Trust and Distrust Across Coalitions: Shapley Value Based Centrality Measures For Signed Networks (Student Abstract Version)

Varun Gangal[1], **Abhishek Narwekar**[1], **Balaraman Ravindran**[1], **Ramasuri Narayanam**[2]
[1] Indian Institute of Technology Madras, Chennai, Tamil Nadu, 600036, India, +91-44-22574350
vgtomahawk@gmail.com, abhisheknkar@gmail.com, ravi@cse.iitm.ac.in
[2] IBM Research India, Bangalore, India
ramasurn@in.ibm.com

Appendix

Shapley Value Centrality - Preliminaries

A coalitional game is defined by a set of agents $A = \{a_1, a_2 \dots a_N\}$ and a characteristic function $\nu() : P(A) \mapsto R$, where $P(A)$ represents the power set of A, and R is the set of real numbers. $\nu(C)$ essentially maps every $C \subseteq A$ to a real number, which represents the payoff of the subset, or the coalition [1]. Generally, $\nu(C)$ is defined such that $\nu(\phi) = 0$, where ϕ is the null set, which corresponds to the coalition with no agents.

Shapley value, first proposed in (Shapley 1952), is a way of distributing the payoff of the grand coalition (the coalition of all agents) amongst each agent. It proposes that the individual payoff of an agent should be determined by considering the marginal contribution of the agent to every possible coalition it is a part of. Such a scheme of distribution is also found to obey certain desirable criteria.

Let $\pi \in \Pi(A)$ be a permutation of A, and let $C_\pi(i)$ denote the coalition of all the predecessors of agent a_i in π, then the Shapley value is defined as

$$SV(a_i) = \frac{1}{|A|!} \sum_{\pi \in \Pi(A)} (\nu(C_\pi(i) \cup a_i) - \nu(C_\pi(i)))$$

For defining a SV based centrality measure on a graph $G = (V, E)$, we consider $A = V$. $\nu(C)$ is defined such that it represents the a measure of power/centrality/influence of the coalition C. The centrality of each node v_i is then given by its Shapley value $SV(v_i)$.

Computing the SV

Earlier methods for SV based centrality, such as (Suri and Narahari 2008) used a Monte-Carlo sampling based approximation to compute $SV(v_i)$ for each v_i. This essentially involves uniformly sampling a large number of permutations from $\Pi(V)$, and then finding the average marginal contribution of v_i as per the definition above. This approach is expensive (since one has to sample a large number of permutations), as well as inexact. Moreover, since the number of permutations grows as $O(n!)$ where $n = |V|$, the number of vertices in the graph.

(Aadithya et al. 2010) first proposed that by defining $\nu(C)$ conveniently, one could compute the SV exactly in closed form. This approach is naturally more preferable than the MC sampling based one, although it requires defining ν, such that SV can be easily derived using arguments from probability and combinatorics.

We now define several SV based centrality measures for directed, signed networks, given by $G = (V, E^+, E^-)$, where each edge $(a, b) \in E^+$ denotes *a trusts b* and each edge $(a, b) \in E^-$ denotes *a distrusts b*. Each measure is based on a different definition of $\nu(C)$. For four of these games, we are able to derive a closed form expression for the measure.

Game Definitions & Closed Form Expressions

Fans Minus Freaks(FMF) This is a simple generalization of the degree centrality measure to directed, signed networks. More formally,

$$FMF(v_i) = d^+_{in}(v_i) - d^-_{in}(v_i)$$

We attempt to generalize this measure to sets of nodes by appropriate definitions of $\nu(C)$, and then compute individual centralities of nodes using the SV of ν.

Net Positive Fringe (NPF) We first define the sets $\nu^+(C)$ and $\nu^-(C)$ [2]. $\nu^+(C)$ is the set of all nodes v_j such that

- v_j has atleast one positive out-neighbor $v_i \in C$ OR
- $v_j \in C$

The second clause results from the intuitive assumption that v_i always trusts itself. $\nu^-(C)$ is the set of all nodes v_j such that v_j has atleast one negative out-neighbor $v_i \in C$. Note that v_j itself may be present inside or outside the coalition, and this does not affect $\nu^-(C)$. The characteristic function $\nu(C)$ is given by

$$\nu(C) = |\nu^+(C)| - |\nu^-(C)|$$

We can think of the $\nu(C)$ as the difference of two characteristic functions, namely $|\nu^+(C)|$ and $|\nu^-(C)|$. Hence

[1] The terms subset and coalition are used here interchangeably

[2] Note the slight overloading of ν here. ν^+ and ν^- denote sets, and not real values

the $SV(v_i)$ of ν will be the difference of the Shapley values $SV^+(v_i)$ and $SV^-(v_i)$ for the characteristic functions $|\nu^+(C)|$ and $|\nu^-(C)|$. Let us refer to these games as Game 1 and Game 2 respectively.

