ADA extension: signed networks

Lorenzo Salmina, Robert Barbaric, Alexandra Korukova
Course of ADA (CS-401)

***Abstract*—When a Wikipedia user is considered for a promotion to the status of admin, the community is able to vote for each candidate. The votes can be either in favour (+1) or in disfavour (-1). Consequently, each vote links the user to the candidate, resulting in a signed network. Furthermore, the Wikipedia voting network offers the unique opportunity to investigate the distribution of positive and negative edges and to compute a vote score for each node. Indeed, each edge representing a positive or negative vote could also be interpreted as a statement of trust/distrust between individuals in the network. To push everything a step forward, we seek to propagate the so called "trust score" along the edges according to specific rules and compare the resulting graph to the original, eventually seeing how different models of propagation affect a network whether.**

I. Introduction

With this extension we aim at analyzing the propagation of the trust and distrust in networks having features which can be described as trust/distrust attributes. Wikipedia dataset, with its information about the admin election, is a perfect candidate for this type of inquiry. Indeed, voting for someone's promotion requires a certain amount of trust in this persons capabilities to take the role. Similarly, casting a negative vote mirrors a distrust of the individual. More specifically, we concentrate our efforts on the modelisation of two different types of propagation: direct trust propagation and transpose trust propagation. This methodology is backed up not only by previous works done in the field [1] but also by the work done in the paper on which this extension is based. In other words, we are combining the propagation of trust models with some rules from status theory.

II. Trust propagation

A. What is trust?

The concept of "trust" or "distrust" underlies in the sign of each edge: for instance, as briefly explained in the introduction, a Wikipedia user during an election can cast a positive (1), neutral (0) and negative (-1) vote. The neutral votes are ignored for the purpose of this extension for two reasons. First, they where not considered in the original study. Second, a neutral vote is not interpretable as a statement of trust/distrust with certainty, since it could also mean that the voter doesn't have an opinion on the trustworthiness of the candidate. Trust can then by assimilated to the concept of interaction's sign, where a positive vote indicates trust and a negative one distrust.

Combining some proprieties of status theory with these premises, it would be possible to assume that the status is in fact the direct consequence of different global trust scores of the network's users. Indeed every individual could have a global trust score which would be the sum of each neighbour's trust statements towards the individual. Higher trust score would indicate higher status.

B. How is trust propagated?

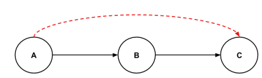


Fig. 1. Directed propagation.

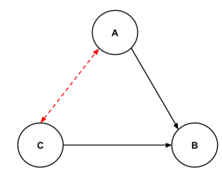


Fig. 2. Transpose trust.

We use two propagation rules to compute the additional trust score of every node of the network due to propagation phenomena: *direct propagation* and *transpose trust* rules. Both *direct propagation* and *transpose trust* involve the triplet of the nodes A, B and C. The former rule states that if the node A trusts B and B trusts C, then there should also be a trust relation from A node towards C (see 1).

The latter one states that if A and C both trust B, than A and B are likely to trust each other (see 2). This kind of model is ambiguous towards the definition of status that we apply. In fact, by following the principle of status, we only know that B has higher status compared to A and C, but nothing can be assumed on the status relation of the latter two. Therefore, we decided to change a little bit the model to fit more accordingly to our assumptions, i.e. by comparing before head the trust scores of A and B and then deciding which individual would trust the other and which one would more likely distrust.

As in both cases the propagated trust that we calculate comes from hypothetical edges that don't exist in the real network, the trust score that comes from these edges should not have the same value as the trust that comes from real edges. Therefore we decided arbitrarily to set the value of hypothetical edges to 0.5.

To include also distrust propagation, we have to define better what score to give in case that we have both negative edges or a mix of positive and negative. Firstly, as before, every hypothetical edge stating distrust is given the value of -0.5. Then, starting with the direct propagation model, if the two existing edges are both negative, then the third edge will also be negative, because by our assumption of status the status of A would be higher then the one of C. On the contrary, if the two existing edges are one positive and one negative the status relation between A and C is ambiguous, therefore we firstly compare the base trust core of each node and assign the hypothetical edge's sign accordingly.

The discussion for the transpose trust model is more complicated. In this model, as stated before, the ambiguous situations arise when both existing edges have the same sign, and in these cases we compare the trust scores of the nodes and move accordingly. In the case where A trusts B and C distrusts B, C is probably of higher status than A, therefore A would hypothetically trust C, and C would distrust A. The opposite situation would happen if A would distrust B and C would trust B.

