Influence of Asymmetry and Structural Roles on Triad Patterns in Undirected Networks

Milos Kudelka
milos.kudelka@vsb.cz
VSB - Technical University of Ostrava

Eliska Ochodkova
eliska.ochodkova@vsb.cz
VSB - Technical University of Ostrava

Jakub Plesnik
jakub.plesnik@vsb.cz
VSB - Technical University of Ostrava

Sarka Zehnalova
sarka.zehnalova@vsb.cz
VSB - Technical University of Ostrava

ABSTRACT

Triads, i.e., variously interconnected triplets of nodes, significantly affect the network structure. Closed triads, for instance, are the building blocks of communities. Our study focuses on the analysis of triads in which the ego is connected to two alters, with the alters not having to be connected; therefore, the triads that are studied need not be closed. The analysis uses two approaches based on asymmetric relationships between pairs of nodes. In the first approach, we work with three different node roles, in which the ego and its alters can appear in triads. We get a total of eighteen different role-based triad patterns. The second approach allows us to work with a total of four different types of ties and six different alter-pair patterns. In experiments with four different types of real-world networks, we show how the properties of these networks differ in terms of role-based triad patterns. In some of these networks, we further show that the triad-based properties remain stable during network growth. The main contribution of our paper is the use of asymmetric relations for the definition of four types of dependency-based tie strengths between nodes and the analysis of their influence on the occurrence of different triad-based patterns in networks.

KEYWORDS

social network, triadic closure, structural role, asymmetric dependency, triad pattern

ACM Reference Format:
Milos Kudelka, Eliska Ochodkova, Jakub Plesnik, and Sarka Zehnalova. 2020. Influence of Asymmetry and Structural Roles on Triad Patterns in Undirected Networks. In *KDD-MLG'20: 16th International Workshop on Mining and Learning with Graphs, Aug 24, 2020, San Diego, CA, USA.* ACM, New York, NY, USA, 8 pages.

1 INTRODUCTION

At present, we can observe the steady growth of various social networks, which is generating a huge amount of data. One way to understand the data is to analyze network structures concerning the properties that are based on the relationships between pairs of nodes in these networks. Here, many challenges and opportunities arise, so it is necessary to come up with new and exact methods, and, because of the complexity of the networks, also heuristics which would help reveal and detect these structural properties.

As it turns out, an important structural characteristic of networks is the tendency (e.g., triadic closure) to create triadic configurations. These triadic configurations are the basis of theoretical considerations leading to experimentally confirmed results, such as the strength of weak ties, tie stability, trust, and structural holes [10, 23]. Social ties are usually divided into three types: strong, weak, or missing ties. In [22], a weak tie hypothesis was formulated: if node A is connected to two nodes B and C, then nodes B and C are very likely to be connected too, leading to a closed triad. Simmel was the first to explain the social tendency to form closed triads as being natural [26].

Weak social ties, as stated in [12], make it possible to address entities (subjects) that are not accessible through strong ties. Weak ties can then be bridges that connect different subgroups of the social network (communities) and thus shorten the distances in the networks. These bridges, however, do not take up the expected connections based on the triadic closure. Therefore, Granovetter assumes that the chance of connecting nodes B and C increases if the ties between A and B and A and C are strong. In [29], the article states that strong ties occur on the shortest paths and therefore shorten distances in networks. However, that is rather in contrast to Granovetter's theory.

2 BACKGROUND

2.1 Strength of ties

An essential part of estimating whether nodes B and C will get connected is, therefore, the method of measuring the tie strength between nodes A and B, and A and C, respectively. The authors of [29] state that there is a positive correlation between the tie strength and degree of nodes in the triad, which can be used for alternative measurements of tie strength. Also related to this are what are called Simmelian ties, i.e., ties that are surrounded by several triads in their neighborhood. In [7], the author states that "The definition of Simmelian ties closely resembles that of a clique; indeed, a perfectly equivalent definition of a Simmelian tie is that it is a tie embedded in a clique." Similarly, the authors of [28] state that their findings "reveal that the distinction between weak and strong bridging ties is not very informative if the dyad forming a bridge is considered independently of the microcontext in which it is embedded." The authors further explain that the bridging strength

KDD-MLG'20, Aug 24, 2020, San Diego, CA, USA

of a tie should be judged by whether or not it is embedded in a dense clique-like structure.

The strong triadic closure (STC) principle, which is based on the assumption that, in the network, there are no nonadjacent nodes with strong ties with a common neighbor, is studied in [27]. The authors formulate a combinatorial optimization problem, and their goal is to find a weak-strong tie labeling that satisfies this property. They also show that although the problem is NP-hard, they are able to identify cases where efficient algorithms with provable approximation guarantees exist. However, as stated in [1], although the STC property is theoretically motivated, real-world networks are noisy and may contain many exceptions to this principle; empirical findings show that the networks that are examined do not have a large number of strong edges. As presented in [9], the STC property seems to be too extreme a principle for large-scale real-world networks; empirical tie strength rather scales linearly with the neighborhood overlap of an edge, as is shown for a "who talks-to-whom" network. In [25], the authors consider weak-strong tie labeling with additional community connectivity constraints; they assume that each community should be interconnected through strong ties which allow these constraints to be met, with a small number of STC violations.

2.2 Asymmetry-based approach

The subject of our experiments is triads with one ego, which has ties to two alters, with there not necessarily being a tie between a pair of alters. As follows from the statements above, the probability of the existence of a tie between alters is influenced by their connections with common neighbors. In our approach, we work with the structural properties of undirected networks and asymmetric relationships between pairs of nodes. This structural asymmetry describes the dependency of one node on another and quantifies the degree of interconnection of both nodes with their common neighbors versus their interconnection with other nodes [15]. With the dependency being used, four different types (strengths) of ties between pairs of adjacent nodes and three roles of nodes are described, with consideration being given to the degree of their dependency on the surroundings. In [11, 13], examples of alternative ways of extracting the structural roles of nodes are presented.

