2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

The Evolution of Ego-Centric Triads: A Microscopic Approach toward Predicting Macroscopic Network Properties

Mina Doroud Prantik Bhattacharyya S. Felix Wu*
Department of Computer Science
University of California, Davis
*Blekinge Institute of Technology, Karlskrona, Sweden
Email: {sdoroud, pbhattacharyya, sfwu}@ucdavis.edu

Diane Felmlee
Department of Sociology
Pennsylvania State University
University of California, Davis
Email: dhfelmlee@ucdavis.edu

***Abstract*—Scalability issues make it time-consuming to estimate even simple characteristics of large scale, online networks, and the constantly evolving qualities of these networks make it challenging to capture a representative picture of a particular networks properties. Here we focus on the evolution of all triads (ties between three nodes) in a graph, as a method of studying change over time in large scale, online social networks. For three month snapshots, we examine, and predict, transitions among all sixteen triad types (i.e., triad census) in a sample of three years of Facebook wall-post interactions. We introduce a new sampling approach for examining triads in online graphs, based on ego-centric networks of random seeds. We examine tendencies in the data toward properties related to balance theory, including structural balance, clusterability, ranked clusters, transitivity, hierarchical clusters, and the presence of "forbidden" triads. In a time series analysis, we successfully predict the evolution over time in the wall post network dataset, with relatively low levels of error. The findings demonstrate the utility of our ego- centric, two-step, random seed sampling approach for studying large scale networks and predicting macroscopic graph properties, as well as the advantages of examining transitions in the complete triad census for an online network.**

I. Introduction

Online social networks have grown enormously over the past decades. Networks have increased both in the number of users present on the networking platform as well as in the number of connections between existing users. The enormous size and the dynamic characteristics of online social networks have created challenging problems in network analysis. Issues related to size make it time-consuming to estimate even simple graph characteristics, and the constantly evolving qualities of online networks make it problematic to capture a representative picture of a particular network's properties.

Network analysts identify at least 800 distinct metrics [1] that can be used to represent the global properties of networks, among which structural balance, clustering, ranked clusters and transitivity are particularly important. These properties can be extracted through either a macroscopic or a microscopic approach. The primary difference between the two approaches lies in scale. A macroscopic approach involves an analysis at a high level of theoretical abstraction, while a microscopic approach focuses on the individual level. In the analysis of large scale online social networks, macroscopic approaches often involve complex computation characteristics, requiring large amounts of memory and time. In contrast, microscopic approaches are inexpensive and time-efficient.

One primary microscopic approach to the analysis of a network is to examine the composition of its component triads. A network triad is defined as the collection of three actors, n_i, n_j and n_k, where $i \neq j \neq k$. Sixteen different triads can be formed among the three actors based on the direction of connection or interaction. A count of each different type of triad that arises in an observed network is defined as a triad census. A triad census can uncover global properties of the network, as will be discussed below [1].

The purpose of this paper, therefore, is to analyze online social network properties using a microscopic approach in which we focus on the evolution of triads. Our contributions can be summarized as follows. We perform a large scale study on Facebook wall-post interaction of $13,922$ users and $121,941$ interactions. We observe the transitions in triad types over time and the key global properties of the network. Additionally, we perform time series analysis to successfully predict network evolution and growth of each triad type over multiple time-stamps. The rest of the paper is organized as follows. In section II, we discuss related work in the areas of network growth modeling. Next we describe our approach to examine changes in triads over time in section III. We then present results from our measurement and analysis of the Facebook data in section IV. Finally we make concluding and suggesting for future work in sectionV.

II. Related Works

An analysis of triadic patterns represents a powerful tool for extracting global graph properties. Studies of triads date back to Simmel where he observed the fundamental properties of three actors rather than two within a network. He pointed out the possible role of the third person in interactions such as mediation and brokerage [2].

978-0-7695-4578-3 2011

DOI

Heider's balance theory maintains that particular configurations of three persons tend toward balance and stability over time, whereas others are imbalanced, stressful, and unstable [3]. A number of theorists have developed variations of balance theory, with transitivity representing one of the most widely used extensions.

In our approach, we examine triads and their evolution over time in an online social network in order to extract certain macroscopic network properties and to make predictions about the subsequent state of the network. Examining the growth of social networks over time has emerged as a challenging problem in the study of social networks. Parameters such as the degree distribution, diameter, and the clustering coefficient are typical metrics obtained in these analyses [5]–[11].

Research in this area can be primarily divided into multiple categories [4]. In the first type, the arrival of new nodes and models of node engagement are used as the basis for analysis. Formally, this method is known as stream mining [5]. In the second category of research, referred to as mining static data, the primary focus is on obtaining the macroscopic statistical properties of the network. In our studies, we focus on microscopic evolution within local groups.

Leskovec et. al [5] had a detailed study on changes in microscopic node behaviours. Their main focus was on node arrival and the edge initiation process as well as edge destination selection. They found that most edge formation is local and proceeds in the transitivity direction of triangle closure. They validated preferential attachment in the large scale networks and modelled node arrival and edge formation based on different statistical functions and triangle closure. Most functions are based on degree, common friends and the last time of activities. In another study by Golder and Yardi [12] a network's tendency toward mutuality and transitivity were observed examining several possible triads' configurations. What makes our approach different from that of previous research is that here we examine all possible triads, as subgroups of three people, and we do not restrict our observation to closing triangles only. In other words, we observe all different configurations of three nodes in the network. In addition, instead of trying to describe and reproduce the network through statistical functions, we monitor the flow of the networks triads and use these analyses to describe aspects of global network properties. Local and global predictions are provided based on the observed evolution of triads in the data.

