The Centrality of Edges Based on their Role in Induced Triads

Lauren Hudson, Roger M. Whitaker, Stuart M. Allen, Liam D. Turner
School of Computer Science & Informatics
Cardiff University,
Cardiff, CF24 3AA, UK
{HudsonL1, WhitakerRM, AllenSM, TurnerL9}@cardiff.ac.uk

Diane Felmlee
Dept of Sociology and Criminology
Penn State University
University Park, PA, 16802, USA
dhf12@psu.edu

Abstract—The prevalence of induced triads play an important role in characterising complex networks, supporting approaches for assessment of dynamic and partially obfuscated scenarios. In this paper we introduce a new local edge-centrality measure that is designed to be deployed in this context for complex networks and is highly scalable. It signifies the importance an edge plays within induced triads for a directed network. We observe that an edge can play one of two roles in providing connectivity within any particular triad, based on whether the edge supports connectivity to the third node or not. We call these alternative states overt and covert. As an edge may play alternative roles in different induced triads, this allows us to assess the local importance of an edge across multiple induced substructures. We introduce theory to count the number of induced triads in which an edge is overt and covert. Using 34 data sets derived from public sources, we show how the presence of overt and covert edges can be used to profile diverse real-world networks. The relationship with global network analysis metrics is examined. We observe that overt and covert edge centrality is useful in further differentiating classes of network, when considered in combination with conventional global network analysis metrics.

Index Terms—centrality, triads, edge-centrality, complex networks, connectivity

I. INTRODUCTION

Triads are important substructures in networks because they are the smallest non-trivial induced substructure after the definition of nodes and edges. Through the concept of the network motif [1], triads have become popular in characterising complex networks (e.g., [2]), with the presence of triads giving insight into the commonality within different classes of network. This has opened up analysis of networks to dynamic and large scenarios where for example, obfuscation may be present. However, less well considered are local approaches to further network characterisation that are solely based on induced triads. These are important because they can be used in tandem with network motif analysis and by definition, they can scale and don't require global network knowledge.

In this paper we focus on only using connected triads to assess the local importance of edges, based on an edges role in providing connectivity within a particular triad. We introduce a new form of edge centrality that is designed to work with a view of a network through the presence of induced triads. It captures the role that an edge plays in providing connectivity relative to a particular triad induced in a directed network. In particular, we note that relative to facilitating a path to the third node, an edge can be characterised in one of two states, which we call *overt* and *covert*.

The presence of an overt edge supports local dissemination (i.e., across a triad), while the reverse is true for a covert edge, thereby allowing us to characterise the local importance of edges to connectivity in the wider network. At the network level, overt edges therefore facilitate contagion while content can be conveyed on a covert edge without making it locally available (i.e., to the third party in a triad). As an edge in a network will typically participate in multiple induced triads, edges may play different roles in different induced triads. This gives a new means to profile the connectivity underlying a network, based on the local role of edges - it allows comparison of networks based on the edge connectivity that their underlying triads provide.

We introduce these concepts and define new forms of centrality to capture the number of induced triads within which each edge of a network is overt or covert. We examine 34 publicly available networks representing a wide range of different scenarios. Overt and covert edge centrality is used to profile them, and to observe the inherent similarities and differences between networks derived from different scenarios. We also consider the relationship between overt and covert centrality, and well known network analysis metrics. The results show that overt and covert centrality play a role in further distinguishing classes of network when considered in combination with global network analysis metrics.

II. RELATED WORK

There are wide-ranging contributions and metrics concerning centrality, and here we highlight a number of those that have relevance to the analysis of networks through induced substructures. The literature is broad and the treatment here is non-exhaustive. Our focus concerns induced triads, which have been

the foundation for assessment of complex networks [1]. The classical definition of the clustering coefficient for a network [3] indicates the number of closed triplets of nodes as a proportion of the total number of triplets, open or closed. This represents a starting point for many generalisations.

For example, Luce and Perry [3] address global clustering in networks by assessing the fraction of closed triples that occur out of all possible triples in an entire network. A triplet is defined as three nodes that satisfy the property of having at least two edges between them. A triplet is defined closed if there are three edges between the nodes, or open otherwise. Alternatively, the authors [4] introduce a local version of clustering coefficient by determining the extent to which a node's neighbours form a clique (i.e., a complete sub-graph). Suppose that a vertex v has k_v neighbours; then at most $k_v(k_v - 1)/2$ edges can exist between them (this occurs when every neighbour of v is connected to every other neighbour of v). Watts and Strogatz calculate the fraction of fraction of these allowable edges that actually exist, thereby measuring the local density of clustering through neighbourhoods.

The starting point for clustering coefficients in directed networks is arguably the concept of triad transitivity. A triad (u, v, w) is transitive if whenever there is a directed edge from u to v and from v to w, then there is a directed edge from u to w [5]. If no edge (u, v) and/or (v, w) exists then a triad is vacuously transitive. Otherwise, it is intransitive. Transitivity is a particularly important in the context of social structures, with clustering metrics defined to this effect such as in [6].