In this paper, while working with triads, we use *patterns*, which take into account the roles that nodes play in the triad. Patterns thus differ from triadic configurations, which do not take into account the types of nodes in the triad but deal with the types of connections between them. Thanks to there being three roles, we get a total of eighteen triad patterns. The individual patterns differ in terms of the roles in which the ego and its two alters appear in the triad.

In the next section of the paper, we will introduce the basic concepts related to dependency and describe the types of connections between nodes and node roles with which we will later work. Furthermore, after twelve datasets have been briefly introduced, experiments aimed at analyzing the frequency and degree of closedness of individual role-based triad patterns and alter-pair patterns will follow. In these experiments, we will show how various types of networks differ in these properties and also that these properties remain stable as the network grows over time. Finally, we summarize the results achieved and formulate questions related to the future use of our approach.

3 DEPENDENCY

Research and experiments with asymmetric structural dependency in artificial and real-world networks have been published in [15]. For the needs of the experiments in this paper, we, therefore, present only basic definitions related to experiments with triad analysis.

Definition 3.1 (Structural dependency). Let x, y be nodes, then dependency $D(x, y)$ of node x on node y is defined as follows:

$$D(x,y) = \frac{w(x,y) + \sum_{v_i \in CN(x,y)} w(x,v_i) \cdot r(x,v_i,y)}{\sum_{v_j \in N(x)} w(x,v_j)} \tag{1}$$

$$r(x,v_i,y) = \frac{w(v_i,y)}{w(x,v_i) + w(v_i,y)}, \tag{2}$$

where $CN(x, y)$ is set of all common neighbors of x, y, $N(x)$ is set of all neighbors of x, $w(x, y)$ is weight of edge between x, y, and $r(x, v_i, y)$ is the coefficient of the dependency of node x on node y via the common neighbor v_i.

Dependency is generally non-symmetric for both weighted and unweighted networks. The reason is that not only are the weights of edges with common neighbors taken into account (which do not differ for unweighted networks), but also the weights of all the other edges that both nodes have with their neighbors (and these may be different for each node). In the illustrative network in Figure 1, the calculated dependencies for the node pairs of an unweighted undirected network are shown. The arrows indicate the direction of dependency with a value of at least 0.5. The exceptions are nodes 4 and 7, which have low mutual dependencies and also low dependencies on their other neighbors.

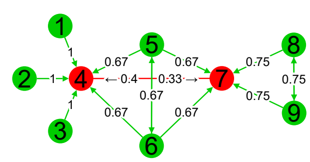

Figure 1: Illustrative network with calculated dependencies.

To simplify the view on the dependency of one node on another, a simple binarization to the *IsDependent* relation can be used.

Definition 3.2 (IsDependent). Let x, y be neighboring nodes, then *IsDependent* relationship is defined as follows:

$$IsDependent(x,y) = \begin{cases} True, & \text{if } D(x,y) \geq 0.5 \\ False, & \text{otherwise} \end{cases}$$

The dependency threshold is set to 0.5 to take into account and reasonably balance a mutual dependency between two neighboring network nodes. The threshold, however, can be both higher and lower than 0.5. At higher values, the number of strong and very strong ties decreases in the network, which can lead to weak ties even in dense clique-like structures. Conversely, a lower threshold value causes a decrease in weak ties, which also results in a decrease in strongly prominent nodes. Moreover, both cases reduce the interpretability of the results.

As shown in experiments with real-world networks [15], estimation of the time complexity of calculating the dependencies between all the node pairs (dependency matrix) of a sparse network is $m \log n$, where n is the number of nodes and m is the number of network edges. Network dependencies are stored in a sparse matrix in the dictionary of keys (DOK) format, network is represented by the adjacency list; computation time of the dependency matrix is given in Table 1.

3.1 Four types of ties between nodes

Because of binarization, regardless of whether the network is weighted or not, there are different dependencies between adjacent nodes u and v. Nodes u and v can be mutually dependent, or mutually independent, or node u may be dependent on node v and vice versa. Below, we will, with regards to the IsDependent relationships, denote the tie between two mutually dependent nodes as *very strong*, the tie between mutually independent nodes as *weak*, and the tie in the two remaining cases as *strong*. If the nodes are not neighbors, then the fourth type of tie will be referred to as *missing*. Note that the dependency-based type (strength) of the tie does not correspond to the strength of the edge in the meaning of weight in weighted networks. A key factor influencing the type of tie between a pair of nodes is the effect of the weights of edges with common neighbors compared to the weights of edges with other neighbors of both nodes. Our exact technical approach is partly close to that of Simmelian ties. The difference is that when determining the strength of the tie, we take into account the surroundings of the pair of nodes that are not part of their common clique-like substructure. The types of ties that are described can be observed in Figure 1 and Figure 2, which shows a Karate Club network with three types of ties based on the directions of mutual dependencies between neighbors. E.g., node 17 is dependent on nodes 6 and 7 and has a strong tie with both, but the tie between nodes 1 and 3 is weak (similarly to the weak tie between nodes 4 and 7 in Figure 1).

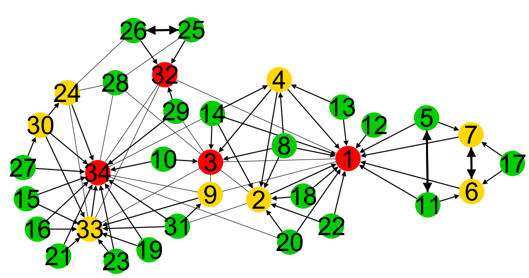

Figure 2: Karate Club: Strengths of ties, IsDependent relationships, and structural roles.