III. Structural Balance Theory and Triads

The triad census has been used in many sociological empirical investigations and social psychological theories, such as balance and transitivity [13]. Heider's balance theory [3], later generalized as structural balance, proposes that people tend to maintain consistency or balance in their cognitions, for example they will like friends of their friends. Balance theory is usually applied to signed complete networks. However, any unsigned network can be mapped easily to a complete signed one by assuming present edges as positive and absent edges as negative. Several variations of balance theory have been developed, such as those of ranked clusters and transitivity.

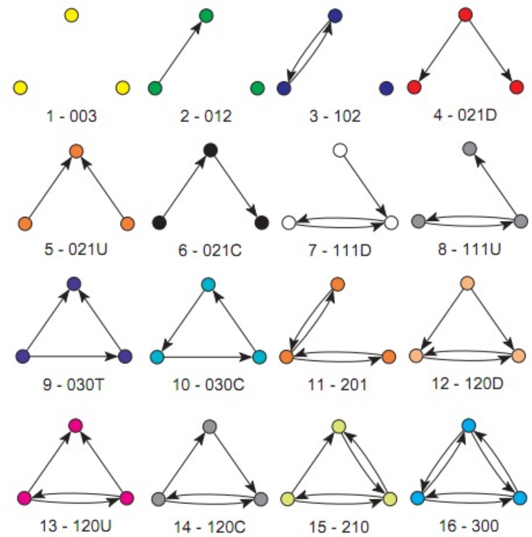


Fig. 1: The triad isomorphism classes (M-A-N labelling) [18]

Local groups of three users and the ties between them define triads. Figure 1 shows sixteen isomorphism classes for the sixty-four different triad states. What makes triads special to study is their ability to link local characteristic to global properties. Empirical and mathematical studies show that global network properties can be derived through information in a triad census [14]–[17]. In other words, instead of observing the whole network in order to derive global properties, it is sufficient to focus only on groups of three. According to the classic labelling scheme M-A-N, each triad type has a label with at most four characters, the number of mutual, asymmetric and null dyads. The fourth character, if present, is used to distinguish further among types. "D" for Down; "U" for up; "T" for transitive; "C" for cyclic.

For the network of size n there will be $\binom{n}{3}$ different triads. The triad census consists of a frequency distribution for the 16 isomorphism triad configurations. the triad census is presented in the format of a vector of length 16, mainly called T. In our method, we use random samples to estimate the census. The estimated vector is usually referred to T'_u. In section IV we show T' is enough for extracting network properties and also for the estimation of actual T.

A. Triad Census and Theories of Balance and Transitivity

A triad census can be used to infer graph properties such as structural balance, clusterability, ranked clusters, and transitivity. In a triad census, different distributions of triad categories represent various theoretical models. There are five main models which can be inferred based on the triad census. The first model is balance, which consists only of triads 102, and 300. This model only allows for symmetric ties within a cluster and no ties between them, and only allows for at most two clusters within a group of three nodes. Also it does not allow any ties between ranks. The next model is clusterability, where triads 102, 300 and 003 are permitted. It has the same properties as the balance model, but it does not have a restriction on the maximum number of clusters. The ranked cluster model is the third model, and it has the same definition as the previous model, but the extra assumption that it allows for asymmetric ties from each vertex to all vertices on higher ranks. It will permit triad types 102, 300, 003, 021D, 021U, 030T, 120D, 120U. The transitivity model relaxes the assumption of null ties between ranks, which includes all mentioned triads in the ranked clusters model beside triad type 012. Finally, hierarchical clusters is the most relaxed model, and it allows for asymmetric ties within a cluster, provided that they are acyclic. It includes all of the previously mentioned triad types, as well as triad types 120C, 210.

The remaining five types of triads, 021C, 111D, 111U, 030C, 201, are not encompassed under any of the models. They are called "forbidden" triads since they contradict all of the five, balance theoretic models. They also violate assumptions regarding symmetry of dyads, on which the overall models are based. If these types of forbidden triads occur frequently, they put into question the assumptions of balance incorporated into the above theoretical, social network models.

T	EF	T	EF	T	EF	T	EF
1	N^3	5	$\frac{3}{4}NA^2$	9	$\frac{3}{4}A^3$	13	$\frac{3}{4}MA^2$
2	$3AN^2$	6	$\frac{3}{2}NA^2$	10	$\frac{1}{4}A^3$	14	$\frac{3}{2}MA^2$
3	$3MN^2$	7	$3MAN$	11	$3NM^2$	15	$3AM^2$
4	$\frac{3}{4}NA^2$	8	$3MAN$	12	$\frac{3}{4}MA^2$	16	M^3

TABLE I: Expected Frequency under M-A-N labelling
Triad's Type(T), Expected Frequency(EF)

Social networks are noisy and rarely represent perfect balance in empirical data. Moreover, it is not the presence of particular triads which define network properties, but their distance from a distribution in a random network with the same number of nodes and links. Table I shows formulas to compute the expected triad types in a random network [19], [20]. Notations are the same as the MAN labelling where M represents the number of mutual dyads, A the number of asymmetric, and N the nulls dyads, where $M+A+N = \frac{n(n-1)}{2}$ and $z^k = z(z-1)\cdots(z-k+1)$. When there is a large discrepancy between the actual number of triads and that expected by chance, a tendency toward a specific model can be assumed. In particular, if the various models of balance help to explain a network's structure, then the forbidden triads should occur less frequently than expected by chance [18], [21].