C. Visualization of the results

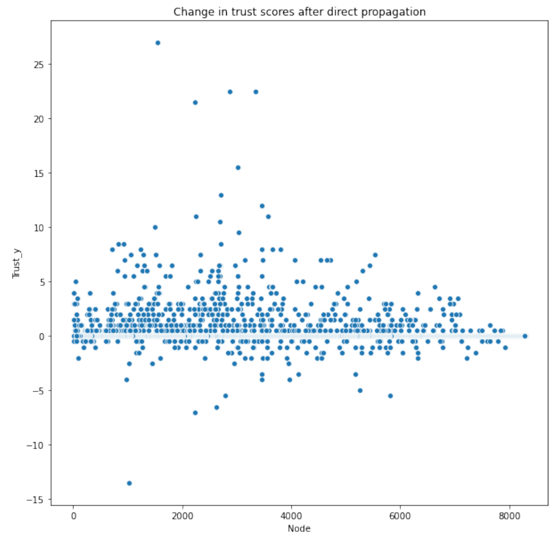
Once the theoretical setup was built, we proceeded to write three functions that we would use to obtain the figures. One function would calculate the trust score of each node, whether the other two would take the wikipedia dataframe and propagate trust on it based on the different models. We were then able to calculate the change in trust score for each node after propagation. To get more insight on the behaviour of propagation we also generated an heatmap representing the various votes given between users before the propagation.

We were unfortunately not able to do the same procedure on the epinions dataframe, which by its own nature has connections that can easily be interpreted as reciprocated trust. This happened mainly because of the enormous size of the latter.

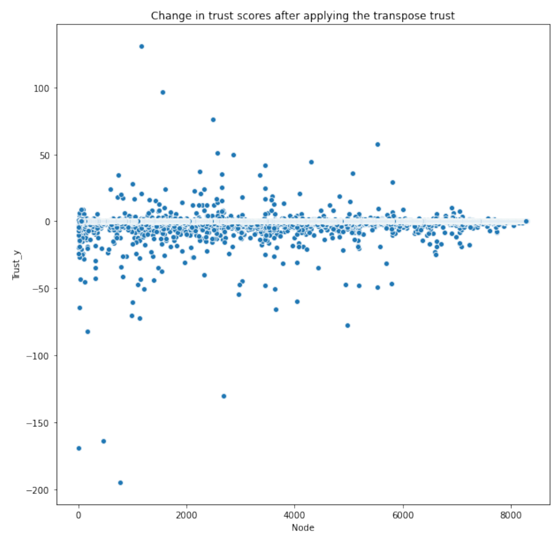
III. Results

To investigate the effects of trust propagation on the trust score we plotted the difference in trust score (trust score after minus before propagation) of each node.

- Direct propagation:



- transpose trust:

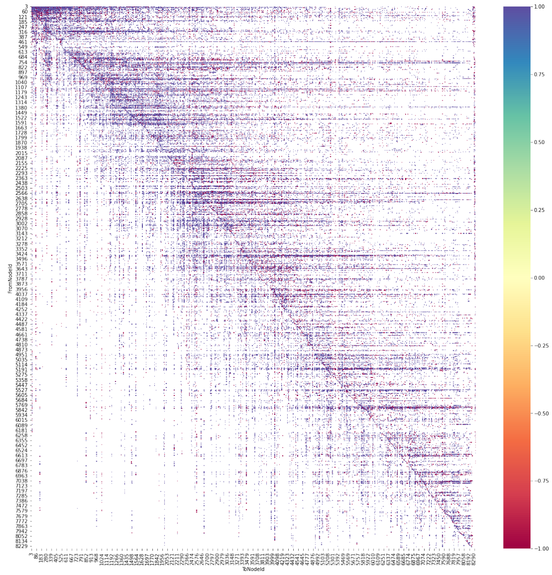


If we compare the two graphs we can observe two striking differences: On is that the scale of the propagation is greater in the case of transpose trust then in the direct propagation (approximately 10 times grater). From this observation we can make the hypothesis that the since the transpose trust is reciprocated the magnitude of the trust score in the regions of high trust become quickly large.

The other is that the transposed trust led to an important increase of the in distrust (some notes got trust score close to -200), on the other hand the direct propagation caused an overall increase in trust.

Also in this case the the fact that the transpose trust is reciprocated could explain the observation: for each positive edge propagated a negative is created in the opposite direction. More in general we can state that that trust is poorly propagated, in both models the trust score does not change significantly (we can see many nodes with value of zero in both graphs).

To backup this last claim we made an heatmap showing the sign of the connection between each nodes.



This representation is very interesting at give us important visual information about the Wikipedia network. For instance the top right triangle of is darker than the bottom left that which shows that there are a lot of people that only take votes without voting. This behaviour leads to the generation of a sparse network which makes the propagation of the trust more difficult.

IV. Conclusion

Working with large signed networks can be very challenging, as it's often hard to get useful visualizations and the computational cost on the networks can quickly go out of hand.

The results seem indicating that the quality and magnitude of trust propagation depends on the model applied to the data. Moreover, it seems that the pre-existing topology of the graph is an influential variable, which can in fact affect the penetrance of trust propagation in the graph.

References

[1] Ramanthan Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, pages 403–412, 2004.