Opsahl et al [7] address clustering in weighted and directed networks by assessing the likelihood of transitive closure of both directed and undirected triads in a network, while taking into account weighted edges. To calculate a generalised global clustering coefficient, C_w, the authors propose

$$C_w = \frac{\text{total weight of closed triplets}}{\text{total weight of triplets}}$$

Similarly to calculate a generalised transitivity measure, T_w, the authors also propose:

$$T_w = \frac{\text{total weight of transitive triplets}}{\text{total weight of triplets}}$$

We note that through this measure, the clustering coefficient assesses the overall triadic structure across the network, rather than the role of the individual edges, which may vary considerably. For example, suppose a triad $T = \{u, v, w\}$ is transitive. Then it is possible to construct a path from u to w. It is not possible to construct a path from v to u, w to u or w to v. Therefore, edge (u, v) plays a significantly different role to (v, w) and (u, w).

Other centrality related metrics focus on defining alternative path-based approaches to measure the importance of edges in different ways. Betweenness centrality [8] is one popular example that considers the number of the shortest paths running through a particular edge. Other alternative approaches considering the role of edges include various forms of network flow (e.g., [9]–[11]). However, global path-based centrality

measures do not scale well, and local approaches to resolve this have received attention such as in [12]. Its also notable that some approaches concerning edges pivot on the inherent relationships held with nodes [13], rather than considering the connectivity that edges fulfil, either locally or globally.

An alternative approach to considering centrality involves taking a complex networks perspective [1], [2], [14], where motif analysis has been effective in characterising networks based on the over representation of particular induced substructures [15]. This approach involves normalising the frequency of induced triads (or other structures) against those which might occur in a sample of random networks, in contrast to computing global metrics for network characterisation and comparison.

Aligned to this network-sampling based perspective, Benson [16] focuses on understanding the higher-order organisation of complex networks through nodes involved in clusters of network motifs based on triads. This brings together motif analysis and network partitioning to reveal organisational patterns at large scale complexity. By focusing on node interactions with induced substructures, Przulj [17] constructs a generalisation of degree distribution from the number of edges being incident with a node to the number of graphlets incident with a node. This is non-trivial due to the number of alternative positions of a particular node within alternative graphlets. Related approaches involve characterising the participation of each node in all subgraphs in a network [18]. More generally, the interplay between non-trivial induced substructures (i.e., beyond edges) has gained further traction as an area of interest in its own right and a comprehensive overview is given in [19].

From the literature, we note that concepts such as subgraph centrality and graphlet degree distribution, as defined for complex networks, are often node-centric. These consider the role of nodes in substructures, rather than the function of the edges themselves within those structures. In contrast we consider all edges in a triad and assess them based on the roles their edges play in connecting to others. Others have defined the characteristics of edges relative to their function in triads, particularly with reference to social interactions [5]. These contributions mainly focus on the importance of particular substructures, rather than the role that an edge may play in *any* given sub-structure, which further motivates the contribution that we make.

III. Defining the role of edges

We assume a graph $G = (V(G), E(G))$ is represented by the set of nodes $V(G)$ and the set of directed edges $E(G)$. For an induced triad T, we refer to its vertices as $V(T) \subseteq V(G)$ and its edges as $E(T) \subseteq E(G)$.
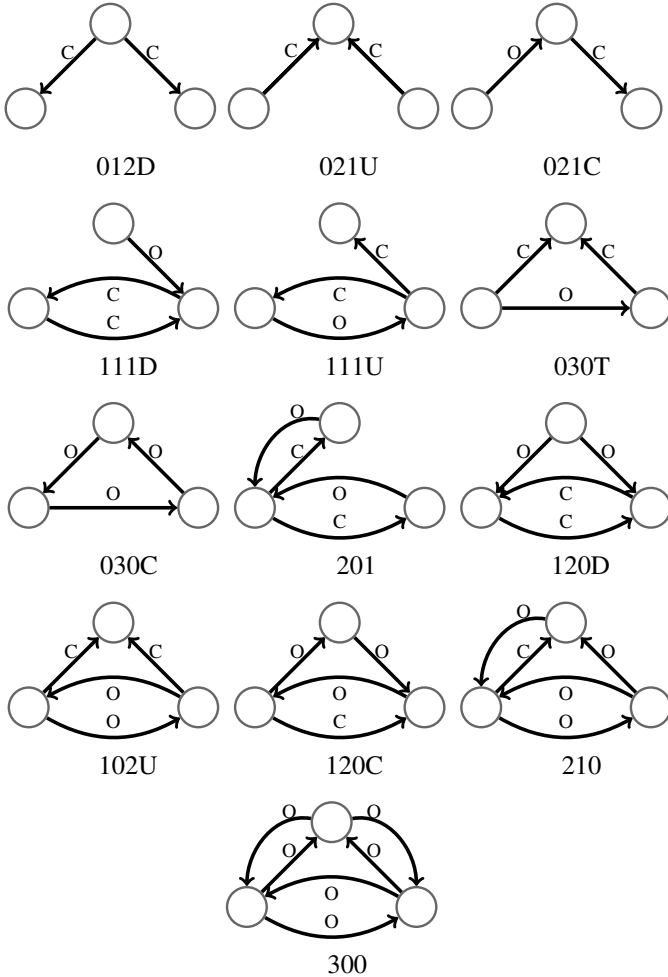
Definition 1: Let T be an induced triad in G with vertex set $V(T) = \{u, v, w\}$ which contains a directed edge from u to v (i.e., $(u, v) \in E(T)$). Edge (u, v) is *overt* with respect to T if and only if $(v, w) \in E(T)$, otherwise edge (u, v) is *covert*.