B. The evolution of a triad census

Any snapshot of a social network is likely to consist of two genres of triads, those that have already reached stability, or a state of equilibrium, and those that are in transition, and progressing towards a more stable state. Although a triad census is a strong tool for revealing network properties, it cannot detect all possible hidden tendencies. By relying only on a triad census, we may not be able to capture deep tendencies toward change. For example, a semi balanced network with a very low level of change could have the same triad census as a highly changing network that consists of many new arriving nodes and ties that are on their way to stability. In a highly evolving network, stability may not occur, simply because not enough time has passed for all triads. This claim is analogous to the problem in statistics referred to as right-censoring, or censor Type I. In order to overcome this disadvantage, here we observed the evolution of triads in a situation in which we could monitor the probability of changes between types. With the emergence of new ties, a triad's status changes. The probability of inter state changes could reveal a lot about network's properties. For example, two different networks with the same triad census could act differently during the time, and what makes them different is not their triad census at a certain snapshot, but probabilities of staying in one state or moving forward. Figure 2 and 3 shows the possible evolution of triads in one or multiple steps.

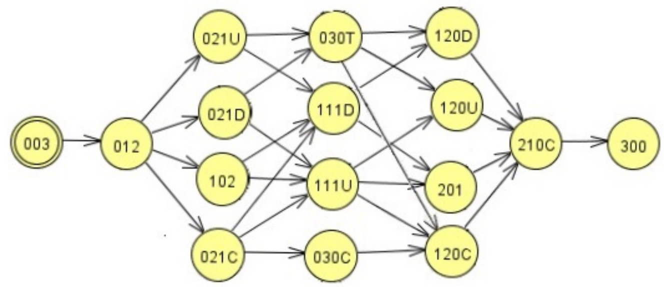


Fig. 2: Triads' evolution within one step

Note that if a particular triad changes by adding or deleting one tie at a time, then it is inevitable that some triads will pass through a forbidden state in order to reach a more stable state. Yet we would expect the transition probabilities to differ between forbidden and non forbidden triads. In previous research, Sorensen and Hallinan [22] studied triad change among a small, network of 28 students using a continuous time discrete state markov chain model relying on the transitivity model. They found that there was a tendency towards more density of ties, with a gradual decrease in triads of types 003 and 012, and an increase in triad type 300.

C. Dataset

To investigate triad evolution and its ability to extract network characteristics, we rely on a Facebook dataset consisting of wall posts. The data set was gathered by Viswanath et. al. [23] and related to 13,922 users and spans interactions from September 2004 to January 2009. We use the network's status at timestamp July 2007 as a basic configuration of the network. It consists of more than 13,922 nodes and 121,941 interactions. We divided the rest of the dataset into 6 different intervals so that we could observe microscopic changes within the network between these snapshots, each of which consists of ties that occurred within three months.

D. Notation

Networks are mainly represented by a graph $G = (V, E)$ in which each edge $e = (u, v) \in E$ represents a social interaction between u and v that took place at a particular time $t(e)$. In the dataset we use, $e = (u, v)$ represents a situation in which the second user v , posted on the first user's wall, u. $G[t, t']$ denotes the sub-graph of G which consist of all the interactions in which the timestamp is greater than t and smaller than t'. Besides the basic configuration of the network, $G[t_0, t_0']$, we divide the social graph into 6 different time intervals, as $t_i < t_i'$ and $t_i' < t_{i+1}$ for $i = 0, \cdots, 6$. We examine the network's characteristics from $G[t_0, t_0']$ to $G[t_4, t_4']$ and then use this information to attempt to predict the network properties and status at $G[t_5, t_5']$ and $G[t_6, t_6']$. In learning language $[t0, t0']$ to $[t4, t4']$ are known as training intervals while $[t_5, t_5']$ and $[t_6, t_6']$ are known as the test intervals. Our focus is on predicting T_5' and T_6' based on T_0' to T_4'.

We attempt to predict the actual Ts with network's ego-centric information. Such a prediction task is important, because by having an accurate estimation of vector T, we will be able to compute and compare the estimation of the actual number of each triad type with the expected frequency. Furthermore, we can examine the network's overall tendency towards a particular type of balance model. Such an approach is especially useful in cases in which the network is either very large or it is subject to substantial change, both of which are common in online social networks.

In most social networks we have mainly two different types of growth, growth in edges and growth in nodes. The former happens as a result of new interaction formation within the network, while the latter is the result of new actors joining the network. In this study we only focus on growth in density, which is the formation of new interactions. Hence, we omit new nodes that join the network following the base configuration. In other words we examine ties among the same set of nodes in all subgraphs G_i for $i = 1 \cdots 6$ as those we have in G_0; by G_i we mean $G[t_i, t_i']$.

