TRANSITIVITY MATRIX OF SOCIAL NETWORK GRAPHS

Mohammad Aghagolzadeh, student member IEEE, Iman Barjasteh, Hayder Radha, IEEE Fellow

Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI, USA
E-mails: aghagol1, barjaste, radha@msu.edu

ABSTRACT

Transitivity in friendship graphs has been well known as a key property of social networks. In this paper, we extend the graph transitivity index by introducing a new characteristic quantity of graphs, namely the transitivity matrix. The transitivity matrix measures the microscopic impact that each link has on the global transitivity index of the graph. We argue that the new transitivity metric is synergetic with key aspects of the social science literature on social network theory, and we show that it can be used as a tool for locating bridges as well as redundant links. This work represents a major departure from the traditional graph transitivity index, which provides a coarse measure for how transitive the overall network is.

Index Terms— transitivity, social networks.

1. INTRODUCTION

A well-known small-scale attribute of friendship graphs is transitivity, which is defined as the tendency among two nodes to be connected if they share a mutual neighbor [1]. The transitivity index is a global metric quantifying the tendency of this small-scale attribute over the entire graph; it is proportional to the ratio of the number of triangles over the total number of connected triples. However, since transitivity is not uniformly distributed throughout the network, the transitivity index alone does not reveal the variation of this attribute along the network. Granovetter in his seminal work of "strength of weak ties" [2, 3] explains that strong social links are transitive and result in redundant social structures like cliques. On the other hand, bridging links that lie between communities are weak and do not follow the transitivity relations. This viewpoint asserts that transitivity, rather than a general feature of social networks, is a function of the type of links. Using this analogy, we assign transitivity scores to social links to reveal their functionality in the social network.

Most prior work in social network analysis have been node-centric; for example, in the leader/follower problem, which has attracted a great deal of attention recently, nodes are categorized as being either influential (leaders) or under the influence (followers). From a link-centric view and in agreement with relevant social science theories [2, 3], one can argue that the influence of a node is a function of the redundancy of its social ties; for example, a node with many redundant links might be less influential than a low-degree node with few links that act as bridges between communities. Therefore, it is arguably beneficial to look into microscopic substructures of networks and measure to what extend their links are redundant—showing transitive attributes—or otherwise strategic bridges in the social capital.

In what follows, we describe the roadmap of this paper. First, we derive a closed-form expression for the transitivity index of general weighted graphs with weight matrix W—we call it the transitivity function $\tau(W)$. The transitivity function reverts back to the original graph transitivity index for binary graphs. To measure the impact of individual links toward the global transitivity index of the graph, we employ a differential analysis. We define the transitivity gradient matrix $\nabla_W \tau$ as the gradient of the transitivity function, which is subsequently used to derive the transitivity matrix T. Each entry $[T]_{ij}$ corresponds to the link transitivity score of the i—j link. Link transitivity can take both positive and negative values-negative if it has a negative effect on the global transitivity index. An important property of T is that the sum of its entries is equal to the graph transitivity index which shows the consistency in our definition of the link transitivity.

In the area of community analysis, we make the following contribution. According to the network theory proposed by Borgatti [3] communities with many redundant links, resulting in high transitivity scores, have strong local cohesion but weak global cohesion. On the other hand, communities with many bridges, resulting in low transitivity scores, have weak local cohesion but strong global cohesion. Therefore, the transitivity matrix can be utilized to assess the local vs. global cohesion of communities. We should note that, in this paper, we do not attempt to detect communities; rather we provide the means for structural analysis of target communities. This paper is organized as follows. In Section 2, we provide our definition of the continuous transitivity function. In Section 3, we derive the transitivity matrix. Section 4 is about the utility of the transitivity matrix for social networks analysis. We conclude this paper in Section 5.

This work is supported by the National Science Foundation.

2. THE TRANSITIVITY FUNCTION

A social network is made up of a set of agents interacting with each other. A network graph may represent the frequency of interactions among its nodes (agents) or it can show the social ties among these agents. Nevertheless, different graphs of a social network are not independent and can be exploited jointly to increase accuracy. Although the techniques developed in this paper are general and can be applied to most social graphs, by the word 'graph', we refer to the friendship graphs unless stated otherwise.

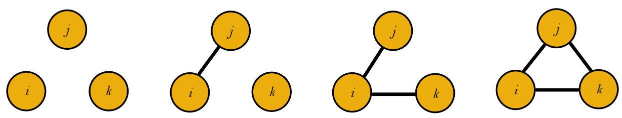

Fig. 1: Triple structures.

