pairs disconnected upon the removal of the node. Furthermore, the comparison with baseline metrics on three network datasets shows that the proposed motif-based metrics capture the spanning potential of nodes and motif substructures. In the future, we plan to extend *mSpan* into a distributed solution that would leverage message passing to identify structural hole spanners without prior knowledge of communities or motif participation of other nodes in large social networks.

Acknowledgments: This work is supported by te NSF projects – CANDY under award # OAC-2104078 and SOTEIRA under award # ECCS-2319995.

REFERENCES

[1] G. Luo, Z. Zhang, and S. Diao, "Empirical analysis and modelling social network user interaction behavior and time characteristics based on selection preference," *Info. Sciences*, vol. 608, pp. 1202–1220, 2022.

[2] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[3] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[4] R. C. Van der Hulst, "Introduction to social network analysis (sna) as an investigative tool," *Trends in Org. Crime*, vol. 12, pp. 101–121, 2009.

[5] R. S. Burt, *Structural holes: The social structure of competition*. Harvard university press, 1995.

[6] T. Lou and J. Tang, "Mining structural hole spanners through information diffusion in social networks," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 825–836.

[7] R. Toivonen, J. Kumpula, J. Saramäki, J. Onnela, J. Kertész, and K. Kaski, "The role of edge weights in social networks: modelling structure and dynamics," in *Noise and Stochastics in Complex Systems and Finance*, vol. 6601. SPIE, 2007, pp. 48–55.

[8] M. S. Granovetter, "The strength of weak ties." *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[9] W. Xu, M. Rezvani, W. Liang, J. X. Yu, and C. Liu, "Efficient algorithms for the identification of top-k structural hole spanners in large social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 5, pp. 1017–1030, 2017.

[10] J. Tang, T. Lou, and J. Kleinberg, "Inferring social ties across heterogenous networks," in *Proceedings of the fifth ACM international conference on Web search and data mining*, 2012, pp. 743–752.

[11] L. Wang, T. Lou, J. Tang, and J. E. Hopcroft, "Detecting community kernels in large social networks," in *2011 IEEE 11th International Conference on Data Mining*. IEEE, 2011, pp. 784–793.

[12] C. Song, W. Hsu, and M. L. Lee, "Mining brokers in dynamic social networks," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 523–532.

[13] M. Rezvani, W. Liang, W. Xu, and C. Liu, "Identifying top-k structural hole spanners in large-scale social networks," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 263–272.

[14] L. Q. Moreno, "Graphlets and motifs in biological networks. encyclopedia of bioinformatics and computational biology," 2019.

[15] K. Juszczyszyn and K. Musial, "Motif analysis and the periodic structural changes in an organizational email-based social network," *International Journal of Virtual Communities and Social Networking (IJVCSN)*, vol. 1, no. 2, pp. 22–35, 2009.

[16] M. Koyama and A. Pujala, "Mutual inhibition of lateral inhibition: a network motif for an elementary computation in the brain," *Current opinion in neurobiology*, vol. 49, pp. 69–74, 2018.

[17] A. Elena *et al.*, "Motifs in regulatory networks and their structural robustness," in *2008 8th IEEE International Conference on BioInformatics and BioEngineering*. IEEE, 2008, pp. 1–6.

[18] K. Gao, G. Yuan, Y. Yang, Y. Fan, and W. Hu, "Mebc: social network immunization via motif-based edge-betweenness centrality," *Knowledge and Information Systems*, vol. 64, no. 5, pp. 1263–1281, 2022.

[19] X. Hong-Lin *et al.*, "Social network analysis based on network motifs," *Journal of Applied Mathematics*, vol. 2014, 2014.

[20] A. Dey, Y. Gel, and V. Poor, "Motif-based analysis of power grid robustness under attacks," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 1015–1019.

[21] S. Roy, M. Raj, P. Ghosh, and S. Das, "Role of motifs in topological robustness of gene regulatory networks," in *IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[22] S. Roy, P. Roy, V. Nadendla, and S. K. Das, "Influence spread control in complex networks via removal of feed forward loops," in *IEEE International Conference on Computer Communications and Networks (ICCCN)*, 2021, pp. 1–9.

[23] U. Roy, S. Roy, R. Khan, P. Ghosh, and N. Ghosh, "MCR: A motif centrality-based distributed message routing for disaster area networks," *IEEE Internet of Things Journal*, 2022.

[24] M. E. Wall, M. J. Dunlop, and W. S. Hlavacek, "Multiple functions of a feed-forward-loop gene circuit," *Journal of molecular biology*, vol. 349, no. 3, pp. 501–514, 2005.

[25] S. Roy, P. Ghosh, N. Ghosh, and S. K. Das, "Transcriptional regulatory network topology with applications to bio-inspired networking: A survey," *ACM Comp. Surveys (CSUR)*, vol. 54, no. 8, pp. 1–36, 2021.

[26] J. Kunegis, "KONECT – The Koblenz Network Collection," in *Proc. Int. Conf. on World Wide Web Companion*, 2013, pp. 1343–1350. [Online]. Available: http://dl.acm.org/citation.cfm?id=2488173

[27] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *AAAI*, 2015. [Online]. Available: https://networkrepository.com

[28] F. Li, Z. Zou, J. Li, Y. Li, and Y. Chen, "Distributed parallel structural hole detection on big graphs," in *Database Systems for Advanced Applications: 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 22–25, 2019, Proceedings, Part I 24*. Springer, 2019, pp. 519–535.

[29] C. L. Staudt, A. Sazonovs, and H. Meyerhenke, "Networkit: A tool suite for large-scale complex network analysis," *Network Science*, vol. 4, no. 4, pp. 508–530, 2016.