- **Game 1**
  Consider a permutation π sampled uniformly from $\Pi(V)$. Let C be the set of nodes preceding v_i in the coalition. Let v_j by any node $\in V$. Note that v_j could also be v_i itself. We denote B_{v_i,v_j} as the random variable denoting the contribution made by v_i through v_j. Here, "v_i through v_j" means that as a result of v_i being added to C, what is the effect on the membership of v_j in $\nu^+(C)$. The SV of v_i, $SV(v_i)$ is then given by $\sum_{v_j \in V} E[B_{v_i,v_j}]$. Note that since we are taking the expectation over a permutation drawn uniformly from $\Pi(V)$, it is equivalent to averaging over every possible permutation. Now, we can easily see that B_{v_i,v_j} can be non-zero only if $v_j \in v_i \cup N^+(v_i)$, since in all other cases, v_j can neither be added or removed from $\nu^+(C)$ as a result of v_i being added. Now, consider the case where $v_j \in v_i \cup N^+_{in}(v_i)$. $E[B_{v_i,v_j}]$ will be equal to the fraction of permutations in which v_i is able to add v_j to the set $\nu^+(C)$. This can happen only if v_j is neither itself in C, nor is any other out-neighbor of v_j. Note that as a result, the event of v_i adding v_j only depends on the ordering of these $d^+_{out}(v_j)+1$ nodes within the permutation. Hence , we can directly consider the ordering of these nodes, ignoring the other nodes in our calculation. v_i must be the first node amongst these $d^+_{out}(v_j)+1$ in the permutation π, for it to contribute v_j. Hence,

$$\begin{aligned} E[B_{v_i,v_j}] &= \frac{(d^+_{out}(v_j))!}{(d^+_{out}(v_j)+1)!} \\ &= \frac{1}{(d^+_{out}(v_j)+1)} \end{aligned}$$

  Now, $SV^+(v_i)$, the Shapley value of Game 1, is consequently given by

$$SV^+(v_i) = \sum_{v_j \in v_i \cup N^+_{in}(v_i)} \frac{1}{(d^+_{out}(v_j)+1)}$$

- **Game 2**
  Using arguments similar to Game 1, we get

$$SV^-(v_i) = \sum_{v_j \in N^-_{in}(v_i)} \frac{1}{(d^-_{out}(v_j))}$$

  Note that the +1 term in the denominator for the $SV^+(v_i)$ expression is not present here, since the node v_j cannot add itself to $\nu^-(C)$

The final expression of $SV(v_i)$ for the NPF game is thus given by

$$SV(v_i) = \sum_{v_j \in v_i \cup N^+_{in}(v_i)} \frac{1}{(d^+_{out}(v_j)+1)} - \sum_{v_j \in \cup N^-_{in}(v_i)} \frac{1}{(d^-_{out}(v_j))}$$

The time taken to compute the NPF, T_{NPF} would be given by

$$T_{NPF} = O(\sum_{v \in V}(1 + d^+_{in}(v) + d^-_{in}(v)))$$
$$T_{NPF} = O(V+E)$$

Fringe Of Absolute Trust (FAT) Note that we refer to $\nu^+(C)$ and $\nu^-(C)$, as defined in the previous game. In this game, a node is included in the coalition's value if it satisfies both the conditions below

- It either belongs to the coalition or has a positive out neighbor in the coalition.
- It does not have a negative out neighbor in the coalition.

If node $v_j \in N^+_{in}(v_i) \cup v_i$, then B_{v_i,v_j} is +1 if v_i is the first of any of the out-neighbours of v_j (positive or negative) or the node v_j itself to occur in the permutation. This is because if a $v_k \in N^+_{out}(v_j) \cup v_j$ is in C, without any negative out neighbor of v_j being in C, then v_j already is such that $v_j \in \nu^+(C) - \nu^-(C)$. Also, if any negative out neighbor of $v_j \in C$, then v_j can never belong to $\nu^+(C) - \nu^-(C)$, hence adding v_i to C would have no effect. This argument holds good even if $v_i = v_j$. Therefore, for $v_j \in N^+_{in}(v_i) \cup v_i$

$$E[B_{v_i,v_j}] = \frac{1}{d^+_{out}(v_j) + d^-_{out}(v_j) + 1}$$

Now consider $v_j \in N^-_{in}(v_i)$. We can see that $B_{v_i,v_j} \neq 0$ iff

- A node $v_k \in N^+_{out}(v_j) \cup v_j$ belongs to C
- No negative out-neighbor v_k of v_j belongs to C.