At a very small scale, the connections between every triple in a binary graph can be one of the four types shown in Fig. 1. The two types from the right, i.e. the triangle and the two-star, are of particular interest. A triangle shows an ideal transitive relationship while the two-star shows the lack of a cohesive relationship among a connected triple. The average transitivity index τ of a social network is defined as the number of triangles divided by the total number of connected triples which is equal to the number of two-stars plus three times the number of triangles:

$$\tau = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}}$$

where the multiplier 3 accounts for the multiple counts of connected triples in each triangle, restricting τ between 0 and 1. While transitivity is well defined on binary graphs, there is not a unique interpretation of the transitivity for general weighted graphs. In the following, we derive a consistent extension of the transitivity index to weighted graphs.

Let the edge weights be normalized between 0 and 1; and let the diagonal entries of the weight matrix W be all zero. Here the edge weights can correspond to the closeness of relationships or to the frequency of interactions. We associate the transitivity of a triangle to the product of its weights as this value would be small even if only one of the links in this triangle has low weight. Similarly, the weight of a connected triple is defined as the product of the two link weights. We express these notions analytically as:

$$\tau = \frac{\sum_{ij} w_{ij} \sum_k w_{ik} w_{jk}}{\sum_{ij} \sum_k w_{ik} w_{jk}} = \frac{\sum_{ij} w_{ij} [W^2]_{ij}}{\sum_{ij} [W^2]_{ij}} \text{ for } i \neq j \quad (1)$$

We should note that Opsahl et al. [4] give a thorough discussion of how the transitivity index-referred to by the name 'clustering coefficient' in there-should be extended into weighted graphs. The 'geometric mean' definition of clustering coefficient in [4] is very close to (1) except that w_{ij} in the numerator of (1) is replaced with 1 for any non-zero w_{ij} and zero otherwise. For two reasons we did not use the expression of [4]: First, for our purposes, a transitive triangle should have edges with significant weights which is why we take the weight of every edge into account. Second, the expression of [4] is not differentiable, which is not suitable for our objective in this paper.

3. THE TRANSITIVITY MATRIX

Using the properties of the trace function, we can write:

$$\tau = \frac{\alpha}{\beta} = \frac{trace(W^3)}{trace(W^2 H)} \quad (2)$$

where $H_n = 1_{n \times n} - I_n$ is an $n \times n$ matrix of all 1 off-diagonal entries and 0 diagonal entries (here, n is the number of nodes). Let $\nabla_W \tau$ denote the derivative of τ with respect to W. The details for the proof of the following lemma are shown in the Appendix ($A \odot B$ denotes the Hadamard (element-wise) product of two matrices A and B):

Lemma 1 *Let W be a symmetric matrix with zero diagonal entries and positive off-diagonal entries, and let $\tau(W)$ be the transitivity function defined as in (2). Then, the transitivity gradient matrix $\nabla_W \tau$ has the following expression:*

$$\nabla_W \tau = \frac{\partial \tau}{\partial W} = \left(\frac{3\beta W^2 - \alpha(WH + HW)}{\beta^2} \right) \odot H \quad (3)$$

As noted in the introduction, we define the transitivity matrix in terms of the gradient of the transitivity function:

Definition 1 *Let W be a symmetric matrix with zero diagonal entries and positive off-diagonal entries between 0 and 1, and let $\tau(W)$ and $\nabla_W \tau$ be the transitivity function and transitivity gradient matrix of W, respectively. Then, the transitivity matrix T of W is defined as the Hadamard product of $\nabla_W \tau$ and W:*

$$T = \nabla_W \tau \odot W$$

Using (3), we can write:

$$T = \frac{3}{\beta} W^2 \odot W - \frac{\alpha}{\beta^2} (WH + HW) \odot W \quad (4)$$

In the following, we explain the local intuition behind (4) in a graphical context.

3.1. Local interpretation of the transitivity matrix

For simplicity, assume a binary W, which reduces the topology into the existence or absence of the edges. The expression for T in (4) is composed of two parts:

- Mutual neighbors: the term $[W^2 \odot W]_{ij}$, for linked nodes i and j, counts the number of nodes that are connected to both i and j (green nodes in Fig. 2). This value has a positive effect on the transitivity score of the i—j link.

- Combined degree: the quantity $[(WH+HW)\odot W]_{ij}$, for connected nodes i and j, is equal to $deg(i) + deg(j) - 2$ which is the total number of links that intersect the dashed curve in Fig. 2. This value has a negative effect on the transitivity score of the i—j link.

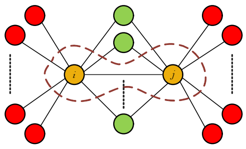

Fig. 2: Computing the transitivity score $[T]_{ij}$ locally.

In the following section, we discuss the utility of the transitivity matrix in social network analysis (SNA).

4. SNA USING THE TRANSITIVITY MATRIX

We begin this section with a critical theorem that endorses the utility of the transitivity matrix for SNA (proof in Appendix).