3.2 Dependency-based roles of nodes

In our approach, the role stems from what the ego network of the node that is examined looks like from the point of view of dependency. We distinguish three different roles; in the experiments, we will denote them by the abbreviations S, W, and N.

A strongly prominent node (S) is not dependent on any of its neighbors, and at least one of its neighbors is dependent on it.

A weakly prominent node (W) has at least one neighbor that is dependent on it, and the node itself is not dependent on that neighbor.

A non-prominent node (N) is a node that is neither strongly nor weakly prominent.

In Figures 1 and 2, the strongly and weakly prominent nodes are shown in red and yellow and the non-prominent nodes in green. It is also possible to observe in Figures 1 and 2 what strengths the ties between which nodes have.

Weak ties typically connect SS and SW pairs of prominent nodes (e.g., 32-34, 1-3, 1-32, but also 3-9, 9-34).

Very strong ties typically connect WW or NN node pairs (e.g., 5-11, 6-7, 25-26). These pairs can be understood with a little exaggeration as mutually dependent weakly or non-prominent twins; these ties are often embedded in clique-like structures (including dyads and triads).

Strong ties typically connect non-prominent or weakly prominent nodes with both types of prominent nodes; they, therefore, include the relations NW, NS, WW, or WS (e.g., most of the green non-prominent neighbors of the prominent red and yellow nodes in Figure 2).

The examples that are given are not, and cannot be understood as, the only possible ones. Networks contain complex substructures that can go beyond common situations. Possible examples in the Karate Club network are weak ties between nodes 20 and 34 or 26 and 24. Another and more interesting example is node 28, which is non-prominent (no other node are dependent on it), but has only weak ties with all of its neighbors.

The structure of the Karate Club in Figure 2 also shows, following Granovetter's theory, the strength of weak ties. Most dependency-based weak ties here represent direct connections between more strongly interconnected network substructures; the connections between substructures are made through strongly and weakly prominent nodes. Therefore, as will be described in our experiments, it makes sense to examine triads with a focus on node roles, because different configurations of roles in a triad can represent a different purpose.

In general, the occurrence of specific situations depends on what the dependencies between the nodes are in a particular network. In [15], the network dependency property (NetDep, see Eq. 3.2) is defined to assess the degree of dependency in the network as a whole, which takes into account the total number of weak ties m^* to all the ties m in the network. If the NetDep value is close to one, then most of the ties between network nodes are strong or very strong, and if it is close to zero, the opposite is true (e.g., NetDep = 0.756 for the Karate Club in Figure 2). Thus, networks with a low NetDep value have a predominant number of weak ties.

$$NetDep = 1 - \frac{m^*}{m}. \tag{3}$$

The next section of the paper describes the networks we used in our experiments. For all these networks, the types (strengths) of ties for all pairs of adjacent nodes and subsequently the roles of all nodes were detected during the preparation of the experiments.

4 DATASETS

In the experiments, we work with a total of twelve networks of four types obtained from the datasets below. For each type (biological, co-authorship, communication, and social), three networks were the subject of our experiments. If a network was weighted, dynamic, signed, or directed, it was straightforwardly converted to an unweighted undirected one. All the calculations are thus based on the same input conditions and allow the interpretation to focus on purely structural properties of the networks that were analyzed.

bio-CE-CX [6] WormNet: a network constructed by integration of all data-type-specific networks (CE-CX, CE-GN, ...) by modified Bayesian integration. Available at http://networkrepository.com/bio-CE-CX.php

ChCh-Miner [31] network of interactions between drugs, which are approved by the U.S. Food and Drug Administration. Available at https://snap.stanford.edu/biodata/datasets/10001/10001-ChCh-Miner.html

PP-Pathways [2] protein-protein interaction network that contains physical interactions between proteins that are experimentally documented in humans. Available at https://snap.stanford.edu/biodata/datasets/10000/10000-PP-Pathways.html

astro-ph [21] weighted network of coauthorships between scientists posting preprints on the Astrophysics E-Print Archive. Available at http://www-personal.umich.edu/~mejn/netdata/

cond-mat-05 update [21] network of coauthorships between scientists posting preprints on the Condensed Matter E-Print Archive. Available at http://www-personal.umich.edu/~mejn/netdata/

coauth-DBLP [4] collaboration graph of authors of scientific papers from the DBLP computer science bibliography. We used a version that is restricted to publications with 25 authors at most. From that, we have created a subset containing the first 200k nodes ordered by the first occurrence from the total 1.9M nodes in the original dataset. Available at https://www.cs.cornell.edu/~arb/data/coauth-DBLP/

anobii [3] two types of networks are available. A network composed of a union of friendships and neighborhood links is the first. The second one is a communication network representing message exchanges. Available (on request) at https://www.icwsm.org/2016/datasets/datasets/

linux-kernel [17] the communication network of the Linux kernel mailing list. Nodes are persons, and each directed edge represents a reply from a user to another. Available at http://networkrepository.com/comm-linux-kernel-reply.php

enron [14, 18] communication network that covers all the email communication within a dataset of around half a million emails. Available at http://networkrepository.com/ia-enron-email-dynamic.php

fb-friends [30] network of friendship where nodes are users and edges between the users represent friendship relations. Available at http://networkrepository.com/fb-wosn-friends.php

musae_git [24] sample of GitHub users created by a random walk sampling algorithm. Available at https://snap.stanford.edu/data/github-social.html

epinions [20] who-trust-whom online social network of a general consumer review site Epinions.com. Available at http://networkrepository.com/soc-epinions-trust-dir.php

Table 1 summarizes the properties of all the networks that were analyzed (type, number of nodes and edges, maximum and average degree, average clustering coefficient, Louvain modularity, assortativity, NetDep, percentages of role occurrences, and computation time for dependency matrix). The biological networks have a high average degree (they are denser). The co-authorship networks differ mainly in their characteristics related to the community structure (modularity and clustering coefficient); they also have a positive assortativity, unlike most other networks. The highest diversity can be seen in the network dependency. The co-authorship networks have the highest NetDep, followed by the communication, social, and biological networks. The reason is that co-authorship and communication interactions often take place in structures close to cliques, in which there are strong dependencies. Besides, for the co-authorship networks, the share of weakly prominent nodes is much higher (see the W% column), which indicates diverse dependencies in small substructures of networks (collaborating teams).