E. Ego-Centric Random Selection

The number of possible triads in a typical, online network can be extremely large. As a result, observing the evolution over time of all possible triads in an online network is inefficient and cumbersome. It is inefficient, because often a very large proportion of triads consist of type 1-003, that is all null dyads, and they will stay there forever. For example in our data set there are 44×10^{10} triads of which 0.998% are triad type 003, null in time t_0' . Out of this number, only 50×10^7 triads move forward out of the null stage, which represents less than 0.01% of all the null triads in the base configuration.Therefore, observing all possible triads in a sparse environment, such as a Facebook wallpost network, is inefficient. To overcome this limitation here we introduced a new, more efficient method of observing a triad census, one that is based on ego-centric networks of random seeds, which we describe in more detail below. Other options for analyzing fewer triads could be selecting random sets of three nodes (n_i, n_j, n_k) from the network. This method is particularly inefficient since the low density of large social networks will still result in the selection of many null triads. The other trivial option for omitting null triads could be to choose a random edge e_{ij} and a random node n_k that are not involved in the edge $k \neq i$ and $k \neq j$. Although much improvement is seen in this approach over the random selection of nodes, it still has substantial disadvantages, mainly because randomness in the selection of nodes usually results in placing node n_k many hops away which will produce many triads of types 012 and 102 that consist of either only one asymmetric dyad, or one mutual dyad without any progress forward. What makes the ego-centric approach special is its particularly efficient way of including those triads that have a relatively high chance of becoming involved in some evolution over time.

Ego-centric random selection begins with a random selection of a particular number of nodes as seeds. Then each seed is used as an ego, as a basis for the ego-centric, two-step local network. The ego-centric network, $ECN(ego)$, consist of an ego and all nodes within two hops of it, using either in or out links and all ties between them. A triad is formed by observing all sets of (ego, n_i, n_j) where $n_i, n_j \in ECN(ego)$. Two nodes n_i and n_j could be as far as four hops from each other, while both of them are at most two hops away from ego. There are several advantages of observing triads within an ego-centric network rather than a random selections of triads. First, this approach enables us to eliminate a large portion of null triads, type 003, in the social network. Instead, we observe change only in those triads that have a greater chance of being involved in a triad transition over time. Moreover, an ego-centric approach is a practical way of identification hidden tendencies within a network, such as tendencies toward mutuality, transitivity or clusterability.

In our ego-centric approach toward triad selection we only focus on $\sum_{ego} \binom{|ECN(ego)|}{2}$ where by $|ECN(ego)|$ we refer to the size of an ego-centric network of node ego. We do this instead of examining all possible triads in a graph $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$. Next we show that this is an efficient way of capturing patterns of triad evolution in large graphs, and that it also can be used as a good predictor of the triad census associated with the entire graph and its future state.

IV. Results

When allowing multiple movements over time, there are 105 possible transitions between different types of triads. These possible transitions are depicted in the transition matrix in Table **??**. In our ego-centric approach toward the triad census, we monitor all possible triads in 6 different snapshots. The probability of a transition from triad type i to j is defined by $P_{ij}^k = \frac{S_{ij}^k}{S_i^k}$ it could be computed through $S_i^k = \sum_j S_{ij}^k$ where S_{ij}^k represents the number of triads that move from type i to type j at snapshot k where S_i^k represents the total

	003	012	102	021D	021U	021C	111D	111U	030T	030C	201	120D	120U	120C	210	300
003	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
012		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
102			x				x	x			x	x	x	x	x	x
021D				x				x		x	x	x	x	x	x	x
021U					x		x		x		x	x	x	x	x	x
021C						x	x	x	x	x	x	x	x	x	x	x
111D							x				x	x			x	x
111U								x			x		x	x	x	x
030T									x			x	x	x	x	x
030C										x				x	x	x
201											x				x	x
120D												x			x	x
120U													x		x	x
120C														x	x	x
210															x	x
300																x

Fig. 3: Possible interstate transition

number of triads of type i at the beginning of the snapshots. The probabilities of diagonal elements depict the networks' tendency toward staying in the same state rather than changing over time. The higher the diagonal probabilities, the less dynamic is the network. The more distant a triad has from the diagonal, the greater the tendency it exhibits to change.

Table III shows the average probability of transition between time stamps for our three different samples. For a large scale social network with a low density we expect a very high staying probability. As we see, the five forbidden triads (6, 7, 8, 10, and 11) are involved in a number of transitions. Passing through forbidden triads as an interim stage is inevitable, if a triad is evolving towards a stable state on a step by step basis. However, the rate of staying in a forbidden state or moving forward reveals a lot of information. Previous research on the same dataset demonstrates that the waiting time is shorter when a triad exits a forbidden or intransitive state than when it moves out of a state that is transitive or not forbidden. In other words, although forbidden triads occur with some frequency, they tend to change relatively quickly [24].