Note that in a triad involving nodes u, v and w, if (u, v) is an overt edge, dissemination of content from u to v may reach w indirectly from u, via v, i.e., u cannot control whether content reaches w. Conversely, when (u, v) is covert, u is assured that

v can not disseminate content directly to w. Therefore local awareness of the status of edges gives nodes some agency in either the containment of information along a path involving that edge (i.e., the edge (u, v) is covert), or the potential for content to be more widely disseminated across a network due to the use of that edge (i.e., the edge (u, v) is overt).

Figure 1 shows all possible connected triads with overt and covert edges indicated. Note that there is high variability on the presence of overt and covert edges, between triads. As shown by example in Figure 2, an edge may be present in multiple triads and may take different roles in each (i.e., overt and covert). The following propositions present counting arguments to characterise this in general. We use d_v^- and d_v^+ to respectively denote the in-degree and out-degree of a node v. We define three neighbourhoods of each vertex v:

$$
\begin{aligned}
N^-(v) &= \{u \in V(G) : (u, v) \in E(G)\} \\
N^+(v) &= \{u \in V(G) : (v, u) \in E(G)\} \\
N(v) &= N^-(v) \cup N^+(v)
\end{aligned}
$$

012D   021U   021C

111D   111U   030T

030C   201   120D

102U   120C   210

300

Fig. 1: *All connected triad types with overt and covert edges indicated by O and C respectively.*

G
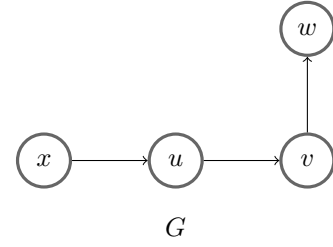
Fig. 2: *A simple network G involving two induced triads: $T_1 = \{x, u, v\}$ and $T_2 = \{u, v, w\}$. Edge (u, v) is involved in both triads T_1 and T_2. (u, v) is covert in T_1 and (u, v) is overt in T_2.*

Proposition 1: Let $(u, v)_p$ denote the number of connected triads in which $(u, v) \in E(G)$ is *present*. Then

$$(u, v)_p = |N(u) \cup N(v) - \{u, v\}|.$$

Proof: Let $A = N(u) \cup N(v) - \{u, v\}$. Suppose $(u, v)_p > |A|$. Then there exists some $q \in V(G)$ such that $\{u, v, w\}$ is a connected triad but $w \notin A$. Then w is not in the neighbourhood of u or v. So either $w \notin \{u, v\}$ or $w \in \{u, v\}$. In both cases $\{u, v, w\}$ is not a connected triad which contradicts the assumptions for w. Hence $(u, v)_p \leq |A|$.

Alternatively suppose $(u, v)_p < |A|$. Then there exists some $w \in A$ such that $\{u, v, w\}$ is not a connected triad. Since $w \in A$ then $w \neq u$ and $w \neq v$ and $w \in N(u) \cup N(v)$. Then this forces $\{u, v, w\}$ to be a connected triad, contradicting the assumptions for w. Hence $(u, v)_p \geq |A|$.

Since $(u, v)_p \leq |A|$ and $(u, v)_p \geq |A|$ then $(u, v)_p = |A|$.

$\square$

Proposition 2: Let $(u, v)_o$ denote the number of triads in which $(u, v) \in E(G)$ is an *overt* edge. Then:

$$(u, v)_o = |N^+(v) - \{u\}|$$

Proof: Let $A = \{w | (v, q) \in E(G) \text{ and } w \neq u, w \neq v\}$. Suppose $(u, v) \in E(G)$ where $(v, u) \notin E(G)$. Then $|A| = d_v^+$.

Suppose $(u, v)_o > d_v^+$. Then $\exists w \in V(G)$ such that $(v, w) \in E(G)$ but $w \notin A$. Then $w = v$ or $w = u$. If $w = v$ then $\{u, v, w\}$ is not a triad hence (u, v) is not overt. Hence $w = u$, but then $(v, u) \in E(G)$, contradicting the assumption that $(v, u) \notin E(G)$. Therefore, $(u, v)_o \leqslant d_v^+$.

Alternatively, suppose $(u, v)_o < d_v^+$. Then $\exists w \in A$ such that $(v, w) \notin E(G)$. This is impossible by definition of A. Hence $(u, v)_o \geqslant d_v^+$.

Since $(u, v)_o \leqslant d_v^+$ and $(u, v)_o \geqslant d_v^+$ then $(u, v)_o = d_v^+$. The proof for $(u, v)_o = d_v^+ - 1$ if $(v, u) \in E(G)$ follows similarly.