The analysis of the roles of nodes in triads will allow us to look at the differences in network types with fresh eyes. Surprisingly, it will be shown how, in terms of various role-based triad patterns, networks of the same type can resemble each other on the one hand, and networks of various types differ on the other.

5 ROLE-BASED TRIAD PATTERNS

It was stated in the introduction that our approach is based on the analysis of triads in which the ego node has ties with two of its alter nodes. If we consider the role of each node of such a triad, we get a total of eighteen different patterns. Individual patterns will be denoted by three abbreviations of role names (W, S, N), with the role of ego in the first place and alter roles arranged from W through S to N.

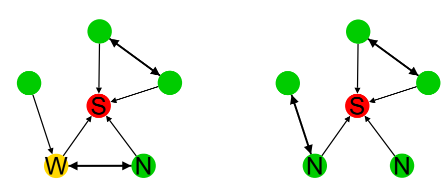

Figure 3: A closed triad with SWN-WSN-NWS patterns on the left, and an open triad of the SNN pattern on the right.

Let us now consider three interconnected nodes of a network. Such a triplet generates three different role-based closed triads because each of them can belong to a different pattern as a result of a different ego and its alters. This situation with patterns SWN, WSN, NWS is shown in Figure 3 on the left. If the triplet is not closed (it has only two edges), then it only belong to one pattern as presented for the SNN pattern in Figure 3 on the right.

Table 2 contains the total counts of detected role-based triads and the percentages of individual patterns. Triads occur in all role-based patterns, at least in units of percentages, only in the co-authorship networks, which distinguishes them from the other

Table 1: Properties of the networks that were analyzed

Network	Type	n	m	k_{max}	k_{avg}	CC	Q_L	r	NetDep	S%	W%	N%	Dep. matrix (s)
bio-CE-CX	BIO	15229	245952	375	**32.30**	0.21	**0.60**	**0.34**	0.02	**24.28**	0.58	75.14	3.08
ChCh-Miner	BIO	1514	48514	443	**64.09**	0.30	0.39	-0.10	0.03	13.01	3.17	**83.82**	0.79
PP-Pathways	BIO	21557	338636	2132	**31.76**	0.13	0.40	-0.04	0.03	16.20	0.21	**83.59**	7.61
astro-ph	COA	160476	121251	360	15.11	**0.67**	**0.73**	**0.24**	**0.50**	**26.63**	**13.25**	60.12	0.81
coauth-DBLP	COA	200000	410775	171	4.11	**0.53**	**0.90**	**0.19**	**0.73**	14.60	**6.44**	78.96	1.65
cond-mat-05	COA	39577	175692	278	8.88	**0.65**	**0.72**	**0.19**	**0.50**	**22.39**	**13.10**	64.50	0.92
anobii	COM	179653	1165929	6559	12.98	0.17	**0.56**	-0.03	**0.12**	**26.82**	2.32	70.85	26.48
linux-kernel	COM	26885	159994	2989	11.90	0.31	0.36	-0.18	**0.23**	15.72	3.50	**80.78**	6.02
enron	COM	86978	297456	1726	6.84	0.12	**0.66**	-0.17	**0.33**	5.48	0.60	**93.93**	5.11
fb-friends	SOC	63731	817090	1098	**25.64**	0.22	**0.61**	**0.18**	0.04	**28.75**	1.79	69.46	8.99
musae_git	SOC	37700	289003	9458	15.33	0.17	0.45	-0.07	0.08	**20.00**	0.90	79.11	17.18
epinions	SOC	131570	711202	3558	10.81	0.13	0.46	-0.07	**0.17**	17.47	1.51	**81.02**	21.92

types of networks. This, with reference to the previous section, again confirms that in the co-authorship networks, dependencies between nodes are much more diverse than in the other types of networks as a result of the high overall network dependency and interconnection of clique-like substructures.

For the biological networks and the fb-friends network, there is a higher proportion of triads in the NSS, NSN, and NNN patterns, in which the ego is in the role of a non-prominent node. As can be seen in Table 1, these networks have a high average degree and low dependencies as a result of the NetDep value. At the same time, they have a small proportion of W nodes and a relatively high proportion of S nodes. Thus, if the ego node is in the role of N in these networks, it is natural that its most probable neighbors in the triad are S and N nodes, and the ties between them are rather weak. Similarly, the proportion of triads is higher in SSN and SNN patterns, but this is also true for most other networks.

The third column of Table 2 shows the percentage of strong triads. By the term *strong triad*, we mean a triad in which the ego has strong or very strong ties with both its neighbors, regardless of the closure of this strong triad. Co-authorship networks have a higher share of strong triads, which is obviously due to the occurrence of clique-like structures. On the other hand, biological and social networks have an almost negligible share of strong triads. This is related to the overall occurrence of strong and very strong ties in the networks, which, as presented in Table 4, has a higher occurrence only in the co-authorship networks.

Another property we investigated in the experiments is the proportion of triads that are closed for a given pattern. The closedness of triads in various contexts seems to us to be important for a more detailed understanding of principles such as triadic closure. Experiments have shown that this property can also help us see in a simple way how networks of the same type resemble each other. In Figure 4, radar charts show the proportions of closed triads in individual role-based patterns in groups according to the network type. At first glance, we can see a significant difference between the co-authorship networks and the other types. The biological networks differ less significantly from the communication and social networks, which, more or less, do not differ from the point of view of the closedness of role-based triad patterns.