In fact, $B_{v_i,v_j} = -1$ if both the conditions above are satisfied. The expectation is given by

$$E[B_{v_i,v_j}] = -\frac{\sum_{x=1}^{x=d^+_{out}(v_j)+1} \binom{d^+_{out}(v_j)+1}{x} x!(d^{total}_{out}(v_j)-x)!}{(d^{total}_{out}(v_j)+1)!}$$

where $d^{total}_{out}(v_j) = d^+_{out}(v_j) + d^-_{out}(v_j)$

Note that $\binom{n}{r}$ represents the number of ways of choosing r distinct things from n distinct things. Since computing factorials for large values can become infeasible in code due to limits on the value of variables, we simplify the expression into a product of fractions form.

$$E[B_{v_i,v_j}] = \frac{-1}{d^{total}_{out}(v_j)+1} \sum_{x=1}^{z=d^+_{out}(v_j)+1} \frac{\prod_{\alpha=1}^{\alpha=x}(d^+_{out}(v_j) - x + \alpha)}{\prod_{\alpha=1}^{\alpha=x}(d^{total}_{out}(v_j) - x + \alpha)}$$

The final expression for $SV(v_i)$ for NPF is given by

$$SV(v_i) = \sum_{v_j \in v_i \cup N^+_{in}(v_i) \cup N^+_{in}(v_i)} B_{v_i,v_j}$$

The complexity of computing FAT, T_{FAT} would be

$$T_{FAT} = O(\sum_{v \in V} d^+_{in}(v_i) + 1 + d{-}_{in}(v_i)(\Delta^+_{out})^2)$$
$$T_{FAT} = O(V + E + E(\Delta^+_{out})^2))$$

where Δ^+_{out} is the maximum positive out degree.

Negated Fringe Of Absolute Distrust (NFADT) This game is in some sense like FAT, but with the roles of distrust and trust reversed. $\nu(C)$ here is given by $-|\nu^-(C) - \nu^+(C)|$. The negative sign is because $|\nu^-(C) - \nu^+(C)|$ would be a measure of disrepute (negative reputation). We omit the expression here for the sake of brevity. The complexity expression of NFADT would be similar to that of T_{FAT}, with Δ^+_{out} replaced by Δ^-_{out}

Net Trust Votes (NTV) Given a coalition C, let E^+ be the set of positive in-edges from a node outside the coalition to a node in the coalition. Similarly, E^- is the set of negative in-edges from a node outside the coalition into the coalition. Note that we do not consider internal edges in either term. In the NTV game, $\nu(C)$ is given by $|E^+| - |E^-|$. Let us now consider the derivation of a closed form expression for this game. Consider the node v_i being added to the C. v_i can contribute to the value of $|E^+|-|E^-|$ in four different ways, as stated below

1. Positive out-edges from v_i to nodes $v_k \in C$. These edges become internal when v_i is added to the coalition, decreasing the value of $|E^+| - |E^-|$ by 1.
2. Negative out-edges from v_i to nodes $v_k \in C$. These edges become internal when v_i is added to the coalition, increasing the value of $|E^+| - |E^-|$ by 1.
3. Positive in-edges from $v_k, v_k \notin C$ to v_i. These edges become a part of E^+, increasing $\nu(C)$ by 1
4. Negative in-edges from $v_k, v_k \notin C$ to v_i. These edges become a part of E^-, decreasing $\nu(C)$ by 1.

Consider case 1. This case only happens for v_i, $v_j \in N^+_{out}(v_i)$. For this event to happen, $v_j \in C$. In other words, it should precede v_i in the permutation. This will happen in exactly half the permutations, and will result in a contribution of $\frac{-1}{2}$. The cumulative contribution as a result of case 1 will be $\sum_{v_j \in N^+_{out}(v_i)} -\frac{1}{2} = -\frac{d^+_{out}(v_i)}{2}$.

Symmetrically, in case 2, the cumulative contribution will be $\sum_{v_j \in N^-_{out}(v_i)} \frac{1}{2} = \frac{d^-_{out}(v_i)}{2}$.