$\square$

Proposition 3: Let $(u, v)_c$ denote the number of connected triads in which $(u, v) \in E(G)$ is a *covert* edge. Then:

$$(u, v)_c = |N(u)^+ \cup N(u)^- \cup N(v)^- - N(v)^+ - \{u, v\}|$$

Proof: Let $A = N(u)^+ \cup N(u)^- \cup N(v)^- - N(v)^+ - \{u, v\}$. Suppose that $(u, v)_c \neq |A|$.

Firstly we assume that $(u, v)_c > |A|$. Then there exists some $w \in V(G)$ such that $\{u, v, w\}$ is a connected triad and (u, v) is a covert edge, but $w \notin A$. Hence $w \in N(v)^+$ or $w \in \{u, v\}$. If $w \in \{u, v\}$ then $w = u$ or $w = v$. Then $\{u, v, w\}$ is not a connected triad, contradicting our definition of w. Hence $w \in N(v)^+$. This implies $(v, w) \in E(G)$ so (u, v) is an overt edge in $\{u, v, w\}$, contradicting our definition of w again. Then no such w exists, hence $(u, v)_c \leq |A|$.

Alternatively suppose $(u, v)_c < |A|$. Then there exists some $w \in A$ such that (u, v) is an overt edge in the connected triad $\{u, v, w\}$ or $\{u, v, w\}$ not a connected triad. Since $w \in A$ then $w \in N(u)$ or $N(v)^-$ and $w \neq u$ and $w \neq v$ so $\{u, v, w\}$ must be a connected triad. Then (u, v) must be overt in the connected triad $\{u, v, w\}$. Then $(v, w) \in E(G)$ but since $w \in A$ then $w \notin N(v)^+$ so this contradicts our definition of w. Hence no such w exists so $(u, v)_c \geq |A|$. Since $(u, v)_c \leq |A|$ and $(u, v)_c \geq |A|$ then $(u, v)_c = |A|$.

$\square$

Note that $(u, v)_p = (u, v)_o + (u, v)_c$. We refer to $(u, v)_o$ as the *overt centrality* of edge (u, v), and $(u, v)_c$ refers to the *covert centrality* of edge (u, v).

An edge (u, v) having high overt centrality indicates that in a large number of triads, edge (u, v) contributes to a path from u to the third node w. Conversely, an edge (u, v) having high covert centrality indicates that in a large number of triads, edge (u, v) is not involved in a path from u to the third node w.

Propositions 1, 2 and 3 mean that an edge's overt and covert counts can be quickly computed using the neighbourhood of nodes belonging to the edge of interest. This highly local approach means that in contrast to some alternative edge centrality metrics, computing overt and covert centrality is highly scalable.

IV. EXPERIMENTATION

To understand the prevalence of overt and covert edges and their role in different classes of network, we take a range networks used in other studies (IV-A), listing their attributes in Table I, and assess the presence of overt and covert edges in Table II (Section IV-B). We conclude by exploring the relationship between the mean overt and covert centrality of edges with existing social network analysis metrics from Section IV-A in Figure 3 and 4 (Section IV-C).

A. Data sets

The networks are collated from a variety of sources, including the Cosin Project [20], Stanford's Snap Datasets [21], NeuroData's Graph Database Connectomes [22], Toreopsahal.com dataset collection [23], amongst others [24], [25],

[26], [27], [28]. All networks are directed. They have been organised into nine classes in Table I according to data type alongside various metrics to represent the size, connectivity and directionality and of the network.

We have chosen these data sets as they represent different natural phenomena well established in the literature. There is diversity within the networks, in particular the number of vertices, from very small organisational networks (between 34 and 77 nodes, excluding the Eva network [28]), and much larger internet networks (up to 10,876 nodes). Further, different structural properties are exhibited by different classes of network. For example, organisational networks (except for Eva [28]) present higher levels of density, clustering and a higher mean degree centrality than other networks, indicating high levels of connectivity. They also have high levels of reciprocity and small diameters. We can interpret that organisational networks are formed of collections of tightly knit groups. In contrast, electrical circuit, internet and neurological networks present low levels of density, clustering, reciprocity and a lower mean degree centrality. These networks display hub-and-spoke structures, where a minority of nodes have a very high degree (the 'hubs') and the majority of nodes a very low degree (the 'spokes'). Other networks, such as food web networks or airport networks, sit somewhere in between: for example, food web networks present a mid level of clustering (higher than in internet networks, lower than in organisational networks) and degree centrality but have lower density (as compared to organisational networks).

B. Profiling Real World Data Sets using Overt and Covert Edge Centrality

Because the computation of overt/covert centricity of an edge (u, v) is dependent on the degree of v, this centrality metric is highly scalable and easily applied to real world data sets. To exemplify this we determine the frequency distribution for overt/covert centrality across 34 test networks from different real world scenarios (Table II).

Interestingly we see that in many cases, the resulting frequency distributions are effective in associating networks originating from same context or domain. For example, the airport flights data sets [29], [30] have long tail distributions with a relatively slow drop off as compared to the peer-to-peer Internet data sets [21]. Some classes of network also exhibit greater weighting for one particular type of centrality, such as the electrical circuit scenarios [26] where covert edges show greater prevalence. Other classes of network, such as the organisational networks [23] exhibit distinctive increasing profiles in frequency of overt centrality. The results also reaffirm that overt and covert centrality are not mutually exclusive. For example, similar distributions are obtained for both overt and covert centrality in numerous categories, such as the *Trust/Social Networks* class and the *Airport flight Networks* class.