Observations (role-based patterns)

(1) All the networks have a very low proportion of closed triads in patterns with ego in the role of S (the lower part of the radar charts). This is because strongly prominent nodes, which usually have a higher degree, are hubs, and thus lie between several otherwise weakly interconnected substructures such as communities (see Figure 2 and experiments in [15]). Therefore, they usually have many pairs of alters in their ego network between which there is a missing tie.

(2) Almost all the networks have a high proportion of closed triads in WWW, WWN, WSW, and WSS patterns (at the top right in the radar charts). This indicates the importance of egos in the W role for the local interconnection of nodes into dense substructures. However, there are two patterns, WSN and WNN, for which this is not the case. In these patterns, the ego, unlike in the previous four patterns, can be, e.g., connected to alters only by weak ties.

(3) It is surprising how small the differences are in the proportions of closed triads in the three co-authorship networks. This is especially evident in patterns with ego in the role of N. These patterns are typical of the co-authorship networks, because one of the characteristics is cooperation in small teams connected into cliques. Therefore, non-prominent nodes connect with nodes from these cliques, regardless of their role.

(4) The biological networks have proportions of closedness that are more balanced than the communication and social networks. However, for all three network types, it can be observed that some patterns have a very low closedness; this is best seen with the NNN pattern. Apparently, the three nodes in the N role in these three types of networks occur in less dependent and more complicated structures than, for example, clique-like substructures in the co-authorship networks.

All these observations lead us to believe that in the study of triads, their closedness, and interconnection into larger substructures, more attention should be paid to asymmetric relationships (regardless of whether the network is directed or undirected) and the roles that nodes in triads play. This is also related to the results of the alter-pairs analysis in the next section.

Table 2: Triad counts and percentage of role-based patterns

Network	Count	Strong	WWW	WWN	WSW	WSS	WSN	WNN	SWW	SWN	SSW	SSS	SSN	SNN	NWW	NWN	NSW	NSS	NSN	NNN
bio-CE-CX	23417777	0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.03	0.02	5.71	17.86	15.88	<0.01	0.02	0.01	**6.89**	**25.90**	**27.67**
ChCh-Miner	6595978	0.29	0.01	0.37	0.11	0.22	**1.55**	**2.79**	0.06	2.05	0.67	2.32	13.91	**21.81**	0.17	**3.26**	**1.49**	**4.38**	**19.93**	**24.90**
PP-Pathways	60401332	0.52	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	0.06	0.05	**18.28**	**37.91**	**24.62**	<0.01	<0.01	0.01	**5.10**	9.05	4.90
astro-ph	5325457	**21.80**	**2.56**	**2.99**	**4.35**	**3.41**	**2.95**	**1.45**	**3.29**	**5.81**	**14.42**	**26.39**	17.50	3.95	**1.00**	**1.65**	**1.62**	1.97	2.29	2.39
coauth-DBLP	4263656	**33.44**	**0.31**	**1.20**	**0.89**	**1.05**	**2.18**	**1.64**	**0.81**	**5.77**	**6.26**	**21.16**	**29.31**	14.77	**0.39**	**1.32**	**1.30**	2.58	4.27	4.80
cond-mat-05	4629273	**13.39**	**0.79**	**1.42**	**2.75**	**3.15**	**2.66**	**0.86**	**2.96**	**7.06**	**16.42**	**27.67**	**22.67**	5.73	**0.37**	0.54	**1.17**	1.77	1.28	0.74
anobii	199783133	**6.46**	<0.01	0.01	0.01	0.01	0.02	0.01	0.06	1.38	0.78	**33.57**	**39.84**	17.40	<0.01	0.01	0.02	3.43	2.84	0.63
linux-kernel	53773227	**5.61**	<0.01	<0.01	<0.01	0.05	0.02	<0.01	0.10	2.25	**3.59**	**40.83**	**39.93**	12.37	<0.01	<0.01	0.02	0.67	0.16	0.02
enron	49424399	**12.13**	<0.01	0.07	0.01	0.08	0.35	**1.02**	0.01	0.85	0.58	16.40	**37.06**	**38.70**	0.03	0.05	0.17	2.63	1.41	0.58
fb-friends	71111352	0.07	<0.01	<0.01	<0.01	0.02	0.02	0.01	<0.01	0.14	0.15	12.72	**28.20**	16.81	<0.01	0.03	0.04	**7.86**	**19.92**	**14.06**
musae_git	127167272	3.06	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	1.03	0.53	12.29	**41.89**	**42.00**	<0.01	<0.01	<0.01	0.98	0.96	0.31
epinions	182188819	1.26	<0.01	<0.01	<0.01	0.05	0.03	0.01	0.01	0.38	1.05	**58.97**	**32.36**	5.74	<0.01	<0.01	0.02	1.14	0.21	0.01

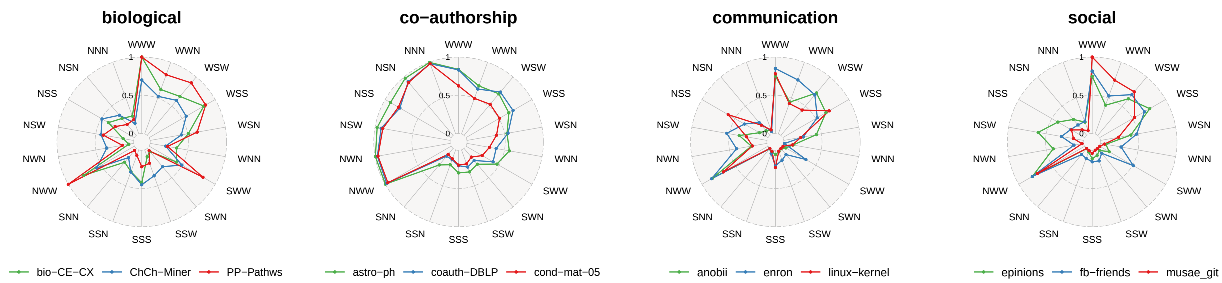

Figure 4: Proportions of closed triads for 18 role-based triad patterns.