Case 3 can only happen for v_i, $v_j \in N^+_{in}(v_i)$. Here v_j should follow v_i in the permutation i.e, it should not be in C. This will happen in exactly half the permutations. Hence, the cumulative contribution will be $\sum_{v_j \in N^+_{in}(v_i)} \frac{1}{2} = \frac{d^+_{in}(v_i)}{2}$. Symetrically in case 4, we have the cumulative contribution given by $\sum_{v_j \in N^-_{in}(v_i)} -\frac{1}{2} = -\frac{d^-_{in}(v_i)}{2}$. Summing over the contributions from the 4 cases we get

$$SV(v_i) = \frac{1}{2}(d^+_{in}(v_j) - d^-_{in}(v_j)) - \frac{1}{2}(d^+_{out}(v_j) - d^-_{out}(v_j))$$

Since this expression involves only the node degrees, $T_{NTV} = O(V)$, provided we maintain the in-degrees and out-degrees separately.

Robustness To Attacks - Evaluation

(Kumar, Spezzano, and Subrahmanian 2014) mentions several types of attacks carried out by trolls in order to boost their centrality. We evaluate the robustness of our proposed measures as opposed to the FMF measure, in the event of two of the attacks mentioned, namely

1. Malicious collectives (MACO) - These are attacks where a pair of trolls endorses each other(add a trust edge from a to b and b to a)

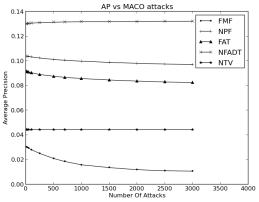
2. Camouflage Behind Good Transcations (CBGT) - These are attacks where a troll adds a trust edge to a benign user. The attack is successful if the benign user reciprocates and adds a trust edge to the troll, increasing the trolls centrality.

We introduce such attacks into the graph by randomly introducing them to the existing Slashdot network, and average the performance of each measure across 100 such random realizations. Each random realization is generated as follows

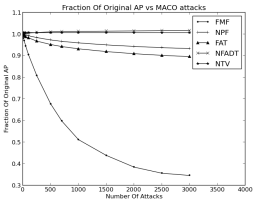
- **MACO**
  Consider all pairs of trolls (a, b) such that there is no positive edge from a to b or b to a. If K attacks are to be added, a set S of K pairs is chosen uniformly at random from these pairs, and a trust edge is added both ways between the vertices in the pair.

- **CBGT**
  Consider all pairs (a, b) such that a is a troll and b is not a troll, and there is no positive edge from a to b or b to a. If K attacks are to be added, a set S of K pairs is chosen uniformly at random from these pairs, and a trust edge is added both ways between the vertices in the pair.

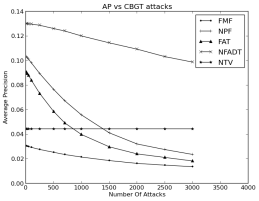
We evaluate each of the approaches for $K \in \{10, 25, 50, 100, 250, 500, 700, 1000, 1500, 2000, 2500, 3000\}$ For each approach, we plot both the absolute variation of AP, as well as the ratio of AP after the attacks to the AP of the approach on the original graph, to consider the relative reduction in AP with increasing number of attacks. We observe that in the case of **MACO** attacks, the **NPF**, **NTV**, **FAT** and **NFADT** approaches show much better robustness in terms of their relative AP as compared to **FMF**, which falls quite sharply. However, in the case of **CBGT** attacks, only **NTV** and **NFADT** perform better than **FMF**, with the relative AP of **NPF** and **FAT** falling more sharply than that of **FMF**.



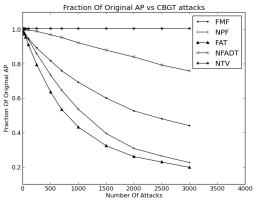
(a) AP for MACO



(b) Relative AP for MACO



(c) AP for CBGT



(d) Relative AP for CBGT

Figure 3: Variation of AP and relative AP with number of attacks

References

Aadithya, K. V.; Ravindran, B.; Michalak, T. P.; and Jennings, N. R. 2010. Efficient computation of the shapley value for centrality in networks. In *Internet and Network Economics*.

Kumar, S.; Spezzano, F.; and Subrahmanian, V. 2014. Accurately detecting trolls in Slashdot Zoo via decluttering. In *ASONAM, 2014*.

Shapley, L. S. 1952. A value for n-person games. Technical report, DTIC Document.

Suri, N. R., and Narahari, Y. 2008. Determining the top-k nodes in social networks using the shapley value. In *AAMAS*.