These results demonstrate that overt centrality and covert centrality offer the potential to characterise complex networks based on the common behaviour of edges with respect to their

TABLE I: *Networks and metrics used to describe them. $|V|$ denotes the number of nodes in the network, $|E|$ the number of edges. Clustering represents the mean global clustering coefficient. Reciprocity is a measure of global reciprocity. RC, DC and BC denote the global reaching centrality, mean degree centrality and mean betweenness centrality respectively. Diameter is taken using the undirected version of the network. Self loops have been removed from the original dataset, so we are testing on the largest connected component of the original dataset. Decimal places are rounded to two significant figures.*

Name	$|V|$	$|E|$	Density	Clustering	Reciprocity	BC	DC	RC	Diameter
Airport Flights Networks									
US Airports [29]	1572	28235	0.011	0.50	0.78	0.0013	0.023	0.059	8
Open Flights [30]	2905	30442	0.0036	0.44	0.97	0.0010	0.0072	0.0072	14
Electrical Circuit Networks									
s420 [26]	252	399	0.0063	0.028	0.00	0.0062	0.013	0.66	13
s838 [26]	512	819	0.0031	0.027	0.00	0.0037	0.0063	0.67	15
Ecological Food Web Networks									
Mangwet [24]	97	1492	0.16	0.26	0.062	0.011	0.32	0.11	3
Baywet [24]	128	2106	0.13	0.18	0.029	0.0091	0.26	0.19	3
Little Rock Lake [29]	183	2476	0.074	0.17	0.034	0.00094	0.15	0.82	4
Ythan [25]	92	414	0.049	0.11	0.00	0.0011	0.099	0.79	4
St. Marks Seagrass [20]	49	223	0.095	0.13	0.00	0.0048	0.19	0.74	4
Grassland [20]	88	137	0.018	0.17	0.00	0.00033	0.036	0.97	6
Peer-to-peer Internet Networks									
p2p-gnutella04 [21]	10876	39994	0.00034	0.0031	0.00	0.00021	0.00068	0.6	10
p2p-gnutella05 [21]	8842	31837	0.00040	0.0036	0.00	0.00024	0.00081	0.6	9
p2p-gnutella06 [21]	8717	31525	0.00041	0.0033	0.00	0.00024	0.00083	0.61	10
p2p-gnutella08 [21]	6299	20776	0.00052	0.0054	0.00	0.00030	0.0010	0.63	9
p2p-gnutella09 [21]	8104	26008	0.00040	0.0048	0.00	0.00024	0.00079	0.65	10
Neurological Networks									
C. Elegans [31]	297	2345	0.027	0.17	0.17	0.01	0.053	0.13	5
Drosophila Medilla 1 [22]	1770	9624	0.0030	0.15	0.15	0.00083	0.0061	0.14	6
Mouse Visual Cortex 2 [22]	193	214	0.0058	0.010	0.00	0.00	0.012	0.24	8
Mouse Retina 1 [22]	1076	90811	0.079	0.30	0.00	0.00046	0.16	0.56	4
Rattus Norvegicus [22]	503	27667	0.11	0.78	0.34	0.0018	0.22	0.0020	3
Organisational Networks									
Cross Parker Consulting [23]	44	521	0.28	0.62	0.77	0.021	0.55	0.047	4
Freemans EIES n48 1 [23]	34	540	0.48	0.69	0.85	0.016	0.96	0.00	2
Freemans EIES n48 2 [23]	34	708	0.63	0.77	0.85	0.012	1.26	0.00	2
Cross Parker Manufacturing [23]	77	1452	0.25	0.67	0.80	0.012	0.50	0.013	3
Eva [28]	4475	4662	0.00023	0.0060	0.0043	0.00	0.00047	0.28	18
Regulatory Networks									
E. coli transcription [26]	328	456	0.0043	0.055	0.00	0.00	0.0085	0.020	13
Yeast transcription [26]	662	1063	0.0024	0.025	0.0019	0.00	0.0049	0.12	15
Trust/Social Networks									
Bitcoin Alpha [21]	3670	22639	0.0017	0.15	0.85	0.00065	0.0034	0.12	10
Bitcoin OTC [21]	5551	32007	0.0010	0.14	0.84	0.00042	0.0020	0.16	14
Email EU Core [21]	986	24929	0.026	0.37	0.71	0.0014	0.051	0.16	7
Prison Inmate [26]	67	182	0.040	0.23	0.44	0.034	0.082	0.34	7
UCIrvine [23]	1893	20292	0.0057	0.088	0.64	0.00080	0.011	0.29	8
WikiVote [21]	7066	103663	0.00208	0.082	0.056	0.00	0.0042	0.089	7
World Wide Web Networks									
Political Blogs[27]	1222	19021	0.013	0.22	0.24	0.0013	0.025	0.13	8

role in induced triads. We note that this is one of the reasons that motif analysis has become a popular tool. However in our case, comparison is based on the role edges play within induced triads, rather than the types of induced triads that are over or under represented. Profiling edges in this way gives insights into the relative prevalence of edges within the network, and determines the status of networks particularly concerning local leakage and containment of content over edges.