Theorem 1 *Let T denote the transitivity matrix of the graph with the weight matrix W as defined in (4). The transitivity function of W is equal to the sum of the entries of T:*

$$\tau = \sum_{ij} [T]_{ij}$$

Theorem 1 states that the transitivity index τ of the graph, which is a macroscopic attribute of the social network, can be decomposed through the definition of the transitivity matrix T into microscopic link attributes. In this way, we can tell which links have positive impacts on the overall transitivity index of the network and which ones have negative impacts. The value of this microscopic metric becomes more visible as in the following paragraphs we connect transitivity with redundancy and strength of ties.

Similar to most natural signals, social network graphs contain redundancy in their structure. For example, assume A is a friend of B and C. In most cases, A could maintain its friendships much easier if B and C where also friends, forming a triangle. This example shows the connection of redundancy and the transitivity attribute of social ties; transitive structures, such as cliques, are easier to maintain and more stable. Thus, redundancy in social structure strengthens and coheres social ties. Consequently, we correlate the 'strength' of a link with the level of redundancy of that link.

The social scientist Granovetter [2], in his seminal work of "the strength of weak ties", connects the microscopic and macroscopic levels of social structure. In his research on recent job changers in a Boston suburb, he found that 'those to whom we are weakly tied are more likely to have access to information other than that which we receive'. Later he states bridges are always weak ties, however, not every weak tie is a bridge. Thus, despite the weakness of bridges, they are crucial parts of the social structure. This can be easily explained using the concept of redundancy: Bridges have low redundancies because they decrease transitivity and are thus less cohesive. Using metrics developed in this paper, redundant links have positive transitivity scores while bridging links have negative transitivity scores. Which one of the two types is of more interest depends on the information that is sought after.

These discussions help us understand more about the utility of the transitivity matrix. For example, if the link transitivity scores are associated with the strength of links, they can be used to analyze the cohesion of the network against weakening forces. A worthy experiment is to remove weak links from the graph of a social network to see what would happen to the connectivity of the network if only strong ties could survive. The purpose of this experiment is to test the cohesion of the social network in desperate conditions. For example, the American college football network [5] in Fig. 3 has high local cohesion and low global cohesion because, as can be seen in the right figure of Fig. 3, strong ties form disjoint clusters which makes it easy to break the social structure. On the contrary, the dolphins network [6] shown in Fig. 4 shows high global cohesion at the cost of low local cohesion; while the backbone of the network is not disconnected, many local side nodes have become loose after removing weak ties.

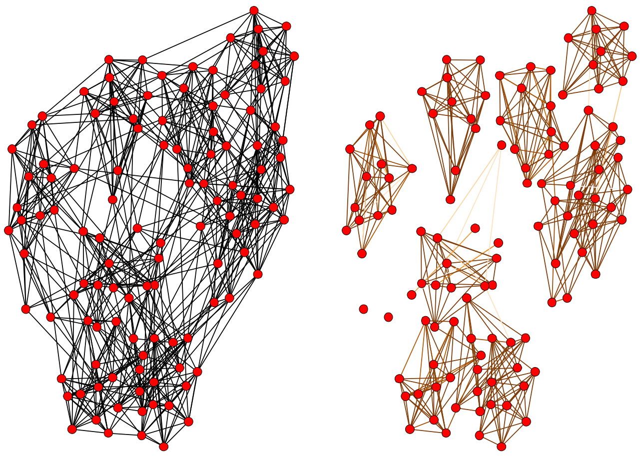

Fig. 3: Left: the football network. Right: showing only edges corresponding to positive link transitivity scores.

One might wonder why there is a need to define a local metric of link transitivity when the transitivity index is closely related to the local clustering coefficient defined in [7]. The next subsection is intended to answer this question.

4.1. Local clustering coefficient vs. link transitivity

In this subsection, we intend to outline the differences between the well-known local clustering coefficient and the proposed link transitivity score. We generalize the clustering coefficient C_i for node i, which quantifies how close i's neighbors are to being a clique to weighted graphs as follows:

$$C_i = \frac{\sum_{jk} w_{ij} w_{ik} w_{jk}}{\sum_{jk} w_{ij} w_{ik}} \quad (5)$$

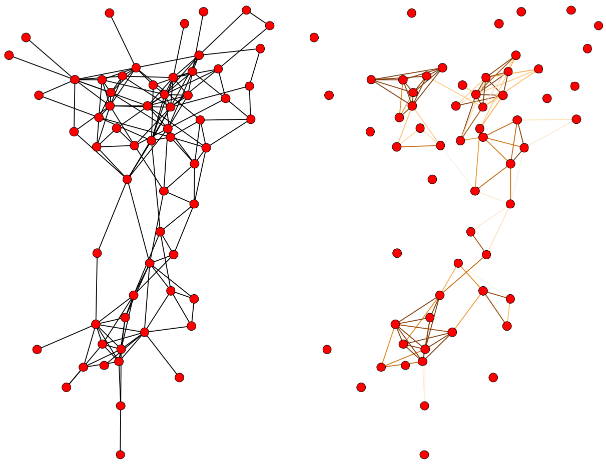

Fig. 4: Left: the dolphins network. Right: showing only edges corresponding to positive link transitivity scores.

which reverts to the original clustering coefficient in the case of binary graphs. The first evident difference is that, unlike link transitivity, the clustering coefficient is defined over nodes. Let's look at both clustering coefficient and link transitivity values on the popular Zachary's karate club network [5] of Fig. 5.