5.1 Alter-pair patterns

If we focus on pairs of alters in role-based triads, then we get a total of six alter-pair patterns. However, the total count of alter-pairs is lower for the networks that were investigated than the total number of role-based triads, because one pair of nodes may have more nodes (egos) in common. Table 3 summarizes the results and also shows the differences between the co-authorship and other networks.

Table 3: Alter-pair counts and percentage of alter-pair patterns

Network	Count	WW	SW	WN	SS	SN	NN
bio-CE-CX	6342068	<0.01	0.05	0.10	9.06	**41.77**	**49.01**
ChCh-Miner	594619	0.14	1.29	6.01	2.74	27.65	**62.17**
PP-Pathways	25351284	<0.01	0.07	0.11	10.56	**46.20**	**43.07**
astro-ph	1779993	**3.29**	**18.19**	**8.83**	**33.05**	29.69	6.95
coauth-DBLP	2943808	**0.82**	**7.19**	**7.42**	**24.12**	**39.30**	**21.16**
cond-mat-05	2572475	**3.11**	**18.01**	**9.31**	**29.78**	31.11	8.68
anobii	138452270	0.08	1.03	1.94	**28.77**	**45.07**	**23.12**
linux-kernel	19770792	0.21	3.87	4.85	11.01	**52.28**	**27.78**
enron	27988775	0.01	0.53	1.08	7.00	**38.28**	**53.10**
fb-friends	31812564	<0.01	0.28	0.30	**18.77**	**48.48**	**32.16**
musae_git	95873484	0.01	0.58	1.23	9.32	**41.53**	**47.33**
epinions	83900931	0.01	1.47	0.57	**39.72**	**48.04**	10.20

Moreover, we can consider two factors; (1) the types of ties between alters may be different (very strong, strong, weak, missing), and (2) the frequency of their occurrence can also vary. Since one alter-pair can be linked to multiple egos, such as alters 6-7 with egos 1 and 17 in Figure 2, the question of how these factors relate is undoubtedly relevant. Table 4 summarizes the results of the analysis focused on these factors. The first part of the table is dedicated to the occurrence of different types of ties between alter-pairs, the second to the number of triads (egos) around an alter-pair, and the third to the degree of closedness for individual alter-pair patterns.

Observations (alter-pair patterns)

(1) The occurrences of very strong, strong, and weak ties in the co-authorship networks are relatively balanced. This is related to the above-mentioned considerations about more diverse dependencies in these networks.
(2) In the biological, communication, and social networks, very strong ties occur very rarely and strong ties only to a small extent. Higher proportions of weak ties occur only in the biological networks and in the fb-friends network; apparently, this is related to their higher density, which also affects the higher degree of closedness of the alter-pair patterns shown in the third part of the table.
(3) A higher proportion of closed alter-pairs for all the patterns in the third part of the table can also be observed for the co-authorship networks. However, here there is another reason; it is related to the large quantity of clique-like substructures in these networks.

The middle part of Table 4 deserves special attention; it shows the average numbers of common egos for the alter-pairs. At first glance,

Table 4: Properties of alter-pairs focusing on tie strengths and closedness

Network	Percentage by tie strengths				Avg count of neighboring egos				Percentage of closed alter-pairs by patterns						STC
	Very strong	Strong	Weak	Missing	Very strong	Strong	Weak	Missing	WW	SW	WN	SS	SN	NN	
bio-CE-CX	<0.01	0.03	**3.59**	96.38	1.30	4.91	**29.56**	2.73	**43.75**	**5.60**	2.29	4.99	3.73	3.27	**57.12**
ChCh-Miner	<0.01	0.24	**7.70**	92.06	**100.50**	**43.76**	**35.91**	**8.93**	14.39	**13.33**	**9.59**	**20.16**	**11.69**	**5.45**	**45.31**
PP-Pathways	<0.01	0.02	1.08	**98.90**	11.79	8.05	**26.08**	2.12	**40.48**	2.14	0.67	3.04	1.21	0.50	4.04
astro-ph	**0.78**	**2.56**	**3.36**	93.30	**23.02**	**19.15**	18.00	1.84	12.38	**6.32**	**7.77**	**6.47**	**5.67**	**9.12**	**79.31**
coauth-DBLP	**3.23**	**5.33**	**3.24**	88.19	3.79	3.57	4.18	1.13	**26.79**	**13.32**	**15.85**	**9.98**	**10.71**	**13.40**	**55.27**
cond-mat-05	**0.64**	**2.67**	**3.29**	93.40	5.48	5.74	7.67	1.45	8.66	**6.29**	**7.35**	**6.58**	**6.41**	**6.52**	**54.96**
anobii	<0.01	0.05	0.55	**99.40**	2.11	2.57	9.20	1.40	1.82	0.93	0.30	1.01	0.56	0.18	0.67
linux-kernel	<0.01	0.11	0.60	**99.28**	1.58	3.83	**46.87**	2.45	0.33	0.66	0.12	3.85	0.48	0.04	0.92
enron	<0.01	0.10	0.64	**99.26**	**37.86**	5.45	18.99	1.65	9.33	2.36	3.13	3.64	1.04	0.07	1.15
fb-friends	<0.01	0.06	**2.25**	97.69	2.10	3.61	14.59	1.95	**27.30**	**4.81**	2.68	2.75	2.30	2.04	**46.27**
musae_git	<0.01	0.01	0.22	**99.77**	1.39	2.44	7.22	1.31	0.12	0.22	0.03	0.83	0.32	0.05	0.19
epinions	<0.01	0.05	0.61	**99.34**	2.14	5.80	**28.32**	2.01	5.85	0.69	0.67	1.21	0.33	0.06	3.57