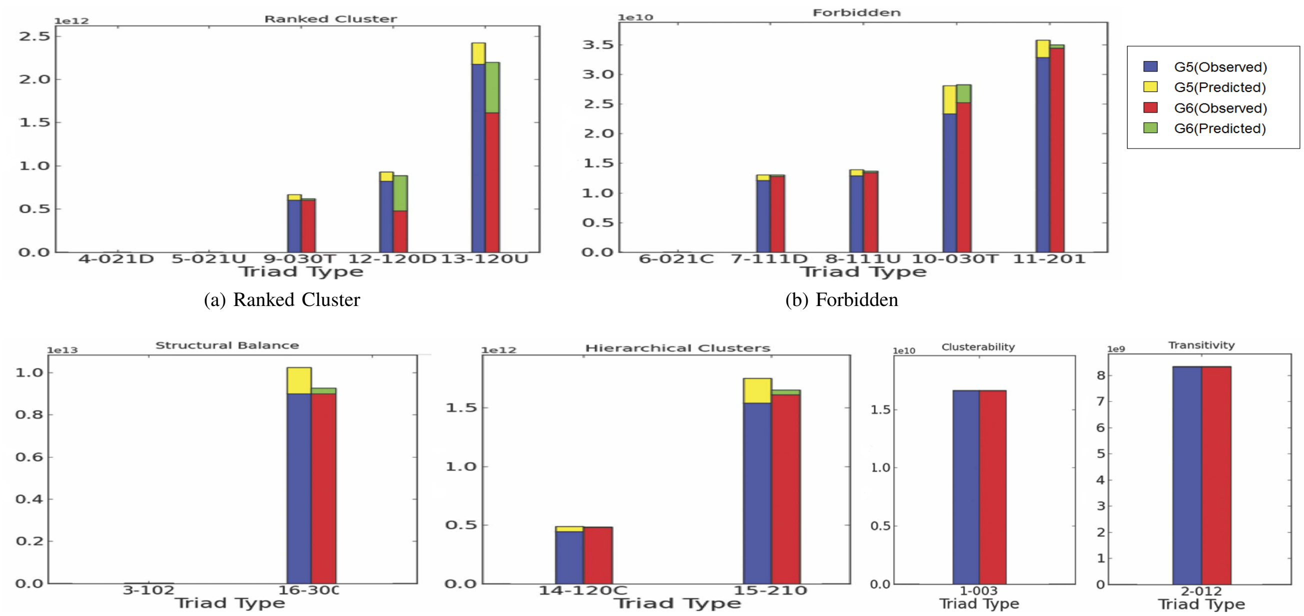
A. *Probability*

P^i_{ij}, the probability of change in triads in transition from G_i to G_{i+1} where $i \in \{0, 1, \cdots 5\}$ is shown in Table III. The triad types, 1 (003) through 16 (300), are defined in Table **??**. This table only contains the probabilities for those movements that had more than a 0.001% chance of change in at least one transition in one of the three samples we used. The rows in bold show the most probable type of move out of each triad type i. For example, as can be seen in the first six columns of Table III, the most frequent transition for a triad consisting of one asymmetric tie (2-012), consisted of a change to a triad with one mutual tie (3-102). The most popular path of evolution for triads in our data set is toward the following triads; $003 \rightarrow 102 \rightarrow 111D \rightarrow 201 \rightarrow 210 \rightarrow 300$. Three of these six triads involve mutual dyads (either one, two, or three), and therefore these findings show evidence of movement towards mutuality. More generally, the results depict that any step in movement towards increasing mutuality is particularly likely. Next, we used linear regression to predict the probability of a transition between snapshots 4 to 5 and snapshots 5 to 6 based on the network's behavior in the prior transitions. The last column depicts the root-mean-squared deviation, RMSD, which can be thought of as a measurement of the size of the error in our predictions based on the regression analysis. Note that the sizes of all the RMSDs are quite small, and are all less than 0.13. In addition, all but two of the possible transitions that originate with Triad 10-030C, are larger than 0.046. The relatively small overall error rate demonstrates that we are able to predict the future configuration of network, its tendency toward change and its triadic configuration. In the next section we show how we use these data to predict the triad census of the larger graph and verify extensions of balance theory within the network by observing a relatively few number of triads.

B. *Census*

As discussed above, a random ego-centric approach is a relatively efficient way of monitoring the evolution of triads in a network. Not only is it successful in reflecting both macroscopic characteristics of the network, as well as microscopic characteristics, but it is also useful in predicting the census of the whole graph. In other words, if we know the hidden tendencies in the network, monitoring the changes over time in a limited sample of nodes is enough to predict the actual triad census of the network at a later time point. As a result, certain overall characteristic of networks, such as their tendency toward mutuality and hierarchy, can be figured out if we have information from an early census of the graph. Computing the triad census for a social graph at one snapshot is not an expensive or difficult task. What remains costly, however, is computing a triad census for dynamic and large, online networks that may change on a daily basis. As we show in Figure 4, surprisingly accurate predictions can be made concerning the actual frequency of triads based on observing only a relatively few of them in a random ego-centric approach. By observing triads within an ego-centric network of only 900 nodes we could estimate the actual frequency of triads in the network with an average RMSD of 0.0353.

The triad census consists of three values of n_i, e_i and the relative difference between these two. Table II shows that we could estimate values of n_i, So we only need values of e_i to be able to come with $\frac{(n_i - e_i)}{e_i}$ and then make judgements about networks properties and the dominant model.



(a) Ranked Cluster

(b) Forbidden

(d) Balance

(e) Hierarchical Clusters

(f) clusterability

(g) Transitivity

Fig. 4: Comparison of observed and predicted values of $\frac{(n_i - e_i)}{e_i}$

Computing e_i is not that difficult since it could be easily computed from triads. We use T_i as the actual frequency and T'_i as an estimation of a triad type i.