C. Comparing with Existing Metrics

To assess whether overt and covert centrality provide additional information about data sets, as compared to standard global metrics for social network analysis, we further consider metrics from Table I. Over each of the 34 assessed real networks, we find the mean overt and covert centrality of edges, by averaging over the total centrality (i.e., overt plus covert centrality) of edges. We normalise this mean count by dividing through by the maximum possible overt/covert centrality of an edge (the number of edges in the network minus 2). We calculate the Spearman rank correlation between metrics using all 34 data sets in order to compare them, as seen in Figure 3. We tested the distribution of each statistic for non-normality using a Shapiro-Wilk Test and in

TABLE II: *The frequency distribution of overt and covert centrality across edges in 34 real-world data sets. $|V|$ indicates the number of nodes, $|E|$ indicates the number of edges, OC indicates the overt centrality frequency distribution, and CC indicates the covert centrality distribution.*

Network	$|V|$	$|E|$	OC	CC	Network	$|V|$	$|E|$	OC	CC
Airport Flights Networks					*Neurological Networks*				
US Airports [29]	1572	28235			Mouse Visual Cortex 2 [22]	193	214		
Open Flights [30]	2905	30442			Mouse Retina 1 [22]	1076	90811		
Electrical Circuit Networks					Rattus Norvegicus [22]	503	27667		
s420 [26]	252	399			*Organisational Networks*				
s838 [26]	512	819			Cross Parker Consulting [23]	44	521		
Ecological Food Web Networks					Freemans EIES n48 1 [23]	34	540		
Baywet [24]	128	2106			Freemans EIES n48 2 [23]	34	708		
Grassland [20]	88	137			Cross Parker Manufacturing [23]	77	1452		
Little Rock Lake [29]	183	2476			Eva [28]	4475	4662		
Mangwet [24]	97	1492			*Regulatory Networks*				
St. Marks Seagrass [20]	49	223			E. coli transcription [26]	328	456		
Ythan [25]	92	414			Yeast transcription [26]	662	1063		
Peer-to-peer Internet Networks					*Trust/Social Networks*				
p2p-gnutella04 [21]	10876	39994			Bitcoin Alpha [21]	3670	22639		
p2p-gnutella05 [21]	8846	31837			Bitcoin OTC [21]	5551	32007		
p2p-gnutella06 [21]	8717	31525			Email EU Core [21]	986	24929		
p2p-gnutella08 [21]	6299	20776			Prison Inmate [26]	67	182		
p2p-gnutella09 [21]	8104	26008			UCIrvine [23]	1893	20292		
Neurological Networks					WikiVote [21]	7066	103663		
C. Elegans [31]	297	2345			*World Wide Web Networks*				
Drosophila Medilla 1 [22]	1706	9624			Political Blogs[27]	1222	19021		

all cases found the data does not follow a normal distribution. Therefore, we use the Spearman rank correlation coefficient to compare statistics. A positive correlation indicates a monotonic increasing relationship, whilst a negative correlation indicates a monotonic decreasing relationship. The closer to 1 (or -1) the value, the stronger the relationship. In particular, a correlation between 0.4 and 0.6 we have classed as moderate, whilst between 0.6 and 0.8 strong and above 0.8 as very strong. We display this in a correlation matrix as shown in Figure 3.

Figure 3 indicates that all significant relationships between overt/covert centrality with existing metrics from Table I are at least moderate, though many are strong or very strong. In particular, we observe very strong correlation between mean covert weight of a network and the network density (0.87), degree centrality (0.87). Mean overt weight of the network follows similarly, with the addition of very strong relationships with betweeness centrality (0.82) and global clustering (0.91). Mean overt and covert weight of networks correlate very
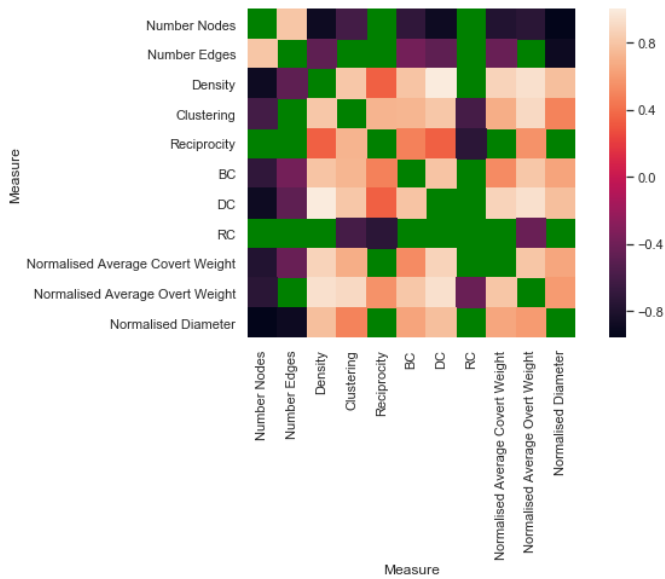
Fig. 3: *Correlation matrix comparing the Spearman correlation of each metric across all 34 data sets, with insignificant correlation results indicated by green cells (p > 0.05)*

strongly with each other (0.8). Interestingly, there was no significant correlation between normalised average covert weight and reciprocity, and global reaching centrality. For overt centrality, there was no significant correlation with number of edges. This indicates that while measure of overt and covert are picking up some similarities of network structure in common with existing metrics, they are some clear differences (i.e., additional network properties) that the concepts are identifying.