it might seem as if the higher the number of common nodes around a pair of alters is, the stronger the tie they have will be. This, however, does not correspond to the research cited in the introduction. More than just the sheer number of common neighbors, it is about local connections around the tie. As illustrated in the Karate Club in Figure 2, the very strong and strong ties which we defined occur in clique-like substructures (from the smallest dyads and triads to larger network substructures). Weak ties can occur, for example, between a pair of nodes in role S. These strongly prominent nodes may have more common neighbors. However, at the same time, each of them can be connected to many other nodes in its neighborhood; it contributes to their independency (see, e.g., nodes 1 and 3 in Figure 2). As can be seen in Table 4, alter-pairs with weak ties and many common neighbors occur in the biological, communication, and social networks.

The last column of Table 4 shows the percentage of closed strong triads, i.e., those that meet the criteria for strong triadic closure (STC). An interesting feature can be observed in the degree of closedness; the shares are either relatively high, at around 50 percent or more (three co-authorship, two biological and one social network), or rather negligible; the figure is 34.80 percent for the Karate Club in Figure 2. The answer to the question of why there are such extreme differences in the closedness may be related to the roles of nodes in triads and the dependencies between them; this hypothesis, however, requires deeper analysis. Moreover, as can be seen in the middle part of the table, alter-pairs that have a common tie have, on average, a higher number of common neighbors than alter-pairs with a missing tie. This observation confirms the results of the empirical findings in [9]. Even with our alternative view on the types of ties, the triadic closure is based on a combination of two factors: neighborhood overlap and the strengths of the ties in the triad.

5.2 Role-based triad patterns in evolving networks

Figure 4 shows that similar types of networks exhibit similar properties according to the analysis of the closedness of role-based triads. We were interested in how much these properties change over time. For five of the networks studied above, their datasets contained information about growth. For each of these networks, we prepared five frames in time, which corresponds to an increase in the number of nodes by twenty percent. We got five networks containing 20, 40, 60, 80, and 100 percent of the nodes for each network; the last frames correspond to the networks analyzed in the previous sections of the paper.

In Figure 5, there are five-frame radar charts for the five networks studied, with the color of the frame corresponding to its age; the oldest frame (20 percent of the nodes) is the lightest, and the last frame of the entire network (100 percent of the nodes) is the darkest. Except for fb-friends, the network properties are very stable from the first frame; this is especially true for the co-authorship dblp network, where the proportions of closed role-based triads do not change substantially during its growth. For the fb-friends network, the properties vary only at 20 and 40 percent, after which they stabilize. So, the degree of triad closedness for individual role-based triad patterns looks like a stable parameter during network growth.

6 CONCLUSIONS AND FUTURE WORK

In the first part of the paper, we described a new perspective on the strength of ties in networks, which corresponds relatively well with the previous research cited in the introduction. We understand the strength of the tie as a consequence of the relationship between what the two nodes have in common structurally and how they are structurally different. However, in our view, the differences in a particular case can cause even two nodes which are part of a densely interconnected clique-like substructure to have a weak tie because of their mutual independency. From this perspective, our alternative approach may seem somewhat controversial in comparison with traditional approaches.

In our experiments, we studied role-based triads and especially the degree of their closedness in various real-world networks. Depending on the type of ties a node has with its neighbors, its role varies; this in turn leads to a new perspective on triads in which the ego and a pair of alters take on different roles. We have shown that looking at a network through role-based triads gives us a simple characteristic of the network type and a property that is relatively stable during network growth.

In our paper, we did not pay attention to directed networks and weights in networks in terms of the intensity of the interaction between nodes, even though structural dependency allows such an approach. Neither did we address the directions of dependencies in

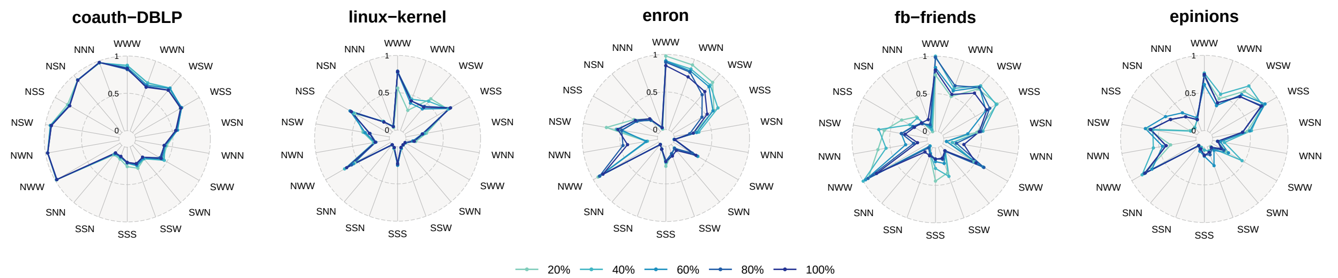

Figure 5: Evolving networks: Proportions of closed triads for 18 role-based triad patterns.

role-based triads. Both lie outside the scope of this paper and will be the subject of further research.

On the basis of the experiments that were performed, we believe that thanks to the application of asymmetry and the use of roles based on structural dependencies between network nodes, we described interesting structural properties with further potential for use. This can be especially true for the area of link prediction [16], which has been investigated intensively recently. Although recent link prediction methods often use ensemble approaches combining structure, heterogeneity and metadata analysis [5, 8, 19], we believe that our purely structural approach can make a significant contribution to this area. Our preliminary experiments so far show a prediction accuracy approaching 80 percent for all the networks studied here.