In the same manner we could define e_i and e'_i. In Table I in Section III we demonstrate how e_i is computed by using the dyad census M,A and N. The estimation of the dyad census is then computed from our frequency estimation n'_i as follows. Since they are estimates, and not the actual values, we depict them by M',A' and N'. They are estimated as follow:

$$A' = \frac{1}{\sum_{i=1}^{n} n' - i - 2} \times \Big(3(T'_9 + T'_{10}) + 2(T'_2 + T'_4 + T'_5 + T'_6 + T'_{12} + T'_{13} + T'_{14}) + T'_{15} + T'_7 + T'_8\Big) \quad (1)$$

$$N' = \frac{1}{\sum_{i=1}^{n} n'_i - 2} \times \Big(3(T'_1) + 2(T'_2 + T'_3) + T'_4 + T'_5 + T'_6 + T'_7 + T'_8 + T'_{11}\Big) \quad (2)$$

$$M' = \frac{1}{\sum_{i=1}^{n} n'_i - 2} \times \Big(3(T'_{16}) + 2(T'_{15} + T'_{11}) + T'_{14} + T'_{13} + T'_{12} + T'_7 + T'_8 + T'_3\Big) \quad (3)$$

Having M',A' and N', e'_i could be easily computed for $i = 1 \cdots 16$. As noted before, what influences the evaluation of the tendency of the network is not its frequency, nor its expected frequency, but how much these two differ. In Figure 4 we compare $\frac{(n_i - e_i)}{e_i}$ and $\frac{(n'_i - e'_i)}{e'_i}$ for testing intervals 5 and 6. As can be seen in Figure 4, the actual and expected values for the various triads are matched extremely closely. The results in these figures provide strong evidence of the accuracy of the ego-centric approach we describe above. Figure 4 demonstrates that our approach extracts almost the identical results as does the actual, empirical triad census. Indeed it is as useful in extracting the global properties of our graph as is the triad census. For example, as both $\frac{(n_i - e_i)}{e_i}$ and $\frac{(n'_i - e'_i)}{e'_i}$ suggest, the wallpost data set shows a high tendency toward balance, and hierarchical clusterability.

V. Concluding Remarks & Future Work

In this work, we examined the evolution of the triad census as a tool to verify network properties of large scale, online social network graphs in an efficient and inexpensive manner. We introduced the use of an ego- centric, two step, random seed, sampling approach to obtain a empirical network in which to examine transitions in the triad census. We observed that the properties predicted by the triad census of individual networks closely matched the triad census of the actual data. We also established that the triad census approach can be used to efficiently extract and verify global properties of the network. Finally, we undertook two sets of time series analyses to predict network changes over time, and we verified that our predictions closely matched the empirical data, with relatively low rates of error. In future work we want to apply the triad census approach, and the ego-centric, random seed sampling method, to compare changes over time in other types of large scale, social network data, such as those obtained from Twitter

Triad	Frequency(Interval 5)	Prediction(Interval 5)	Frequency(Interval 6)	Prediction(Interval 6)	rmsd
1-003	448533859933.0	448475403926.0	448460849133.0	448480239428.0	0.0001
2-012	614176381.0	637783022.123	652592306.0	656659850.05	0.0430
3-102	485143509.0	507366527.293	519524096.0	524312987.09	0.0584
4-021D	147667.0	146561.223782	164196.0	160624.069092	0.0393
5-021U	296705.0	294489.083538	328490.0	325478.044129	0.0214
6-021C	152339.0	153013.825315	171627.0	169341.982803	0.0251
7-111D	346903.0	349448.342403	395650.0	394576.229461	0.0150
8-111U	371269.0	372238.246191	415453.0	414842.342445	0.0055
9-030T	10258.0	10338.9471599	11376.0	11312.9076131	0.0161
10-030C	144.0	134.08797654	174.0	158.134897361	0.2100
11-201	377538.0	378269.62461	423261.0	423643.790876	0.0043
12-120D	5641.0	5644.54013356	6525.0	6550.67919087	0.0085
13-120U	14726.0	14909.40809	16168.0	16277.3625223	0.0266
14-120C	2968.0	3061.0227511	3555.0	3613.57305893	0.0770
15-210	16869.0	16854.9554356	19379.0	19379.2867832	0.0018
16-300	12990.0	13054.3109807	14451.0	14394.3415757	0.0139

TABLE II: Census Regression

and email correspondence. A comparison of different networks will demonstrate the types of online interactions that can be predicted most, and least, accurately using triad census information. More generally, we plan to apply our findings to the important task of network link prediction.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for their insightful comments. This work was supported by the National Science Foundation FIND (Future Internet Design) program under Grant No. 0832202, GENI, MURI under ARO (Army Research Office), and was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053.