The existence of significant, directed correlations motivates further investigation through direct comparison of the mean overt/covert centrality with each metric for each data set, the results for which are found in Figure 4. We observe in Figure 4 that results are different depending on comparison with mean overt or covert centrality. For example, when comparing reaching centrality with mean covert weight the social networks category of data points cluster together. When comparing reaching centrality with mean overt weight, the data points are more dispersed along the x axis, pulling apart this cluster of nodes. The same effect can also be seen when comparing diameter with the average overt and covert weight in social networks.

These results show that overt and covert centrality measures have a role to play in differentiating the alternative network classes. Without overt and covert centrality, the data points in Figure 4 would project onto the y-axis, with little or no differentiation between some classes of network in many cases. One such example is the Clustering measure, where average overt centrality is able to distinguish alternative classes of network with similar clustering values (i.e., separate in the x-axis). This is particularly the case for low clustering levels (e.g., 0.3 or lower).
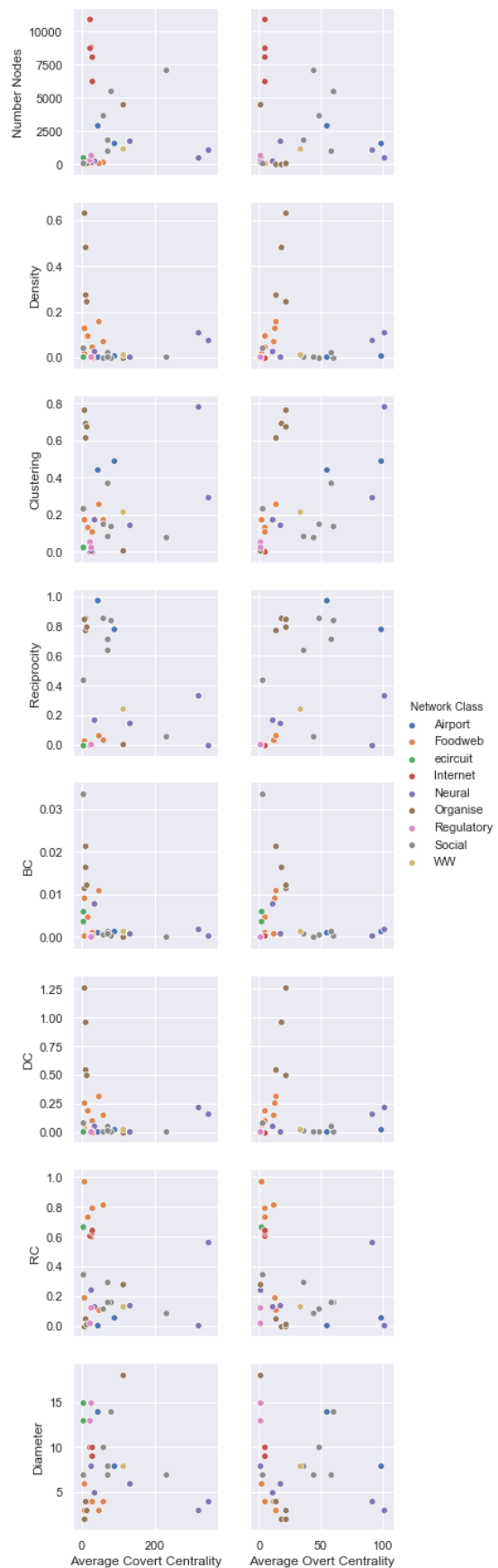
Fig. 4: *Scatterplots showing the relationship between standard network metrics and mean covert and overt edge centrality.*

V. Discussion and Conclusions

We have introduced a novel and fundamental approach to characterising the centrality of edges in networks based on their role in supporting connectivity within induced triads. This constitutes a binary classification of edges with respect to a particular induced triad within which they participate. In other words, we classify an edge as either overt or covert. This concept is important because triads are the smallest possible (non-trivial) induced substructure in a network. Therefore our definition of overt and covert centrality with respect to edges captures the lowest level at which non-trivial network connectivity can be considered, beyond the concept of an edge itself. Using counting arguments based on in and out degrees within a triad, we have derived formula to determine the both overt and covert edge centrality. This means that overt and covert centrality can be determined without recourse to lengthy search - the concept is locally defined and therefore scalable.