ACKNOWLEDGMENTS

This work is supported by SGS, VSB - Technical University of Ostrava, under grant no. SP2020/161, and Ministry of Health of the Czech Republic under grant no. NU20-06-00269.

REFERENCES

[1] Florian Adriaens, Tijl De Bie, Aristides Gionis, Jefrey Lijffijt, Antonis Matakos, and Polina Rozenshtein. 2020. Relaxing the strong triadic closure problem for edge strength inference. *Data Mining and Knowledge Discovery* (2020), 1–41.

[2] Monica Agrawal, Marinka Zitnik, Jure Leskovec, et al. 2018. Large-scale analysis of disease pathways in the human interactome. In *Pacific Symposium on Biocomputing*, Vol. 23. World Scientific, 111–122.

[3] Luca Maria Aiello, Martina Deplano, Rossano Schifanella, and Giancarlo Ruffo. 2012. People Are Strange When You're a Stranger: Impact and Influence of Bots on Social Networks. (2012), 10–17.

[4] Austin R Benson, Rediet Abebe, Michael T Schaub, Ali Jadbabaie, and Jon Kleinberg. 2018. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences* 115, 48 (2018), E11221–E11230.

[5] Hongxu Chen, Hongzhi Yin, Weiqing Wang, Hao Wang, Quoc Viet Hung Nguyen, and Xue Li. 2018. PME: projected metric embedding on heterogeneous networks for link prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1177–1186.

[6] Ara Cho, Junha Shin, Sohyun Hwang, Chanyoung Kim, Hongseok Shim, Hyojin Kim, Hanhae Kim, and Insuk Lee. 2014. WormNet v3: a network-assisted hypothesis-generating server for Caenorhabditis elegans. *Nucleic acids research* 42, W1 (2014), W76–W82.

[7] DJ Dekker. 2006. Measures of Simmelian tie strength, Simmelian brokerage, and, the Simmelianly brokered. (2006).

[8] Liang Duan, Charu Aggarwal, Shuai Ma, Renjun Hu, and Jinpeng Huai. 2016. Scaling up link prediction with ensembles. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. 367–376.

[9] David Easley, Jon Kleinberg, et al. 2010. *Networks, crowds, and markets*. Vol. 8. Cambridge university press Cambridge.

[10] Katherine Faust. 2010. A puzzle concerning triads in social networks: Graph constraints and the triad census. *Social Networks* 32, 3 (2010), 221–233.

[11] Zakariya Ghalmane, Chantal Cherifi, Hocine Cherifi, and Mohammed El Hassouni. 2020. Exploring Hubs and Overlapping Nodes Interactions in Modular Complex Networks. *IEEE Access* 8 (2020), 79650–79683.

[12] Mark S Granovetter. 1977. The strength of weak ties. In *Social networks*. Elsevier, 347–367.

[13] Keith Henderson, Brian Gallagher, Tina Eliassi-Rad, Hanghang Tong, Sugato Basu, Leman Akoglu, Danai Koutra, Christos Faloutsos, and Lei Li. 2012. Rolx: structural role extraction & mining in large graphs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1231–1239.

[14] Bryan Klimt and Yiming Yang. 2004. The Enron corpus: A new dataset for email classification research. (2004), 217–226.

[15] Milos Kudelka, Eliska Ochodkova, Sarka Zehnalova, and Jakub Plesnik. 2019. Ego-zones: non-symmetric dependencies reveal network groups with large and dense overlaps. *Applied Network Science* 4, 1 (2019), 81.

[16] Ajay Kumar, Shashank Sheshar Singh, Kuldeep Singh, and Bhaskar Biswas. 2020. Link prediction techniques, applications, and performance: A survey. *Physica A: Statistical Mechanics and its Applications* 553 (2020), 124289.

[17] Jérôme Kunegis. 2013. Konect: the koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*. 1343–1350.

[18] Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney. 2009. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* 6, 1 (2009), 29–123.

[19] Tiancheng Lou, Jie Tang, John Hopcroft, Zhanpeng Fang, and Xiaowen Ding. 2013. Learning to predict reciprocity and triadic closure in social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 7, 2 (2013), 1–25.

[20] Paolo Massa and Paolo Avesani. 2005. Controversial users demand local trust metrics: An experimental study on epinions. com community. In *AAAI*. 121–126.

[21] Mark EJ Newman. 2001. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences* 98, 2 (2001), 404–409.

[22] Anatol Rapoport. 1977. Contribution to the theory of random and biased nets. In *Social Networks*. Elsevier, 389–409.

[23] Burt Ronald. 1992. Structural holes: The social structure of competition. *Cambridge: Harvard* (1992).

[24] Benedek Rozemberczki, Carl Allen, and Rik Sarkar. 2019. Multi-scale Attributed Node Embedding. *arXiv preprint arXiv:1909.13021* (2019).

[25] Polina Rozenshtein, Nikolaj Tatti, and Aristides Gionis. 2017. Inferring the strength of social ties: a community-driven approach. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1017–1025.

[26] George Simmel. 2010. *Conflict and the web of group affiliations*. Simon and Schuster.

[27] Stavros Sintos and Panayiotis Tsaparas. 2014. Using strong triadic closure to characterize ties in social networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1466–1475.

[28] Marco Tortoriello and David Krackhardt. 2010. Activating cross-boundary knowledge: The role of Simmelian ties in the generation of innovations. *Academy of Management Journal* 53, 1 (2010), 167–181.

[29] Marco Van der Leij and Sanjeev Goyal. 2011. Strong ties in a small world. *Review of Network Economics* 10, 2 (2011).

[30] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. 2009. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks*. 37–42.

[31] David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, et al. 2017. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* 46, D1 (2017), D1074–D1082.