REFERENCES

[1] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences), 1st ed., ser. Structural analysis in the social sciences, 8. Cambridge University Press, Nov. 1994.
[2] G. Simmel and K. H. Wolff, The Sociology of Georg Simmel. Free Press, Aug. 1964.
[3] F. Heider, Attitudes and cognitive organization, J Psych, vol. 21, pp. 107-112, 1946.
[4] M. Spiliopoulou, Evolution in Social Networks: A Survey, Social Network Data Analytics, Springer US, 2011, pp. 149-175
[5] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, Microscopic evolution of social networks, in Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ser. KDD '08. New York, NY, USA: ACM, 2008, pp. 462-470.
[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationships of the internet topology, in Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication, ser. SIGCOMM '99, vol. 29, no. 4. New York, NY, USA: ACM, Oct. 1999, pp. 251-262.
[7] S. Strogatz. Exploring complex networks. Nature, 410, 2001.
[8] R. Albert and A.-L. Barabsi. Statistical mechanics of complex networks. Reviews of Modern Physics, 74, 47, 2002
[9] M. Newman. The structure and function of complex networks. SIAM Review, 45, 2:167256, 2003.
[10] S. Dorogovtsev and J. Mendes. Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford Univ. Press, 2003.
[11] A. Broder, S. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. L. Wiener. Graph structure in the web. Computer Networks/WWW, 33(1-6):309320, 2000.
[12] Golder, S., Yardi, S. (2010). Structural Predictors of Tie Formation in Twitter: Transitivity and Mutuality. In Proceedings of the IEEE Conference on Social Computing (SocialCom'10). Minneapolis, MN, Aug 20-22, 2010.
[13] Felmlee, Diane H. 2003. AInteraction in Social Networks.@ PP. 389-409 in J. Delamater (Ed.) /Handbook of Social Psychology/. New York: Kluwer/Plenum.
[14] J. A. Davis, Clustering and structural balance in graphs. Human Relations, 1967 20:181-187
[15] E. C. Johnsen, Network macrostructure models for the Davis-Leinhardt set of empirical socio-matrices. Social Networks 7:203-224. 1985.
[16] D. N. Wouter, The sign of affection: Balance-theoretic models and incomplete signed digraphs, Social Networks, Volume 21, Issue 3, July 1999, PP 269-286
[17] J. A. Davis and S. Leinhardt, The structure of positive interpersonal relations in small groups, in J. Berger (ed.) Sociological Theories in Progress, Vol. 2. 1970 Boston: Houghton Mifflin.
[18] W. de Nooy, A. Mrvar, and V. Batagelj, Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Sciences), illustrated edition ed. Cambridge University Press, Jan. 2005.
[19] P. W. Holland and S. Leinhardt, A method for detecting structure in sociometric data, The American Journal of Sociology, vol. 76, no. 3, pp. 492-513, 1970.
[20] P.W. Holland, S.Leinhardt, Local Structure in Social Networks. Sociological Methodology 1976, 7:145.
[21] K. Faust, A puzzle concerning triads in social networks: Graph constraints and the triad census, Social Networks, Apr. 2010.
[22] A. B. Sorensen and M. T. Hallinan, A Stochastic Model for Change in Group Structure, Social Science Research 5, 43-61. 1976.
[23] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, On the evolution of user interaction in facebook, in Proceedings of WOSN'09, Aug. 2009.
[24] M. Doroud, D. Felmlee, S. F. Wu, Mutuality and Triadic Development in Two Years of Facebook Interactions, American Sociological Association annual meeting 2011