To demonstrate these concepts on 'real-world' scenarios, we have applied it to data derived from 34 public data sets. From profiling the overt and covert centrality distributions across these networks, we have shown that the frequency of overt and covert edges is useful in presenting further insights both within and between different classes of network. These concern the relative dependency that networks have on the prevalence of edges with key roles concerning connectivity. On the one hand, overt edges play a role in potential containment of content transferred through an edge, since the edge cannot support dissemination to the third node in the triad. Alternatively, overt edges align with the potential for onward local dissemination, since by definition an overt edge supports a path to the third node in the triad. This provides an alternative lens through which global network connectivity can be considered at a local level.

We have additionally compared overt and covert centrality with a number of conventional network analysis metrics, including global reach centrality, mean degree centrality, between-ness centrality, density, clustering, reciprocity and diameter. Potential significant correlations have been considered with a view to understanding relationships with these metrics. By examining the relationship between these metrics and overt/covert centrality across 34 networks, we have found that overt/covert centrality can provide additional insights. These relate to distinguishing between classes of network, which the global network metrics would fail to achieve in isolation, for numerous cases.

In summary, we observe that overt and covert centrality are important concepts for characterising a network's structural potential for dissemination and contagion, based on consideration of a network's underlying building blocks. These issues are further being pursued through ongoing research.

References

[1] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[2] U. Alon, "Network motifs: theory and experimental approaches," *Nature Reviews Genetics*, vol. 8, no. 6, pp. 450–461, 2007.

[3] R. D. Luce and A. D. Perry, "A method of matrix analysis of group structure," *Psychometrika*, vol. 14, no. 2, pp. 95–116, 1949.

[4] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, p. 440, 1998.

[5] S. Wasserman, K. Faust *et al.*, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.

[6] P. W. Holland and S. Leinhardt, "Transitivity in structural models of small groups," *Comparative group studies*, vol. 2, no. 2, pp. 107–124, 1971.

[7] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social networks*, vol. 31, no. 2, pp. 155–163, 2009.

[8] 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.

[9] L. C. Freeman, S. P. Borgatti, and D. R. White, "Centrality in valued graphs: A measure of betweenness based on network flow," *Social networks*, vol. 13, no. 2, pp. 141–154, 1991.

[10] S. P. Borgatti, "Centrality and network flow," *Social networks*, vol. 27, no. 1, pp. 55–71, 2005.

[11] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Social Networks*, vol. 30, no. 2, pp. 136–145, 2008.

[12] P. De Meo, E. Ferrara, G. Fiumara, and A. Ricciardello, "A novel measure of edge centrality in social networks," *Knowledge-based systems*, vol. 30, pp. 136–150, 2012.

[13] D. Koschützki, K. A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski, "Centrality indices," in *Network analysis*. Springer, 2005, pp. 16–61.

[14] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon, "Topological generalizations of network motifs," *Physical Review E*, vol. 70, no. 3, p. 031909, 2004.

[15] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, "Superfamilies of evolved and designed networks," *Science*, vol. 303, no. 5663, pp. 1538–1542, 2004.

[16] A. R. Benson, D. F. Gleich, and J. Leskovec, "Higher-order organization of complex networks," *Science*, vol. 353, no. 6295, pp. 163–166, 2016.

[17] N. Pržulj, "Biological network comparison using graphlet degree distribution," *Bioinformatics*, vol. 23, no. 2, pp. e177–e183, 2007.

[18] E. Estrada and J. A. Rodriguez-Velazquez, "Subgraph centrality in complex networks," *Physical Review E*, vol. 71, no. 5, p. 056103, 2005.

[19] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, "Networks beyond pairwise interactions: structure and dynamics," *Physics Reports*, vol. 874, pp. 1–92, 2020.

[20] COSIN, "The cosin network data and analysis." [Online]. Available: http://www.cosinproject.eu/extra/data/foodwebs/WEB.html

[21] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," 2014. [Online]. Available: http://snap.stanford.edu/data

[22] NeuroData, "Animal connectomes." [Online]. Available: https://neurodata.io/project/connectomes/

[23] T. Opsahl, "Network datasets." [Online]. Available: https://toreopsahl.com/datasets/

[24] V. Batagelj and A. Mrvar, "Pajek datasets." [Online]. Available: http://vlado.fmf.uni-lj.si/pub/networks/data/bio/foodweb/foodweb.htm

[25] A. R. Cirtwill and A. Eklöf, "Data from: Feeding environment and other traits shape species' roles in marine food webs," 2018. [Online]. Available: https://doi.org/10.5061/dryad.1mv20r6

[26] U. A. Lab, "Collection of complex networks." [Online]. Available: https://www.weizmann.ac.il/mcb/UriAlon/download/collection-complex-networks

[27] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 us election: divided they blog," in *Proceedings of the 3rd international workshop on Link discovery*, 2005, pp. 36–43.

[28] G. L. Kim Norlen, M. Gebbie, and J. Chuang, "Eva: Extraction, visualization and analysis of the telecommunications and media ownership network," in *Proceedings of International Telecommunications Society 14th Biennial Conference*. Citeseer, 2002.

[29] J. Kunegis, "Konect: the koblenz network collection," in *Proceedings of the 22nd International Conference on World Wide Web*, 2013, pp. 1343–1350.

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

[31] J. D. Johnson, "Ucinet: A software tool for network analysis," 1987.