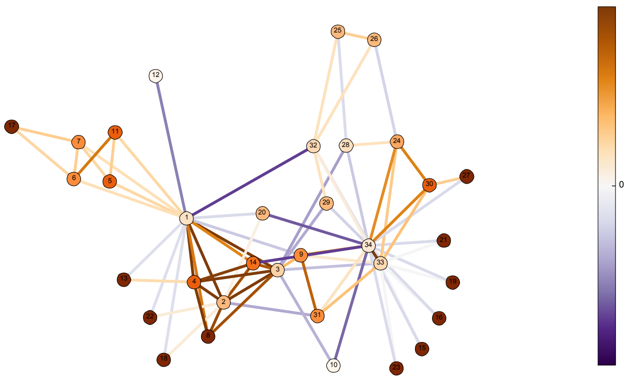

Fig. 5: The karate network (should be viewed in color).

In Fig. 5, colors of nodes correspond to values of their corresponding clustering coefficients, and colors of edges correspond to values of link transitivity scores (the purple color shows negative values). As can be seen in this figure, the side nodes with low degrees have high clustering coefficients while link transitivity scores show that the edges attached to these nodes have negative impact on the overall transitivity score of the network. This counterexample shows that there is not necessarily a meaningful correlation between these two local measures. In other words, both measures provide different microscopic information about the social graph.

5. CONCLUSION

In this paper, we presented a first step towards a matrix algebra theory for the analysis of the transitivity structure of social networks. Nonuniform and singular distribution of transitivity throughout the network shows the existence of closed communities with higher than average transitivities. We described a characteristic quantity of networks that captures the variation of the transitivity throughout the entire network.

6. APPENDIX

6.1. Proof of Lemma 1

Define $\tau(W)$ as in (2). The derivative of τ with respect to diagonal entries w_{ii} will be zero because τ is not a function of w_{ii}. From the basic matrix calculus, we can write (for off-diagonal entries of W):

$$\frac{\partial \alpha}{\partial w_{ij}} = \frac{\partial trace(W^3)}{w_{ij}} = 3[W^2]_{ij} \text{ for } i \neq j$$

and

$$\frac{\partial \beta}{\partial w_{ij}} = \frac{\partial trace(W^2 H)}{w_{ij}} = [WH + HW]_{ij} \text{ for } i \neq j$$

Hence (3).

6.2. Proof of Theorem 1

$$\begin{aligned}
\sum_{ij}[T]_{ij} &= \sum_{ij}[\nabla_W \tau \odot W]_{ij} = trace((\nabla_W \tau)^T W) \\
&= trace((\frac{3\beta W^2 - \alpha(WH + HW)}{\beta^2})W) \\
&= trace((\frac{3\beta W^2}{\beta^2})W - (\frac{\alpha(WH + HW)}{\beta^2})W)
\end{aligned}$$

$$= \frac{3}{\beta} trace(W^3) - \frac{\alpha}{\beta^2}[trace(WHW) + trace(HW^2)]$$

$$= 3\frac{\alpha}{\beta} - \frac{\alpha}{\beta^2}[2\beta] = \frac{\alpha}{\beta} = \tau$$

where we used $trace(ABC) = trace(CAB)$.

7. REFERENCES

[1] Newman M.J. and Park J., "Why social networks are different from other types of networks," in *Physical Review*, 2003, vol. 68, p. 036122.

[2] Granovetter M.S., "The strength of weak ties," in *American Journal of Sociology*, 1973, vol. 78, pp. 1360–1380.

[3] S.P. Borgatti and D.S. Halgin, "On network theory," in *Organization Science*, 2011, vol. 22(5), pp. 1168–1181.

[4] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," in *Social Networks*, 2009, vol. 31(2), pp. 155–163.

[5] M. Girvan and M.J. Newman, "Community structure in social and biological networks," in *Proc. Natl Acad. Sci.*, 2002, vol. 99, pp. 7821–7826.

[6] Lusseau D. et al., "The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: can geographic isolation explain this unique trait?," in *Behav. Ecol. Sociobiol.*, 2003, vol. 54, pp. 396–405.

[7] Watts D.J. and Strogatz S.H., "The collective dynamics of small-world networks," in *Nature*, 1998, vol. 393, pp. 440–442.