Transition	T_0-T_1	T_1-T_2	T_2-T_3	T_3-T_4	T_4-T_5	$Prediction_{T_4-T_5}$	T_5-T_6	$Prediction_{T_5-T_6}$	rmsd
1(003)-1(003)	0.986	0.989	0.991	0.991	0.991	0.9935	0.991	0.9952	0.0035
1(003)-2(012)	**0.008**	**0.006**	**0.005**	**0.005**	**0.005**	**0.0035**	**0.005**	**0.0025**	**0.0021**
1(003)-3(102)	0.005	0.003	0.002	0.002	0.002	0.0005	0.001	-0.0005	0.0015
2(012)-2(012)	0.945	0.945	0.952	0.957	0.957	0.9605	0.964	0.9648	0.0025
2(012)-3(102)	**0.037**	**0.041**	**0.036**	**0.032**	**0.032**	**0.0315**	**0.026**	**0.0295**	**0.0025**
2(012)-4(021D)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.0000
2(012)-5(021U)	0.006	0.005	0.003	0.003	0.003	0.0015	0.003	0.0004	0.0021
2(012)-6(021C)	0.003	0.002	0.001	0.002	0.002	0.001	0.002	0.0006	0.0012
2(012)-7(111D)	0.002	0.002	0.001	0.001	0.001	0.0005	0.001	0.0001	0.0007
2(012)-8(111U)	0.002	0.001	0.001	0.001	0.001	0.0005	0.001	0.0002	0.0007
2(012)-11(201)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
3(102)-3(102)	0.984	0.987	0.99	0.99	0.99	0.993	0.991	0.9951	0.0036
3(102)-7(111D)	**0.005**	**0.004**	**0.002**	**0.003**	**0.003**	**0.0015**	**0.002**	**0.0007**	**0.0014**
3(102)-8(111U)	0.004	0.003	0.002	0.002	0.002	0.001	0.002	0.0003	0.0014
3(102)-11(201)	**0.006**	**0.004**	**0.003**	**0.003**	**0.003**	**0.0015**	**0.002**	**0.0005**	**0.0015**
4(021D)-4(021D)	0.912	0.901	0.904	0.928	0.928	0.924	0.937	0.9291	0.0063
4(021D)-8(111U)	**0.064**	**0.079**	**0.075**	**0.054**	**0.054**	**0.0595**	**0.048**	**0.0561**	**0.0069**
4(021D)-9(030T)	0.007	0.004	0.005	0.006	0.006	0.005	0.005	0.0048	0.0007
4(021D)-11(201)	0.01	0.012	0.009	0.007	0.007	0.0065	0.005	0.0053	0.0004
4(021D)-12(120D)	0.003	0.001	0.002	0.001	0.001	0.0005	0.001	0.0	0.0008
4(021D)-13(120U)	0.001	0.0	0.001	0.0	0.0	0.0	0.0	-0.0002	0.0001
5(021U)-5(021U)	0.959	0.955	0.956	0.958	0.958	0.9565	0.964	0.9563	0.0055
5(021U)-7(111D)	**0.032**	**0.035**	**0.037**	**0.033**	**0.033**	**0.0355**	**0.028**	**0.036**	**0.0059**
5(021U)-9(030T)	0.003	0.002	0.001	0.002	0.002	0.001	0.002	0.0006	0.0012
5(021U)-11(201)	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.0029	0.0001
5(021U)-13(120U)	0.001	0.001	0.001	0.0	0.0	0.0	0.0	-0.0003	0.0002
6(021C)-6(021C)	0.903	0.894	0.904	0.916	0.916	0.9165	0.929	0.9214	0.0054
6(021C)-7(111D)	**0.046**	**0.041**	**0.04**	**0.043**	**0.043**	**0.04**	**0.031**	**0.039**	**0.0060**
6(021C)-8(111U)	**0.04**	**0.054**	**0.045**	**0.032**	**0.032**	**0.0345**	**0.034**	**0.0312**	**0.0027**
6(021C)-9(030T)	0.005	0.004	0.004	0.003	0.003	0.0025	0.001	0.0019	0.0007
6(021C)-11(201)	0.002	0.003	0.003	0.001	0.001	0.0015	0.001	0.0012	0.0004
6(021C)-13(120U)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
6(021C)-14(120C)	0.0	0.0	0.0	0.001	0.001	0.001	0.0	0.0013	0.0009
7(111D)-7(111D)	0.947	0.944	0.949	0.962	0.962	0.963	0.967	0.968	0.0010
7(111D)-11(201)	**0.046**	**0.049**	**0.046**	**0.032**	**0.032**	**0.032**	**0.028**	**0.0275**	**0.0004**
7(111D)-12(120D)	0.002	0.002	0.001	0.001	0.001	0.0005	0.002	0.0001	0.0014
7(111D)-14(120C)	0.0	0.0	0.0	0.0	0.0	0.0	0.001	0.0	0.0007
7(111D)-15(210)	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.0009	0.0001
8(111U)-8(111U)	0.95	0.951	0.954	0.955	0.955	0.957	0.963	0.9588	0.0033
8(111U)-11(201)	0.039	0.039	0.039	0.037	0.037	0.037	0.03	0.0364	0.0045
8(111U)-13(120U)	**0.006**	**0.004**	**0.003**	**0.003**	**0.003**	**0.0015**	**0.003**	**0.0005**	**0.0021**
8(111U)-14(120C)	0.001	0.0	0.0	0.001	0.001	0.0005	0.0	0.0005	0.0005
8(111U)-15(210)	**0.002**	**0.002**	**0.001**	**0.001**	**0.001**	**0.0005**	**0.001**	**0.0001**	**0.0007**
9(030T)-9(030T)	0.916	0.914	0.91	0.904	0.904	0.901	0.923	0.897	0.0185
9(030T)-12(120D)	0.031	0.028	0.036	0.033	0.033	0.0355	0.03	0.0369	0.0052
9(030T)-13(120U)	**0.038**	**0.044**	**0.036**	**0.038**	**0.038**	**0.037**	**0.032**	**0.0362**	**0.0031**
9(030T)-14(120C)	**0.008**	**0.005**	**0.01**	**0.015**	**0.015**	**0.016**	**0.012**	**0.0186**	**0.0047**
9(030T)-15(210)	**0.005**	**0.006**	**0.006**	**0.007**	**0.007**	**0.0075**	**0.002**	**0.0081**	**0.0043**
9(030T)-16(300)	0.002	0.0	0.0	0.0	0.0	-0.001	0.0	-0.0016	0.0013
10(030C)-10(030C)	0.923	0.85	0.74	0.892	0.892	0.8005	0.937	0.7802	0.1284
10(030C)-14(120C)	**0.076**	**0.15**	**0.259**	**0.107**	**0.107**	**0.1985**	**0.062**	**0.2187**	**0.1283**
10(030C)-15(210)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
11(201)-11(201)	0.991	0.993	0.993	0.994	0.994	0.995	0.994	0.9959	0.0015
11(201)-15(210)	**0.004**	**0.004**	**0.003**	**0.003**	**0.003**	**0.0025**	**0.003**	**0.0021**	**0.0007**
11(201)-16(300)	0.003	0.002	0.002	0.002	0.002	0.0015	0.001	0.0012	0.0004
12(120D)-12(120D)	0.912	0.94	0.912	0.944	0.944	0.944	0.944	0.9508	0.0048
12(120D)-15(210)	**0.081**	**0.057**	**0.082**	**0.048**	**0.048**	**0.0485**	**0.053**	**0.0411**	**0.0084**
12(120D)-16(300)	0.005	0.001	0.004	0.007	0.007	0.0065	0.002	0.0074	0.0038
13(120U)-13(120U)	0.967	0.963	0.952	0.959	0.959	0.9515	0.958	0.948	0.0088
13(120U)-15(210)	**0.028**	**0.033**	**0.044**	**0.036**	**0.036**	**0.044**	**0.039**	**0.0475**	**0.0083**
13(120U)-16(300)	**0.004**	**0.003**	**0.002**	**0.003**	**0.003**	**0.002**	**0.001**	**0.0016**	**0.0008**
14(120C)-14(120C)	0.912	0.868	0.883	0.883	0.883	0.8685	0.936	0.8613	0.0538
14(120C)-15(210)	**0.087**	**0.128**	**0.109**	**0.11**	**0.11**	**0.121**	**0.062**	**0.126**	**0.0459**
14(120C)-16(300)	0.0	0.002	0.006	0.005	0.005	0.008	0.001	0.0099	0.0066
15(210)-15(210)	0.94	0.924	0.93	0.943	0.943	0.938	0.961	0.9395	0.0156
15(210)-16(300)	**0.059**	**0.075**	**0.069**	**0.056**	**0.056**	**0.061**	**0.038**	**0.0595**	**0.0156**
16(300)-16(300)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0000

TABLE III: Transition